

ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ
ΟΙΚΟΝΟΜΙΚΩΝ ΚΑΙ ΚΟΙΝΩΝΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΕΦΑΡΜΟΣΜΕΝΗΣ ΠΛΗΡΟΦΟΡΙΚΗΣ

Αρχιτεκτονικές και Αλγόριθμοι
Τεχνητών Νευρωνικών Δικτύων
για την Επίλυση Συστημάτων Γραμμικών Εξισώσεων
και Προβλημάτων Ταξινόμησης και Προσαρμογής

Γουλιάνας Κώστας

Διδακτορική Διατριβή
που υποβλήθηκε στο Τμήμα Εφαρμοσμένης Πληροφορικής
του Πανεπιστημίου Μακεδονίας Οικονομικών και Κοινωνικών Επιστημών

ΕΠΙΒΛΕΠΟΥΣΑ ΤΡΙΜΕΛΗΣ ΕΠΙΤΡΟΠΗ

ΕΠΙΒΛΕΠΩΝ : **ΚΩΝΣΤΑΝΤΙΝΟΣ – ΚΛΑΥΔΙΟΣ ΤΣΟΥΡΟΣ**, ΚΑΘΗΓΗΤΗΣ
ΠΑΝΕΠΙΣΤΗΜΙΟΥ ΜΑΚΕΔΟΝΙΑΣ, ΤΜΗΜΑΤΟΣ
ΕΦΑΡΜΟΣΜΕΝΗΣ ΠΛΗΡΟΦΟΡΙΚΗΣ

ΣΥΜΒΟΥΛΕΥΤΙΚΗ ΕΠΙΤΡΟΠΗ

ΚΩΝΣΤΑΝΤΙΝΟΣ ΜΑΡΓΑΡΙΤΗΣ, ΚΑΘΗΓΗΤΗΣ ΠΑΝΕΠΙΣΤΗΜΙΟΥ
ΜΑΚΕΔΟΝΙΑΣ, ΤΜΗΜΑΤΟΣ
ΕΦΑΡΜΟΣΜΕΝΗΣ ΠΛΗΡΟΦΟΡΙΚΗΣ

ΧΑΡΑΛΑΜΠΟΣ ΦΡΑΓΚΑΚΙΣ, ΚΑΘΗΓΗΤΗΣ ΠΟΛΥΤΕΧΝΙΚΗΣ ΣΧΟΛΗΣ ΤΟΥ
ΑΡΙΣΤΟΤΕΛΕΙΟΥ ΠΑΝΕΠΙΣΤΗΜΙΟΥ
ΘΕΣΣΑΛΟΝΙΚΗΣ

Θεσσαλονίκη 2000

ΠΕΡΙΕΧΟΜΕΝΑ

| | | |
|-----------|--|----|
| <u>1.</u> | Εισαγωγή | 1 |
| 1.1 | Γενικά | 1 |
| 1.2 | Επαναληπτικές Μέθοδοι Επίλυσης Συστημάτων Γραμμικών Εξισώσεων και Τεχνητά Νευρωνικά Δίκτυα | 2 |
| 1.3 | Κίνητρα και Στόχοι | 2 |
| 1.4 | Επισκόπηση | 3 |
| <u>2.</u> | Γραμμική Άλγεβρα και Ανάλυση | 4 |
| 2.1 | Βασικές Έννοιες Γραμμικής Άλγεβρας και Ανάλυσης | 4 |
| 2.2 | Σύνολα, Διανύσματα, Υποχώροι, Γραμμική Ανεξαρτησία | 4 |
| 2.2.1 | Norms Διανυσμάτων | 5 |
| 2.2.2 | Ακολουθίες Διανυσμάτων | 6 |
| 2.3 | Μήτρες και Γραμμικά Συστήματα | 7 |
| 2.3.1 | Πράξεις με Μήτρες | 7 |
| 2.3.2 | Γραμμικά Συστήματα Εξισώσεων | 8 |
| 2.3.3 | Τάξη Μήτρας | 8 |
| 2.3.4 | Norms Πινάκων | 9 |
| 2.3.5 | Ακολουθίες Πινάκων | 9 |
| 2.3.6 | Ιδιοτιμές και Ιδιοδιανύσματα Μήτρας | 10 |
| 2.3.7 | Συμμετρικές και Θετικά Ορισμένες Μήτρες | 11 |
| 2.4 | Παράγωγοι και Κυρτότητα | 12 |
| 2.4.1 | Κυρτότητα | 13 |
| <u>3.</u> | Επαναληπτικές Μέθοδοι Επίλυσης Συστημάτων Γραμμικών Εξισώσεων | 14 |
| 3.1 | Γενικά | 14 |
| 3.2 | Επαναληπτικές Μέθοδοι Επίλυσης Γραμμικών Συστημάτων | 14 |
| 3.2.1 | Η Μέθοδος Jacobi | 15 |
| 3.2.2 | Η Μέθοδος Gauss-Seidel | 16 |
| 3.2.3 | Η Μέθοδος Jacobi Overrelaxation (JOR) | 16 |
| 3.2.4 | Η Μέθοδος Successive Overrelaxation (SOR) | 16 |
| 3.2.5 | Η Μέθοδος Richardson | 17 |
| 3.2.6 | Η Μέθοδος Richardson-Gauss-Seidel (RGS) | 17 |
| 3.3 | Σύγκλιση των Επαναληπτικών Μεθόδων | 17 |
| 3.3.1 | Ανάλυση της Σύγκλισης με Τετραγωνική Συνάρτηση Κόστους | 19 |
| 3.3.2 | Ανάλυση της Σύγκλισης με Ελάττωση της Παραγώγου (Gradient Descent) | 20 |
| 3.3.3 | Γενίκευση της Μεθόδου Jacobi | 21 |
| 3.3.4 | Γενίκευση της Μεθόδου Gauss-Seidel | 22 |
| 3.3.5 | Γενίκευση της Μεθόδου Jacobi Overrelaxation (JOR) | 22 |
| 3.3.6 | Γενίκευση της Μεθόδου Successive Overrelaxation (SOR) | 22 |
| 3.3.7 | Γενίκευση της Μεθόδου Richardson | 23 |
| 3.3.8 | Γενίκευση της Μεθόδου Richardson- Gauss-Seidel | 23 |
| 3.4 | Gradient Descent Αλγόριθμοι Εκπαίδευσης | 23 |
| 3.4.1 | Η μέθοδος Newton | 24 |
| 3.4.2 | Η Μέθοδος Adaptive (Optimal) Steepest Descent (ASD) | 25 |
| 3.5 | Conjugate-Gradient Μέθοδοι | 26 |
| 3.5.1 | Η Μέθοδος Polak-Ribiere (PR) | 26 |
| 3.5.2 | Η Μέθοδος Fletcher-Reeves (FR) | 27 |
| 3.5.3 | Η Μέθοδος Optimal Fletcher-Reeves (OFR) | 27 |

| | | |
|-----------|---|-----------|
| 4 | Τεχνητά Νευρωνικά Δίκτυα | 28 |
| 4.1 | Εισαγωγή | 28 |
| 4.2 | Ιστορική Αναδρομή | 29 |
| 4.3 | Εγκέφαλος : Το Βιολογικό Πρότυπο | 29 |
| 4.4 | Χαρακτηριστικά Νευρωνικών Δικτύων | 32 |
| 4.4.1 | Ο Τεχνητός Νευρώνας | 32 |
| 4.4.2 | Συναρτήσεις Ενεργοποίησης | 33 |
| 4.4.3 | Εκπαίδευση Νευρωνικών Δικτύων | 34 |
| 4.4.4 | Ανάκληση - Recall | 35 |
| 4.5 | Παραδείγματα Τεχνητών Νευρωνικών Δικτύων | 36 |
| 4.5.1 | Συσχετιστική Μνήμη (Associative Memory) | 36 |
| 4.5.1.1 | Γραμμική Συσχετιστική Μνήμη (Linear Associative Memory-LAM) | 37 |
| 4.5.1.2 | Εκμάθηση με τον Κανόνα του Hebb | 37 |
| 4.5.1.3 | Βέλτιστη Γραμμική Συσχετιστική Μνήμη (<i>Optimal Linear Associative Memory- OLAM</i>) | 39 |
| 4.5.2 | Νευρωνικά Δίκτυα Μονής Κατεύθυνσης (Feed Forward) | 41 |
| 4.5.2.1 | Μηχανή Αντίληψης (Perceptron) | 42 |
| 4.5.2.2 | Κανόνες Εκπαίδευσης | 43 |
| 4.5.2.3 | Γραμμικό Προσαρμοστικό Στοιχείο (Adaline) | 44 |
| 4.5.2.3.1 | Ο Αλγόριθμος Εκπαίδευσης α -LMS | 44 |
| 4.5.2.3.2 | Ο Αλγόριθμος Εκπαίδευσης μ -LMS (Least Mean Square) | 45 |
| 4.5.2.3.3 | Ο Αλγόριθμος Εκπαίδευσης Batch-LMS | 47 |
| 4.5.2.4 | Αλγόριθμοι Κλιμακωτής Μείωσης (Gradient Descent) | 47 |
| 4.5.2.5 | Γενίκευση του Αλγορίθμου Εκπαίδευσης μ -LMS | 48 |
| 4.5.2.6 | Γενίκευση του Αλγορίθμου Εκπαίδευσης Batch-LMS | 49 |
| 4.5.2.7 | Back-Error Propagation | 51 |
| 4.5.2.8 | Ο Κανόνας του Δέλτα για μη Γραμμικές Συναρτήσεις Ενεργοποίησης σε Δίκτυα Μονής Κατεύθυνσης | 52 |
| 4.5.2.9 | Συντελεστής Εκπαίδευσης και Ορμή (Momentum) | 54 |
| 4.5.2.10 | Η Μέθοδος RPROP | 56 |
| 4.5.3 | Αναδρομικά Δίκτυα (Hopfield) | 57 |
| 4.5.3.1 | Ο Αλγόριθμος του Δικτύου Hopfield | 58 |
| 4.5.3.2 | Η Επιφάνεια της Ενέργειας | 59 |
| 4.5.3.3 | Αποθήκευση Προτύπων - Συνειρμική Μνήμη | 60 |
| 4.5.3.4 | Ανάκληση Διπολικών Προτύπων | 61 |
| 4.5.3.5 | Ανάκληση Συνεχών Προτύπων | 62 |
| 5 | Νευρωνικά Δίκτυα για την Επίλυση Γραμμικών Συστημάτων | 63 |
| 5.1 | Αρχιτεκτονικές και Αλγόριθμοι Δικτύων FeedForward | 63 |
| 5.2 | Αρχιτεκτονικές και Αλγόριθμοι Δικτύων Hopfield | 66 |
| 5.2.1 | Δίκτυα τύπου Hopfield για την Εύρεση Αντιστρόφου Πίνακα | 71 |
| 6 | Νέες Αρχιτεκτονικές και Αλγόριθμοι Νευρωνικών Δικτύων για την Επίλυση Γραμμικών Συστημάτων με Επαναληπτικές Μεθόδους | 72 |
| 6.1 | Αρχιτεκτονικές και Αλγόριθμοι Δικτύων FeedForward | 72 |
| 6.1.1 | Προσομοίωση της μεθόδου SOR | 73 |
| 6.1.2 | Προσομοίωση της μεθόδου JOR | 74 |
| 6.1.3 | Εξισώσεις της μορφής $AX=B$ και Αντιστροφή Πίνακα | 75 |
| 6.2 | Αρχιτεκτονικές και Αλγόριθμοι για επιτάχυνση της Batch-LMS | 76 |
| 6.2.1 | Αρχιτεκτονική του Δικτύου | 76 |
| 6.2.2 | Εκπαίδευση του Δικτύου | 77 |
| 6.2.2.1 | Ο Αλγόριθμος Heuristic Steepest Descent (HSD) | 77 |
| 6.2.2.2 | Ο Αλγόριθμος Adaptive Steepest Descent (ASD) | 79 |
| 6.2.3 | Εξισώσεις της μορφής $AX=B$ και Αντιστροφή Πίνακα | 80 |
| 6.3 | Γενίκευση των Επαναληπτικών Μεθόδων | 81 |

| | | |
|----------|---|------------|
| 6.3.1 | Προσομοίωση της μεθόδου LMS | 84 |
| 6.3.2 | Προσομοίωση της μεθόδου Batch-LMS (B-LMS) | 84 |
| 6.3.3 | Προσομοίωση της μεθόδου JOR..... | 84 |
| 6.3.4 | Προσομοίωση της μεθόδου SOR..... | 85 |
| 6.3.5 | Προσομοίωση της μεθόδου Richardson-Gauss-Seidel (RGS)..... | 85 |
| 6.3.6 | Προσομοίωση της μεθόδου Adaptive Steepest Descent (ASD) | 86 |
| 6.3.7 | Προσομοίωση της μεθόδου Polak-Ribiere (PR)..... | 87 |
| 6.3.8 | Προσομοίωση της μεθόδου Fletcher-Reeves (FR) | 87 |
| 6.3.9 | Προσομοίωση της μεθόδου Optimal Fletcher-Reeves (OFR) | 88 |
| 6.3.10 | Προσομοίωση της μεθόδου RPROP..... | 88 |
| 6.3.11 | Σύγκλιση των μεθόδων 6.3.1-6.3.10..... | 89 |
| 7 | Απόδοση των Προτεινομένων Αλγορίθμων στην Επίλυση Γραμμικών Συστημάτων και Προβλημάτων Ταξινόμησης και Προσαρμογής | 90 |
| 7.1 | Αρχιτεκτονική του Σχήματος 6.1 με Συμμετρική και Θετικά Ορισμένη Μήτρα A | 90 |
| 7.2 | Αρχιτεκτονική του Σχήματος 6.2 - Αλγόριθμοι HSD και ASD | 93 |
| 7.2.1 | Συμπεράσματα | 97 |
| 7.3 | Εφαρμογή της Γενικευμένης Αρχιτεκτονικής 6.3 σε Προβλήματα Ταξινόμησης και Προσαρμογής..... | 99 |
| 7.3.1 | Περιγραφή των Προβλημάτων Ταξινόμησης | 100 |
| 7.3.1.1 | Διάγνωση Καρκίνου Μαστού | 100 |
| 7.3.1.2 | Έγκριση/Απόρριψη Πιστωτικής Κάρτας | 100 |
| 7.3.1.3 | Διάγνωση Διαβήτη | 100 |
| 7.3.1.4 | Κατάταξη Γονιδίων | 101 |
| 7.3.1.5 | Τύποι Γυαλιού..... | 101 |
| 7.3.1.6 | Διάγνωση Καρδιοπάθειας | 101 |
| 7.3.1.7 | Διάγνωση Κολικού Αλόγου | 101 |
| 7.3.1.8 | Αρρώστιες Φασολιών..... | 102 |
| 7.3.1.9 | Διάγνωση υπέρ-θυρεοειδή ή Υπο-θυρεοειδή..... | 102 |
| 7.3.2 | Περιγραφή των Προβλημάτων Προσαρμογής..... | 102 |
| 7.3.2.1 | Οικονομία Ενέργειας Κτιρίων..... | 102 |
| 7.3.2.2 | Ηλιακές Κηλίδες | 102 |
| 7.3.2.3 | Διάγνωση Καρδιοπάθειας | 103 |
| 7.3.3 | Δεδομένα Εκπαίδευσης - Εγκυρότητας - Ανάκλησης | 103 |
| 7.3.4 | Απόδοση των Αλγορίθμων | 104 |
| 7.3.4.1 | Παρατηρήσεις στα Αποτελέσματα των Πινάκων (7.1)-(7.10)..... | 117 |
| 7.3.4.2 | Παρατηρήσεις στα Αποτελέσματα των Πινάκων (7.11)-(7.14)..... | 123 |
| 7.3.4.3 | Παρατηρήσεις στα Αποτελέσματα των Πινάκων (7.15)-(7.24)..... | 135 |
| 7.3.4.4 | Παρατηρήσεις στα Αποτελέσματα των Πινάκων (7.25)-(7.28)..... | 141 |
| 7.3.4.5 | Παρατηρήσεις στα Διαγράμματα 7.10-7.19..... | 153 |
| 7.3.4.6 | Παρατηρήσεις στα Διαγράμματα 7.20-7.23..... | 158 |
| 7.3.4.7 | Παρατηρήσεις στα Διαγράμματα 7.24-7.33..... | 169 |
| 7.3.4.8 | Παρατηρήσεις στα Διαγράμματα 7.34-7.37..... | 174 |
| 8 | Συμπεράσματα | 175 |
| | ΑΝΑΦΟΡΕΣ..... | 177 |

ΔΗΜΟΣΙΕΥΣΕΙΣ

- [1] “**Structured Artificial Neural Networks for Fast Batch LMS Algorithms**“, Goulianas K., Adamopoulos M., Margaritis K.G..
Δημοσιεύτηκε στο περιοδικό *Neural, Parallel & Scientific Computations* Vol. 5, No. 4, (1997), pp. 549-562.
- [2] “**On Multi-Property Set Operations using Neural Networks and Systolic Arrays**“, Margaritis K.G., Goulianas K., Adamopoulos M., Tsouros K.K. and Evans D.J..
Δημοσιεύτηκε στο περιοδικό *Neural, Parallel & Scientific Computations* Vol. 3, Number 1 (1995), pp. 1-14.
- [3] “**Artificial Neural Networks and Iterative Linear Algebra Methods**“, Margaritis KG, Adamopoulos M., Goulianas K., and Evans D.J..
Δημοσιεύτηκε στο περιοδικό *Parallel Algorithms and Applications*, Vol. 3 (1994), pp. 31-44.
- [4] “**Linear Neural Networks and Algorithms for Real-World Benchmark Problems**“, Goulianas K., Adamopoulos M., Fragakis Ch..

Προς Δημοσίευση

- [5] “**Επίλυση Γραμμικών Συστημάτων με Ελαχιστοποίηση της Συνάρτησης Ενέργειας Νευρομορφικού Δικτύου**“, Μαργαρίτης Κ., Αδαμόπουλος Μ., Γουλιάνας Κ..
Δημοσιεύτηκε στην *Επιστημονική Επετηρίδα του Πανεπιστημίου Μακεδονίας*, 1993, Τόμος XI, (Σελίδες 501-525).
- [6] “**Επίλυση Γραμμικών Συστημάτων με τη χρήση Δικτύων τύπου Hopfield και Back-Propagation**“, Αδαμόπουλος Μ., Μαργαρίτης Κ., Γουλιάνας Κ..
Δημοσιεύτηκε στην *Επετηρίδα του Τομέα Ηλεκτρονικής*, του Τμήματος Ηλεκτρολόγων του Αριστοτελείου Παν/μίου Θεσ/νίκης, 1992, (Σελ. 7.1-7.10).

ΣΥΜΜΕΤΟΧΕΣ ΣΕ ΣΥΝΕΔΡΙΑ

- [1] “**Searching for the Occurences of a String in a Text using Artificial Neural Networks**“, Tsouros K. K., Adamopoulos M., Goulianas K. and Margaritis K.G..
Παρουσιάστηκε και Δημοσιεύτηκε στα Πρακτικά του *2nd Balkan Conference on Operational Research*, 18-21 October 1993, Thessaloniki, pp. 784-794.
- [2] “**Information Processing with Artificial Neural Networks**“, Margaritis K.G., Adamopoulos M. and Goulianas K..
Παρουσιάστηκε και Δημοσιεύτηκε στα Πρακτικά (Proceedings) του *2nd BIX Conference of South-East European Society for Neurology and Psychiatry*, 23-26 September 1992, Thessaloniki.
- [3] “**Ανίχνευση Συνόλων Πολλαπλών Ιδιοτήτων με τη χρήση Νευρομορφικών Δικτύων**“, Αδαμόπουλος Μ., Γουλιάνας Κ., Μαργαρίτης Κ. Γ..
Παρουσιάστηκε και Δημοσιεύτηκε στο *4^ο Πανελλήνιο Συνέδριο Πληροφορικής της Ε. Π. Υ.*, Πάτρα, Νοέμβριος 1993, ΤΟΜΟΣ II, (Σελίδες 495-506).
- [4] “**Structured Neural Networks for Iterative Solution of Linear Systems**“, Margaritis K.G., Adamopoulos M., Goulianas K..
Παρουσιάστηκε και Δημοσιεύτηκε στο 1^ο Ελληνικό Συνέδριο Μαθηματικών και Πληροφορικής *HERMIS*, 22 - 26 September 1992 (Σελίδες 499-508).

Περίληψη

Η πρόοδος που έχει σημειωθεί τα τελευταία χρόνια στον τομέα των Τεχνητών Νευρωνικών Δικτύων (ΤΝΔ) και των VLSI κυκλωμάτων, μας δίνει τη δυνατότητα να σχεδιάσουμε αρχιτεκτονικές και να αναπτύξουμε αλγορίθμους εκπαίδευσης που εκμεταλλεύονται τη δυνατότητα παραλληλοποίησης που παρέχουν τα ΤΝΔ.

Έτσι, σε πολλούς τομείς επιστημονικών εφαρμογών, όπως η επεξεργασία σήματος και εικόνας, ρομποτική, συστήματα αυτομάτου ελέγχου, κ.λ.π., όπου τα πιο πολλά προβλήματα αυτού του είδους καταλήγουν σε ένα σύστημα γραμμικών εξισώσεων, καθίσταται επιτακτική η εκμετάλλευση των πλεονεκτημάτων των νευρωνικών δικτύων (παραλληλοποίηση, δυνατότητα εκπαίδευσης) και των πλεονεκτημάτων των επαναληπτικών μεθόδων της Αριθμητικής Ανάλυσης (ταχύτητα, ακρίβεια της λύσης).

Οι επαναληπτικές μέθοδοι της Αριθμητικής Ανάλυσης είναι πολύ εύκολο να υλοποιηθούν με τη χρήση Τεχνητών Νευρωνικών Δικτύων (ΤΝΔ), επειδή οι πράξεις που απαιτούνται περιέχουν συνήθως παλλαπλασιασμό πίνακα επί διάνυσμα.

Στόχος της παρούσης διατριβής είναι η παρουσίαση νέων αλγορίθμων εκπαίδευσης Γραμμικών Νευρωνικών Δικτύων που προσομοιώνουν κλασσικές μεθόδους της Αριθμητικής Ανάλυσης όπως είναι η μέθοδος Jacobi Extrapolated (JOR), η μέθοδος Successive OverRelaxation (SOR), η μέθοδος Richardson, η μέθοδος Richardson Gauss Seidel (RGS), μεθόδους κλιμακωτής μείωσης, όπως είναι η Optimal Steepest Descent (OSD) και μεθόδους συζυγούς κλίσης, όπως είναι η μέθοδος Polak-Ribiere και η μέθοδος Fletcher-Reeves. Αυτό γίνεται δυνατό με την προσαρμογή γνωστών αλγορίθμων εκπαίδευσης, όπως είναι οι αλγόριθμοι εκπαίδευσης των Widrow-Hoff (LMS learning) και ο αλγόριθμος Batch-LMS. Τα πλεονεκτήματα των νέων αλγορίθμων είναι η δυνατότητα παραλληλοποίησής τους, η ταχύτητα των υπολογισμών και η ακρίβεια της λύσης τους. Ο δεύτερος στόχος είναι η παρουσίαση νέων αρχιτεκτονικών, ώστε να μειωθεί το υπολογιστικό κόστος των αλγορίθμων εκπαίδευσης.

Η εκπαίδευση των δικτύων αυτής της μορφής ανάγεται σε ένα πρόβλημα βελτιστοποίησης, όπου ζητείται να βρεθεί ο κατάλληλος συνδυασμός συνάψεων και παραμέτρων, έτσι ώστε η αρχιτεκτονική να παράγει την καλύτερη απεικόνιση μεταξύ των δεδομένων.

Τα δίκτυα που προτείνονται με τους αντίστοιχους αλγόριθμους χρησιμοποιούνται για την επίλυση Συστημάτων Γραμμικών Εξισώσεων κάθε είδους (Τετραγωνικά, Υπερ-ορισμένα και Αόριστα), όπως επίσης και στην Αντιστροφή Πίνακα και ελέγχεται η απόδοσή τους (ταχύτητα σύγκλισης στη λύση του συστήματος και ακρίβεια της λύσης) σε σχέση με τους κλασσικούς αλγόριθμους εκπαίδευσης Γραμμικών Νευρωνικών Δικτύων, LMS και Batch-LMS.

Με τη χρήση των δικτύων σαν ταξινομητών προτύπων (pattern classifiers) εξετάζεται η συμπεριφορά και η απόδοση των παραπάνω μεθόδων στην επίλυση προβλημάτων με πραγματικά δεδομένα, όπως είναι η συλλογή προβλημάτων Proben1, η οποία περιλαμβάνει προβλήματα που εφαρμόζονται στην Ιατρική, τη Γενετική και τη Βιομηχανία, από όπου εξάγονται χρήσιμα συμπεράσματα.

1

Εισαγωγή

1.1 Γενικά

Η επίλυση συστημάτων γραμμικών εξισώσεων θεωρείται σαν ένα απ' τα βασικότερα προβλήματα σε πολλούς τομείς επιστημονικών εφαρμογών, όπως η επεξεργασία σήματος και εικόνας, ρομποτική, συστήματα αυτομάτου ελέγχου, στατιστική, φυσική, ιατρική κ.λ.π. [Ande88], [GoVL83], [Koho89], [MaHP90], [Stur88], [WiWB87], [WiWi88], [WiSt85]. Τα πιο πολλά προβλήματα αυτού του είδους καταλήγουν σε ένα σύστημα γραμμικών εξισώσεων (συνήθως υπερ-ορισμένο) του οποίου ζητείται η λύση ή η αντιστροφή ενός πίνακα σε πραγματικό χρόνο. Μια συλλογή από προβλήματα αυτού του τύπου είναι και η συλλογή Proben1, η οποία περιέχει πραγματικά δεδομένα (real world data) για 12 προβλήματα ταξινόμησης (pattern classification) και προσαρμογής συναρτήσεων (function approximation), εφαρμογές που μπορούν να χρησιμοποιηθούν για την εκτίμηση αλγορίθμων νευρωνικών δικτύων και τη σύγκριση των αποτελεσμάτων. Η χρήση ενός συμβατικού Ηλεκτρονικού Υπολογιστή για την επίλυση των παραπάνω προβλημάτων προσκρούει στην ταχύτητα των υπολογισμών και στο κόστος τους.

Τα τελευταία χρόνια έχουν αναπτυχθεί πολλές παράλληλοι μέθοδοι επίλυσης συστημάτων γραμμικών εξισώσεων και έχουν προταθεί και οι αντίστοιχες αρχιτεκτονικές. Οι πιο πολλές απ' αυτές τις μεθόδους παραλληλοποιούν έναν υπάρχοντα σειριακό αλγόριθμο ή αναπτύσσουν σειριακούς αλγόριθμους που είναι εύκολο να παραλληλοποιηθούν, αλλά πάσχουν από βασικά προβλήματα υλοποίησης. Συνήθως οι υπολογισμοί ακολουθούν μια σειρά βημάτων που εκτελούνται παράλληλα, αλλά υπάρχει χάσιμο χρόνου στο συγχρονισμό όλων των βημάτων επίλυσης του προβλήματος. Με την ανάπτυξη που υπάρχει στον τομέα των VLSI κυκλωμάτων η πλήρης παραλληλοποίηση επιτυγχάνεται με τη χρήση των Τεχνητών Νευρωνικών Δικτύων (ΤΝΔ). Η πρόοδος που έχει σημειωθεί στον τομέα των ΤΝΔ και των VLSI κυκλωμάτων, μας δίνει τη δυνατότητα να σχεδιάσουμε αρχιτεκτονικές και να αναπτύξουμε αλγορίθμους εκπαίδευσης που εκμεταλλεύονται τη δυνατότητα παραλληλοποίησης που παρέχουν τα ΤΝΔ.

1.2 Επαναληπτικές Μέθοδοι Επίλυσης Συστημάτων Γραμμικών Εξισώσεων και Τεχνητά Νευρωνικά Δίκτυα

Ένα πλεονέκτημα των επαναληπτικών μεθόδων είναι η δυνατότητα παραλληλοποίησής τους (αφού οι πράξεις που απαιτούνται περιέχουν συνήθως παλλαπλασιασμό πίνακα επί διάνυσμα), πράγμα που παρέχει τη δυνατότητα υλοποίησής τους με τη χρήση Τεχνητών Νευρωνικών Δικτύων (ΤΝΔ). Επειδή οι επαναληπτικές μέθοδοι δεν τερματίζουν σε συγκεκριμένο χρόνο, ένα μέτρο σύγκρισης των επαναληπτικών αλγορίθμων είναι η ταχύτητα σύγκλισης στη λύση του συστήματος και η ακρίβεια της λύσης.

Οι διάφορες αρχιτεκτονικές ΤΝΔ που έχουν προταθεί για την επίλυση συστημάτων γραμμικών εξισώσεων και γενικά προβλημάτων γραμμικής άλγεβρας, ανάλογα με το είδος του ΤΝΔ που χρησιμοποιείται, μπορούν να χωριστούν σε δύο βασικές κατηγορίες: σε αρχιτεκτονικές ΤΝΔ με προώθηση πληροφοριών προς τα εμπρός (*Feedforward Neural Networks-FNN*) και αρχιτεκτονικές αναδρομικών δικτύων (*Recurrent Neural Networks-RNN*).

Η εκπαίδευση των FNN ανάγεται σε ένα πρόβλημα βελτιστοποίησης, όπου ζητείται να βρεθεί ο κατάλληλος συνδυασμός συνάψεων και παραμέτρων, έτσι ώστε η αρχιτεκτονική να παράγει την καλύτερη απεικόνιση μεταξύ των δεδομένων. Η εκπαίδευση αυτών των δικτύων γίνεται με επίβλεψη (*Supervised*), ενώ ο αλγόριθμος εκπαίδευσης στηρίζεται βασικά στον κανόνα εκπαίδευσης των Widrow-Hoff (*LMS learning*) ή σε παραλλαγές του, όπως η μέθοδος *batch-LMS*, ισοδύναμη με τη μέθοδο *Richardson*.

Στα αναδρομικά δίκτυα, η συγκεκριμένη αρχιτεκτονική καθορίζεται απ' το συγκεκριμένο πρόβλημα προς επίλυση, για το οποίο ζητείται να ελαχιστοποιηθεί η *Συνάρτηση Ενέργειας (Liapunov Function)* του δικτύου. Οι τιμές των συνάψεων μεταξύ των νευρώνων αποβλέπουν σ' αυτό το σκοπό, ενώ οι μέθοδοι ανάκλησης διαφοροποιούνται ανάλογα με τον τρόπο λειτουργίας των νευρώνων που μπορεί να είναι σύγχρονος ή ασύγχρονος.

1.3 Κίνητρα και Στόχοι

Σκοπός της παρούσας διατριβής είναι η ενοποίηση των αρχιτεκτονικών ΤΝΔ για την επίλυση συστημάτων γραμμικών εξισώσεων και η προσαρμογή του αλγορίθμου εκπαίδευσης, ώστε να αξιοποιηθούν τα πλεονεκτήματα των μεθόδων της αριθμητικής ανάλυσης. Οι αλγόριθμοι εκπαίδευσης τροποποιούνται, ώστε να προσομοιάσουν κλασσικές μεθόδους της Αριθμητικής Ανάλυσης όπως η *Jacobi Extrapolated (JOR)*, η οποία προσφέρει τη δυνατότητα παραλληλοποίησης, η μέθοδος *Successive OverRelaxation (SOR)*, η μέθοδος *Richardson* (ισοδύναμη της *Batch-LMS*), η μέθοδος *Richardson Gauss Seidel (RGS)*, μεθόδους κλιμακωτής μείωσης, όπως η *Adaptive Steepest Descent (ASD)* και μεθόδους συζυγούς κλίσης, όπως η *Polak-Ribiere* και *Fletcher-Reeves*. Η συμπεριφορά και απόδοση των παραπάνω μεθόδων εξετάζεται στην επίλυση των προβλημάτων ταξινόμησης και προσαρμογής της συλλογής *Proben1*.

1.4 Επισκόπηση

Η διατριβή χωρίζεται σε πέντε τμήματα: στο πρώτο τμήμα παρουσιάζεται το βασικό μαθηματικό υπόβαθρο (Κεφάλαιο 2) και οι κλασσικές επαναληπτικές μέθοδοι επίλυσης συστημάτων γραμμικών εξισώσεων, όπου εξετάζεται η σύγκλιση και η απόδοσή τους (Κεφάλαιο 3).

Στο δεύτερο τμήμα παρουσιάζονται βασικές έννοιες ΤΝΔ και οι αντίστοιχες αρχιτεκτονικές δικτύων με επίβλεψη και αναδρομικών δικτύων καθώς και οι αλγόριθμοι εκπαίδευσης και ανάκλησης. Εξετάζεται επίσης η σύγκλιση και η απόδοση αυτών των αλγορίθμων (Κεφάλαιο 4). Οι αρχιτεκτονικές και οι αλγόριθμοι που έχουν προταθεί για την επίλυση συστημάτων γραμμικών εξισώσεων παρουσιάζονται στο Κεφάλαιο 5.

Στο τρίτο τμήμα (Κεφάλαιο 6) προτείνεται μια αρχιτεκτονική για τη γενίκευση των παραπάνω μεθόδων, η οποία γίνεται με την παρουσίαση αλγορίθμων εκπαίδευσης, οι οποίοι προσομοιώνουν τη λειτουργία των παραπάνω μεθόδων. Συζητούνται επίσης θέματα που αφορούν τη σύγκλιση των προτεινόμενων αλγορίθμων. Η εφαρμογή των αλγορίθμων στην επίλυση προβλημάτων ταξινόμησης (classification) προσφέρει τη δυνατότητα να λειτουργούν τα παραπάνω δίκτυα σαν συσχετιστικές μνήμες (associative memories).

Στο τέταρτο τμήμα (Κεφάλαιο 7) εξετάζεται η απόδοση των αλγορίθμων, όσον αφορά τον αριθμό επαναλήψεων που απαιτούνται για την εύρεση λύσης και τη μεταβολή του μέσου τετραγωνικού σφάλματος (Mean Square Error-MSE), με τη χρήση παραδειγμάτων συστημάτων γραμμικών εξισώσεων, ενώ στα προβλήματα ταξινόμησης εξετάζεται η επιτυχία των μεθόδων στην κατάταξη των προτύπων.

Στο πέμπτο τμήμα (Κεφάλαιο 8) εξάγονται κάποια γενικά συμπεράσματα και δίνονται κατευθυντήριες γραμμές για περαιτέρω έρευνα στο χώρο.

2 Γραμμική Άλγεβρα και Ανάλυση

2.1 Βασικές Έννοιες Γραμμικής Άλγεβρας και Ανάλυσης

Η επίλυση των γραμμικών συστημάτων και η χρήση των ΤΝΔ απαιτούν κάποιο υπόβαθρο μαθηματικών, γραμμικής άλγεβρας και ανάλυσης και πιο συγκεκριμένα εννοιών όπως διανύσματα, χώροι, εσωτερικά γινόμενα και μήτρες. Το κεφάλαιο που ακολουθεί χρησιμοποιείται σαν μια εισαγωγή στη γραμμική άλγεβρα και ανάλυση, όπου δίνονται οι βασικοί ορισμοί και τα θεωρήματα που απαιτούνται για την παρουσίαση των επαναληπτικών αριθμητικών μεθόδων και των αρχιτεκτονικών ΤΝΔ. Αποδείξεις θεωρημάτων μπορούν να βρεθούν στα [BeTs96], [BeTs89], [DaBi74], [Atki89], [DeMa76].

2.2 Σύνολα, Διανύσματα, Υποχώροι, Γραμμική Ανεξαρτησία

ΟΡΙΣΜΟΣ 2.1: Ένα σύνολο S ορίζεται σαν $S = \{x/x \text{ ικανοποιεί την ιδιότητα } P\}$.

Παραδείγματα συνόλων είναι το σύνολο των πραγματικών αριθμών \mathfrak{R} και το σύνολο των μιγαδικών αριθμών C , ενώ το \mathfrak{R}^n (αντίστοιχα το C^n) είναι το σύνολο των n -διάστατων πραγματικών (αντίστοιχα μιγαδικών) διανυσμάτων $x \in \mathfrak{R}^n$ με $x_i, i=1,2,\dots,n$ τις συντεταγμένες του. Τα διανύσματα θεωρούνται σαν διανύσματα-στήλες, ενώ με x^T (αντίστοιχα x^*) συμβολίζουμε το ανάστροφο διάνυσμα-γραμμή του $x \in \mathfrak{R}^n$ (αντίστοιχα για $x \in C^n$). Το εσωτερικό γινόμενο δύο διανυσμάτων $x, y \in \mathfrak{R}^n$ ορίζεται από τη σχέση $\langle x, y \rangle = \sum_{i=1}^n x_i y_i = x^T y = y^T x$, ενώ για $x, y \in C^n$ έχουμε $\langle x, y \rangle = \sum_{i=1}^n x_i \overline{y_i} = \overline{\langle y, x \rangle}$. Δύο διανύσματα $x, y \in \mathfrak{R}^n$ θα λέγονται ορθογώνια αν $x^T y = 0$.

ΟΡΙΣΜΟΣ 2.2: Ονομάζουμε Διανυσματικό Χώρο (Vector Space) ένα σύνολο n διανυσμάτων, στο οποίο έχουν οριστεί οι πράξεις της πρόσθεσης και του αριθμητικού (scalar) πολλαπλασιασμού, δηλαδή για $x, y \in S \subseteq C^n$ και $a, b \in C$ θα ισχύει $ax + by \in S$. Παραδείγματα διανυσματικών χώρων είναι ο $V = \mathfrak{R}^n$ που περιέχει το σύνολο των n -άδων $(x_1, x_2, \dots, x_n), x_i \in \mathfrak{R}$ και ο $V = C^n$ με το σύνολο των n -άδων $(y_1, y_2, \dots, y_n), y_i \in C$.

ΟΡΙΣΜΟΣ 2.3: Τα n διανύσματα v_1, v_2, \dots, v_n θα λέγονται γραμμικώς εξαρτημένα, εάν υπάρχει ένα σύνολο αριθμών $\alpha_1, \alpha_2, \dots, \alpha_n$, με έναν τουλάχιστον εκ των $\alpha_i \neq 0$, ώστε να ισχύει $\alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n = 0$ ή $v_i = -\frac{\alpha_1}{\alpha_i} v_1 - \frac{\alpha_2}{\alpha_i} v_2 - \dots - \frac{\alpha_n}{\alpha_i} v_n, \alpha_i \neq 0 \forall i=1,2,\dots,n$ και τα διανύσματα v_i ονομάζονται τότε γραμμικός συνδυασμός των v_1, v_2, \dots, v_n .

ΟΡΙΣΜΟΣ 2.4: Τα διανύσματα v_1, v_2, \dots, v_n θα ονομάζονται *γραμμικά ανεξάρτητα*, αν δεν υπάρχει ένα σύνολο συντελεστών $\alpha_1, \alpha_2, \dots, \alpha_n$, τέτοιοι ώστε $\alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n = 0$, εκτός αν $\alpha_1 = \alpha_2 = \dots = \alpha_n = 0$.

ΟΡΙΣΜΟΣ 2.5: Ένα σύνολο διανυσμάτων $\{v_1, v_2, \dots, v_n\}$ θα είναι μια *βάση* του διανυσματικού χώρου V , εάν $\forall v \in V, \exists$ μια μοναδική επιλογή αριθμών (scalars) $\alpha_1, \alpha_2, \dots, \alpha_n$, ώστε να ισχύει $v = \alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n$. Τα διανύσματα v_1, v_2, \dots, v_n είναι γραμμικά ανεξάρτητα και τότε λέμε ότι ο χώρος έχει *διάσταση πεπερασμένη n* . Εάν V είναι ένας διανυσματικός χώρος και $\{v_1, v_2, \dots, v_n\}$ μια βάση αυτού του χώρου, τότε κάθε βάση του χώρου V περιέχει ακριβώς n διανύσματα, και ο αριθμός n ονομάζεται *διάσταση* του χώρου. Π.χ. ο χώρος \mathfrak{R}^n έχει ως μία βάση την $\{e_1, e_2, \dots, e_n\}$ όπου $e_i = \{0, 0, \dots, 0, 1, 0, \dots, 0\}$ με τη μονάδα στην i θέση. Η βάση αυτή ονομάζεται *κανονική* και τα διανύσματά της *μοναδιαία*. Εάν τα διανύσματα v_i είναι κάθετα ανά δύο, τότε η βάση ονομάζεται *ορθοκανονική*, όπως συμβαίνει π.χ. στο τρισσορθογώνιο ορθοκανονικό σύστημα.

2.2.1 Norms Διανυσμάτων

ΟΡΙΣΜΟΣ 2.6: Norm $\|\cdot\|$ του διανύσματος x ενός διανυσματικού χώρου $V \subseteq \mathfrak{R}^n$ είναι μια απεικόνιση ενός πραγματικού αριθμού σε κάθε $x \in V$ που έχει τις παρακάτω ιδιότητες:

- a) $\|x\| \geq 0 \quad \forall x \in V$
- b) $\forall x \in V$ και $a \in \mathfrak{R} \Rightarrow \|ax\| = |a|\|x\|$
- (2.1)
- c) $\|x\| = 0 \Leftrightarrow x = 0$
- d) $\forall x, y \in V \Rightarrow \|x + y\| \leq \|x\| + \|y\|$

- Στην πράξη χρησιμοποιούνται τρεις κυρίως τύποι Norm που προκύπτουν από τον παρακάτω γενικό τύπο

$$\|x\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}}, \quad p=1,2,\infty$$

- Έτσι έχουμε :

- 1) $p = 1$: Η ℓ_1 -norm $\|\cdot\|_1$ ορίζεται σαν $\|x\|_1 = \sum_{i=1}^n |x_i|$.

- 2) $p = 2$.. Το *Ευκλείδειο Μέτρο* (ℓ_2 ή *Euclidean Norm*) $\|\cdot\|_2$ του διανύσματος $x \in \mathfrak{R}^n$

$$\text{ή } x \in C^n \text{ ορίζεται από τη σχέση } \|x\|_2 = \sqrt{\sum_{i=1}^n |x_i|^2}$$

$$3) p \rightarrow \infty: \text{ Η maximum norm ή } \ell_{\infty} \text{ ορίζεται σαν } \|x\|_{\infty} = \lim_{p \rightarrow \infty} \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}} = \max_i |x_i|.$$

- Εάν $x \in V$ με $V \in \mathfrak{R}^n$ ή $V \in C^n$, τότε το Ευκλείδειο Μέτρο διανύσματος έχει τις παρακάτω ιδιότητες:

$$a) \forall x, y \in V \Rightarrow \left| \langle x, y \rangle \right|^2 \leq \langle x, x \rangle \langle y, y \rangle = \|x\|_2 \|y\|_2 \quad (\text{Ανισότητα Cauchy})$$

$$b) \forall x, y \in V \Rightarrow \|x + y\|_2 \leq \|x\|_2 + \|y\|_2 \quad (\text{Τριγωνική ιδιότητα})$$

$$c) \forall \text{ τετραγωνική μήτρα } A, \text{ τάξεως } n \text{ και } \forall x, y \in C^n \Rightarrow \langle Ax, y \rangle = \langle x, A^T y \rangle.$$

ΟΡΙΣΜΟΣ 2.7: Για κάθε $\forall x, y \in \mathfrak{R}^n$ η γωνία ω μεταξύ των διανυσμάτων x, y δίνεται

$$\text{από τη σχέση } \omega = \cos^{-1} \left[\frac{\langle x, y \rangle}{\|x\|_2 \|y\|_2} \right] \Leftrightarrow \langle x, y \rangle = \|x\|_2 \|y\|_2 \cos \omega.$$

ΟΡΙΣΜΟΣ 2.8: Δύο διανύσματα x, y θα λέγονται *ορθογώνια* αν $\langle x, y \rangle = 0$. Εάν $x = \{x^{(1)}, x^{(2)}, \dots, x^{(n)}\}$ είναι μια βάση στο χώρο \mathfrak{R}^n ή C^n και ισχύει $\langle x^i, x^j \rangle = 0$ $\forall i \neq j, 1 \leq i, j \leq n$ τότε η βάση ονομάζεται *ορθογώνια*. Εάν ακόμη ισχύει $\langle x^i, x^i \rangle = 1$, $\forall i = 1, 2, \dots, n$, η βάση ονομάζεται *ορθοκανονική*.

2.2.2 Ακολουθίες Διανυσμάτων

Μια ακολουθία $\{x^{(k)} | k = 1, 2, \dots\}$, $x^{(k)} \in \mathfrak{R}$ συγκλίνει στον πραγματικό αριθμό x , αν για κάθε $\varepsilon > 0$ υπάρχει κάποιο K τέτοιο ώστε $|x^{(k)} - x| < \varepsilon$ για κάθε $k \geq K$. Το x λέγεται *όριο* του $x^{(k)}$ και η σύγκλιση συμβολίζεται με το $x^{(k)} \rightarrow x \Leftrightarrow \lim_{k \rightarrow \infty} x^{(k)} = x$. Μια ακολουθία $\{x^{(k)}\}$ συγκλίνει *γεωμετρικά* στον πραγματικό αριθμό x^* , αν για κάθε $\varepsilon > 0$ υπάρχουν σταθερές $A \geq 0$ και $0 \leq \alpha < 1$ τέτοιες ώστε $|x^{(k)} - x^*| \leq A \alpha^k$ για κάθε k . Μια ακολουθία $\{x^{(k)}\}$ λέγεται *ακολουθία Cauchy*, αν για κάθε $\varepsilon > 0$ υπάρχει K τέτοιο ώστε $|x^{(k)} - x^{(m)}| < \varepsilon$ για κάθε $k \geq K$ και $m \geq K$. Μια ακολουθία $\{x^{(k)} | k = 1, 2, \dots\}$, $x^{(k)} \in \mathfrak{R}^n$ συγκλίνει στο διάνυσμα $x \in \mathfrak{R}^n$, αν για κάθε $i = 1, 2, \dots, n$ το $x_i^{(k)}$ συγκλίνει στο x_i και συμβολίζεται $x^{(k)} \rightarrow x \Leftrightarrow \lim_{k \rightarrow \infty} x_i^k = x_i, \forall i = 1, 2, 3, \dots, n$.

ΘΕΩΡΗΜΑ 2.1 : Μια ακολουθία $\{x^{(k)}\}$, $x^{(k)} \in \mathfrak{R}^n$ συγκλίνει στο x , όταν και μόνο όταν $\lim_{k \rightarrow \infty} \|x^{(k)} - x\| = 0$. Μια ακολουθία $\{x^{(k)}\}$, $x^{(k)} \in \mathfrak{R}^n$ συγκλίνει *γεωμετρικά* στο x , όταν και μόνο όταν το $\|x^{(k)} - x\|$ συγκλίνει γεωμετρικά στο 0 .

2.3 Μήτρες και Γραμμικά Συστήματα

Ονομάζουμε *Μήτρα* ή *Πίνακα* τάξεως $m \times n$ μια ορθογώνια διάταξη στοιχείων, πραγματικών ή μιγαδικών αριθμών, που η γενική μορφή της είναι

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \quad (2.2)$$

με $a_{ij} \in \mathfrak{R}$ ή $a_{ij} \in \mathfrak{C}$. Οι μήτρες συνήθως συμβολίζονται με τα κεφαλαία γράμματα του Λατινικού αλφαβήτου και τα στοιχεία τους με μικρά, (συνήθως τα αντίστοιχα του ονόματος της μήτρας), με δύο δείκτες, όπου ο πρώτος δείκτης ορίζει τη γραμμή και ο δεύτερος τη στήλη του στοιχείου. Εάν $A \in \mathfrak{R}^{m \times n}$, ονομάζουμε *ανάστροφη* μήτρα της A τη μήτρα $C = A^T \in \mathfrak{R}^{n \times m}$, της οποίας τα στοιχεία ορίζονται από την σχέση $c_{ij} = a_{ji}$. Εάν $A^T = A$, η μήτρα A καλείται *συμμετρική*. Ως *μηδενική* μήτρα ορίζουμε τη μήτρα $0 \in \mathfrak{R}^{m \times n}$ ή $0 \in \mathfrak{C}^{m \times n}$: $0_{ij} = 0$ ώστε $0 + A = A + 0 = A$, $\forall A \in \mathfrak{R}^{m \times n}$, ενώ ως *ταυτοτική* ή *μοναδιαία* μήτρα $I \in \mathfrak{R}^{n \times n}$ ορίζουμε τη μήτρα που τα στοιχεία της ορίζονται από την

$$\text{σχέση } \delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}, \text{ όπου το σύμβολο } \delta_{ij} \text{ ονομάζεται και } \textit{σύμβολο του Kronecker}$$

(*Kronecker Delta Function*). Εάν η μήτρα A είναι τετραγωνικής τάξεως n και υπάρχει μια μήτρα B , τάξεως n τέτοια ώστε $AB = BA = I$, τότε η μήτρα B ονομάζεται *αντίστροφη* της A και γράφουμε $B = A^{-1}$. Μια μήτρα ονομάζεται *Hermitian*, αν $A^* = A$, $A \in \mathfrak{C}^{n \times n}$, όπου A^* η *συζυγής* μήτρα της A . Μια μήτρα A ονομάζεται *αντι-συμμετρική* (*Skew-Symmetric*), αν $A^T = -A$. Οι μήτρες των μορφών αυτών είναι φανερό ότι πρέπει να είναι τετραγωνικές. Μια μήτρα $A \in \mathfrak{C}^{n \times n}$ ονομάζεται *ορθογώνια* (*unitary*) εάν ισχύει $A^* A = AA^* = I$, όπου A^* η *συζυγής* μήτρα της A . Αν $A \in \mathfrak{R}^{n \times n}$ και ισχύει $A^T A = AA^T = I$, η μήτρα ονομάζεται επίσης *ορθογώνια* (*orthogonal*).

2.3.1 Πράξεις με Μήτρες

Αν $A \in \mathfrak{R}^{m \times n}$, $B \in \mathfrak{R}^{m \times n}$ τότε το *άθροισμα* $A+B$ ορίζει στο διανυσματικό χώρο $\mathfrak{R}^{m \times n}$ μια νέα μήτρα C με στοιχεία c_{ij} , τα οποία δίνονται από τη σχέση $c_{ij} = a_{ij} + b_{ij}$. Εάν $A \in \mathfrak{R}^{m \times n}$ και $\lambda \in \mathfrak{R}$ τότε ο αριθμητικός *πολλαπλασιασμός* λA ορίζει στο διανυσματικό χώρο $\mathfrak{R}^{m \times n}$ μια νέα μήτρα C , της οποίας τα στοιχεία c_{ij} ορίζονται από την σχέση $c_{ij} = \lambda a_{ij}$. Εάν $A \in \mathfrak{R}^{m \times n}$, $B \in \mathfrak{R}^{n \times p}$ ορίζουμε ως *γινόμενο* των μητρών A, B τη μήτρα $C \in \mathfrak{R}^{m \times p}$, της οποίας τα στοιχεία ορίζονται από τη σχέση $c_{ij} = \sum_{k=1}^n a_{ik} b_{kj}$, $i = 1 \dots m$, $j = 1 \dots p$.

- Οι παρακάτω ιδιότητες μπορούν να αποδειχτούν χωρίς ιδιαίτερη δυσκολία.

$$A + B = B + A \quad (A + B) + C = A + (B + C)$$

$$A(B + C) = AB + AC \quad A(BC) = (AB)C$$

$$(A + B)^T = A^T + B^T \quad (AB)^T = B^T A^T$$

- Οι διανυσματικοί χώροι \mathbb{R}^n και C^n ταυτίζονται με τα σύνολα των στηλών τάξεως $n \times 1$ με στοιχεία πραγματικούς ή μιγαδικούς αντίστοιχα.

2.3.2 Γραμμικά Συστήματα Εξισώσεων

- Το γραμμικό σύστημα της μορφής

$$a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1$$

$$a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2$$

.....

$$a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n = b_m$$

(2.3)

θα μπορούσε να γραφεί σαν

$$Ax = b \quad (2.4)$$

όπου ο πίνακας A δίνεται από τον τύπο (2.2), $x = [x_1, x_2, \dots, x_n]^T$ και $b = [b_1, b_2, \dots, b_m]^T$.

2.3.3 Τάξη Μήτρας

ΟΡΙΣΜΟΣ 2.9: Εάν A είναι μια $m \times n$ μήτρα, ο αριθμός των γραμμικώς ανεξαρτήτων στηλών της ονομάζεται *τάξη της μήτρας Κατά Στήλες (Column Rank)*, ο δε αριθμός των γραμμικώς ανεξαρτήτων γραμμών της ονομάζεται *Τάξη της μήτρας Κατά Γραμμές*. Αποδεικνύεται ότι οι δύο αυτοί αριθμοί συμπίπτουν και η κοινή τους τιμή ονομάζεται *Τάξη της μήτρας*.

2.3.4 Norms Πινάκων

Τα σύνολα των $n \times n$ πινάκων με στοιχεία πραγματικούς ή και μιγαδικούς αριθμούς μπορούν να θεωρηθούν ως ισοδύναμα των διανυσματικών χώρων $\mathfrak{R}^{n \times n}$ και $C^{n \times n}$ αντίστοιχα.

ΟΡΙΣΜΟΣ 2.10: Ως *Norm* του πίνακα A ορίζεται το μέγεθος $\|A\|$ που πληροί τις παρακάτω σχέσεις

- 1) $\forall A \in V \Rightarrow \|A\| \in \mathfrak{R}^+$
- 2) $\forall A \in V$ και $\lambda \in \mathfrak{R} \Rightarrow \|\lambda A\| = |\lambda| \cdot \|A\|$
- 3) $\forall A, B \in V \Rightarrow \|A + B\| \leq \|A\| + \|B\|$
- 4) $\forall A, B \in V \Rightarrow \|A \cdot B\| \leq \|A\| \cdot \|B\|$
- 5) $\forall x \in V_1$ και $A \in V_2 \Rightarrow \|Ax\| \leq \|A\| \cdot \|x\|$ με $V_2 = C^{n \times n}$ ή $V_2 = \mathfrak{R}^{n \times n}$, $V_1 = C^n$ ή $V_1 = \mathfrak{R}^n$.

- Συνήθως, όταν δίνεται ένας διανυσματικός χώρος με μία Norm $\|x\|$ και ένας πίνακας A , ορίζουμε ως συναφή προς τον πίνακα A Norm την ισοδύναμη με την ιδιότητα 5

$$\|A\| = \max \frac{\|Ax\|}{\|x\|} \quad (2.5)$$

- Χρησιμοποιώντας την (2.5), μπορούμε να αποδείξουμε ότι για τις Norms του πίνακα ισχύουν

$$\begin{aligned} 1) \|A\|_1 &= \max_j \sum_{i=1}^n \|a_{ij}\| \\ 2) \|A\|_2 &= [\rho(A^T A)]^{1/2} \\ 3) \|A\|_\infty &= \max_i \sum_{j=1}^n \|a_{ij}\| \end{aligned} \quad (2.6)$$

όπου ρ είναι η φασματική ακτίνα του γινομένου $A^T A$.

2.3.5 Ακολουθίες Πινάκων

- 1) Αν $A^{(k)} \in \mathfrak{R}^{n \times n}$ ή $A^{(k)} \in C^{n \times n}$, τότε $A^{(k)} \rightarrow 0 \Leftrightarrow \lim_{k \rightarrow \infty} \{a_{ij}^{(k)}\} = 0$, $i, j = 1, 2, 3, \dots, n$
- 2) Αν $A^{(k)}, A \in \mathfrak{R}^{n \times n}$ ή $A^{(k)}, A \in C^{n \times n}$ τότε $A^{(k)} \rightarrow A \Leftrightarrow \lim_{k \rightarrow \infty} \|A^{(k)} - A\| = 0$.

ΘΕΩΡΗΜΑ 2.2:

- 1) $\lim_{k \rightarrow \infty} \|A^k - A\| = 0 \Leftrightarrow \rho(A) < 1$
- 2) $\forall A \in \mathfrak{R}^{n \times n}, A \in C^{n \times n} \Rightarrow \|A^m\| \leq \|A\|^m, m \in \mathbb{N}$
- 3) $\lim_{k \rightarrow \infty} \|A^m\| = 0 \Leftrightarrow \|A\| < 1, m \in \mathbb{N}$
- 4) $\forall A \in \mathfrak{R}^{n \times n}, A \in C^{n \times n} \Rightarrow \rho(A) \leq \|A\|$

2.3.6 Ιδιοτιμές και Ιδιοδιανύσματα Μήτρας

ΟΡΙΣΜΟΣ 2.11: Μια τετραγωνική μήτρα λέγεται *ιδιάζουσα* (*singular*) αν η ορίζουσά της είναι μηδέν. Διαφορετικά λέγεται *μη-ιδιάζουσα* (*nonsingular*) ή *αντιστρέψιμη*.

ΘΕΩΡΗΜΑ 2.3: Για μια τετραγωνική μήτρα $A \in \mathfrak{R}^{n \times n}$, $b \in \mathfrak{R}^n$, $x \in \mathfrak{R}^n$ (C^n), οι παρακάτω προτάσεις είναι ισοδύναμες:

- α) Η μήτρα A είναι *μη-ιδιάζουσα*
- β) Η μήτρα $A^T A$ είναι *μη-ιδιάζουσα*
- γ) $\forall x \neq 0 \Rightarrow Ax \neq 0, x \in \mathfrak{R}^n$
- δ) $Ax = 0 \Rightarrow x = 0$.
- ε) $\forall b \in \mathfrak{R}^n$ ή $b \in C^n$, υπάρχει ένα μοναδικό $x \in \mathfrak{R}^n$ ή $x \in C^n$ έτσι ώστε $Ax = b$.
- ζ) Υπάρχει μια μήτρα B , έτσι ώστε $AB = I = BA$.
- η) Η Ορίζουσα της μήτρας A , $\det(A) \neq 0$.
- θ) Η Τάξη της μήτρας A είναι $r(A) = n$.

- Η μήτρα B της πρότασης ζ) λέγεται *αντίστροφη* της A , συμβολίζεται με A^{-1} και είναι μοναδική.
- Για κάθε τετραγωνικές μήτρες A, B με την ίδια διάσταση θα ισχύει $(AB)^{-1} = B^{-1}A^{-1}$.

ΟΡΙΣΜΟΣ 2.12: Εάν $A \in \mathfrak{R}^{n \times n}$, $x \in \mathfrak{R}^n$, $\lambda \in \mathfrak{R}$ και $Ax = \lambda x$, τότε ο πραγματικός αριθμός λ ονομάζεται *ιδιοτιμή* της μήτρας A και το διάνυσμα x το αντίστοιχο *ιδιοδιάνυσμα* στην ιδιοτιμή λ .

ΟΡΙΣΜΟΣ 2.13 : Εάν $A \in \mathfrak{R}^{n \times n}$, $x \in \mathfrak{R}^n$ και $\lambda \in \mathfrak{R}$, τότε το *χαρακτηριστικό πολυώνυμο* φ της μήτρας A ορίζεται σαν $\varphi(\lambda) = \det(\lambda I - A)$, με I τη μοναδιαία μήτρα και λ πραγματικό αριθμό. Οι n ρίζες του φ ονομάζονται *ιδιοτιμές* της μήτρας A και τα διανύσματα που πληρούν τη σχέση $Ax = \lambda x$ τα *ιδιοδιανύσματα* που αντιστοιχούν στην ιδιοτιμή λ .

- Το γραμμικό σύστημα που ορίζει η (2.3) είναι ισοδύναμο με το

$$\begin{aligned}(a_{11} - \lambda)x_1 + a_{12}x_2 + \dots + a_{1n} &= 0 \\ a_{21}x_1 + (a_{22} - \lambda)x_2 + \dots + a_{2n} &= 0 \\ \dots & \\ a_{n1}x_1 + a_{n2}x_2 + \dots + (a_{nn} - \lambda)x_n &= 0\end{aligned}$$

- Το σύστημα αυτό θέλουμε να έχει και άλλες λύσεις εκτός της μηδενικής, γι' αυτό ισχύει η σχέση

$$\det \begin{bmatrix} a_{11} - \lambda & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} - \lambda & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} - \lambda \end{bmatrix} = 0$$

ή

$$\begin{aligned}\varphi(\lambda) &= (a_{11} - \lambda)(a_{22} - \lambda) \dots (a_{nn} - \lambda) + P_{n-2}(\lambda) \\ &= (-1)^n \lambda^n + (-1)^{n-1} (a_{11} + a_{22} + \dots + a_{nn}) \lambda^{n-1} + P_{n-2}(\lambda)\end{aligned}$$

όπου P_{n-2} ένα πολυώνυμο ως προς λ βαθμού $\leq n-2$.

ΟΡΙΣΜΟΣ 2.14: Το ίχνος της τετραγωνικής μήτρας A ορίζεται (από το συντελεστή του λ^{n-1}) σαν το άθροισμα των διαγωνίων στοιχείων της $\text{tr}(A) = a_{11} + a_{22} + \dots + a_{nn}$.

ΟΡΙΣΜΟΣ 2.15: Η μέγιστη εκ των ιδιοτιμών ονομάζεται φασματική ακτίνα της μήτρας A , είναι δηλαδή $\rho(A) = \max\{|\lambda_1|, |\lambda_2|, |\lambda_3|, \dots, |\lambda_n|\}$.

2.3.7 Συμμετρικές και Θετικά Ορισμένες Μήτρες

ΘΕΩΡΗΜΑ 2.4: Για μια πραγματική συμμετρική $n \times n$ μήτρα A ισχύουν τα παρακάτω:

- Οι ιδιοτιμές της A είναι πραγματικές
- Η μήτρα A έχει n πραγματικά ορθογώνια ανά δύο και διάφορα του μηδενός ιδιοδιανύσματα $x^{(1)}, x^{(2)}, \dots, x^{(n)}$ που αντιστοιχούν στις ιδιοτιμές $\lambda_1, \lambda_2, \dots, \lambda_n$.
- Αν για τα παραπάνω ιδιοδιανύσματα ισχύει $\|x^{(i)}\|_2 = 1, i = 1, 2, \dots, n$ τότε

$$A = \sum_{i=1}^n \lambda_i x^{(i)} (x^{(i)})^T.$$

ΟΡΙΣΜΟΣ 2.16: Μια τετραγωνική μήτρα $A \in \mathfrak{R}^{n \times n}$ λέγεται *θετικά ορισμένη* αν $x^T A x > 0 \quad \forall x \in \mathfrak{R}^n, x \neq 0$. Αν $x^T A x \geq 0 \quad \forall x \in \mathfrak{R}^n$ η μήτρα λέγεται *μη-αρνητικά ορισμένη*.

ΘΕΩΡΗΜΑ 2.5: Για κάθε πραγματική μήτρα A , η μήτρα $A^T A$ είναι συμμετρική και μη-αρνητικά ορισμένη. Αν δε είναι *μη-ιδιάζουσα*, τότε η $A^T A$ είναι και θετικά ορισμένη.

ΘΕΩΡΗΜΑ 2.6: Η αντίστροφη μιας συμμετρικής και θετικά ορισμένης μήτρας είναι συμμετρική και θετικά ορισμένη.

2.4 Παράγωγοι και Κυρτότητα

Για μια συνάρτηση $f: \mathfrak{R}^n \rightarrow \mathfrak{R}$ και $x \in \mathfrak{R}^n$ ορίζουμε σαν *μερική παράγωγο* της f στο x το όριο $\lim_{\alpha \rightarrow 0} \frac{f(x + \alpha e_i) - f(x)}{\alpha}$ και συμβολίζεται σαν $\frac{\partial f(x)}{\partial x_i}$. Αν υπάρχουν οι μερικές παράγωγοι για κάθε i , το διάνυσμα με στοιχεία τις μερικές παραγώγους $\frac{\partial f(x)}{\partial x_i}$ συμβολίζεται με $\nabla f(x)$ και ονομάζεται *κλίση* της f .

Μια συνάρτηση $f: \mathfrak{R}^n \rightarrow \mathfrak{R}^m$ είναι *παραγωγίσιμη* αν κάθε στοιχείο f_i της f είναι παραγωγίσιμο. Τότε με $\nabla f(x)$ ορίζουμε την $n \times m$ μήτρα με στήλες τα αντίστοιχα $\nabla f_i(x)$, δηλαδή $\nabla f(x) = [\nabla f_1(x), \nabla f_2(x), \dots, \nabla f_m(x)]$. Η ανάστροφη της $\nabla f(x)$ λέγεται *Ιακωβιανή (Jacobian)* της f . Αν όλες οι μερικές παράγωγοι $\frac{\partial f(x)}{\partial x_i}$ μιας συνάρτησης

$f: \mathfrak{R}^n \rightarrow \mathfrak{R}$ είναι συνεχείς και διαφορίσιμες ορίζουμε σαν $\frac{\partial^2 f(x)}{\partial x_i \partial x_j}$ την i -οστή μερική παράγωγο δεύτερης τάξης της $\frac{\partial f(x)}{\partial x_j}$ στο $x \in \mathfrak{R}^n$. Οι μερικές παράγωγοι $\frac{\partial^2 f(x)}{\partial x_i \partial x_j}$ που συμβολίζονται $\nabla_{ij}^2 f(x)$ για $i = j$ λέγονται *άμεσες* και για $i \neq j$ *σταυροειδείς* μπορούν να θεωρηθούν σαν στοιχεία μιας $n \times n$ μήτρας που λέγεται *Hessian* της f , η οποία συμβολίζεται $\nabla^2 f(x)$ και είναι συμμετρική (δηλαδή $\frac{\partial^2 f(x)}{\partial x_i \partial x_j} = \frac{\partial^2 f(x)}{\partial x_j \partial x_i}$).

ΘΕΩΡΗΜΑ 2.7: (Descent Lemma) Αν $f: \mathfrak{R}^n \rightarrow \mathfrak{R}$ συνεχής και παραγωγίσιμη και $\forall x, y \in \mathfrak{R}^n$ ισχύει $\|\nabla f(x) - \nabla f(y)\|_2 \leq K \|x - y\|_2$ τότε $f(x + y) \leq f(x) + y^T \nabla f(x) + \frac{K}{2} \|y\|_2^2$.

ΟΡΙΣΜΟΣ 2.17: Αν $V \subset \mathfrak{R}^n$ και $f:V \mapsto \mathfrak{R}$, το διάνυσμα $x \in V$ λέγεται *τοπικό ελάχιστο* (*local minimum*) της f , αν υπάρχει $\varepsilon > 0$ τέτοιο ώστε $f(y) \geq f(x)$ για κάθε $y \in V$ με $\|x - y\|_2 \leq \varepsilon$. Αν $f(y) \geq f(x)$ για κάθε $y \in V$, το διάνυσμα $x \in V$ λέγεται *ολικό ελάχιστο* (*global minimum*) της f .

ΘΕΩΡΗΜΑ 2.8 : Αν $f:\mathfrak{R}^n \mapsto \mathfrak{R}$ είναι συνεχής και παραγωγίσιμη, τότε αν το $x \in \mathfrak{R}^n$ είναι *τοπικό ελάχιστο* της f , θα ισχύει $\nabla f(x) = 0$.

2.4.1 Κυρτότητα

ΟΡΙΣΜΟΣ 2.18: Το υποσύνολο $V \subset \mathfrak{R}^n$ λέγεται *κυρτό* (*convex*) αν $\alpha x + (1 - \alpha)y \in V$ για κάθε $x, y \in V$ και $\alpha \in [0, 1]$. Μια συνάρτηση $f:V \mapsto \mathfrak{R}$ λέγεται *κυρτή* αν $f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y)$, $\forall x, y \in V$ και $\alpha \in [0, 1]$. Η f λέγεται *αυστηρώς κυρτή* (*strictly convex*) αν $\forall x, y \in V$ με $x \neq y$ ισχύει $f(\alpha x + (1 - \alpha)y) < \alpha f(x) + (1 - \alpha)f(y)$ $\forall \alpha \in [0, 1]$.

ΘΕΩΡΗΜΑ 2.9: Για ένα κυρτό σύνολο $V \subset \mathfrak{R}^n$ θα ισχύει

- a) Αν το $x \in \mathfrak{R}^n$ είναι τοπικό ελάχιστο της $f:V \mapsto \mathfrak{R}$, θα είναι και ολικό ελάχιστο.
- b) Αν η $f:V \mapsto \mathfrak{R}$ είναι αυστηρώς κυρτή, τότε υπάρχει το πολύ ένα ολικό ελάχιστο της f .

ΘΕΩΡΗΜΑ 2.10: Για μια παραγωγίσιμη κυρτή συνάρτηση $f:\mathfrak{R}^n \mapsto \mathfrak{R}$ θα ισχύει $\nabla f(x) = 0$ αν και μόνον εάν το $x \in \mathfrak{R}^n$ είναι ολικό ελάχιστο της f .

ΘΕΩΡΗΜΑ 2.11: Για μια διπλά παραγωγίσιμη συνεχή συνάρτηση $f:\mathfrak{R}^n \mapsto \mathfrak{R}$ και μια πραγματική συμμετρική $n \times n$ μήτρα A θα ισχύει

- a) Η συνάρτηση f είναι κυρτή αν και μόνον εάν η $\nabla^2 f(x)$ είναι μη-αρνητικά ορισμένη για κάθε $x \in \mathfrak{R}^n$.
- b) Αν η $\nabla^2 f(x)$ είναι θετικά ορισμένη για κάθε $x \in \mathfrak{R}^n$, τότε η f είναι αυστηρώς κυρτή.
- c) Η συνάρτηση $f(x) = x^T A x$ είναι κυρτή αν και μόνον εάν η μήτρα A είναι μη-αρνητικά ορισμένη.
- d) Η συνάρτηση $f(x) = x^T A x$ είναι αυστηρώς κυρτή αν και μόνον εάν η μήτρα A είναι θετικά ορισμένη.

3 Επαναληπτικές Μέθοδοι Επίλυσης Συστημάτων Γραμμικών Εξισώσεων

3.1 Γενικά

Δίνεται η μήτρα $A \in \mathbb{R}^{n \times n}$ και το διάνυσμα $b \in \mathbb{R}^n$ και ζητείται να βρεθεί ένα διάνυσμα $x \in \mathbb{R}^n$, έτσι ώστε $Ax = b$. Οι μέθοδοι για την επίλυση του παραπάνω συστήματος μπορεί να είναι άμεσες ή επαναληπτικές. Οι άμεσες μέθοδοι βρίσκουν την ακριβή λύση με ένα πεπερασμένο αριθμό πράξεων, συνήθως $O(n^3)$. Οι επαναληπτικές μέθοδοι δε βρίσκουν την ακριβή λύση σε συγκεκριμένο χρόνο, αλλά συγκλίνουν ασυμπτωτικά στη λύση του συστήματος. Η λύση αυτή είναι μέσα σε κάποια αποδεκτή ακρίβεια με μικρό αριθμό επαναλήψεων, γι' αυτό και προτιμούνται απ' τις άμεσες μεθόδους. Αυτό συμβαίνει όταν το n είναι πολύ μεγάλο, όπως για παράδειγμα πίνακες που προκύπτουν κατά την αριθμητική επίλυση διαφορικών εξισώσεων με μερικές παραγώγους. Χρησιμοποιούνται επίσης για την επίλυση προβλημάτων όπου ο πίνακας των συντελεστών των αγνώστων είναι μεγάλης τάξης και αραιός, δηλαδή τα περισσότερα στοιχεία του είναι μηδέν. Σ' αυτές τις περιπτώσεις το πλήθος των πράξεων που απαιτούνται για την εύρεση της λύσης με την επαναληπτική μέθοδο είναι πολύ μικρότερο απ' το αντίστοιχο πλήθος μιας άμεσης μεθόδου. Ένα πλεονέκτημα των επαναληπτικών μεθόδων είναι η δυνατότητα παραλληλοποίησής τους (αφού οι πράξεις που απαιτούνται περιέχουν συνήθως πολλαπλασιασμό πίνακα επί διάνυσμα), πράγμα που παρέχει τη δυνατότητα υλοποίησής τους με τη χρήση Τεχνητών Νευρωνικών Δικτύων. Επειδή οι επαναληπτικές μέθοδοι δεν τερματίζουν σε συγκεκριμένο χρόνο, είναι δύσκολο να εξετάσουμε την πολυπλοκότητα και την αποτελεσματικότητα των αλγορίθμων. Ένα μέτρο σύγκρισης των επαναληπτικών αλγορίθμων είναι η ταχύτητα σύγκλισης στη λύση του συστήματος και η ακρίβεια της λύσης.

Στις παραγράφους που ακολουθούν θα παρουσιάσουμε τις κλασικές επαναληπτικές μεθόδους, τις μεθόδους κλιμακωτής μείωσης και απότομης κατάβασης (*gradient* και *steepest descent*), συζυγούς κλίσης (*conjugate gradient*) και τη γενίκευσή τους με τη χρήση βελτιστοποίησης χωρίς περιορισμούς (*unconstrained optimization*).

3.2 Επαναληπτικές Μέθοδοι Επίλυσης Γραμμικών Συστημάτων

Δίνεται η μήτρα $A \in \mathbb{R}^{n \times n}$ και το διάνυσμα $b \in \mathbb{R}^n$ και ζητείται να βρεθεί ένα διάνυσμα $x \in \mathbb{R}^n$, έτσι ώστε

$$Ax = b$$

(3.1)

Υποθέτουμε ότι η μήτρα A έχει αντίστροφο, έτσι ώστε το σύστημα (3.1) να έχει μια και μοναδική λύση. Η i -οστή εξίσωση του συστήματος (3.1) μπορεί να γραφεί με τη μορφή

$$\sum_{j=1}^n a_{ij} x_j = b_i \quad (3.2)$$

με a_{ij} τους συντελεστές της i -οστής γραμμής της μήτρας A , x_j και b_i τους συντελεστές των διανυσμάτων x και b αντίστοιχα. Αν υποθέσουμε ότι $a_{ii} \neq 0$ και λύσουμε ως προς x_i θα έχουμε

$$x_i = \frac{1}{a_{ii}} \left[b_i - \sum_{j \neq i}^n a_{ij} x_j \right] \quad (3.3)$$

Αν έχουμε βρει κάποιες προσεγγιστικές τιμές για τα x_i , η εξίσωση (3.3) θα μας δώσει μια προσεγγιστική τιμή για το x_i . Έτσι μπορούμε να βρούμε προσεγγιστικές τιμές $\forall x_i$, $i = 1, 2, \dots, n$ ταυτόχρονα ή με τη σειρά. Ανάλογα με τον τρόπο υπολογισμού των τιμών του x μπορούμε να έχουμε τις μεθόδους που ακολουθούν.

3.2.1 Η Μέθοδος Jacobi

Αρχίζοντας με ένα αυθαίρετο διάνυσμα $x^{(0)} \in \mathbb{R}^n$, υπολογίζουμε τα $x_i^{(t+1)}$, $i = 1, 2, \dots, n$ και $t = 0, 1, 2, \dots$, σύμφωνα με τον τύπο

$$x_i^{(t+1)} = \frac{1}{a_{ii}} \left[b_i - \sum_{\substack{j=1 \\ j \neq i}}^n a_{ij} x_j^{(t)} \right] \quad (3.4)$$

Ο αλγόριθμος παράγει μια ακολουθία διανυσμάτων $\{x^{(0)}\}$, η οποία, αν συγκλίνει στο x^* , το x^* είναι η λύση του συστήματος (3.1). Όπως φαίνεται απ' τον τύπο (3.4) για τον υπολογισμό των συντελεστών του x στο βήμα $t+1$, χρησιμοποιούνται οι τιμές των συντελεστών του x στο βήμα t . Αν για τον υπολογισμό του $x_i^{(t+1)}$ χρησιμοποιήσουμε τις τιμές $x_j^{(t+1)}$, $j < i$, των συντελεστών του x , οι οποίοι έχουν ήδη υπολογισθεί στο βήμα $t+1$ προκύπτει η μέθοδος των Gauss-Seidel.

3.2.2 Η Μέθοδος Gauss-Seidel

Αρχίζοντας με ένα αυθαίρετο διάνυσμα $x^{(0)} \in \mathbb{R}^n$, υπολογίζουμε τα $x_i^{(t+1)}$, $i = 1, 2, \dots, n$ και $t = 0, 1, 2, \dots$, σύμφωνα με τον τύπο

$$x_i^{(t+1)} = \frac{1}{a_{ii}} \left[b_i - \sum_{j=1}^{i-1} a_{ij} x_j^{(t+1)} - \sum_{j=i+1}^n a_{ij} x_j^{(t)} \right] \quad (3.5)$$

3.2.3 Η Μέθοδος Jacobi Overrelaxation (JOR)

Ένας τρόπος επιτάχυνσης της σύγκλισης των επαναληπτικών μεθόδων είναι η χρήση ενός θετικού αριθμού που λέγεται *παράμετρος χαλάρωσης (relaxation parameter)*, η οποία, αν χρησιμοποιηθεί στη μέθοδο Jacobi, δίνει τη *μέθοδο ταυτόχρονης υπερχαλάρωσης του Jacobi (Jacobi Overrelaxation ή JOR)*, και η εξίσωση (3.4) γίνεται

$$x_i^{(t+1)} = (1 - \omega)x_i^{(t)} + \frac{\omega}{a_{ii}} \left[b_i - \sum_{\substack{j=1 \\ j \neq i}}^n a_{ij} x_j^{(t)} \right] \quad (3.6)$$

Αν $0 < \omega < 1$, η νέα τιμή του x_i στην εξίσωση (3.6) είναι ένας συνδιασμός της παλιάς τιμής του x_i και της τιμής του x_i που θα παίρναμε με τη μέθοδο Jacobi. Με την επιλογή καλής τιμής για το ω , ο αλγόριθμος (3.6) συγκλίνει πιο γρήγορα απ' τον (3.4). Αν $\omega = 1$, ο τύπος (3.6) οδηγεί στη μέθοδο Jacobi.

3.2.4 Η Μέθοδος Successive Overrelaxation (SOR)

Αν η *παράμετρος χαλάρωσης (relaxation parameter)* χρησιμοποιηθεί στη μέθοδο Gauss-Seidel, δίνει τη *μέθοδο διαδοχικής υπερχαλάρωσης (Successive Overrelaxation ή SOR)*, και η εξίσωση (3.5) γίνεται

$$x_i^{(t+1)} = (1 - \omega)x_i^{(t)} + \frac{\omega}{a_{ii}} \left[b_i - \sum_{j=1}^{i-1} a_{ij} x_j^{(t+1)} - \sum_{j=i+1}^n a_{ij} x_j^{(t)} \right] \quad (3.7)$$

Με την επιλογή καλής τιμής για το ω , $0 < \omega < 2$, ο αλγόριθμος (3.7) συγκλίνει πιο γρήγορα απ' τον (3.5). Για ω , $0 < \omega < 1$, η μέθοδος λέγεται *Μέθοδος Διαδοχικής Υποχαλάρωσης (Successive Underrelaxation)*, ενώ για $\omega = 1$, ο τύπος (3.7) οδηγεί στη μέθοδο Gauss-Seidel.

3.2.5 Η Μέθοδος Richardson

Αν η εξίσωση $Ax = b$ γραφεί με τη μορφή $x = x + \omega(b - Ax)$ προκύπτει η μέθοδος Richardson, όπου οι συντελεστές $x_i^{(t+1)}$, $i = 1, 2, \dots, n$ και $t = 0, 1, 2, \dots$, δίνονται απ' τον τύπο

$$x_i^{(t+1)} = x_i^{(t)} + \omega \left[b_i - \sum_{j=1}^n a_{ij} x_j^{(t)} \right] \quad (3.8)$$

3.2.6 Η Μέθοδος Richardson-Gauss-Seidel (RGS)

Αν ο υπολογισμός των συντελεστών $x_i^{(t+1)}$ της προηγούμενης μεθόδου γίνει όπως και στη μέθοδο Gauss-Seidel, προκύπτει η μέθοδος Richardson-Gauss-Seidel, όπου οι συντελεστές $x_i^{(t+1)}$, $i = 1, 2, \dots, n$ και $t = 0, 1, 2, \dots$, δίνονται απ' τον τύπο

$$x_i^{(t+1)} = x_i^{(t)} + \omega \left[b_i - \sum_{j=1}^{i-1} a_{ij} x_j^{(t+1)} - \sum_{j=i}^n a_{ij} x_j^{(t)} \right] \quad (3.9)$$

3.3 Σύγκλιση των Επαναληπτικών Μεθόδων

Οι επαναληπτικές μέθοδοι που παρουσιάστηκαν στην προηγούμενη ενότητα, παράγουν μια άπειρη ακολουθία προσεγγιστικών τιμών του x , τη $\{x^{(t)}\}$, η οποία έχει την ιδιότητα, αν συγκλίνει, να συγκλίνει στη λύση του συστήματος (3.1). Αυτό είναι εύκολο να αποδειχθεί για μια επαναληπτική μέθοδο, π.χ. τη μέθοδο Jacobi. Αν πάρουμε τα όρια των δύο όρων στη σχέση (3.4) καθώς το t τείνει στο ∞ , κάθε x_i πληροί τη σχέση (3.3), το οποίο συμβαίνει μόνο αν το x είναι η λύση του συστήματος (3.1). Υπάρχουν όμως και περιπτώσεις που οι παραπάνω αλγόριθμοι δεν συγκλίνουν. Σ' αυτή την ενότητα θα εξετάσουμε τις συνθήκες κάτω απ' τις οποίες οι αλγόριθμοι συγκλίνουν, αφού πρώτα τους εντάξουμε σε μια γενική μέθοδο, τη *γενική επαναληπτική μέθοδο*.

Αν D είναι η μήτρα που περιέχει τα διαγώνια στοιχεία της μήτρας A και $B = A - D$, η μήτρα με τα στοιχεία του A και μηδενικά τα διαγώνια στοιχεία, αν τα διαγώνια στοιχεία της μήτρας A είναι $\neq 0$, και αντικαταστήσουμε το A με το $B + D$, το σύστημα (3.1) γίνεται $(B + D)x = b$ ή $Bx + Dx = b$ ή $Dx = b - Bx$. Πολλαπλασιάζοντας από αριστερά με το D^{-1} , το σύστημα γίνεται $x = D^{-1}b - D^{-1}Bx$, και ο αλγόριθμος του Jacobi μπορεί να γραφεί

$$x^{(t+1)} = -D^{-1}Bx^{(t)} + D^{-1}b \quad (3.10)$$

ενώ ο αλγόριθμος JOR γίνεται

$$x^{(t+1)} = [(I - \omega)I - \omega D^{-1}B]x^{(t)} + \omega D^{-1}b \quad (3.11)$$

με I το μοναδιαίο πίνακα. Με παρόμοιο τρόπο μπορούμε να παραστήσουμε τους αλγόριθμους Gauss-Seidel και SOR σε μορφή πινάκων, αν η μήτρα A εκφρασθεί σαν $A=L+D+U$, όπου L αυστηρά κάτω τριγωνική, D διαγώνια και U αυστηρά άνω τριγωνική μήτρα της A . Έτσι, ο αλγόριθμος Gauss-Seidel μπορεί να γραφεί

$$x^{(t+1)} = D^{-1} [b - Lx^{(t+1)} - Ux^{(t)}]$$

που δίνει τον επαναληπτικό τύπο σε μορφή πινάκων

$$x^{(t+1)} = -(I + D^{-1}L)^{-1} D^{-1} Ux^{(t)} + (I + D^{-1}L)^{-1} D^{-1} b \quad (3.12)$$

ενώ ο αλγόριθμος SOR γίνεται

$$x^{(t+1)} = (1 - \omega)x^{(t)} + \omega D^{-1} [b - Lx^{(t+1)} - Ux^{(t)}]$$

που δίνει τον επαναληπτικό τύπο σε μορφή πινάκων

$$x^{(t+1)} = (I + \omega D^{-1}L)^{-1} [(1 - \omega)I - \omega D^{-1}U]x^{(t)} + \omega (I + \omega D^{-1}L)^{-1} D^{-1} b \quad (3.13)$$

Η μέθοδος Richardson μπορεί να παρασταθεί σε μορφή πινάκων σαν

$$x^{(t+1)} = (I - \omega A)x^{(t)} + \omega b \quad (3.14)$$

Είναι φανερό ότι οι εξισώσεις (3.10) - (3.14) μπορούν να γραφούν στη μορφή

$$x^{(t+1)} = Mx^{(t)} + Gb \quad (3.15)$$

με M , G πίνακες που έχουν προκύψει απ' τη μήτρα A , ανάλογα με τον αλγόριθμο. Ειδικά η μήτρα M λέγεται *επαναληπτικός πίνακας* του επαναληπτικού αλγορίθμου και ο αλγόριθμος (3.15) *γενική επαναληπτική μέθοδος*. Οι προϋποθέσεις σύγκλισης της παραπάνω μεθόδου μπορούν να εξειδικευθούν για τους αλγόριθμους (3.10)-(3.14).

Πρόταση 3.1. Αν ο πίνακας $I-M$ έχει αντίστροφο και το μοναδικό διάνυσμα x^* ικανοποιεί τη σχέση $x^* = Mx^* + Gb$, η ακολουθία των διανυσμάτων $\{x^{(t)}\}$ που ορίζεται απ' τον αλγόριθμο $x^{(t+1)} = Mx^{(t)} + Gb$ συγκλίνει στο x , όταν το t τείνει στο άπειρο ($\lim_{t \rightarrow \infty} x^{(t)} = x^*$, $\forall x^{(0)}$), αν και μόνο αν $\rho(M) < 1$, όπου $\rho(M)$ η φασματική ακτίνα του πίνακα M , σύμφωνα με τον ορισμό 2.14.

Απόδειξη. Αν συμβολίσουμε με $\varepsilon(t)$ το σφάλμα της t επανάληψης που ορίζεται απ' τη σχέση

$$\varepsilon^{(t)} = x^{(t)} - x^* \quad (3.16)$$

το σφάλμα στην $t+1$ επανάληψη με τη χρήση της εξίσωσης (3.15) θα δίνεται απ' τη σχέση

$$\varepsilon^{(t+1)} = x^{(t+1)} - x^* = Mx^{(t)} + Gb - (Mx^* + Gb) = M(x^{(t)} - x^*) = M\varepsilon^{(t)} \quad (3.17)$$

και επαγωγικά προκύπτει ότι

$$\varepsilon^{(t)} = M^t \varepsilon^{(0)} \quad | \quad t = 0, 1, 2, \dots, \quad (3.18)$$

Επειδή θέλουμε ο αλγόριθμος (3.15) να συγκλίνει στη λύση του συστήματος (3.1), θα πρέπει $\lim_{t \rightarrow \infty} x^{(t)} = x^*$ ή $\lim_{t \rightarrow \infty} (x^{(t)} - x^*) = 0$ ή $\lim_{t \rightarrow \infty} \varepsilon^{(t)} = 0$ ή απ' τη σχέση (3.18) $\lim_{t \rightarrow \infty} M^t \varepsilon^{(0)} = 0$. Αυτό όμως συμβαίνει όταν και μόνον όταν

$$\rho(M) < 1 \quad (3.19)$$

(Θεώρημα 2.4, περίπτωση 1), όπου $\rho(M)$ η φασματική ακτίνα της μήτρας M , η μεγαλύτερη ιδιοτιμή της μήτρας M . Ο υπολογισμός των ιδιοτιμών όμως είναι αρκετά δύσκολος, γι' αυτό χρησιμοποιούνται συνήθως *norms* για την απόδειξη και τα κριτήρια σύγκλισης. Όταν όμως η μήτρα A είναι *συμμετρική και θετικά ορισμένη*, η ανάλυση της σύγκλισης γίνεται με μια *συνάρτηση κόστους* που είναι *τετραγωνική μορφή*.

3.3.1 Ανάλυση της Σύγκλισης με Τετραγωνική Συνάρτηση Κόστους

Όταν η μήτρα A είναι *συμμετρική και θετικά ορισμένη* (SPD) έχει πάντα αντίστροφο, οπότε το σύστημα (3.1) έχει μια και μοναδική λύση. Αν η μήτρα A δεν είναι SPD, πολλαπλασιάζουμε το σύστημα (3.1) από αριστερά με τη μήτρα A^T και παίρνουμε το αντίστοιχο σύστημα κανονικών εξισώσεων

$$A^T A x = A^T b \quad (3.20)$$

ή

$$C x = d \quad (3.21)$$

με $C = A^T A$, και $d = A^T b$, όπου $C \in \mathcal{R}^{n \times n}$ συμμετρική και θετικά ορισμένη (δηλαδή $x^T C x > 0$, \forall μη μηδενικό $x \in \mathcal{R}^n$). Γι' αυτά τα συστήματα, η επίλυση του συστήματος (3.20) είναι ισοδύναμη με την ελαχιστοποίηση της συνάρτησης κόστους $E(x) = \frac{1}{2} x^T C x - d^T x$. Αφού η μήτρα C είναι SPD, η συνάρτηση κόστους E είναι *αυστηρά κλειστή* (*strictly convex*) σύμφωνα με το Θεώρημα 2.11, οπότε ένα διάνυσμα x^* ελαχιστοποιεί την E , αν και μόνον αν

$$\nabla E(x^*) = 0 \quad (3.22)$$

Αλλά, $\nabla E(x) = Cx - d$, οπότε το x^* είναι λύση της (3.1), όταν και μόνον όταν το x^* ελαχιστοποιεί την $E(x)$. Η ελάχιστη τιμή της $E(x)$ είναι μοναδική και ισούται με $-\frac{1}{2}d^T C^{-1}d$, ενώ επιτυγχάνεται θέτοντας $x = C^{-1}d$. Έτσι, η ελαχιστοποίηση του $E(x)$ και η επίλυση του συστήματος (3.1) είναι ισοδύναμα προβλήματα.

3.3.2 Ανάλυση της Σύγκλισης με Ελάττωση της Παραγώγου (Gradient Descent)

Η επίλυση του συστήματος (3.20) μπορεί να αναχθεί στην ελαχιστοποίηση μιας συνεχούς και διαφορίσιμης συνάρτησης $E: \mathcal{R}^n \rightarrow \mathcal{R}$. Αν χρησιμοποιήσουμε τη συνάρτηση του Μέσου Τετραγωνικού Σφάλματος, ή τη Συνάρτηση Κόστους

$$E(x) = \sum_{i=1}^m E_i(x) = \sum_{i=1}^m \frac{1}{2} (a_i^T x - b_i)^2 = \frac{1}{2} \sum_{i=1}^m (a_i^T x - b_i)^2 = \frac{1}{2} (Ax - b)^T (Ax - b) = \frac{1}{2} \|Ax - b\|^2 \quad (3.23)$$

σαν κριτήριο βελτιστοποίησης, τα ακρότατα (μέγιστο και ελάχιστο) της συνάρτησης $E(x)$ είναι λύσεις της εξίσωσης που προκύπτει χρησιμοποιώντας παραγώγους για την ελαχιστοποίηση συνάρτησης

$$\nabla E(x) = A^T (Ax - b) = 0 \quad (3.24)$$

το οποίο μας δίνει το αντίστοιχο σύστημα κανονικών εξισώσεων (3.20) ή (3.21).

Η βέλτιστη λύση ελαχίστων τετραγώνων του συστήματος (3.1) ή του ισοδύναμου (3.20), (3.21), με τη χρήση του γενικευμένου αντιστρόφου του Moore-Penrose A^+ [DaBi74], [Koho84], [Atki89], [GiMW91] ορίζεται σαν

$$x = A^+ b \quad (3.25)$$

Για το υπερορισμένο σύστημα $Ax = b$, με A μια $(m \times n)$ μήτρα, η λύση του συστήματος (3.1) είναι η Λύση Ελαχίστων Τετραγώνων. Αν $r(A) = n$, τότε η λύση είναι μοναδική, και δίνεται απ' τον τύπο

$$x = A^+ b \quad (3.26)$$

με

$$A^+ = (A^T A)^{-1} A^T \quad (3.27)$$

όπου A^+ είναι ο ψευδο-αντίστροφος του $(m \times n)$ πίνακα A , ο οποίος πληροί τις συνθήκες των Moore-Penrose.

Για το τετραγωνικό σύστημα $Ax = b$, με τη μήτρα A τετραγωνική $(n \times n)$ με ορίζουσα $\text{Det}(A) \neq 0$, η λύση του συστήματος (3.1) είναι μοναδική και ο γενικευμένος

αντίστροφος της μήτρας A , A^+ που ορίστηκε στην (3.27) ισούται με τον αντίστροφο A^{-1} .

Για το αόριστο σύστημα $Ax = b$, με A μια $(m \times n)$ μήτρα, η λύση του συστήματος (3.1) είναι μια απ' τις άπειρες Ελαχίστων Τετραγώνων. Αν $r(A) = m$, τότε η λύση δεν είναι μοναδική, και δίνεται απ' τον τύπο

$$x = G^+ b \quad (3.28)$$

όπου G^+ ο γενικευμένος αντίστροφος της μήτρας A , ο οποίος δεν είναι μοναδικός [DaBi74], [Koho84],[Atki89], [DeMa76].

Είναι εύκολο να αποδειχθεί ότι η λύση x της εξίσωσης (3.1, 3.20, 3.21) είναι το *καθολικό (global) ελάχιστο* της συνάρτησης $E(x)$, όπως ορίστηκε στην (3.23), αρκεί να δειχθεί ότι η δεύτερη παράγωγος (*Hessian*) της $E(x)$

$$\nabla^2 E(x) = \left(\frac{\partial^2 E(x)}{\partial x_i \partial x_j} \right) = A^T A \quad (3.29)$$

είναι συμμετρική και θετικά ορισμένη (SPD). Αυτό όμως ισχύει, αφού η μήτρα $A^T A = C$ είναι συμμετρική και θετικά ορισμένη. Οι επαναληπτικές μέθοδοι (3.4)-(3.9) για το σύστημα (3.19) ή (3.20) με τη χρήση της μήτρας $C = A^T A$ θα έχουν την παρακάτω μορφή:

3.3.3 Γενίκευση της Μεθόδου Jacobi

Η γενίκευση του αλγόριθμου του Jacobi για το σύστημα (3.21), όταν η μήτρα A δεν είναι SPD, θα έχει τη μορφή

$$\begin{aligned} x_i^{(t+1)} &= \frac{1}{c_{ii}} \left[d_i - \sum_{\substack{j=1 \\ j \neq i}}^n c_{ij} x_j^{(t)} \right] = x_i^{(t)} + \frac{1}{c_{ii}} \left[d_i - \sum_{j=1}^n c_{ij} x_j^{(t)} \right] = \\ &= x_i^{(t)} - \frac{1}{\nabla_{ii}^2 E(x^{(t)})} \nabla_i E(x^{(t)}) \end{aligned} \quad (3.30)$$

όπου $\nabla_i E(x)$ το i ($i = 1, 2, \dots, n$) στοιχείο του διανύσματος $\nabla E(x)$, όπως ορίστηκε στην (3.23) και $\nabla_{ii}^2 E(x)$ το διαγώνιο στοιχείο της μήτρας A . Σε μορφή πινάκων η (3.30) γίνεται

$$x^{(t+1)} = x^{(t)} - \frac{1}{\nabla^2 E(x^{(t)})} \nabla E(x^{(t)}) \quad (3.31)$$

3.3.4 Γενίκευση της Μεθόδου Gauss-Seidel

Η γενίκευση του αλγόριθμου των Gauss-Seidel θα έχει τη μορφή

$$\begin{aligned} x_i^{(t+1)} &= \frac{1}{c_{ii}} \left[d_i - \sum_{j=1}^{i-1} c_{ij} x_j^{(t+1)} - \sum_{j=i+1}^n c_{ij} x_j^{(t)} \right] = x_i^{(t)} + \frac{1}{c_{ii}} \left[d_i - \sum_{j=1}^{i-1} c_{ij} x_j^{(t+1)} - \sum_{j=i}^n c_{ij} x_j^{(t)} \right] = \\ &= x_i^{(t)} - \frac{\nabla_i E(z)}{\nabla_{ii}^2 E(z)} \end{aligned} \quad (3.32)$$

με $z = (x_1^{(t+1)}, \dots, x_{i-1}^{(t+1)}, x_i^{(t)}, \dots, x_n^{(t)})$. Σε μορφή πινάκων η (3.32) γίνεται

$$x^{(t+1)} = x^{(t)} - \frac{1}{\nabla^2 E(z)} \nabla E(z) \quad (3.33)$$

3.3.5 Γενίκευση της Μεθόδου Jacobi Overrelaxation (JOR)

Η γενίκευση του αλγόριθμου JOR θα έχει τη μορφή

$$\begin{aligned} x_i^{(t+1)} &= (1 - \omega)x_i^{(t)} + \frac{\omega}{c_{ii}} \left[d_i - \sum_{\substack{j=1 \\ j \neq i}}^n c_{ij} x_j^{(t)} \right] = x_i^{(t)} + \frac{\omega}{c_{ii}} \left[d_i - \sum_{j=1}^n c_{ij} x_j^{(t)} \right] = \\ &= x_i^{(t)} - \frac{\omega}{\nabla_{ii}^2 E(x)} \nabla_i E(x^{(t)}) \end{aligned} \quad (3.34)$$

όπου ω η παράμετρος χαλάρωσης. Σε μορφή πινάκων η (3.34) γίνεται

$$x^{(t+1)} = x^{(t)} - \frac{\omega}{\nabla^2 E(x^{(t)})} \nabla E(x^{(t)}) \quad (3.35)$$

3.3.6 Γενίκευση της Μεθόδου Successive Overrelaxation (SOR)

Η γενίκευση του αλγόριθμου SOR θα έχει τη μορφή

$$x_i^{(t+1)} = (1 - \omega)x_i^{(t)} + \frac{\omega}{c_{ii}} \left[d_i - \sum_{j=1}^{i-1} c_{ij} x_j^{(t+1)} - \sum_{j=i+1}^n c_{ij} x_j^{(t)} \right] = x_i^{(t)} - \omega \frac{\nabla_i E(z)}{\nabla_{ii}^2 E(z)} \quad (3.36)$$

Σε μορφή πινάκων η (3.36) γίνεται

$$x^{(t+1)} = x^{(t)} - \frac{\omega}{\nabla^2 E(z)} \nabla E(z) \quad (3.37)$$

3.3.7 Γενίκευση της Μεθόδου Richardson

Η γενίκευση του αλγόριθμου Richardson θα έχει τη μορφή

$$x_i^{(t+1)} = x_i^{(t)} + \omega \left[d_i - \sum_{j=1}^n c_{ij} x_j^{(t)} \right] = x_i^{(t)} - \omega \nabla E(x^{(t)}) \quad (3.38)$$

3.3.8 Γενίκευση της Μεθόδου Richardson- Gauss-Seidel

Η γενίκευση του αλγόριθμου Richardson- Gauss-Seidel θα έχει τη μορφή

$$x_i^{(t+1)} = x_i^{(t)} + \omega \left[d_i - \sum_{j=1}^{i-1} c_{ij} x_j^{(t+1)} - \sum_{j=i}^n c_{ij} x_j^{(t)} \right] = x_i^{(t)} - \omega \nabla E(z^{(t)}) \quad (3.39)$$

3.4 Gradient Descent Αλγόριθμοι Εκπαίδευσης

Οι αλγόριθμοι που ορίστηκαν στις προηγούμενες παραγράφους ανήκουν στην κατηγορία των *gradient descent* αλγορίθμων, οι οποίοι ορίζονται ως εξής:

ΟΡΙΣΜΟΣ 3.1. Κάθε αλγόριθμος που αρχίζει από ένα τυχαίο σημείο $x^{(0)}$ και δημιουργεί μια ακολουθία από βελτιωμένες προσεγγίσεις $x^{(1)}, x^{(2)}, \dots$, τέτοια ώστε, για $t \geq 0$, $\nabla E(x^{(t)}) \neq 0$, κατά μήκος μιας κατεύθυνσης κατάβασης (*descent direction*) λέγεται *gradient descent* αλγόριθμος.

ΟΡΙΣΜΟΣ 3.2. Για κάποιο $x \in R^n$, τέτοιο ώστε $\nabla E(x^{(t)}) \neq 0$, κάθε διάνυσμα $d \in R^n$ για το οποίο ισχύει $d^T \nabla E(x^{(t)}) < 0$ ορίζεται σαν κατεύθυνση κατάβασης (*descent direction*).

Αφού το $d^T \nabla E(x^{(t)})$ είναι η παράγωγος κατεύθυνσης (*directional derivative*) κατά μήκος της κατεύθυνσης d , για κάποιο ρ μικρό θετικό αριθμό, ο οποίος λέγεται βήμα (*stepsize*), θα ισχύει $E(x^{(t+1)}) = E(x^{(t)} + \rho d^{(t)}) < E(x^{(t)})$.

Η κατεύθυνση κατάβασης (*descent direction*) στους προηγούμενους αλγόριθμους είναι $d^{(t)} = -\nabla E(x^{(t)})$ και θα ισχύει $d^T \nabla E(x^{(t)}) = -\nabla E(x^{(t)}) \nabla E(x^{(t)}) = -(\nabla E(x^{(t)}))^2 < 0$. Ειδικότερα, ο αλγόριθμος Richardson, στον οποίο $\rho = \omega$, λέγεται και *steepest descent*.

Ο αλγόριθμος Gauss-Seidel χρησιμοποιεί σαν βήμα το $\frac{1}{\nabla_{ii}^2 E(z)}$, ο αλγόριθμος SOR το

$\frac{\omega}{\nabla_{ii}^2 E(z)}$, ο αλγόριθμος του Jacobi το $\frac{1}{\nabla_{ii}^2 E(x)}$, ενώ ο αλγόριθμος του JOR το

$\frac{\omega}{\nabla_{ii}^2 E(x)}$. Όλοι οι προηγούμενοι αλγόριθμοι χρησιμοποιούν σταθερό βήμα ω , το οποίο

καθορίζεται με *ευρετικό (heuristic)* τρόπο. Αν χρησιμοποιήσουμε μεταβλητό βήμα, έχουμε μια άλλη κατηγορία αλγορίθμων που συγκλίνουν πιο γρήγορα.

Η αντικειμενική συνάρτηση $E(x)$ της εξίσωσης (3.23) είναι τετραγωνική μορφή ως προς x . Η $E(x)$ ορίζει ένα *κυρτό (convex)* υπερπαραβολοειδές με ένα ολικό ελάχιστο x^* . Αν το βήμα ρ είναι πολύ μικρό, η διαδικασία της *φθίνουσας αναζήτησης (gradient search)* που υλοποιείται με τις εξισώσεις (3.31)-(3.39), θα συγκλίνει ασυμπτωτικά στη λύση x^* , ανεξάρτητα απ' τις αρχικές τιμές του x .

3.4.1 Η μέθοδος Newton

Η χρήση τιμών παραγώγων δυο συνεχόμενων βημάτων στην εξίσωση (3.38) για την επιτάχυνση της σύγκλισης στηρίζεται σε προσεγγίσεις μεθόδων αναζήτησης δευτέρας τάξης, όπως η μέθοδος Newton. Η μέθοδος του Newton [BeTs86], [Batt92] στηρίζεται σε μια τετραγωνική μορφή $\tilde{E}(x)$ της αντικειμενικής συνάρτησης $E(x)$, η οποία χρησιμοποιεί μόνο τους 3 πρώτους όρους του αναπτύγματος σε σειρά Taylor του $E(x)$ στο διάνυσμα x

$$\tilde{E}(x + \Delta x) = E(x) + \nabla E(x)^T \Delta x + \frac{1}{2} \Delta x^T \nabla^2 E(x) \Delta x \quad (3.40)$$

Το ελάχιστο της συνάρτησης $\tilde{E}(x + \Delta x)$ είναι η λύση της εξίσωσης

$$\nabla \tilde{E}(x + \Delta x) = 0 \quad \text{ή} \quad \nabla E(x) + \nabla^2 E(x) \Delta x = 0, \quad \text{η οποία δίνει τη μέθοδο Newton}$$

$$x^{(t+1)} = x^{(t)} - \frac{\nabla E(x^{(t)})}{\nabla^2 E(x^{(t)})} = -[H(x^{(t)})]^{-1} \nabla E(x^{(t)}) \quad (3.41)$$

όπου H είναι η *Hessian* μήτρα με στοιχεία $H_{ij} = \frac{\partial^2 E(x)}{\partial x_i \partial x_j}$. Αναλυτικά, οι συντελεστές

$x_i, i = 1, 2, \dots, n$ δίνονται απ' τον τύπο

$$x_i^{(t+1)} = x_i^{(t)} - \frac{\partial E(x)}{\partial x_i} \left(\frac{\partial^2 E(x)}{\partial x_i \partial x_j} \right)^{-1} \quad (3.42)$$

Η μέθοδος του Newton υπολογίζει επαναληπτικά τις αλλαγές Δx και έχει καλά αποτελέσματα, όταν οι αρχικές τιμές βρίσκονται μέσα σε μια κυρτή (convex) περιοχή της $E(x)$. Είναι όμως μια μέθοδος ασύμφορη υπολογιστικά, γιατί ο υπολογισμός της απαιτεί $O(n^3)$ πράξεις σε κάθε επανάληψη, όπου N , η διάσταση του χώρου. Εκτός αυτού, απαιτεί και τον υπολογισμό του αντιστρόφου H^{-1} της μήτρας $H = A^T A$.

3.4.2 Η Μέθοδος Adaptive (Optimal) Steepest Descent (ASD)

Μια άλλη προσέγγιση είναι να βρεθεί μια κατεύθυνση αναζήτησης, η οποία είναι τοπικά βέλτιστη. Αυτό γίνεται με τη χρήση ενός συντελεστή εκπαίδευσης στο βήμα t , ο οποίος ελαχιστοποιεί την αντικειμενική συνάρτηση $E(x)$ στο βήμα $t+1$, δηλαδή την $E(x^{(t+1)}) = E[x^{(t)} - \rho \nabla E(x^{(t)})]$. Το βέλτιστο $\rho = \rho_{opt}(t)$ ορίζεται σαν η μικρότερη θετική ρίζα της εξίσωσης

$$\frac{\partial E(x^{(t+1)})}{\partial \rho} = \frac{\partial}{\partial \rho} E[x^{(t)} - \rho \nabla E(x^{(t)})] = 0 \quad (3.43)$$

και δίνεται προσεγγιστικά σαν

$$\rho_{opt}(t) \approx \frac{\|\nabla E(x^{(t)})\|^2}{\nabla E(x^{(t)})^T \nabla^2 E(x^{(t)}) \nabla E(x^{(t)})} \quad (3.44)$$

Αναλυτικά, οι συντελεστές $x_i, i = 1, 2, \dots, n$ δίνονται απ' τον τύπο

$$x_i^{(t+1)} = x_i^{(t)} + \rho_{opt}(t) \nabla E(x)^{(t)} \quad (3.45)$$

Η μέθοδος που προκύπτει είναι γνωστή σαν optimal steepest descent [Tsy71], [Hass93]. Η σύγκλιση του παραπάνω αλγορίθμου αποδεικνύεται με το θεώρημα που ακολουθεί.

Πρόταση 3.2. Με τη χρήση του βήματος $\rho_{opt}(t)$, όπως ορίστηκε στην (3.44) το $E(x)$ ελαχιστοποιείται, δηλαδή $E(x^{(t+1)}) = E[x^{(t)} - \rho \nabla E(x^{(t)})] < E(x^{(t)})$.

Απόδειξη

Η επιλογή του $\rho_{opt}(t)$ για την ελαχιστοποίηση της $E(x^{(t+1)}) = E[x^{(t)} - \rho \nabla E(x^{(t)})]$ δίνεται απ' τη λύση της εξίσωσης $\nabla E(\rho_{opt}(t)) = 0$. Με τη χρήση των (3.24), (3.29) έχουμε

$$\begin{aligned} \nabla E(x^{(t+1)}) &= \nabla E(x^{(t)} - \rho \nabla E(x^{(t)})) \\ &= \nabla E(x^{(t)} - \rho \nabla E(x^{(t)})) \nabla E(x^{(t)}) \\ &= (C(x^{(t)} - \rho \nabla E(x^{(t)})) - d) \nabla E(x^{(t)}) = (\nabla E(x^{(t)}) - \rho C \nabla E(x^{(t)})) \nabla E(x^{(t)}) \\ &= (\nabla E(x^{(t)})^T \nabla E(x^{(t)}) - \rho \nabla E(x^{(t)})^T C \nabla E(x^{(t)})) \nabla E(x^{(t)}) \end{aligned}$$

Θέτοντας $\nabla E(\rho_{opt}(t)) = 0$, και με τη χρήση της (3.29), το $\rho_{opt}(t)$ δίνεται απ' τη σχέση (3.47), **Ο.Ε.Δ.**

3.5 Conjugate-Gradient Μέθοδοι

Μια απ' τις δυσκολίες των steepest-descent μεθόδων είναι ότι η ελαχιστοποίηση σε μια κατεύθυνση που ακολουθείται από μια ελαχιστοποίηση στην επόμενη κατεύθυνση δε σημαίνει ότι η συνάρτηση ελαχιστοποιείται στον υποχώρο που ορίζεται απ' τις δύο τελευταίες κατευθύνσεις, γιατί η ελαχιστοποίηση στην τελευταία κατεύθυνση μπορεί να καταστρέψει την ελαχιστοποίηση που έγινε στην προηγούμενη κατεύθυνση. Αν όμως δύο κατευθύνσεις είναι συζυγείς, ο αλγόριθμος συγκλίνει σε N βήματα στο ελάχιστο της συνάρτησης. Δύο κατευθύνσεις λέγονται *αμοιβαία συζυγείς* (*mutually conjugate*) όταν ισχύει $(d^{(t-1)})^T H^{(t-1)} d^{(t)} = 0$, με H τη *Hessian* μήτρα που πρέπει να είναι συμμετρική και θετικά ορισμένη. Μετά την ελαχιστοποίηση στην κατεύθυνση $d^{(t-1)}$ η παράγωγος της συνάρτησης ελαχιστοποίησης θα είναι κάθετη στο $d^{(t-1)}$. Αν η επόμενη ελαχιστοποίηση είναι στην κατεύθυνση $d^{(t)}$ η αλλαγή στην παράγωγο της συνάρτησης ελαχιστοποίησης ως προς τη νέα κατεύθυνση θα είναι $\nabla E(x^{(t)}) - \nabla E(x^{(t-1)}) = \alpha H d^{(t)}$, η οποία είναι κάθετη στην προηγούμενη κατεύθυνση $d^{(t-1)}(\nabla E(x^{(t)}) - \nabla E(x^{(t-1)})) = 0$, επομένως η αλλαγή στο νέο σημείο παραμένει κάθετος στο $d^{(t-1)}$ και η προηγούμενη ελαχιστοποίηση δεν καταστρέφεται.

3.5.1 Η Μέθοδος Polak-Ribiere (PR)

Μια άλλη προσέγγιση που αποτελεί και τη βάση των Conjugate-Gradient Μεθόδων είναι η νέα κατεύθυνση αναζήτησης να είναι ένας συνδιασμός της πρώτης παραγώγου $\nabla E(x^{(t)})$ και της προηγούμενης κατεύθυνσης αναζήτησης $d^{(t-1)}$, δηλαδή

$$d^{(t)} = -\nabla E(x^{(t)}) + \beta d^{(t-1)} \quad (3.46)$$

με $d^{(0)} = -\nabla E(x^{(0)})$, όπου ο συντελεστής β επιλέγεται έτσι ώστε να ελαχιστοποιεί την παρέκκλιση απ' την προηγούμενη κατεύθυνση αναζήτησης. Σ' αυτή την περίπτωση, η νέα κατεύθυνση αναζήτησης είναι *συζυγής* (*conjugate*) ως προς τη μήτρα H με την προηγούμενη κατεύθυνση αναζήτησης. Η αναγκαία συνθήκη γι' αυτό είναι $(d^{(t-1)})^T H^{(t-1)} d^{(t)} = 0$, η οποία ικανοποιείται όταν η μήτρα H είναι SPD. Έτσι με τη χρήση του συντελεστή β

$$\beta = \beta(t) = \frac{[\nabla E(x^{(t)}) - \nabla E(x^{(t-1)})]^T \nabla E(x^{(t)})}{\|\nabla E(x^{(t-1)})\|^2} \quad (3.47)$$

θα έχουμε

$$d^{(t)} = -\nabla E(x^{(t)}) + \beta d^{(t-1)} = -\nabla E(x^{(t)}) + \frac{[\nabla E(x^{(t)}) - \nabla E(x^{(t-1)})]^T \nabla E(x^{(t)})}{\|\nabla E(x^{(t-1)})\|^2} d^{(t-1)} \quad (3.48)$$

και τελικά

$$x^{(t+1)} = x^{(t)} + \rho d^{(t)} = x^{(t)} - \rho \nabla E(x^{(t)}) + \frac{[\nabla E(x^{(t)}) - \nabla E(x^{(t-1)})]^T \nabla E(x^{(t)})}{\|\nabla E(x^{(t-1)})\|^2} \Delta x^{(t-1)} \quad (3.49)$$

3.5.2 Η Μέθοδος Fletcher-Reeves (FR)

Η Μέθοδος Fletcher-Reeves διαφέρει απ' την προηγούμενη στο γεγονός ότι χρησιμοποιεί στον αριθμητή του συντελεστή β μόνο την παρούσα παράγωγο

$$\beta = \beta(t) = \frac{\|\nabla E(x^{(t)})\|^2}{\|\nabla E(x^{(t-1)})\|^2} \quad (3.50)$$

και ο τύπος (3.48) γίνεται

$$d^{(t)} = -\nabla E(x^{(t)}) + \beta d^{(t-1)} = -\nabla E(x^{(t)}) + \frac{\|\nabla E(x^{(t)})\|^2}{\|\nabla E(x^{(t-1)})\|^2} d^{(t-1)} \quad (3.51)$$

οπότε, με τη χρήση της εξίσωσης (3.50), αντικαθιστώντας έχουμε

$$x^{(t+1)} = x^{(t)} + \rho d^{(t)} = x^{(t)} - \rho \nabla E(x^{(t)}) + \frac{\|\nabla E(x^{(t)})\|^2}{\|\nabla E(x^{(t-1)})\|^2} \Delta x^{(t-1)} \quad (3.52)$$

3.5.3 Η Μέθοδος Optimal Fletcher-Reeves (OFR)

Ένας συνδυασμός των Μεθόδων Adaptive Steepest Descent και Fletcher-Reeves με τους συντελεστές ρ και β , όπως ορίστηκαν στις (3.44), (3.50) δίνει τη μέθοδο Optimal Fletcher-Reeves, όπου η κάθε προσέγγιση $x^{(t+1)}$ δίνεται απ' τον τύπο

$$x^{(t+1)} = x^{(t)} + \rho_{opt}(t) d^{(t)} = x^{(t)} - \rho_{opt}(t) \nabla E(x^{(t)}) + \frac{\|\nabla E(x^{(t)})\|^2}{\|\nabla E(x^{(t-1)})\|^2} \Delta x^{(t-1)} \quad (3.53)$$

4

Τεχνητά Νευρωνικά Δίκτυα

4.1 Εισαγωγή

Τα Τεχνητά Νευρωνικά Δίκτυα (ΤΝΔ) ή Νευρωνικά Συστήματα (ΝΣ) έχουν μεγάλη ανάπτυξη τα τελευταία χρόνια. Η προσπάθεια να επιτευχθεί απόδοση παρόμοια με την ανθρώπινη σε τομείς της γνωστικής επιστήμης και η έρευνα σε βάθος γύρω από τη λειτουργία του Βιολογικού Νευρώνα γέννησε την ιδέα της δημιουργίας συστημάτων που αποτελούνται από υπολογιστικά στοιχεία συνδεδεμένα μεταξύ τους, των οποίων η αρχιτεκτονική βασίζεται στις γνώσεις μας για τα βιολογικό νευρωνικό σύστημα. Το βασικό υπολογιστικό στοιχείο των Τεχνητών Νευρωνικών Δικτύων, είναι ο *Τεχνητός Νευρώνας (artificial neuron)*, ο οποίος κατέχει μερικά από τα κύρια χαρακτηριστικά του Βιολογικού προτύπου, όπως: την ικανότητα να εκτελεί υπολογισμούς και την ικανότητα να συγκρίνει και να αποφασίζει. Τα τελευταία χρόνια έχουν αναπτυχθεί πολλοί αλγόριθμοι ΤΝΔ, με τους οποίους το δίκτυο μαθαίνει να εκτελεί κάποια συγκεκριμένη λειτουργία, (συνήθως μια μη-γραμμική απεικόνιση), μέσω ενός συνόλου κανόνων που αποθηκεύονται στις συνάψεις των νευρώνων. Έτσι, τα ΤΝΔ *μαθαίνουν* απ' την πείρα, *γενικεύουν* από προηγούμενα παραδείγματα και *αφαιρούν* βασικά χαρακτηριστικά από εισόδους που περιέχουν περιττά δεδομένα. Παρ' όλες αυτές τις ομοιότητες των ΤΝΔ με το ανθρώπινο μυαλό, τα ΤΝΔ δεν μπορούν να το υποκαταστήσουν. Ένας πιο τυποποιημένος ορισμός των Τεχνητών Νευρωνικών Δικτύων θα μπορούσε να είναι ο εξής:

ΟΡΙΣΜΟΣ 4.1. Με τον όρο Τεχνητά Νευρωνικά Δίκτυα εννοούμε τα μαθηματικά μοντέλα θεωρητικής νοητικής δραστηριότητας που βασικό στόχο έχουν την εξερεύνηση και αναπαραγωγή της ανθρώπινης επεξεργασίας πληροφοριών σχετικές με τη γνώση, τη συμπίεση δεδομένων, την επίλυση προβλημάτων βελτιστοποίησης, την αναγνώριση προτύπων κ.λ.π..

4.2 Ιστορική Αναδρομή

Η συστηματική έρευνα στον τομέα των Νευρωνικών Δικτύων άρχισε τη δεκαετία του '40 με τις εργασίες των *McCulloch* και *Pitts* [McPi43], όπου οι νευρώνες παριστάνονται σαν απλές συσκευές λογικής στάθμης (*linear threshold devices*), οι οποίες μπορούσαν να σχηματίσουν υπολογιστικά κυκλώματα. Η συμπεριφορά των νευρώνων έγινε κατανοητή με τη συμβολή του D. O. Hebb [Hebb49], ο οποίος περιγράφει τη διαδικασία της τροποποίησης των συνάψεων μεταξύ των συνδεδεμένων νευρώνων, όταν επηρεάζονται από ένα εξωτερικό ερέθισμα. Ο Roswblatt [Rose59] προτείνει μια συσκευή αναγνώρισης που την αποκαλεί *perceptron* ικανή για ταξινόμηση προτύπων. Οι περιορισμοί όμως που είχε αυτό το μοντέλο παρουσιάστηκαν στο βιβλίο των Minsky-Papert [MiPa69], "*Perceptron*", στο οποίο αποδεικνύονταν ότι το δίκτυο ήταν ανίκανο να επιλύσει μερικά πολύ απλά προβλήματα, όπως π.χ. να υλοποιήσει την πύλη XOR. Οι απόψεις αυτές προκάλεσαν απογοήτευση σε πολλούς ερευνητές, με αποτέλεσμα να υπάρξει μια περίοδος περιορισμένης ερευνητικής δραστηριότητας. Παρ' όλα αυτά με τη συνεισφορά ερευνητών όπως οι Hopfield [Hopf82], Rumelhart και MacLeland [RuML86], Kohonen [Koho89] και την τεχνολογία των VLSI κυκλωμάτων, ο τομέας των ΤΝΔ γνώρισε μια νέα ανάπτυξη, νέα μοντέλα προτάθηκαν με πολλές και ποικίλες εφαρμογές σε τομείς όπως *Αναγνώριση Προτύπων*, *Τεχνητή Ευφυΐα*, *Συνειρμική Μνήμη*, *Ρομποτική*, *Αναγνώριση Σημάτων*, ακόμη και σε τομείς όπως *Πρόβλεψη Μετοχών* και *Καιρού* άρχισαν ήδη να υλοποιούνται.

4.3 Εγκέφαλος : Το Βιολογικό Πρότυπο

Η δομή των ΤΝΔ έχει σχεδιαστεί σύμφωνα με την οργάνωση του ανθρώπινου εγκεφάλου. Τα ΤΝΔ έχουν χαρακτηριστικά παρόμοια με το ανθρώπινο μυαλό, χαρακτηριστικά όπως την ικανότητα να μαθαίνουν εμπειρικά, να γενικεύουν, να κάνουν λάθη κ.λ.π.. Αν και η επιστήμη έχει προχωρήσει πολύ στην εξερεύνηση του ανθρώπινου εγκεφάλου, υπάρχουν ακόμη πολλοί ανεξερεύνητοι τομείς. Μέρα τη μέρα, παίρνουμε περισσότερες πληροφορίες σχετικά με την ηλεκτροχημική συμπεριφορά των νευρώνων, χωρίς όμως να έχουμε πλήρη εικόνα για τη λειτουργία τους.

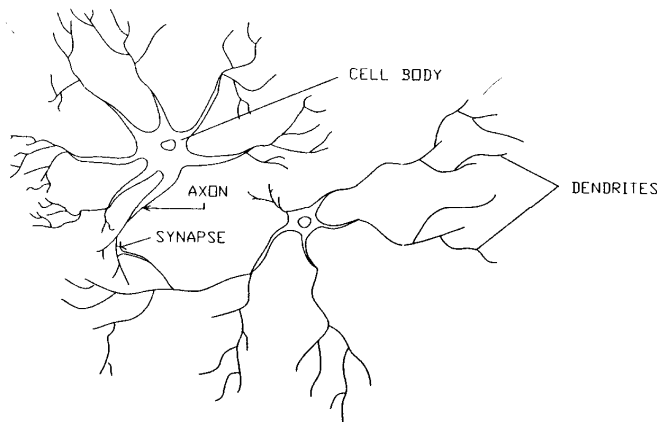
Παρ' όλα αυτά, χρησιμοποιώντας τα ΤΝΔ, τα οποία προσομοιώνουν τη λειτουργία του ανθρώπινου εγκεφάλου, έχουμε τη δυνατότητα να αναπαράγουμε τις πιο πολλές απ' αυτές τις λειτουργίες.

Ο ανθρώπινος εγκέφαλος περιλαμβάνει πάνω από 100 δισεκατομμύρια υπολογιστικά στοιχεία, τους *νευρώνες* (*neurons*). Οι νευρώνες επικοινωνούν μεταξύ τους με *νευρικές ίνες* (περίπου 100 τρισεκατομμύρια), οι οποίες αποκαλούνται *συνάψεις* (*synapsis*) και μ' αυτό τον τρόπο σχηματίζουν ένα πολύπλοκο δίκτυο. Αυτό το δίκτυο των νευρώνων είναι υπεύθυνο για τα φαινόμενα που ονομάζουμε σκέψη, αντίληψη και γνώση και για την εκτέλεση πολλών άλλων αυτόνομων λειτουργιών. Ο ακριβής τρόπος με τον οποίο επιτυγχάνεται αυτό δεν είναι πλήρως κατανοητός, αλλά η φυσιολογική τους δομή έχει ερευνηθεί καθώς και πολλά άλλα που έχουν σχέση με τη λειτουργία τους.

Ο εγκέφαλος περιλαμβάνει επίσης ένα πυκνό δίκτυο από αρτηρίες, το οποίο τροφοδοτεί με οξυγόνο και άλλα θρεπτικά συστατικά τους νευρώνες. Αυτό το δίκτυο, συνδέεται με το κυκλοφοριακό σύστημα με ένα σύστημα φιλτραρίσματος, το *εγκεφαλικό φράγμα αίματος (blood-brain barrier)*, ένα μηχανισμό προστασίας, ο οποίος απομονώνει τον εγκέφαλο από τη διείσδυση τοξικών ουσιών που κυκλοφορούν στο αίμα και που επιτρέπει μόνο ελάχιστες θρεπτικές ουσίες να διαπεράσουν το εγκεφαλικό φράγμα. Είναι απαραίτητο για την ασφάλεια του μυαλού αλλά είναι και πολύ περίπλοκο και εμποδίζει τους ερευνητές που επιθυμούν να παρατηρήσουν τις λειτουργίες που συντελούνται στον εγκέφαλο.

Όπως είναι γνωστό, ο εγκέφαλος είναι ο μεγαλύτερος καταναλωτής ενέργειας του ανθρώπινου σώματος. Αν και αποτελεί το 2% του βάρους του σώματος, χρησιμοποιεί το 20% του συνολικού οξυγόνου του οργανισμού. Ακόμη και όταν κοιμόμαστε, ο εγκέφαλος εξακολουθεί να καταναλώνει οξυγόνο. Η δε ισχύς του είναι 20 watts, ενώ οι Η.Υ. καταναλώνουν χιλιάδες watts, και απαιτούν προστασία από υπερθέρμανση.

Ο *νευρώνας (neuron)* αποτελεί και το βασικό δομικό στοιχείο του νευρικού συστήματος. Είναι όπως όλα τα κύτταρα του οργανισμού, αλλά διαθέτει ειδικά χαρακτηριστικά που του προσφέρουν τη δυνατότητα να εκτελεί υπολογιστικές λειτουργίες. Αποτελείται από 3 στοιχεία: το *σώμα*, τους *δενδρίτες* και τον *άξονα*, όπως φαίνεται και στο σχήμα 4.1.



Σχήμα 4.1 Δομή Βιολογικών Νευρώνων

Οι δενδρίτες λαμβάνουν σήματα από άλλα κύτταρα μέσω συνδέσεων που αποκαλούνται συνάψεις. Από εκεί, τα σήματα περνούν στο σώμα, όπου προστίθενται σε άλλα παρόμοια σήματα. Αν ο μέσος όρος σε κάποια χρονική στιγμή είναι αρκετά μεγάλος, το κύτταρο *εκπυρσοκροτεί (fires)*, παράγοντας έναν παλμό, ο οποίος μέσω του άξονα και των συνάψεων μεταδίδεται στους επόμενους νευρώνες.

Η παραπάνω απλή λειτουργία, η οποία στην ουσία είναι μια πολύπλοκη ηλεκτροχημική λειτουργία, έχει σαν συνέπεια τις γνωστές λειτουργίες του εγκεφάλου.

Οι νευρώνες στον εγκέφαλο των ενηλίκων δεν αναπαράγονται, αλλά κρατάνε μια ζωή. Αυτό σημαίνει πως όλα τα στοιχεία ανανεώνονται για την απρόσκοπτη λειτουργία του νευρώνα. Οι περισσότερες απ' αυτές τις δραστηριότητες διατήρησης γίνονται στο *σώμα* το οποίο είναι ένα πραγματικό χημικό εργοστάσιο, το οποίο παράγει μια πολύ

μεγάλη ποικιλία πολύπλοκων μορίων. Ακόμα, διαχειρίζεται την ενεργειακή οικονομία του νευρώνα και ρυθμίζει ένα μεγάλο πλήθος δραστηριοτήτων μέσα στο κύτταρο. Η εξωτερική μεμβράνη του σώματος έχει τη μοναδική ικανότητα να εκπέμπει παλμούς. Είναι δηλαδή μια ζωτική συνάρτηση του νευρικού συστήματος με υπολογιστικές δυνατότητες.

Το σώμα μορφολογικά είναι διαφορετικό ανάλογα με τον τύπο του νευρώνα. Σήμερα η αναγνώριση των λειτουργιών των διαφόρων κυττάρων απασχολεί έντονα τους ερευνητές και είναι απαραίτητη για την κατανόηση των μηχανισμών επεξεργασίας που υπάρχουν στον εγκέφαλο.

Τα σήματα από άλλους νευρώνες φτάνουν στο νευρώνα μέσω των δενδριτών, διακλαδώσεις που ξεκινούν απ' το σώμα και συνδέονται με τους άξονες άλλων νευρώνων μέσω των συνάψεων.

Σε αντίθεση με τα ηλεκτρονικά κυκλώματα, δεν υπάρχει φυσική ή ηλεκτρονική σύνδεση στις συνάψεις. Αντίθετα, μια στενή λωρίδα, η *συναπτική σχισμή* (*synaptic cleft*) χωρίζει τους δενδρίτες από τον άξονα μεταφοράς. Διάφορες χημικές ουσίες που εκλύονται από τον άξονα στη συναπτική σχισμή, διαπερνούν τους δενδρίτες. Αυτές οι χημικές ουσίες που καλούνται *νευρο-μεταφορείς*, περνούν μέσω των δενδριτών στο σώμα. Το σώμα, στη συνέχεια, συνδέει όλα τα σήματα που λαμβάνει από τους δενδρίτες και στην περίπτωση που το σήμα εξόδου ξεπερνά κάποιο κατώφλι, παράγεται ένας παλμός, ο οποίος μέσω του άξονα περνάει σε άλλους νευρώνες.

Το μήκος του άξονα ποικίλει από 0.1 mm μέχρι 1 μέτρο. Κοντά στο άκρο του, ο άξονας έχει πολλούς κλάδους που ο καθένας καταλήγει σε μια σύναψη, από όπου το σήμα μεταφέρεται σε άλλους νευρώνες μέσω των δενδριτών και σε μερικές περιπτώσεις κατ' ευθείαν στο σώμα άλλων νευρώνων. Μ' αυτό τον τρόπο, ένας νευρώνας μπορεί να παράγει ένα σήμα, το οποίο να θέσει σε λειτουργία εκατοντάδες ή χιλιάδες άλλους νευρώνες, οι οποίοι, με τη σειρά τους, μέσω των δενδριτών τους μπορεί να ενεργήσουν σε χιλιάδες άλλους νευρώνες. Έτσι, είναι αυτή η υψίστου βαθμού συνδεσμολογία μάλλον και όχι η πολύπλοκη λειτουργία που κάνει το νευρώνα να έχει μεγάλες υπολογιστικές δυνατότητες.

Το άκρο του άξονα είναι μια μικρή καμπύλη απόληξη που αποτελείται από σφαιρικές δομές, τις *συναπτικές φυσαλίδες* (*synaptic vesicles*), οι οποίες περιλαμβάνουν ένα μεγάλο αριθμό νευρο-μεταφορικών μορίων. Όταν ένας νευρικός παλμός φτάνει στον άξονα, μερικές απ' αυτές τις φυσαλίδες απελευθερώνουν το περιεχόμενό τους στη συναπτική σχισμή που συνδέει τον άξονα με τους άλλους δενδρίτες, αρχίζοντας έτσι την επικοινωνία μεταξύ των νευρώνων.

Η επικοινωνία στον εγκέφαλο είναι δύο τύπων: χημικά σήματα μέσω των συνάψεων και ηλεκτρικά σήματα μέσα στους νευρώνες. Η κυτταρική μεμβράνη, είναι αυτή που δίνει τη δυνατότητα στο νευρώνα να παράγει και να μεταφέρει αυτά τα δύο είδη σημάτων.

Η κυτταρική μεμβράνη έχει πάχος 5 nm περίπου και αποτελείται από δύο στρώματα, μεταξύ των οποίων υπάρχουν πολλές πρωτεΐνες, οι οποίες λειτουργούν σαν κανάλια, αντλίες, αισθητήρια, ένζυμα κ.λ.π..

Το πάχος της μεμβράνης είναι περίπου 5 nanometers και αποτελείται από δύο στρώματα λιπιδίων. Ακόμα εμπεριέχονται σ' αυτήν ειδικές πρωτεΐνες που

κατατάσσονται σε 5 κατηγορίες: αντλίες, διάυλοι, υποδοχείς, ένζυμα και δομικές πρωτεΐνες.

Οι αντλίες κινούν τα ιόντα δια μέσου της μεμβράνης ώστε ο βαθμός συγκέντρωσής του να παραμένει σταθερός. Οι διάυλοι βοηθούν στην επιλεκτική διόδο των ιόντων και ρυθμίζουν τη ροή τους μέσω της μεμβράνης. Μερικοί διάυλοι ανοίγουν ή κλείνουν ανάλογα με το δυναμικό της μεμβράνης. Άλλοι ελέγχονται χημικά, αλλάζοντας τη διεύθυνσή τους σύμφωνα με τα χημικά μηνύματα που λαμβάνουν.

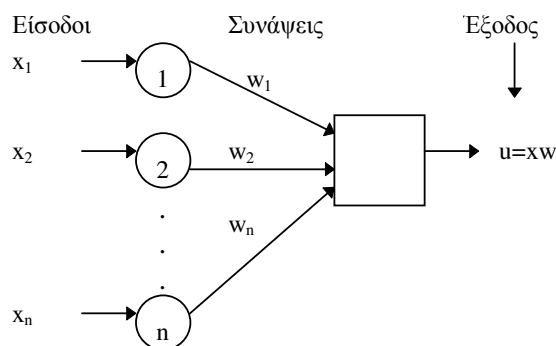
Οι υποδοχείς είναι πρωτεΐνες που έχουν την ικανότητα ν' αναγνωρίζουν και να συγκερατούν στο περιβάλλον του κυττάρου πολλούς χρήσιμους τύπους μορίων. Τα ένζυμα μέσα ή κοντά στη μεμβράνη επιταχύνουν την εκτέλεση των χημικών αντιδράσεων. Οι δομικές πρωτεΐνες ενώνουν μεταξύ τους τα κύτταρα και φροντίζουν να διατηρείται η δομή του κυττάρου σταθερή.

4.4 Χαρακτηριστικά Νευρωνικών Δικτύων

Παρά τη μεγάλη ποικιλία των Τεχνητών Νευρωνικών Δικτύων που έχει αναπτυχθεί, μπορούμε να ξεχωρίσουμε πολλά κοινά χαρακτηριστικά σχετικά με τη δομή και τον τρόπο λειτουργίας τους, τα οποία παρουσιάζονται στις παραγράφους που ακολουθούν.

4.4.1 Ο Τεχνητός Νευρώνας

Ο τεχνητός νευρώνας σχεδιάστηκε με σκοπό να μιμηθεί τα χαρακτηριστικά του βιολογικού νευρώνα. Ένα σύνολο εισόδων που η κάθε μια τους αποτελεί την έξοδο ενός άλλου νευρώνα, εισάγονται σε κάθε νευρώνα. Κάθε είσοδος πολλαπλασιάζεται με την αντίστοιχη σύναψη (ανάλογα με τη συναπτική δύναμη) και όλα αυτά τα γινόμενα αθροίζονται, ώστε να αποτελέσουν το *επίπεδο ενεργοποίησης* του νευρώνα όπως φαίνεται στο σχήμα 4.2 που ακολουθεί:



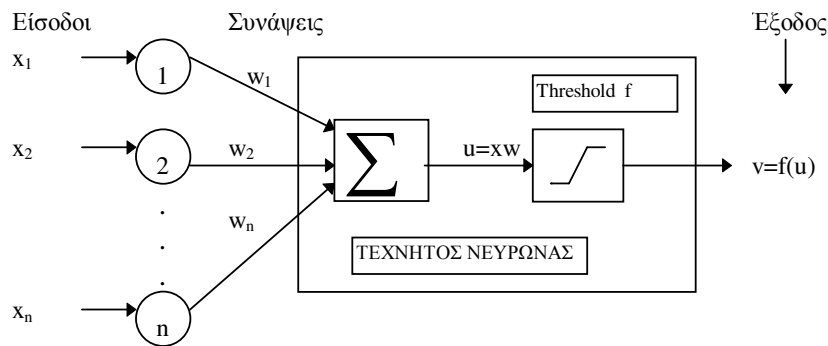
Σχήμα 4.2. Λειτουργία Τεχνητού Νευρώνα

Στο παραπάνω σχήμα, ένα σύνολο εισόδων, οι $x = (x_1, x_2, \dots, x_n)$, εισέρχονται σε κάποιο νευρώνα και αντιστοιχούν στα σήματα του βιολογικού νευρώνα. Κάθε σήμα x_i ,

$i = 1, 2, \dots, n$ πολλαπλασιάζεται με την αντίστοιχη σύναψη w_i , $w = (w_1, w_2, \dots, w_n)$, πριν αθροιστεί για να αποτελέσει το *προσθετικό φράγμα*. Το προσθετικό φράγμα αντιστοιχεί στο κυτταρικό σήμα, ενώ όλα τα γινόμενα προστίθενται αλγεβρικά παράγοντας μια έξοδο, τη u . Τα παραπάνω μπορούν να συμβολιστούν με τη χρήση του εσωτερικού γινομένου των διανυσμάτων x και w :

$$u = \sum_{i=1}^n w_i \cdot x_i = \bar{w} \cdot \bar{x} \quad (4.1)$$

Στο προσθετικό φράγμα u που συγκεντρώνεται απ' τις εισόδους σε κάποιο νευρώνα εφαρμόζεται μια συνάρτηση, η *συνάρτηση ενεργοποίησης*, όπως ονομάζεται, $f()$, η οποία και παράγει το σήμα εξόδου $y = f(u)$. Η λειτουργία της συνάρτησης ενεργοποίησης φαίνεται στο σχήμα 4.3 που ακολουθεί



Σχήμα 4.3. Λειτουργία Νευρώνα με Συνάρτηση Ενεργοποίησης

4.4.2 Συναρτήσεις Ενεργοποίησης

Η συνάρτηση ενεργοποίησης f μπορεί να δίνεται απ' τους παρακάτω τύπους :

$$v = f(u) = au \quad (\text{Γραμμική Συνάρτηση}) \quad (4.2)$$

$$v = f(u) = \begin{cases} u, & \text{αν } u > 0 \\ 0, & \text{αν } u \leq 0 \end{cases} \quad (\text{Συνάρτηση Λογικής Στάθμης}) \quad (4.3)$$

$$v = f(u) = \begin{cases} 1, & \text{αν } u > 0 \\ 0, & \text{αν } u \leq 0 \end{cases} \quad (\text{Συνάρτηση Μοναδιαίου Βήματος}) \quad (4.4)$$

$$v = f(u) = \begin{cases} 1, & \text{αν } u > 0 \\ -1, & \text{αν } u \leq 0 \end{cases} \quad (\text{Συνάρτηση Πρόσημου - sign function}) \quad (4.5)$$

$$v = f(u) = \frac{1}{1 + e^{-ku}} \quad (\text{Σιγμοειδής Συνάρτηση}) \quad (4.6)$$

$$v = f(u) = \tanh(ku) = \frac{e^{ku} - e^{-ku}}{e^{ku} + e^{-ku}} \quad (\text{Συνάρτηση Υπερβολικής Εφαπτομένης}) \quad (4.7)$$

Οι συναρτήσεις (4.2), (4.6) και (4.7) είναι συνεχείς, διαφορίσιμες, ενώ οι (4.6) και (4.7) είναι μη φθίνουσες και μονότονες, ενώ ο συντελεστής k ελέγχει την κλίση τους.

Τα ΤΝΔ είναι το αποτέλεσμα της συνδεσμολογίας πολλών νευρώνων μεταξύ τους και της οργάνωσής τους σε στρώματα (*layers*). Οι συνδεσμολογίες αυτές έχουν διάφορους τύπους και σχήματα. Έτσι, μπορούμε να έχουμε συνάψεις μεταξύ των νευρώνων του ίδιου στρώματος (*intra-field*), συνάψεις μεταξύ των νευρώνων διαφορετικών στρωμάτων (*inter-field*) και αναδρομικές (*recurrent*) συνάψεις που ξεκινούν από κάποιο νευρώνα και επιστρέφουν στον ίδιο. Τα σήματα των συνάψεων μεταξύ των νευρώνων διαφορετικών στρωμάτων ανάλογα με τον τρόπο που προωθούνται μπορεί να είναι σήματα που προωθούνται στα επόμενα στρώματα (*feedforward*) ή σήματα που προωθούνται και προς τις δύο κατευθύνσεις, στα επόμενα, τα προηγούμενα στρώματα και ίσως να είναι και αναδρομικά (*feedback*). Ανάλογα με τον τρόπο ροής των σημάτων (πληροφοριών), κάθε στρώμα (*field* ή *layer*) μπορεί να είναι στρώμα εισόδου (*input layer*), το οποίο δέχεται σήματα (πληροφορίες) απ' το εξωτερικό περιβάλλον, στρώμα εξόδου (*output layer*), το οποίο στέλνει σήματα στο εξωτερικό περιβάλλον ή κρυφό ή εσωτερικό στρώμα (*middle* ή *hidden layer*), κάθε στρώμα μεταξύ των στρωμάτων εισόδου-εξόδου, το οποίο δεν έχει άμεση επαφή με το εξωτερικό περιβάλλον.

4.4.3 Εκπαίδευση Νευρωνικών Δικτύων

Ένα απ' τα πιο ενδιαφέροντα χαρακτηριστικά των ΤΝΔ είναι η ικανότητά τους να "μαθαίνουν". Η εκπαίδευσή τους έχει πολλά κοινά χαρακτηριστικά με την πνευματική ανάπτυξη των ανθρώπων. Παρ' όλα όσα έχουν γίνει όμως σ' αυτό τον τομέα, υπάρχουν ακόμη πολλά προβλήματα και η εκπαίδευση των ΤΝΔ γίνεται με περιορισμούς σε σχέση με την εκπαίδευση των ανθρώπων.

Με τον όρο "εκπαίδευση" ή "εκμάθηση" ενός ΤΝΔ εννοούμε την εύρεση κάποιων τιμών των συνάψεων, ώστε ένα σύνολο εισόδων σ' αυτό να παράγει το επιθυμητό (ή τουλάχιστον να προσεγγίζει το επιθυμητό) σύνολο εξόδων. Η εκπαίδευση γίνεται με τη σειριακή εισαγωγή εισόδων, ενώ ταυτόχρονα οι συνάψεις τροποποιούνται σύμφωνα με κάποιον προκαθορισμένο κανόνα. Κατά τη διάρκεια της εκπαίδευσης, οι συνάψεις, σταδιακά, συγκλίνουν σε τέτοιες τιμές, ώστε κάθε διάνυσμα εισόδου να παράγει την επιθυμητή έξοδο.

Για το σκοπό αυτό, έχουν αναπτυχθεί διάφοροι Αλγόριθμοι Εκπαίδευσης, οι οποίοι χωρίζονται σε 2 κατηγορίες: αλγόριθμοι εκπαίδευσης με επίβλεψη (*supervised learning*) και αλγόριθμοι εκπαίδευσης χωρίς επίβλεψη (*unsupervised learning*). Στην εκπαίδευση με επίβλεψη θεωρούμε τα ζεύγη των διανυσμάτων εισόδου με ένα διάνυσμα-στόχο (*target vector*), το οποίο αποτελεί την επιθυμητή έξοδο. Το δίκτυο εκπαιδεύεται μ' έναν αριθμό από τέτοια ζεύγη ως εξής: εισάγεται πρώτα στο δίκτυο ένα διάνυσμα εισόδου, υπολογίζεται η έξοδος του δικτύου, συγκρίνεται με την αντίστοιχη έξοδο-στόχο και η διαφορά (*διόρθωση* ή *σφάλμα*) ανατροφοδοτείται στο δίκτυο τροποποιώντας τις συνάψεις, σύμφωνα με κάποιον αλγόριθμο, ο οποίος ελαχιστοποιεί το σφάλμα. Η προηγούμενη διαδικασία επαναλαμβάνεται με άλλα ζεύγη εισόδων-στόχων, μέχρι να ελαχιστοποιηθεί το σφάλμα για όλα τα ζεύγη εισόδων-στόχων. Αν και η εκπαίδευση με επίβλεψη θεωρείται σ' ένα βαθμό επιτυχημένη, είναι απίθανο μια τέτοια διαδικασία να συμβαίνει στον ανθρώπινο εγκέφαλο, αφού ο εγκέφαλος δεν μπορεί να ξέρει, εκ των προτέρων, ποια είναι η επιθυμητή έξοδος, για να τη συγκρίνει με την αντίστοιχη πραγματική έξοδο και να κάνει τις κατάλληλες διορθώσεις στις αντίστοιχες συνάψεις. Επίσης, με την προηγούμενο μέθοδο, δεν θα μπορούσαμε να εξηγήσουμε την ικανότητα αυτο-οργάνωσης που έχει αποδειχθεί πως υπάρχει στα βρέφη κατά τα πρώτα στάδια της ανάπτυξής τους. Γι' αυτό και η εκμάθηση χωρίς επίβλεψη θεωρείται πολύ πιο πιθανό μοντέλο για την κατανόηση της λειτουργίας του βιολογικού μοντέλου. Η εκμάθηση με επίβλεψη αναπτύχθηκε απ' τον Kohonen [Koh84] και άλλους και δεν απαιτεί εξόδους-στόχους, ούτε συγκρίσεις των επιθυμητών με τις πραγματικές εξόδους. Απλώς, εισάγουμε ένα σύνολο διανυσμάτων εισόδου. Ο αλγόριθμος εκπαίδευσης τροποποιεί τις συνάψεις του δικτύου, έτσι ώστε τα διανύσματα εισόδου να είναι συνεπή, δηλαδή η είσοδος δύο παρόμοιων διανυσμάτων εισόδου να παράγει την ίδια έξοδο. Στην ουσία, η διαδικασία εκμάθησης αφαιρεί τις βασικές ιδιότητες των διανυσμάτων εισόδου και ομαδοποιεί παρόμοια διανύσματα εισόδου σε κατηγορίες, οι οποίες έχουν παρόμοια χαρακτηριστικά. Εισάγοντας στο δίκτυο ένα διάνυσμα εισόδου που ανήκει σε μια δεδομένη κατηγορία θα παραχθεί μια έξοδος. Δεν υπάρχει όμως τρόπος να βρούμε -πριν την εκπαίδευση- ποια έξοδος θα παραχθεί από μια δεδομένη κατηγορία διανυσμάτων εισόδου.

4.4.4 Ανάκληση - Recall

Με τον όρο *ανάκληση (Recall)* εννοούμε τον τρόπο με τον οποίο το δίκτυο επεξεργάζεται ένα διάνυσμα εισόδου, το οποίο εισάγεται στο στρώμα εισόδου και παράγει μια έξοδο στο στρώμα εξόδου. Υπάρχουν δύο μηχανισμοί ανάκλησης : α) *Nearest-neighbor Recall*, η οποία βρίσκει το αποθηκευμένο πρότυπο που ταιριάζει πιο πολύ με το διάνυσμα εισόδου παράγοντας την αντίστοιχη έξοδο. Π.χ. όταν δοθεί ένα πρότυπο A, απ' όλα τα αποθηκευμένα πρότυπα θα παράγει σαν έξοδο το διάνυσμα A' που απέχει τη μικρότερη απόσταση (*Ευκλείδεια* ή απόσταση *Hamming*) απ' το A. β) *Ανάκληση με Παρεμβολή - Interpolative Recall*, η οποία δέχεται μια άγνωστη είσοδο και *παρεμβάλλει* (πιθανώς με γραμμική παρεμβολή) απ' το σύνολο των αποθηκευμένων ζευγών - προτύπων, παράγοντας την αντίστοιχη έξοδο. Π.χ. για τα αποθηκευμένα ζεύγη

προτύπων (A, A') και (C, C'), όταν δοθεί σαν είσοδος ένα πρότυπο B μεταξύ των A και C, θα παράγει σαν έξοδο το B' μεταξύ των A' και C'.

4.5 Παραδείγματα Τεχνητών Νευρωνικών Δικτύων

Η εκπαίδευση ενός ΤΝΔ μπορεί να έχει τις παρακάτω μορφές :

- **Αυτο-Συσχετιστής (Linear Associator).** Το δίκτυο μαθαίνει πώς να παραστήσει εσωτερικά πρότυπα που εισάγονται σ' αυτό πολλές φορές. Με την εισαγωγή ενός αγνώστου προτύπου, το δίκτυο θα πρέπει να “ανακτήσει” το αποθηκευμένο πρότυπο που βρίσκεται πιο κοντά στο πρότυπο ανάκλησης.
- **Συσχετιστής Προτύπων (Pattern Associator).** Το δίκτυο αποθηκεύει ζεύγη προτύπων που εισάγονται σ' αυτό, όπου το ένα πρότυπο αποτελεί το *κλειδί (key)* και το άλλο το *πρότυπο που σχετίζεται μ' αυτό (associated pattern)*. Με την εισαγωγή ενός ατελούς ή θολού προτύπου, το δίκτυο θα πρέπει να “ανακτήσει” το πρότυπο που σχετίζεται με το πρότυπο εισαγωγής.
- **Ταξινομητής Προτύπων (Pattern Classifier).** Σ' αυτό το είδος δικτύου εισάγονται πρότυπα, τα οποία το δίκτυο πρέπει να ταξινομήσει σε διάφορες κατηγορίες. Συνήθως η διαδικασία ελέγχεται από ένα δάσκαλο (*supervised learning*) που εισάγει τα πρότυπα και την κατηγορία στην οποία ανήκει το καθένα. Μετά την εκπαίδευση, το δίκτυο πρέπει να είναι σε θέση να κατατάσσει παρόμοια πρότυπα στη σωστή κατηγορία.

4.5.1 Συσχετιστική Μνήμη (Associative Memory)

Τα νευρωνικά δίκτυα που ασχολούνται με τη συσχετιστική εκμάθηση και την ανάκτηση πληροφοριών (σε διανυσματική μορφή) και συνήθως αναφέρονται σαν *συσχετιστικές νευρωνικές μνήμες (associative neural memories)* ή *συσχετιστικές μνήμες (associative memories)* αντιπροσωπεύουν μια κατηγορία τεχνητών νευρωνικών δικτύων που έχει μελετηθεί πολύ αναλυτικά [Amar77], [Amar89], [Ande88], [Hopf82]. Αυτά τα μοντέλα μπορούν να ταξινομηθούν με διάφορους τρόπους, ανάλογα με την αρχιτεκτονική τους (στατικά ή αναδρομικά), τον τρόπο ανάκλησης πληροφοριών (σύγχρονος ή ασύγχρονος), τη φύση της αποθήκευσης των συσχετίσεων (αυτο-συσχετιστικά ή έτερο-συσχετιστικά), τους αλγόριθμους αποθήκευσης και ανάκλησης πληροφοριών κ.λ.π..

4.5.1.1 Γραμμική Συσχετιστική Μνήμη (Linear Associative Memory-LAM)

Ένα απ' τα πρώτα μοντέλα συσχετιστικής μνήμης αποτελεί ο γραμμικός συσχετιστής (*linear associator*) που μελετήθηκε απ' τους Anderson [Ande72], Kohonen [Koho72], [Koho89], Nakano [Naka72]. Αποτελείται από ένα στρώμα εισόδου με n νευρώνες πλήρως συνδεδεμένο με ένα στρώμα από k γραμμικούς νευρώνες εξόδου. Το δίκτυο συσχετίζει m διανύσματα εισόδου $x^i \in \mathcal{R}^n$, $i=1,2,\dots,m$ με τα αντίστοιχα m διανύσματα εξόδου $y^i \in \mathcal{R}^k$, $i=1,2,\dots,m$, σύμφωνα με την εξίσωση μεταφοράς

$$y^i = Wx^i \quad (4.8)$$

όπου $\{x^i, y^i\}$, $i=1,2,\dots,m$, είναι μια συλλογή απ' τις επιθυμητές συσχετίσεις και W μία $n \times k$ μήτρα συνάψεων με αντίστοιχα διανύσματα τα $w_1^T, w_2^T, \dots, w_k^T$, όπου $w_j = (w_{j1}, w_{j2}, \dots, w_{jn})^T$, $j=1,2,\dots,k$. Επειδή η παραπάνω συνειρμική μνήμη χαρακτηρίζεται απ' τη χρήση πολλαπλασιασμού μήτρας επί διάνυσμα, αναφέρεται σαν *γραμμική συσχετιστική μνήμη* (*Linear Associative Memory-LAM*). Αν τα διανύσματα x^i, y^i είναι διαφορετικά μεταξύ τους λέγεται *έτερο-συσχετιστική μνήμη* (*heteroassociative memory*), ενώ αν $y^i = x^i$ για κάθε $i=1,2,\dots,m$, τότε λέγεται *αυτο-συσχετιστική μνήμη* (*autoassociative memory*).

4.5.1.2 Εκμάθηση με τον Κανόνα του Hebb

Για να μπορέσει το δίκτυο να συσχετίσει σωστά το διάνυσμα x^i με το y^i , θα πρέπει να βρεθεί η ιδανικότερη μήτρα W που θα εξυπηρετεί αυτό το σκοπό. Με τη χρήση του κανόνα του Hebb μπορούμε να εισάγουμε τις m συσχετίσεις στο δίκτυο εισάγοντας τα m διανύσματα x^i, y^i . Αν ξεκινήσουμε με τη μηδενική μήτρα W , εισάγοντας το ζεύγος προτύπου-στόχου x^i, y^i , και εφαρμόζοντας τον κανόνα του Hebb, η διόρθωση των συνάψεων w_{jl} , $j=1,2,\dots,k$, $l=1,2,\dots,n$ θα έχει τη μορφή

$$w_{jl}^{new} = w_{jl}^{old} + y_{ij}x_{il} \quad (4.9)$$

ή σε μορφή πινάκων, η μήτρα W παίρνει τη μορφή

$$W^{new} = W^{old} + y^i (x^i)^T \quad (4.10)$$

όπου $y^i (x^i)^T$ το εξωτερικό γινόμενο των διανυσμάτων y^i και x^i . Προσθέτοντας και το τελευταίο ζεύγος προτύπου-στόχου x^i, y^i , η μήτρα W θα έχει τη μορφή

$$W = y^1(x^1)^T + y^2(x^2)^T + \dots + y^m(x^m)^T = \sum_{i=1}^m y^i(x^i)^T \quad (4.11)$$

και αποτελείται απ' το άθροισμα των εξωτερικών γινομένων $y^i(x^i)^T$. Ο προηγούμενος τύπος μπορεί να γραφεί

$$W = YX^T \quad (4.12)$$

όπου

$$Y = [y^1, y^2, \dots, y^m] \text{ και } X = [x^1, x^2, \dots, x^m]$$

Ένα σημαντικό πλεονέκτημα του παραπάνω τρόπου αποθήκευσης των συσχετίσεων είναι η ευκολία αποθήκευσης νέων συσχετίσεων ή η διαγραφή παλιών. Έτσι, μετά την αποθήκευση m συσχετίσεων, αν θέλουμε να αποθηκεύσουμε μια νέα συσχέτιση $\{x^{m+1}, y^{m+1}\}$, το μόνο που έχουμε να κάνουμε είναι να προσθέσουμε στη μήτρα W το εξωτερικό γινόμενο $y^{m+1}(x^{m+1})^T$. Με παρόμοιο τρόπο μπορούμε να διαγράψουμε μια συσχέτιση $\{x^i, y^i\}$, αφαιρώντας απ' τη μήτρα W το εξωτερικό γινόμενο $y^i(x^i)^T$.

Στην ανάκτηση όμως των αποθηκευμένων συσχετίσεων η παραπάνω μέθοδος απαιτεί κάποιες προϋποθέσεις που πρέπει να διέπουν τα διανύσματα x^i . Αν π.χ. το διάνυσμα x^h είναι κάποιο απ' τα πρότυπα x^i , $i=1,2,\dots,m$, το πρότυπο y^h που μπορεί να ανακληθεί θα έχει τη μορφή

$$\tilde{y}^h = Wx^h = \left[\sum_{i=1}^m y^i(x^i)^T \right] x^h = \|x^h\|^2 y^h + \sum_{\substack{i=1 \\ i \neq h}}^m y^i(x^i)^T x^h \quad (4.13)$$

Ο πρώτος όρος της εξίσωσης (4.13) είναι ανάλογος του y^h ως προς το τετράγωνο της norm του διανύσματος-κλειδιού x^h , ενώ ο δεύτερος όρος λέγεται 'cross-talk' μεταξύ του κλειδιού x^h και των υπολοίπων $m-1$ προτύπων x^i . Η απαραίτητη προϋπόθεση να ανακληθεί το διάνυσμα y^h με την είσοδο του διανύσματος-κλειδιού x^h είναι

$$\tilde{y}^h = Wx^h = y^h$$

πράγμα που σημαίνει ότι η norm του x^h στην εξίσωση (4.13) είναι ίση με τη μονάδα, ενώ ο δεύτερος όρος της εξίσωσης (4.13) ισούται με το μηδέν. Αυτό συμβαίνει μόνο στην περίπτωση που τα διανύσματα x^i , $i=1,2,\dots,m$ είναι ορθοκανονικά (*orthonormal*), δηλαδή

$$x^i x^j = (x^i)^T x^j = (x^j)^T x^i = \delta_{ij}, \text{ όπου } \delta_{ij} = \begin{cases} 1, & \text{αν } i = j \\ 0, & \text{αν } i \neq j \end{cases}$$

Σ' αυτή την περίπτωση η εξίσωση (4.13) γίνεται

$$\tilde{y}^h = Wx^h = \left[\sum_{i=1}^m y^i (x^i)^T \right] x^h = \|x^h\|^2 y^h + \sum_{\substack{i=1 \\ i \neq h}}^m y^i (x^i)^T x^h = I \cdot y^h + 0 = y^h \quad (4.14)$$

και το διάνυσμα που ανακαλείται με την είσοδο του διανύσματος-κλειδιού x^h είναι το αποθηκευμένο πρότυπο της αντίστοιχης συσχέτισης y^h . Το δίκτυο ανακαλεί τις m αποθηκευμένες συσχετίσεις με τους παρακάτω περιορισμούς: Πρώτον, ο αριθμός των m συσχετίσεων που μπορούν να αποθηκευθούν πρέπει να είναι $\leq n$, επειδή ο μέγιστος αριθμός των ορθοκανονικών διανυσμάτων στο διανυσματικό χώρο με διάσταση n μπορεί να είναι n . Επίσης, τα διανύσματα x^i , $i=1,2,\dots,m$ θα πρέπει να έχουν μήκος τη μονάδα.

4.5.1.3 Βέλτιστη Γραμμική Συσχετιστική Μνήμη (*Optimal Linear Associative Memory- OLAM*)

Αν τα διανύσματα x^i , $i=1,2,\dots,m$ δεν είναι ορθοκανονικά, η καλύτερη μήτρα W μπορεί να βρεθεί με τη χρήση της γραμμικής άλγεβρας. Αν τα διανύσματα x^i , $i=1,2,\dots,m$ είναι γραμμικά ανεξάρτητα, μια καλύτερη μέθοδος εγγυάται την τέλεια ανάκληση των διανυσμάτων y^i με την είσοδο του προτύπου-κλειδιού x^i , $i=1,2,\dots,m$. Η τεχνική αυτή λέγεται Βέλτιστη Γραμμική Συσχετιστική Μνήμη (*Optimal Linear Associative Memory- OLAM*) [Koho74], [Koho89]. Για την τέλεια ανάκληση όλων των m προτύπων θα πρέπει να ισχύει

$$Y = WX \quad (4.15)$$

όπου οι μήτρες X , Y ορίστηκαν στην εξίσωση (4.12). Αν ορίσουμε το Μέσο Τετραγωνικό Σφάλμα (*Mean Squared Error*) του γραμμικού συσχετιστή

$$\begin{aligned} E(W) &= \sum_{i=1}^m E_i(W) = \sum_{i=1}^m |y^i - Wx^i|^2 = \sum_{i=1}^m \sum_{j=1}^k \frac{1}{2} (y_{ij} - \sum_{l=1}^n w_{lj} x^l)^2 = \\ &= \frac{1}{2} (Y - WX)^T (Y - WX) = \frac{1}{2} \|Y - WX\|^2 \end{aligned} \quad (4.16)$$

σαν το κριτήριο βελτιστοποίησης, τα ακρότατα (μέγιστο και ελάχιστο) της συνάρτησης $E(W)$ είναι λύσεις της εξίσωσης που προκύπτει χρησιμοποιώντας παραγώγους για την ελαχιστοποίηση συνάρτησης

$$\nabla E(W) = X^T (XW - Y) = 0 \quad (4.17)$$

το οποίο μας δίνει το αντίστοιχο σύστημα κανονικών εξισώσεων

$$X^T XW = X^T Y \quad (4.18)$$

ή

$$CW = D \quad (4.19)$$

με $C = X^T X$, και $D = X^T Y$, όπου $C \in \mathcal{R}^{n \times n}$ συμμετρική και θετικά ορισμένη.

Η βέλτιστη λύση ελαχίστων τετραγώνων του συστήματος (4.15) ή του ισοδύναμου (4.19) με τη χρήση του γενικευμένου αντιστρόφου του Moore-Penrose A^+ ορίζεται σαν

$$W = YX^+ \quad (4.20)$$

με

$$X^+ = (X^T X)^{-1} X^T$$

Αν $m=n$, η λύση του συστήματος (4.19) είναι μοναδική και ο γενικευμένος αντίστροφος της μήτρας X , X^+ που ορίστηκε στην (3.27) ισούται με τον αντίστροφο X^{-1} .

Αν $m < n$, η λύση του συστήματος (4.19) δεν είναι μοναδική. Μπορούμε να διαλέξουμε οποιαδήποτε απ' τις άπειρες λύσεις που ικανοποιούν την εξίσωση (4.19). Η καλύτερη λύση με το μικρότερο σφάλμα είναι η λύση Ελαχίστων Τετραγώνων που έχει τη μορφή

$$W = Y(X^T X)^{-1} X^T \quad (4.21)$$

Αν $m > n$, η λύση του συστήματος (4.19) είναι η Λύση Ελαχίστων Τετραγώνων. Αν $r(A)=n$, τότε η λύση είναι μοναδική, και δίνεται απ' τον τύπο

$$W = YX^+ \text{ με } X^+ = (X^T X)^{-1} X^T \quad (4.22)$$

όπου X^+ είναι ο ψευδο-αντίστροφος του $(m \times n)$ πίνακα X , ο οποίος πληροί τις συνθήκες των Moore-Penrose του X^+ .

$$XX^+ X = X$$

$$X^+ XX^+ = X^+$$

$$XX^+ = (X X^+)^T$$

$$X^+ X = (X^+ X)^T$$

Είναι εύκολο να αποδειχθεί ότι η λύση x της εξίσωσης (4.19) είναι το ελάχιστο της συνάρτησης $E(W)$ όπως ορίστηκε στην (4.16), αρκεί να δειχθεί ότι η δεύτερη παράγωγος (Hessian) της $E(W)$

$$\nabla \nabla E(W) = \left[\frac{\partial^2 E(W)}{\partial w^i \partial w^j} \right] = \partial \left[\frac{\partial E(W)}{\partial w^i \partial w^j} \right] = \partial \frac{X^T (WX - Y)}{\partial w^i \partial w^j} = X^T X \quad (4.23)$$

είναι συμμετρική και θετικά ορισμένη (SPD), πράγμα που ισχύει, αφού η μήτρα $X^T X = C$ είναι συμμετρική και θετικά ορισμένη.

Ο γενικευμένος αντίστροφος X^+ του Moore-Penrose μπορεί να υπολογισθεί με τη χρήση π.χ. του αλγορίθμου του Greville αν $m \geq n$, και στοιχεία του $W=X^+Y$ τοποθετούνται σαν συνάψεις μεταξύ των νευρώνων εισόδου και του νευρώνα εξόδου, έτσι ώστε η εισαγωγή κάθε προτύπου x^i , $i=1,2,\dots,m$. να δίνει το αντίστοιχο y^i , $i=1,2,\dots,m$.

Με τη χρήση του αλγορίθμου του Greville, αν $W^{old} = Y^{old} (X^{old})^+$ είναι η μήτρα των συνάψεων που αποτελείται απ' τα πρώτα $(m \times (k-1))$ και $((n \times (k-1)))$ μήτρες των πρώτων $(k-1)$ διανυσμάτων y^i και x^i , όταν προστεθεί το νέο διάνυσμα k το W^{new} θα έχει τη μορφή

$$W^{new} = W^{old} + (y^k - W^{old} x^k)(p^k)^T \quad (4.24)$$

με

$$p^k = \begin{cases} \frac{[I - X^{old} (X^{old})^+] x^k}{|I - X^{old} (X^{old})^+ x^k|^2}, & \text{αν ο παρονομαστής είναι } \neq 0 \\ \frac{(X^{old})^T (X^{old})^+ x^k}{I + |(X^{old})^+ x^k|^2}, & \text{αν ο παρονομαστής είναι } = 0 \end{cases}$$

Το δίκτυο αυτό είναι πού εύκολο να υλοποιηθεί, απαιτείται όμως εξωτερικός υπολογισμός του ψευδο-αντίστροφου, και το δίκτυο δεν εκπαιδεύεται, αλλά απλώς μπορεί να ανακαλέσει τις αποθηκευμένες πληροφορίες. Εν τούτοις, ο ψευδο-αντίστροφος μπορεί να υπολογισθεί προσεγγιστικά με τη βοήθεια δικτύων της επόμενης παραγράφου, όπου οι m γραμμές της μήτρας X εισάγονται στο δίκτυο κυκλικά, ενώ για την τροποποίηση των συνάψεων χρησιμοποιείται ο επαναληπτικός αλγόριθμος εκπαίδευσης *LMS* [WiHo60], [WiLe90].

4.5.2 Νευρωνικά Δίκτυα Μονής Κατεύθυνσης (Feed Forward)

Τα Νευρωνικά Δίκτυα με αρχιτεκτονική προώθησης πληροφοριών προς τα εμπρός αποτελούνται από δύο ή περισσότερα στρώματα νευρώνων, οι οποίοι συνδέονται με τους νευρώνες του επομένου στρώματος, ενώ η έξοδος κάθε στρώματος χρησιμοποιείται σαν είσοδος στο επόμενο στρώμα. Χρησιμοποιούνται συνήθως σαν δίκτυα *ταξινόμησης* (*classification*), οπότε σαν είσοδοι χρησιμοποιούνται τα διανύσματα που πρόκειται να ταξινομηθούν και σαν έξοδοι-στόχοι οι αντίστοιχες κατηγορίες-κλάσεις. Στις πιο πολλές περιπτώσεις χρησιμοποιούνται αλγόριθμοι εκπαίδευσης με επίβλεψη.

4.5.2.1 Μηχανή Αντίληψης (Perceptron)

Η πρώτη συστηματική μελέτη των Τεχνητών Νευρωνικών Δικτύων στην Αναγνώριση Προτύπων ήταν μια απλή συσκευή αντίληψης που την αποκάλεσαν perceptron. Αποτελείται από ένα στρώμα νευρώνων εισόδου που συνδέονται με I νευρώνα εξόδου (single perceptron neuron), όπως στο Σχήμα 4.1, ο οποίος δέχεται σαν είσοδο το διάνυσμα x και παράγει την έξοδο v με τη χρήση των εξισώσεων (4.1) και (4.5). Η έξοδος μπορεί να είναι $+I$, αν το διάνυσμα x ανήκει στην κλάση A^+ και $-I$, αν το διάνυσμα x ανήκει στην κλάση A^- . Αν υπάρχουν περισσότερες κλάσεις από δύο, μπορεί να χρησιμοποιηθεί ένα δίκτυο με τόσους νευρώνες εξόδου, όσες και οι κλάσεις (ένας νευρώνας για κάθε κλάση). Το διάνυσμα των συνάψεων w ορίζει μια ευθεία (υπέρ-επίπεδο για περισσότερες κλάσεις από δύο), η οποία χωρίζει το χώρο των διανυσμάτων εισόδου σε δύο κλάσεις, ενώ με τον όρο *εκπαίδευση* (*learning* ή *training*) εννοούμε τη δυνατότητα να βρούμε έναν αλγόριθμο τροποποίησης των συνάψεων, ώστε να βρούμε το διάνυσμα των συνάψεων που χωρίζει το χώρο των διανυσμάτων εισόδου σε δύο κλάσεις. Η εκπαίδευση του perceptron γίνεται με την εισαγωγή ενός συνόλου διαδικών προτύπων στο στρώμα εισόδου, ένα κάθε φορά, και την τροποποίηση των συνάψεων μέχρι που να πάρουμε την επιθυμητή έξοδο για το κάθε πρότυπο. Σαν αλγόριθμος εκπαίδευσης [Rose59] χρησιμοποιείται ο ακόλουθος:

- Ένα σύνολο προτύπων εισάγεται στο δίκτυο (ένα πρότυπο κάθε φορά).
- Αν το δίκτυο δώσει την επιθυμητή έξοδο, εισάγουμε το επόμενο πρότυπο.
- Αν το δίκτυο δεν δίνει την επιθυμητή έξοδο και είναι $-I$ αντί για I , τροποποιούμε τις συνάψεις προσθέτοντας σε κάθε σύναψη την αντίστοιχη είσοδο, η οποία προκάλεσε τη λάθος έξοδο και εισάγουμε το επόμενο πρότυπο.
- Αν το δίκτυο δεν δίνει την επιθυμητή έξοδο και είναι I αντί για $-I$, τροποποιούμε τις συνάψεις αφαιρώντας από κάθε σύναψη την αντίστοιχη είσοδο, η οποία προκάλεσε τη λάθος έξοδο και εισάγουμε το επόμενο πρότυπο.
- Συνεχίζουμε την προηγούμενη διαδικασία, μέχρις ότου το δίκτυο να δίνει την επιθυμητή έξοδο για κάθε πρότυπο.

Όπως αποδείχτηκε απ' τον Rosenblatt [Rose59] το Perceptron μπορεί να εκπαιδευτεί (ο αλγόριθμος εκπαίδευσης να συγκλίνει σε συγκεκριμένο αριθμό βημάτων) όταν οι δυο κλάσεις A^+ και A^- είναι γραμμικά διαχωρίσιμες. Γενικά, με τον όρο *γραμμική διαχωρισιμότητα* εννοούμε τη δυνατότητα εύρεσης κάποιου υπερ-επιπέδου $n-1$ διαστάσεων, όπου n ο αριθμός των εισόδων (ευθεία για $n = 2$, επίπεδο για $n = 3$ κ.λ.π.) το οποίο διαχωρίζει το σύνολο των σημείων απεικόνισης μιας συνάρτησης (που αντιστοιχούν στις εισόδους) στο επίπεδο, τον κύβο και γενικά το χώρο n διαστάσεων σε 2 κλάσεις.

Ένα απ' τα συμπεράσματα του Minsky [MiPa69], το οποίο είχε σαν αποτέλεσμα να σταματήσει το ενδιαφέρον των επιστημόνων για τα Τεχνητά Νευρωνικά Δίκτυα ήταν το γεγονός πως ένα perceptron με μια έξοδο δε μπορούσε να παραστήσει την πύλη XOR 2 εισόδων.

Μια γενίκευση του αλγορίθμου εκπαίδευσης του perceptron για συνεχείς εισόδους αποτελεί ο *Κανόνας του Δέλτα (Delta Rule)* [RuML86]. Η εκπαίδευση του perceptron γίνεται με την εισαγωγή ενός συνόλου προτύπων στο στρώμα εισόδου, ένα κάθε φορά, και την τροποποίηση των συνάψεων μέχρι που να πάρουμε την επιθυμητή έξοδο για το κάθε πρότυπο. Σαν αλγόριθμος εκπαίδευσης χρησιμοποιείται ο Κανόνας του Δέλτα (Delta Rule) όπου T είναι ο στόχος (η επιθυμητή έξοδος) και v η πραγματική έξοδος:

- 1) Εισάγεται ένα πρότυπο και υπολογίζεται η έξοδος v .
- 2) Τροποποιούνται οι συνάψεις w_i σύμφωνα με τον τύπο

$$w_i^{(t+1)} = w_i^{(t)} + n(T-v)x_i$$

με n το συντελεστή εκπαίδευσης, $0 < n < 1$.

- 3) Επαναλαμβάνονται τα βήματα 1,2, μέχρι να συγκλίνει ο αλγόριθμος.

4.5.2.2 Κανόνες Εκπαίδευσης

Οι επαναληπτικοί αλγόριθμοι που περιγράφηκαν διαχωρίζονται σε δύο κατηγορίες, ανάλογα με τον τρόπο διόρθωσης των σφαλμάτων.

- Οι *Κανόνες Διόρθωσης Σφάλματος (Error Correction Rules)*, οι οποίοι μεταβάλουν τα βάρη του δικτύου, με σκοπό να διορθώσουν το λάθος που παρουσιάζεται στην έξοδο και αντιστοιχεί σε μια συγκεκριμένη είσοδο.
- Οι *Κανόνες Κλιμακωτής Μείωσης (Gradient Descent Rules)*, οι οποίοι μεταβάλουν τα βάρη του δικτύου κατά τη διάρκεια της εκπαίδευσης κάθε προτύπου μειώνοντας την κλίση με το μέσο τετράγωνο του λάθους που είναι ο μέσος όρος όλων των προτύπων εκπαίδευσης.

Οι κανόνες διόρθωσης λαθών συνήθως χρησιμοποιούνται όταν ο σκοπός της εκπαίδευσης δεν είναι εύκολο να προσδιοριστεί ποσοτικά ή όταν ένα πρόβλημα δεν οδηγεί σε ικανοποιητική ανάλυση, όπως τα νευρωνικά δίκτυα με μη συνεχείς συναρτήσεις.

4.5.2.3 Γραμμικό Προσαρμοστικό Στοιχείο (Adaline)

Μια επέκταση του παραπάνω αλγορίθμου αποτελεί ο αλγόριθμος LMS ή ο αλγόριθμος των Widrow-Hoff [WiHo60], [WiLe90]. Ο αλγόριθμος αρχικά χρησιμοποιήθηκε για το *Adaline*, μια συσκευή παρόμοια με το Perceptron, αλλά με γραμμική συνάρτηση ενεργοποίησης, ενώ εισάγεται και η έννοια του σφάλματος, το οποίο καθορίζει και την απόδοση του δικτύου. Με την παρουσίαση κάθε προτύπου το σφάλμα του ορίζεται σαν η διαφορά της εξόδου απ' την επιθυμητή έξοδο, ενώ σκοπός του αλγορίθμου είναι η ελαχιστοποίηση του αθροίσματος των τετραγώνων αυτών των σφαλμάτων, δηλαδή το μέσο τετραγωνικό σφάλμα (Mean Square Error), απ' όπου προκύπτει και το όνομα του αλγορίθμου LMS (Least Mean Square). Το βασικό πλεονέκτημα του αλγορίθμου είναι ότι βρίσκει πάντα μια λύση, ακόμη κι αν οι κλάσεις δεν είναι γραμμικά διαχωρίσιμες, ενώ το μειονέκτημα είναι ότι το σφάλμα αυτό δεν είναι ελάχιστο.

Το adaline είναι μια μηχανή που αποτελείται από ένα μόνο νευρώνα. Ο όρος adaline είναι ακρώνυμο, αν και η σημασία του άλλαξε κάπως με τα χρόνια. Αρχικά ονομάστηκε ADA-ptive LI-near NE-uron και αργότερα έγινε ADA-ptive LIN-ear E-lement, όταν τα νευρωνικά δίκτυα απέτυχαν να λύσουν το πρόβλημα XOR στα τέλη της δεκαετίας του 60. Το adaline (και το πολλαπλό adaline, το οποίο καλείται Madaline) ανακαλύφθηκε απ' το Bernard Widrow [WiHo60] στο Πανεπιστήμιο του Stanford και χρησιμοποιήθηκε στη μεταφορά σημάτων. Η μεταφορά σημάτων είναι μια θεωρία της μηχανικής και έχει να κάνει κυρίως με την εφαρμογή φίλτρων για την εξάλειψη ή τη μείωση ανεπιθύμητων συχνοτήτων από το σήμα της πληροφορίας.

Το adaline (ή το ALC) είναι προσαρμοστικό (ADaptive) με την έννοια ότι υπάρχει εκ των προτέρων μια καλά ορισμένη διαδικασία για την προσαρμογή των βαρών έτσι ώστε να επιτρέπει στο adaline να δίνει τις σωστές τιμές στην έξοδο για τις συγκεκριμένες τιμές της εισόδου. Είναι επίσης γραμμικό (LINEar), διότι η έξοδος είναι μια απλή γραμμική συνάρτηση των τιμών της εισόδου.

4.5.2.3.1 Ο Αλγόριθμος Εκπαίδευσης α -LMS

Ο αλγόριθμος εκπαίδευσης α -LMS είναι ένας κανόνας διόρθωσης σφάλματος που χρησιμοποιείται στην εύρεση λύσεων σε καλά ορισμένα προβλήματα και εύκολα γραμμικά προβλήματα, ενώ κάνει διόρθωση λαθών ανάλογη με το λάθος του κάθε προτύπου και ακολουθεί τα παρακάτω βήματα:

1. Αρχικά, οι συνάψεις w_j , $j = 0, 1, 2, \dots, n$ παίρνουν τυχαίες τιμές στο διάστημα $(-1, 1)$.
2. Όταν εισάγεται στο στρώμα εισόδου το p -στό πρότυπο $x^p = (x_1^p, x_2^p, \dots, x_n^p)$, $1 \leq p \leq M$, υπολογίζεται στο νευρώνα εξόδου η αντίστοιχη έξοδος

$$u^p = w_0 + \sum_{i=1}^n w_i x_i^p = \sum_{i=0}^n w_i x_i^p \quad (4.25)$$

όπου w_i η σύναψη μεταξύ του i -στού νευρώνα εισόδου και του νευρώνα εξόδου και x_0^p , η είσοδος της διέγερσης (*bias*) για κάθε πρότυπο p είναι $+1$.

3. Υπολογίζεται η διαφορά μεταξύ της επιθυμητής εξόδου d^p και της πραγματικής εξόδου u^p , $\varepsilon^p = d^p - u^p$, η οποία δίνει τη διαφορά των συνάψεων

$$\Delta w_i = \alpha \frac{\varepsilon^p x_i^p}{\|x^p\|^2}, \quad i=0,1,2,\dots,n \quad (4.26)$$

4. Μετά την εισαγωγή του p -στού προτύπου, η διόρθωση των συνάψεων θα έχει τη μορφή

$$w_i = w_i + \Delta w_i = w_i + \alpha \frac{\varepsilon^p x_i^p}{\|x^p\|^2}, \quad i=0,1,2,\dots,n \quad (4.27)$$

Η παραπάνω διαδικασία επαναλαμβάνεται με την είσοδο του επόμενου προτύπου, μέχρις ότου το σφάλμα μεταξύ των επιθυμητών και πραγματικών εξόδων είναι μέσα σε αποδεκτά όρια.

Ο αλγόριθμος κάνει κανονικοποίηση από μόνος του, αφού διαιρεί με τη νόρμα (*norm*) του προτύπου x^p , ενώ η επιλογή του α δεν εξαρτάται από την τιμή των προτύπων εισόδου.

Ο αλγόριθμος α -LMS είναι ένας γραμμικός κανόνας εκπαίδευσης που κάνει διόρθωση λαθών ανάλογη με το λάθος. Είναι γνωστό ότι μερικές φορές ο γραμμικός κανόνας πιθανόν να αποτύχει να διαχωρίσει τα πρότυπα εκπαίδευσης, ακόμη κι αν είναι γραμμικά διαχωρίσιμα. Σ' αυτές τις περιπτώσεις μπορούν να χρησιμοποιηθούν μη γραμμικοί κανόνες εκπαίδευσης, οι οποίοι κάνουν αλλαγές στο διάνυσμα βαρών, αλλαγές που βασίζονται σε γραμμικό λάθος, αλλά που δεν είναι απ' ευθείας ανάλογο με αυτό.

4.5.2.3.2 Ο Αλγόριθμος Εκπαίδευσης μ -LMS (Least Mean Square)

Ο αλγόριθμος μ -LMS [WiHo60], [WiLe90] είναι ο πιο αναλυτικός και φημισμένος αλγόριθμος εκπαίδευσης. Αν και πρωτο-εφαρμόστηκε σε δίκτυα με ένα νευρώνα εξόδου, μπορεί να επεκταθεί και σε δίκτυα με πολλούς νευρώνες εξόδου. Χρησιμοποιήθηκε αρχικά για την εκπαίδευση των γραμμικών μονάδων ADALINE. Ο αλγόριθμος εκπαίδευσης ανήκει στην κατηγορία των αλγορίθμων κλιμακωτής μείωσης (*gradient descent*), στους οποίους ορίζεται μια κατάλληλη αντικειμενική συνάρτηση που πρέπει να ελαχιστοποιηθεί και για το σκοπό αυτό χρησιμοποιείται ένας επαναληπτικός αλγόριθμος βελτιστοποίησης. Ο αλγόριθμος ελαχιστοποιεί το άθροισμα των τετραγώνων των λαθών κατά τη διάρκεια της εκπαίδευσης. Σαν αντικειμενική συνάρτηση ορίζεται η συνάρτηση του Μέσου Τετραγωνικού Σφάλματος, ή Συνάρτηση Κόστους

$$E^p(w) = \frac{1}{2}(\varepsilon^p)^2 = \frac{1}{2}(d^p - \sum_{j=0}^n x_j^p w_j)^2 \quad (4.28)$$

με $E^p(w)$ το τετραγωνικό σφάλμα για το πρότυπο p .

Λειτουργία του Αλγόριθμου

1. Αρχικά, οι συνάψεις w_j , $j=0,1,2,\dots,n$ παίρνουν τυχαίες τιμές στο διάστημα $(-1, 1)$.
2. Εισάγεται στο στρώμα εισόδου του δικτύου το p -στό πρότυπο $x^p = (x_1^p, x_2^p, \dots, x_n^p)$, $1 \leq p \leq M$ και υπολογίζεται στο νευρώνα εξόδου η αντίστοιχη έξοδος $u^p = \sum_{i=0}^n w_i x_i^p$ όπως και στην (4.25).
3. Υπολογίζεται η διαφορά μεταξύ της επιθυμητής εξόδου d^p και της πραγματικής εξόδου u^p , $\varepsilon^p = d^p - u^p$, η οποία δίνει τη διαφορά των συνάψεων

$$\Delta w_i = \mu \varepsilon^p x_i^p \quad i=0,1,2,\dots,n \quad (4.29)$$

4. Μετά την εισαγωγή του p -στού προτύπου, η διόρθωση των συνάψεων w_i , $i=0,1,2,\dots,n$ θα έχει τη μορφή

$$w_i = w_i + \Delta w_i = w_i + \mu \varepsilon^p x_i^p \quad (4.30)$$

όπου μ ο συντελεστής εκπαίδευσης ή μέγεθος βήματος, ο οποίος μπορεί να είναι σταθερός ή να τροποποιείται σε κάθε επανάληψη. Ο αλγόριθμος επαναλαμβάνεται μέχρις ότου το σφάλμα μεταξύ των επιθυμητών και πραγματικών εξόδων είναι μέσα σε αποδεκτά όρια, ενώ οι τιμές w_i , $i=0,1,2,\dots,n$ του διανύσματος w συγκλίνουν σε μια προσεγγιστική λύση του συστήματος

$$Xw = d \quad (4.31)$$

όπου $X = [x^1, x^2, \dots, x^m]^T$ ο πίνακας των προτύπων εκπαίδευσης και $d = [d^1, d^2, \dots, d^m]$ το διάνυσμα των αντίστοιχων στόχων. Έχει αποδειχθεί [Koho74] ότι, αν ο συντελεστής εκπαίδευσης μ είναι πολύ μικρός η ακολουθία των διανυσμάτων w που δημιουργείται με τον αλγόριθμο LMS συγκλίνουν πολύ κοντά στη βέλτιστη λύση.

4.5.2.3 Ο Αλγόριθμος Εκπαίδευσης Batch-LMS

Αν στον προηγούμενο αλγόριθμο χρησιμοποιήσουμε την ολική παράγωγο

$$E(w) = \sum_{p=1}^M E^{(p)}(w) = \sum_{p=1}^M \frac{1}{2} (\varepsilon^{(p)})^2 = \frac{1}{2} \sum_{p=1}^M (d^{(p)} - u^{(p)})^2 = \frac{1}{2} \sum_{p=1}^M (d^{(p)} - \sum_{i=0}^n x_i^{(p)} w_i)^2 \quad (4.32)$$

αντί της προσεγγιστικής $\nabla E^p(w)$, παίρνουμε τον *ομαδικό αλγόριθμο LMS (batch-LMS)*. Η ολική παράγωγος $\nabla E(w)$ υπολογίζεται απ' το άθροισμα των μερικών παραγώγων $\nabla E^p(w)$ για τα M διαφορετικά πρότυπα. Η διόρθωση των συνάψεων w_i $i=0,1,2,\dots,n$ θα έχει τη μορφή

$$w_i = w_i + \Delta w_i = w_i + \mu \sum_{p=1}^M \varepsilon^p x_i^p \quad (4.33)$$

Ο αλγόριθμος επαναλαμβάνεται μέχρις ότου το σφάλμα μεταξύ των επιθυμητών και πραγματικών εξόδων είναι μέσα σε αποδεκτά όρια, ενώ οι τιμές w_i $i=0,1,2,\dots,n$ του διάνυσματος w συγκλίνουν σε μια προσεγγιστική λύση του συστήματος (4.31).

Η σύγκλιση των δυο προηγούμενων μεθόδων έχει σχέση με το συντελεστή εκπαίδευσης μ . Αν το μ είναι μικρό, η σύγκλιση είναι πολύ αργή, εξ αιτίας του μεγάλου αριθμού των βημάτων που απαιτούνται για την εύρεση ενός τοπικού ελαχίστου. Αν το μ είναι μεγάλο, η σύγκλιση στην αρχή είναι πολύ γρήγορη, αλλά στη συνέχεια ο αλγόριθμος παλινδρομεί, χωρίς να συγκλίνει σε κάποιο τοπικό ελάχιστο. Γενικά, το καλύτερο είναι να έχουμε μεγάλα βήματα, όταν το σημείο αναζήτησης είναι μακριά απ' το ελάχιστο και μικρότερα βήματα, όταν το σημείο αναζήτησης πλησιάζει το ελάχιστο.

4.5.2.4. Αλγόριθμοι Κλιμακωτής Μείωσης (Gradient Descent)

Οι δύο προηγούμενοι αλγόριθμοι (μ -LMS και *batch-LMS*) είναι αλγόριθμοι *κλιμακωτής μείωσης (gradient descent)*. Οι αλγόριθμοι *gradient-descent* αρχίζοντας σε ένα τυχαίο σημείο $w^{(0)}$ δημιουργούν μια ακολουθία από βελτιωμένες προσεγγίσεις $w^{(1)}, w^{(2)}, \dots$, τέτοια ώστε, για $t \geq 0$,

$$w^{(t+1)} = w^{(t)} + \mu s^{(t)} \quad (4.34)$$

όπου $s^{(t)}$ είναι η *κατεύθυνση κατάβασης (descent direction)* και μ ο συντελεστής εκπαίδευσης ή μέγεθος βήματος. Ο συντελεστής εκπαίδευσης μπορεί να είναι σταθερός ή να τροποποιείται σε κάθε επανάληψη. Η μέθοδος αυτή εφαρμόζεται μόνο σε συναρτήσεις που μπορούν να παραγωγηθούν. Η προσαρμοστικότητα του δικτύου με τη μέθοδο *steepest descent* αρχίζει δίνοντας αυθαίρετες αρχικές τιμές στο διάνυσμα των

βαρών. Υπολογίζεται η κλίση της συνάρτησης MSE και το διάνυσμα των βαρών μεταβάλλεται στη διεύθυνση που αντιστοιχεί με την αρνητική κλίση που υπολογίσθηκε. Αυτή η διαδικασία επαναλαμβάνεται, έτσι ώστε το μέσο τετραγωνικό σφάλμα (MSE) να ελαττώνεται σταδιακά κατά μέσο όρο, αναγκάζοντας το διάνυσμα των βαρών να πλησιάσει μια τοπικά ελάχιστη τιμή.

Οι κανόνες steepest descent για ένα μόνον στοιχείο λέμε ότι είναι γραμμικοί εάν οι αλλαγές των βαρών είναι ανάλογες του γραμμικού λάθους, τη διαφορά μεταξύ της επιθυμητής και της γραμμικής εξόδου του στοιχείου.

4.5.2.5 Γενίκευση του Αλγορίθμου Εκπαίδευσης μ -LMS

Αποδεικνύεται ότι στην περίπτωση της (4.30) η κατεύθυνση κατάβασης $s^{(i)}$ ισούται με $-\nabla E^p(w)$, με το $E^p(w)$ όπως ορίστηκε στην (4.27) και η (4.29) γίνεται

$$w^{(i+1)} = w^{(i)} + \mu(-\nabla E^{(p)}(w^{(i)})) \quad (4.35)$$

Ο αλγόριθμος μ -LMS ελαχιστοποιεί τα τετράγωνα των διαφορών μεταξύ των πραγματικών και των επιθυμητών τιμών εξόδου για το κάθε πρότυπο και υλοποιεί μια κλιμακωτή μείωση σε μία επιφάνεια της διάστασης του βάρους, έτσι ώστε το ύψος της σε οποιοδήποτε σημείο να είναι ίσο με την τιμή του σφάλματος. Αυτό σημαίνει ότι η παράγωγος του μέτρου του σφάλματος ως προς το κάθε βάρος είναι ανάλογη με την μεταβολή του βάρους που ορίστηκε από τον κανόνα εκπαίδευσης Δέλτα ή μ -LMS, με μία αρνητική σταθερά αναλογικότητας. Αν $E^p(w)$, όπως ορίστηκε στην (4.27), είναι το μέτρο του σφάλματος του p προτύπου εισόδου/εξόδου και $E(w)$, όπως ορίστηκε στην (4.31), είναι το συνολικό μέτρο του σφάλματος, θα πρέπει να αποδειχθεί ότι

$$-\frac{\partial E^{(p)}(w)}{\partial w_i} = \varepsilon^{(p)} x_i^{(p)} \quad (4.36)$$

το οποίο είναι ανάλογο με το $\Delta_p w_i$ όπως καθορίστηκε από τον Κανόνα Δέλτα. Αν γράψουμε την παράγωγο σαν γινόμενο δύο παραγόντων, της μερικής παραγώγου του σφάλματος σε σχέση με την έξοδο του νευρώνα επί την παράγωγο της εξόδου σε σχέση με το βάρος

$$\frac{\partial E^{(p)}(w)}{\partial w_i} = \frac{\partial E^{(p)}(w)}{\partial u^{(p)}} \frac{\partial u^{(p)}}{\partial w_i} \quad (4.37)$$

το πρώτο μέρος δηλώνει με ποιον τρόπο μεταβάλλεται το σφάλμα με την έξοδο του νευρώνα, μετά την είσοδο του p προτύπου, και το δεύτερο μέρος δηλώνει κατά πόσο η μεταβολή του w_i διαφοροποιεί την έξοδο. Από την εξίσωση (4.28) έχουμε

$$\frac{\partial E^{(p)}(w)}{\partial u^p} = -(d^{(p)} - u^{(p)}) = -\varepsilon^{(p)} \quad (4.38)$$

Η συνεισφορά του νευρώνα στο σφάλμα είναι ανάλογη με το $\varepsilon^{(p)}$. Αν παραγωγίσουμε τη σχέση (4.25) έχουμε

$$\frac{\partial u^{(p)}}{\partial w_i} = x_i^{(p)} \quad (4.39)$$

Αντικαθιστώντας στην Εξίσωση (4.37), παίρνουμε τη σχέση (4.36) και η (4.30) γίνεται

$$w_i = w_i + \mu \varepsilon^{(p)} x_i^{(p)} = w_i + \mu (d^{(p)} - u^{(p)}) x_i^{(p)} = w_i - \mu \frac{\partial E^{(p)}(w)}{\partial w_i} \quad (4.40)$$

ή σε μορφή πινάκων

$$w = w - \mu \nabla E^{(p)}(w) \quad (4.41)$$

όπου $\nabla E^{(p)}(w)$ η μερική παράγωγος της $E^{(p)}$, όπως ορίστηκε στην (4.27).

4.5.2.6 Γενίκευση του Αλγορίθμου Εκπαίδευσης Batch-LMS

Στην περίπτωση της (4.33) η κατεύθυνση κατάβασης $s^{(i)}$ ισούται με $-\nabla E(w)$, με το $E(w)$ όπως ορίστηκε στην (4.33). Ο αλγόριθμος batch-LMS ελαχιστοποιεί τα τετράγωνα των διαφορών μεταξύ των πραγματικών και των επιθυμητών τιμών εξόδου για όλα τα πρότυπα και υλοποιεί μια κλιμακωτή μείωση στο $E(w)$. Αν $E(w)$, όπως ορίστηκε στην (4.33), είναι το συνολικό μέτρο του σφάλματος όλων των προτύπων εισόδου/εξόδου θα δείξουμε ότι

$$-\frac{\partial E(w)}{\partial w_i} = \sum_{p=1}^M \varepsilon^{(p)} x_i^{(p)} \quad (4.42)$$

Αν χρησιμοποιήσουμε την (4.37), η (4.42) γίνεται

$$\frac{\partial E(w)}{\partial w_i} = \frac{\partial E(w)}{\partial u^{(p)}} \frac{\partial u^{(p)}}{\partial w_i} \quad (4.43)$$

Αλλά

$$\frac{\partial E(w^{(l)})}{\partial u^{(p)}} = -\sum_{p=1}^M (d^{(p)} - u^{(p)}) = -\sum_{p=1}^M \varepsilon^{(p)} \quad (4.44)$$

οπότε, με τη χρήση της (4.38), η (4.42) γίνεται

$$\frac{\partial E(w)}{\partial w_i} = \frac{\partial E(w)}{\partial u^{(p)}} \frac{\partial u^{(p)}}{\partial w_i} = -\sum_{p=1}^M \varepsilon^{(p)} x_i^{(p)} \quad (4.45)$$

Αντικαθιστώντας την (4.44) στην (4.32) έχουμε

$$w_i = w_i - \mu \frac{\partial E(w)}{\partial w_i} \quad (4.46)$$

ή σε μορφή πινάκων

$$w = w - \mu \nabla E(w) \quad (4.47)$$

Όπως φαίνεται απ' τις σχέσεις (4.40) και (4.46), η καθαρή αλλαγή στο w_i $i=0,1,2,\dots,n$, μετά την εισαγωγή ενός προτύπου είναι ανάλογη προς την παράγωγο και ως εκ τούτου ο κανόνας δέλτα υλοποιεί μια κλιμακωτή μείωση στο E . Στην πραγματικότητα, αυτό ισχύει, μόνο αν οι τιμές των βαρών δεν αλλάζουν κατά τη διάρκεια αυτού του κύκλου, όπως γίνεται στη μέθοδο batch-LMS. Αλλάζοντας τα βάρη μετά την εισαγωγή κάθε προτύπου, όπως γίνεται στη μέθοδο μ -LMS, παρεκκλίνουμε σε μικρό βαθμό από την κλιμακωτή μείωση στο E . Παρ' όλα αυτά, εφ' όσον ο συντελεστής εκπαίδευσης είναι αρκετά μικρός, αυτή η παρέκκλιση θα είναι αμελητέα και ο Κανόνας του Δέλτα θα πραγματοποιήσει μια πολύ καλή προσέγγιση της κλιμακωτής μείωσης του σφάλματος του αθροίσματος των τετραγώνων. Συγκεκριμένα με αρκετά μικρό συντελεστή εκπαίδευσης, ο Κανόνας του Δέλτα θα βρει ένα σύνολο βαρών ελαχιστοποιώντας τη συνάρτηση του σφάλματος.

Η αντικειμενική συνάρτηση $E(w)$ είναι τετραγωνική μορφή σε σχέση με τις συνάψεις, εξ αιτίας της γραμμικής σχέσης μεταξύ u^p και w , όπως φαίνεται στην εξίσωση (4.31). Η $E(w)$ ορίζει ένα *κυρτό* (convex) υπερπαραβολοειδές που ποτέ δεν γίνεται αρνητική με ένα ολικό ελάχιστο w^* . Αν ο συντελεστής εκπαίδευσης μ είναι πολύ μικρός, η κατάβαση κατά μήκος αυτής της επιφάνειας με τη διαδικασία της φθίνουσας αναζήτησης (gradient search) που υλοποιείται με τις εξισώσεις (4.24) και (4.31), θα συγκλίνει ασυμπτωτικά στη λύση w^* , ανεξάρτητα απ' τις αρχικές τιμές του w .

Είναι εύκολο να αποδειχθεί ότι η λύση x της εξίσωσης (4.31) είναι το ελάχιστο της συνάρτησης $E(w)$ όπως ορίστηκε στην (4.32), αρκεί να δειχθεί ότι η δεύτερη παράγωγος (Hessian) της $E(w)$

$$\nabla \nabla E(w) = \left[\frac{\partial^2 E(w)}{\partial w_i \partial w_j} \right] = \frac{\partial}{\partial w_i} \left[\frac{\partial E(w)}{\partial w_j} \right] = \frac{\partial}{\partial w_i} \frac{X^T (Xw - d)}{\partial w_i \partial w_j} = X^T X \quad (4.48)$$

είναι συμμετρική και θετικά ορισμένη (SPD), πράγμα που ισχύει, αφού η μήτρα $X^T X = C$ είναι συμμετρική και θετικά ορισμένη.

4.5.2.7 Back-Error Propagation

Το πρόβλημα με το Perceptron και το Adaline είναι οι περιορισμένη τους δυνατότητα να χωρίζουν το χώρο των εισόδων σε δύο μέρη. Αν και πολλά προβλήματα ανήκουν σ' αυτή την κατηγορία, οι περισσότερες εφαρμογές απαιτούν την εύρεση πιο πολύπλοκων περιοχών αποφάσεων. Παράδειγμα η πύλη XOR, όπου ο χωρισμός σε δύο κλάσεις με κάποια απ' τα προηγούμενα δίκτυα ενός στρώματος (το στρώμα εισόδου δεν υπολογίζεται σαν ξεχωριστό στρώμα) είναι αδύνατος [MiPa69]. Μια λύση είναι η χρήση των κρυφών στρωμάτων (hidden layers) συνδέοντας τις εξόδους του κάθε στρώματος με τις εισόδους του επόμενου στρώματος και η τροποποίηση των συνάψεων με τρόπο ώστε να απομονωθούν πολύπλοκες περιοχές στα ανώτερα στρώματα με ενώσεις και τομές περιοχών των χαμηλότερων στρωμάτων. Η αρχιτεκτονική αυτή, γνωστή σαν *πολλαπλό δίκτυο απλής κατεύθυνσης (multi-layer feed-forward network)*, έχει αποδειχθεί επιτυχής στην επίλυση προβλημάτων ταξινόμησης, αναγνώρισης σημάτων, αναγνώρισης προτύπων κ.λ.π.. Έχει αποδειχθεί ότι με τη χρήση δύο κρυφών στρωμάτων οποιαδήποτε περιοχή μπορεί να διαχωριστεί. Πρέπει να τονισθεί ότι η δύναμη αυτών των δικτύων οφείλεται στη μη-γραμμικότητα των νευρώνων. Σε ένα δίκτυο με γραμμικούς νευρώνες (όπως στο Adaline), το αποτέλεσμα ενός δικτύου με ενδιάμεσα στρώματα μπορεί να προκύψει με τη χρήση ενός ισοδύναμου δικτύου με ένα στρώμα. Η χρήση των μη-γραμμικών συναρτήσεων ενεργοποίησης (4.3) και (4.4) που είναι συνεχείς, μονότονες, φθίνουσες και διαφορίσιμες προσφέρει το πλεονέκτημα να χρησιμοποιηθούν διάφορες μέθοδοι ελαχιστοποίησης. Ο αλγόριθμος Back-Propagation (ή *γενικευμένος κανόνας Δέλτα*) [Werb74], [RuHW86] χρησιμοποιεί όλα τα παραπάνω πλεονεκτήματα και αργά, αλλά σταθερά μετακινεί τα διανύσματα βαρών στις σωστές τιμές.

Στα προηγούμενα παραδείγματα χρησιμοποιήθηκε μόνο ένας νευρώνας εξόδου. Στη γενική περίπτωση μπορούν να υπάρχουν πολλοί νευρώνες εξόδου και το δίκτυο παράγει μια απεικόνιση ενός πολυδιάστατου χώρου σε έναν άλλο πολυδιάστατο χώρο. Αν π.χ. υπάρχουν m ζεύγη προτύπων/στόχων (X^p, D^p) , με $X^p = [x_1^p, x_2^p, \dots, x_n^p]$ και $D^p = [d_1^p, x_2^p, \dots, x_k^p]$, με $p = 1, 2, \dots, m$, τα διανύσματα (X^p, D^p) αποτελούν τις εισόδους του δικτύου, οι οποίες παράγουν τις αντίστοιχες εξόδους $V(W, X^p) = [v_1^p, v_2^p, \dots, v_k^p]$, όπου $V(W, X^p)$ είναι η έξοδος του δικτύου και W το διάνυσμα των βαρών, το οποίο ελαχιστοποιεί το *Μέσο Τετραγωνικό Σφάλμα*

$$E(w) = \sum_{p=1}^m E^{(p)}(w) = \frac{1}{2} \sum_{p=1}^m (D^{(p)} - V(W, X^{(p)}))^2 \quad (4.49)$$

4.5.2.8 Ο Κανόνας του Δέλτα για μη Γραμμικές Συναρτήσεις Ενεργοποίησης σε Δίκτυα Μονής Κατεύθυνσης

Έχουμε δείξει πως ο τυπικός Κανόνας του Δέλτα ουσιαστικά υλοποιεί την κλιμακωτή μείωση στο άθροισμα των τετραγώνων του σφάλματος για γραμμικές συναρτήσεις ενεργοποίησης. Σ' αυτήν την περίπτωση, χωρίς κρυφούς νευρώνες, έχει το σχήμα καμπύλης με μόνο ένα ελάχιστο, έτσι ώστε η κλιμακωτή μείωση εγγυάται να βρει το καλύτερο σύνολο βαρών. Με τους κρυφούς νευρώνες, ωστόσο, δεν είναι τόσο φανερός ο τρόπος υπολογισμού της παραγώγου, και η επιφάνεια του σφάλματος δεν είναι κοίλη προς τα πάνω, οπότε υπάρχει ο κίνδυνος να κολλήσουμε σε τοπικό ελάχιστο. Υπάρχει ένας αποτελεσματικός τρόπος υπολογισμού της παραγώγου.

Αν υποθεθεί ότι το δίκτυο αποτελείται απ' το στρώμα εισόδου με $n = N_0$ νευρώνες, $M-1$ κρυφά στρώματα με N_l , $l=1,2,\dots,M-1$ και το στρώμα εξόδου με $k = N_M$ νευρώνες, με την είσοδο του προτύπου $X^p = [x_1^p, x_2^p, \dots, x_n^p]$, $p=1,2,\dots,m$, στο στρώμα εισόδου, οι νευρώνες του κάθε στρώματος υπολογίζουν την έξοδό τους $V_l^{(p)} = [v_{1,l}^{(p)}, v_{2,l}^{(p)}, \dots, v_{N_l,l}^{(p)}]$, $l=1,2,\dots,M-1$, την οποία προωθούν στο επόμενο στρώμα. Η έξοδος του νευρώνα j του στρώματος εξόδου του δικτύου θα είναι

$$v_{l,j}^{(p)} = f(\text{net}_{l,j}^{(p)}) = f\left(\sum_{i=1}^{l-1} w_{l,j,i} v_{l-1,i}^{(p)}\right) \quad (4.50)$$

όπου $v_j^{(p)} = x_j^{(p)}$, αν ο νευρώνας j είναι νευρώνας εισόδου και $f()$ μια μη γραμμική συνάρτηση ενεργοποίησης διαφορίσιμη και μη φθίνουσα, όπως οι (4.3), (4.4). Για να πάρουμε τη σωστή γενίκευση του κανόνα του δέλτα πρέπει να θέσουμε

$$\Delta_p w_{ji} = -\frac{\partial E^p(w)}{\partial w_{l,j,i}} \quad (4.51)$$

όπου E είναι η ίδια συνάρτηση των αθροισμάτων των τετραγώνων του σφάλματος που ορίστηκε προηγούμενα. Όπως και στον κανόνα του δέλτα μπορούμε να δούμε αυτήν την παράγωγο σαν το γινόμενο δύο παραγόντων, της μεταβολής του σφάλματος ως συνάρτηση της μεταβολής της τελικής εισόδου στο νευρώνα και το αποτέλεσμα της μεταβολής ενός συγκεκριμένου βάρους πάνω στην τελική είσοδο. Έτσι, μπορούμε να γράψουμε

$$\frac{\partial E^p(w)}{\partial w_{M,j,i}} = \frac{\partial E^p(w)}{\partial \text{net}_{M,j}^{(p)}} \frac{\partial \text{net}_{M,j}^{(p)}}{\partial w_{M,j,i}} = \frac{\partial E^p(w)}{\partial v_{M,j}^{(p)}} \frac{dv_{M,j}^{(p)}}{d\text{net}_{M,j}^{(p)}} \frac{\partial \text{net}_{M,j}^{(p)}}{\partial w_{M,j,i}} = -(d_j^{(p)} - v_{M,j}^{(p)}) f^{\circ}(\text{net}_{M,j}^{(p)}) v_{M-1,j}^{(p)}$$

Αν ορίσουμε σαν $\delta_{M,j}^{(p)}$, $j=1,2,\dots,N_M$ το

$$\delta_{M,j}^{(p)} = (d_j^{(p)} - v_{M,j}^{(p)}) f^{\circ}(\text{net}_{M,j}^{(p)}) \quad (4.52)$$

θα έχουμε

$$\frac{\partial \mathcal{E}^p(w)}{\partial w_{M,j,i}} = -\delta_{M,j}^{(p)} v_{M-1,j}^{(p)} \quad (4.53)$$

Το τελικό σφάλμα στο στρώμα εξόδου επηρεάζεται από τους νευρώνες των ενδιάμεσων στρωμάτων l μόνο μέσω του επόμενου στρώματος $l+1$. Έτσι, για τα ενδιάμεσα στρώματα θα έχουμε

$$\frac{\partial \mathcal{E}^p(w)}{\partial w_{l,j,i}} = \frac{\partial \mathcal{E}^p(w)}{\partial v_{l,j}^{(p)}} \frac{dv_{l,j}^{(p)}}{dnet_{l,j}^{(p)}} \frac{\partial net_{l,j}^{(p)}}{\partial w_{l,j,i}} \quad (4.54)$$

Αλλά

$$\frac{\partial \mathcal{E}^p(w)}{\partial v_{l,j}^{(p)}} = \sum_{j=1}^{N_{l+1}} \frac{\partial \mathcal{E}^p(w)}{\partial v_{l+1,j}^{(p)}} \frac{dv_{l+1,j}^{(p)}}{dnet_{l+1,j}^{(p)}} \frac{\partial net_{l+1,j}^{(p)}}{\partial v_{l,j}^{(p)}} = \sum_{j=1}^{N_{l+1}} \delta_{l+1,j}^{(p)} w_{l+1,j,i}$$

οπότε θα έχουμε

$$\frac{\partial \mathcal{E}^p(w)}{\partial w_{l,j,i}} = \sum_{j=1}^{N_{l+1}} -\delta_{l+1,j}^{(p)} w_{l+1,j,i} f'(net_{l,j}^{(p)}) v_{l,j}^{(p)} \quad (4.55)$$

Αν ορίσουμε σαν $\delta_{l,j}^{(p)}$, $j=1,2,\dots,N_l$, $l=M-1,\dots,1$ το

$$\delta_{l,j}^{(p)} = f'(net_{l,j}^{(p)}) \sum_{j=1}^{N_{l+1}} \delta_{l+1,j}^{(p)} w_{l+1,j,i} \quad (4.56)$$

θα έχουμε

$$\frac{\partial \mathcal{E}^p(w)}{\partial w_{l,j,i}} = -\delta_{l,j}^{(p)} v_{l-1,j}^{(p)} \quad (4.57)$$

Η προηγούμενη διαδικασία συνεχίζεται μέχρι το πρώτο κρυφό στρώμα και το σφάλμα κάθε στρώματος προωθείται στο προηγούμενο στρώματος, εξ ου και το όνομα *προώθηση προς τα πίσω* (*back-propagation*). Αφού βρεθούν όλα τα σφάλματα, αρχίζοντας απ' το τελευταίο στρώμα γίνεται η διόρθωση των συνάψεων σύμφωνα με τον τύπο

$$w_{l,j,i}^{(t+1)} = w_{l,j,i}^{(t)} - \rho \frac{\partial \mathcal{E}^p(w^{(t)})}{\partial w_{l,j,i}^{(t)}} = w_{l,j,i}^{(t)} - \rho \delta_{l,j}^{(p)} v_{l-1,j}^{(p)} \quad (4.58)$$

όπου ρ ο συντελεστής εκπαίδευσης, ο οποίος μπορεί να είναι σταθερός ή μεταβλητός. Ο αλγόριθμος που περιγράφηκε αποκαλείται *on line back-propagation*, γιατί η

τροποποίηση των συνάψεων γίνεται αμέσως μετά την εισαγωγή του κάθε προτύπου, ενώ στόχος είναι η ελαχιστοποίηση του μερικού σφάλματος του κάθε προτύπου $E^p(w)$. Αν θέλουμε να ελαχιστοποιήσουμε το συνολικό σφάλμα $E(w)$, όπως ορίστηκε στην εξίσωση (4.32) προκύπτει η μέθοδος *batch back-propagation*, και οι σχέσεις (4.55)-(4.57) γίνονται

$$\sum_{p=1}^m \frac{\partial E^p(w)}{\partial w_{M,j i}} = \sum_{p=1}^m -(d_j^{(p)} - v_{M,j}^{(p)}) f^{\odot}(net_j^{(p)}) v_{M-1,j}^{(p)} \quad (4.59)$$

και

$$\frac{\partial E(w)}{\partial w_{M,j i}} = \sum_{p=1}^m -\delta_{M,j}^{(p)} v_{M-1,j}^{(p)} \quad (4.60)$$

για το στρώμα εξόδου

$$\sum_{p=1}^m \frac{\partial E^p(w)}{\partial w_{l,j i}} = f'(net_{l,j}^{(p)}) \sum_{j=l+1}^{N_{l+1}} -\delta_{l+1,j}^{(p)} w_{l+1,j i} \quad (4.61)$$

για τα υπόλοιπα στρώματα, ενώ η διόρθωση των συνάψεων θα γίνει σύμφωνα με τον τύπο

$$w_{l,j i} = w_{l,j i} - \rho \frac{\partial E(w)}{\partial w_{l,j i}} = w_{l,j i} - \rho \sum_{p=1}^m \delta_{l,j}^{(p)} v_{l-1,j}^{(p)} \quad (4.62)$$

4.5.2.9 Συντελεστής Εκπαίδευσης και Ορμή (Momentum)

Η σύγκλιση των δυο προηγούμενων μεθόδων έχει σχέση με το συντελεστή εκπαίδευσης ρ . Αν το ρ είναι μικρό, η σύγκλιση είναι πολύ αργή, εξ αιτίας του μεγάλου αριθμού των βημάτων που απαιτούνται για την εύρεση ενός τοπικού ελαχίστου. Αν το ρ είναι μεγάλο, η σύγκλιση στην αρχή είναι πολύ γρήγορη, αλλά στη συνέχεια ο αλγόριθμος παλινδρομεί, χωρίς να συγκλίνει σε κάποιο τοπικό ελάχιστο. Γενικά, το καλύτερο είναι να έχουμε μεγάλα βήματα, όταν το σημείο αναζήτησης είναι μακριά απ' το ελάχιστο και μικρότερα βήματα, όταν το σημείο αναζήτησης πλησιάζει το ελάχιστο. Μια τέτοια μέθοδος χρησιμοποιείται απ' τον Luo [LuZQ91], ο οποίος χρησιμοποιεί σαν συντελεστή εκπαίδευσης στον αλγόριθμο LMS σε κάθε επανάληψη το $\rho(t) = \frac{I}{t}$, όπου t = αριθμός επανάληψης. Η παραπάνω μέθοδος είναι πολύ αργή. Οι Darken and Moody [DaMo91] προτείνουν τη μέθοδο *search then converge*, η οποία χρησιμοποιεί σαν συντελεστή εκπαίδευσης σε κάθε επανάληψη $\rho(t) = \rho(0)/(1 + \frac{t}{\tau})$, t = αριθμός επανάληψης. Έτσι, το $\rho(t)$ έχει μεγάλες τιμές για το χρόνο τ , στον οποίο οι συνάψεις θα προσεγγίσουν το ελάχιστο και μικρές για $t \gg \tau$.

Ο βέλτιστος συντελεστής εκπαίδευσης για γρήγορη σύγκλιση είναι το αντίστροφο της μεγαλύτερης ιδιοτιμής λ_{max} της (*Hessian*) μήτρας H της αντικειμενικής συνάρτησης $E(x)$, η οποία υπολογίζεται στο σημείο x . Ο υπολογισμός όμως της *Hessian* μήτρας και της μεγαλύτερης της ιδιοτιμής απαιτεί πολλές πράξεις και μνήμη. Μπορούμε όμως να προσεγγίσουμε την τιμή της λ_{max} [LCSP93] με μια μέθοδο που βασίζεται στην προσέγγιση του γινομένου της μήτρας H επί ένα τυχαίο διάνυσμα z μέσω του αναπτύγματος Taylor: $H_z = (1/\alpha)[\nabla E(x + \alpha z) - \nabla E(x)]$, όπου α μικρή θετική σταθερά. Χρησιμοποιώντας τη μέθοδο της δυνάμεως για τον υπολογισμό ιδιοτιμών $z \leftarrow \frac{Hz}{\|z\|} = \frac{1}{\alpha} \left[\nabla E(w + \alpha \frac{z}{\|z\|}) - \nabla E(w) \right]$ το διάνυσμα z συγκλίνει στο $|\lambda_{max}|/c_{max}$, όπου c_{max} το κανονικοποιημένο ιδιοδιάνυσμα του H που αντιστοιχεί στο λ_{max} . Έτσι, η norm του διανύσματος z δίνει μια καλή προσέγγιση του $|\lambda_{max}|$ και ο αντίστροφός του μπορεί να χρησιμοποιηθεί σαν συντελεστής εκπαίδευσης. Ένας τέτοιος on-line αλγόριθμος παρουσιάζεται απ' τον Le Cun (LCSP93).

Ένας άλλος τρόπος επιτάχυνσης είναι η πρόσθεση μιας ορμής, ενός επί πλέον όρου σε κάθε w_i , το οποίο επιταχύνει την κάθοδο στο τοπικό ελάχιστο, αντί να παλινδρομεί με κάθε μεταβολή του προσήμου της παραγώγου $\frac{\partial E(w^{(t)})}{\partial w_i}$. Η πρόσθεση της ορμής έχει σαν αποτέλεσμα ο τύπος (4.58) να γίνει

$$w_{l,j i}^{(t+1)} = w_{l,j i}^{(t)} - \rho \frac{\partial E^p(w^{(t)})}{\partial w_{l,j i}^{(t)}} + \alpha \Delta w_{l,j i}^{(t)} = w_{l,j i}^{(t)} - \rho \delta_{l,j}^{(p)} v_{l-1,j}^{(p)} + \alpha \Delta w_{l,j i}^{(t)} \quad (4.63)$$

όπου α είναι ο συντελεστής ορμής με τιμή μεταξύ 0 και 1, $\Delta w_{l,j i}^{(t)} = w_{l,j i}^{(t)} - w_{l,j i}^{(t-1)}$, ενώ για την περίπτωση της ομαδικής διόρθωσης θα είναι

$$w_{l,j i}^{(t+1)} = w_{l,j i}^{(t)} - \rho \nabla \mathcal{E}(w^{(t)}) + \alpha \Delta w_{l,j i}^{(t)} \quad (4.64)$$

Ο συντελεστής ορμής είναι ένας τρόπος να αυξήσουμε το συντελεστή εκπαίδευσης σε επίπεδες περιοχές της επιφάνειας του σφάλματος και να διατηρούμε το συντελεστή εκπαίδευσης κοντά στο ρ ($0 < \rho < 1$) σε περιοχές της επιφάνειας του σφάλματος με παλινδρομήσεις. Αν το σημείο αναζήτησης βρίσκεται σε επίπεδη περιοχή, η μερική παράγωγος $\frac{\partial E(w)}{\partial w_i}$ θα παραμείνει ίδια σε κάθε επανάληψη, οπότε η ορμή αυξάνει το συντελεστή εκπαίδευσης κατά $1/(1-\alpha)$, ενώ για περιοχές με παλινδρομήσεις ο παράγοντας της ορμής δεν έχει αποτέλεσμα στο συντελεστή εκπαίδευσης.

Πολλοί αλγόριθμοι έχουν προταθεί για την αυτόματη προσαρμογή του συντελεστή εκπαίδευσης $\rho(t)$. Οι Chan and Fallside [ChFa87] προτείνουν έναν κανόνα τροποποίησης του συντελεστή εκπαίδευσης ανάλογα με το συνημίτονο της γωνίας μεταξύ των παραγώγων $\nabla E(x^{(t)})$ και $\nabla E(x^{(t-1)})$, όπου t ο αριθμός επανάληψης. Ο Sutton [Sut86] παρουσιάζει μια μέθοδο, η οποία αυξάνει ή ελαττώνει κάθε $\rho_i(t)$ για κάθε σύναψη w_i , σύμφωνα με τις φορές που αλλάζει πρόσημο η μερική παράγωγος

$\frac{\partial E(w)}{\partial w_i}$. Ο Franzini [Fran87] προτείνει μια μέθοδο που αυξάνει τον $\rho(t)$ κάθε φορά που το $\nabla E(x^{(t)})$ είναι κοντά στο $\nabla E(x^{(t-1)})$, διαφορετικά τον ελαττώνει. Οι Silva and Almeida [SiAl90] προτείνουν μια μέθοδο όπου για κάθε σύναψη w_i ο συντελεστής εκπαίδευσης $\rho_i(t)$ παίρνει την τιμή $\alpha \rho_i(t)$, αν τα $\frac{\partial E(w)^{(t)}}{\partial w_i}$ και $\frac{\partial E(w)^{(t-1)}}{\partial w_i}$ έχουν το ίδιο πρόσημο, με $\alpha > 1$. Αν οι μερικές παράγωγοι έχουν διαφορετικά πρόσημα, χρησιμοποιείται ο συντελεστής εκπαίδευσης $\beta \rho_i(t)$, όπου $0 < \beta < 1$. Μια παρόμοια μέθοδος [Pfl90] είναι να θέσουμε $\rho(t) = \rho(t-1)$, αν τα $\nabla E(x^{(t)})$ και $\nabla E(x^{(t-1)})$ έχουν το ίδιο πρόσημο, και $\rho(t) = \rho(t-1)/2$, αν τα $\nabla E(x^{(t)})$ και $\nabla E(x^{(t-1)})$ έχουν διαφορετικά πρόσημα. Ο αλγόριθμος SuperSAB [Toll90] αποτελεί μια τροποποίηση των προηγούμενων μεθόδων, όπου οι τροποποιήσεις των συνάψεων γίνονται δεκτές, μόνο αν τα $\nabla E(x^{(t)})$ και $\nabla E(x^{(t-1)})$ έχουν το ίδιο πρόσημο, διαφορετικά ακυρώνονται. Ο Fahlman [Fahl88] πρότεινε έναν *ευρετικό* (*heuristic*) αλγόριθμο back-propagation, τον Quickprop, ο οποίος χρησιμοποιεί μεταβλητή ορμή, έτσι ώστε, αν η παρούσα κλίση (slope) είναι μικρότερη απ' την προηγούμενη, αλλά έχει το ίδιο πρόσημο, η ορμή να είναι θετική και η σύγκλιση να επιταχύνεται, ενώ αν η παρούσα κλίση έχει διαφορετικό πρόσημο απ' την προηγούμενη, δηλαδή το διάνυσμα των βαρών φτάνει σε ελάχιστο, η ορμή να είναι αρνητική και η σύγκλιση επιβραδύνεται. Για τον υπολογισμό της ορμής έχει χρησιμοποιηθεί και η μέθοδος Polak-Ribiere [KrSV89], όπου η ορμή είναι ο συντελεστής $\beta(t)$, όπως ορίστηκε στην (3.47). Μια τεχνική Scaled Conjugate Gradient, όπου ο συντελεστής εκπαίδευσης προσαρμόζεται ανάλογα με τη μείωση του σφάλματος παρουσιάστηκε απ' το Moller [Moll93].

4.5.2.10 Η Μέθοδος RPROP

Οι Riedmiller και Braun [RiBr92] χρησιμοποιούν έναν αλγόριθμο, όπου η κάθε σύναψη w_{ij} τροποποιείται ξεχωριστά, με τη χρήση των συντελεστών Δ_{ij} , οι οποίοι παίρνουν κάποιες μικρές αρχικές τιμές και στη συνέχεια τροποποιούνται σύμφωνα με τον τύπο

$$\Delta_{ij}^{(t+1)} = \begin{cases} \Delta_{ij}^{(t)} * u, & \alpha \nu \frac{\partial E(w^{(t)})}{\partial w_{ij}} * \frac{\partial E(w^{(t-1)})}{\partial w_{ij}} > 0 \\ \Delta_{ij}^{(t)} * d, & \alpha \nu \frac{\partial E(w^{(t)})}{\partial w_{ij}} * \frac{\partial E(w^{(t-1)})}{\partial w_{ij}} < 0 \\ \Delta_{max}, & \alpha \nu \Delta_{ij}^{(t)} > \Delta_{max} \\ \Delta_{min}, & \alpha \nu \Delta_{ij}^{(t)} \leq \Delta_{min} \end{cases} \quad (4.65)$$

όπου $\Delta_{max} = 50$, $\Delta_{min} = 0.000001$, $u = 1.2$ και $d = 0.5$. Στη συνέχεια, αν ισχύει $\frac{\partial E(w^{(t)})}{\partial w_{ij}} * \frac{\partial E(w^{(t-1)})}{\partial w_{ij}} \geq 0$, τροποποιούνται οι συνάψεις σύμφωνα με τον τύπο

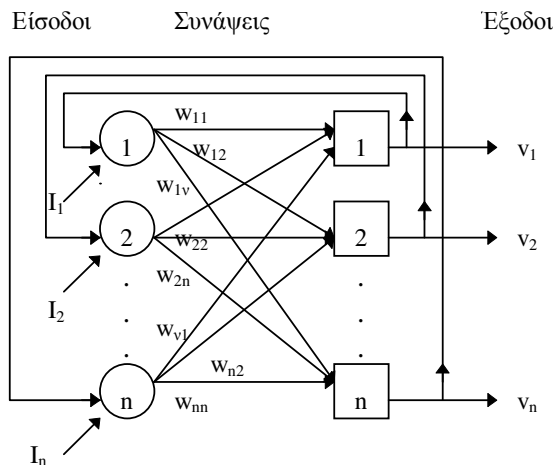
$$\Delta w_{ij}^{(t+1)} = \begin{cases} -\Delta_j^{(t)} & \alpha \nu \frac{\partial \mathcal{E}(w^{(t)})}{\partial w_{ij}} > 0 \\ +\Delta_j^{(t)} & \alpha \nu \frac{\partial \mathcal{E}(w^{(t)})}{\partial w_{ij}} < 0 \end{cases} \quad (4.66)$$

ενώ, αν ισχύει $\frac{\partial \mathcal{E}(w^{(t)})}{\partial w_{ij}} * \frac{\partial \mathcal{E}(w^{(t-1)})}{\partial w_{ij}} < 0$, οι συνάψεις παίρνουν τις παλιές τιμές, δηλαδή

$$w_{ij}^{(t+1)} = w_{ij}^{(t-1)} \quad (4.67)$$

4.5.3 Αναδρομικά Δίκτυα (Hopfield)

Μία από τις μεγαλύτερες συνεισφορές στην περιοχή των νευρωνικών δικτύων έγινε στις αρχές της δεκαετίας του 1980 από τον John Hopfield [Hopf82], ο οποίος μελέτησε ένα αυτο-συνδεδεμένο δίκτυο, το οποίο συσχέτισε με άλλα φυσικά συστήματα, αναπτύσσοντας τη χρήση μιας *Συνάρτησης Ενέργειας (Liarunon Function)*. Το δίκτυο Hopfield αποτελείται από έναν αριθμό νευρώνων, ο καθένας απ' τους οποίους συνδέεται με κάθε άλλο νευρώνα: είναι ένα πλήρως συνδεδεμένο δίκτυο, όπως φαίνεται στο Σχήμα 4.4. Είναι επίσης ένα συμμετρικά σταθμισμένο δίκτυο, αφού τα βάρη στη σύναψη από τον έναν νευρώνα στον άλλο είναι τα ίδια και προς τις δύο κατευθύνσεις. Κάθε νευρώνας έχει ένα κατώφλι, όπως το perceptron ενός στρώματος και μία συνάρτηση ενεργοποίησης. Οι νευρώνες υπολογίζουν το σταθμισμένο άθροισμα των εισόδων τους μείον την τιμή του κατωφλιού και περνούν το αποτέλεσμα στη συνάρτηση για να καθορίσουν την κατάσταση εξόδου τους. Αυτό που πραγματικά διακρίνει το δίκτυο Hopfield από τα άλλα δίκτυα είναι ο τρόπος με τον οποίο παράγει μία λύση.



Σχήμα 4.4 Το δίκτυο Hopfield

Το Σχήμα 4.4 δείχνει πώς είναι ένα πλήρως συνδεδεμένο δίκτυο. Δεν υπάρχουν προφανείς συνάψεις εισόδου ή εξόδου. Κάθε νευρώνας είναι ίδιος με οποιοδήποτε άλλο.

Αυτό είναι το σημαντικότερο χαρακτηριστικό του δικτύου Hopfield και αυτή η διαφορά στην αρχιτεκτονική σημαίνει ότι το δίκτυο λειτουργεί με διαφορετικό τρόπο. Οι είσοδοι στο δίκτυο εφαρμόζονται σε όλους τους νευρώνες ταυτόχρονα και αποτελούνται από ένα σύνολο αρχικών τιμών. Το δίκτυο τότε αφήνεται μόνο του και εκτελεί ανακυκλώσεις διαμέσου διαδοχικών καταστάσεων, μέχρι να συγκλίνει σε μια σταθερή λύση, στην οποία φτάνει όταν οι τιμές των νευρώνων δεν αλλάζουν. Η τελική έξοδος του δικτύου είναι οι τιμές όλων των νευρώνων, όταν το δίκτυο έχει φτάσει σε μια σταθερή κατάσταση. Ένας απλός τρόπος να απεικονίσουμε το σύστημα είναι να θεωρήσουμε ότι επειδή κάθε νευρώνας συνδέεται με κάθε άλλο νευρώνα, η τιμή του κάθε νευρώνα επηρεάζει την τιμή όλων των άλλων. Η αρχική κατάσταση παριστάνει πολλές διαφορετικές τιμές που η κάθε μία προσπαθεί να επηρεάσει τις άλλες. Αυτό μπορεί να δημιουργήσει αστάθεια, επειδή μία τιμή μπορεί να προσπαθεί να ενεργοποιήσει άλλους νευρώνες, ενώ μία άλλη προσπαθεί να τους απενεργοποιήσει. Καθώς το δίκτυο μεταβαίνει διαμέσου διαδοχικών καταστάσεων, προσπαθεί να φτάσει σε ένα συμβιβασμό μεταξύ όλων των τιμών του δικτύου, και η τελική κατάσταση παριστάνει τη λύση του "καλύτερου συμβιβασμού" που μπορεί να βρει το δίκτυο. Σ' αυτή τη κατάσταση, υπάρχουν τόσες είσοδοι που προσπαθούν να ενεργοποιήσουν μία μονάδα όσες είναι αυτές που προσπαθούν να την απενεργοποιήσουν. Έτσι, η μονάδα παραμένει στην ίδια κατάσταση.

Η λειτουργία του δικτύου είναι ριζικά διαφορετική από αυτή ενός συστήματος perceptron, στο οποίο οι είσοδοι εφαρμόζονται και το δίκτυο παράγει μία έξοδο που παριστάνει τη λύση. Στο δίκτυο Hopfield, αυτή η πρώτη έξοδος λαμβάνεται σαν νέα είσοδος, η οποία παράγει μία νέα έξοδο, κ.ο.κ.. Φτάνουμε στη λύση όταν δεν υπάρχει καμία αλλαγή από τη μία ανακύκλωση στην άλλη. Η διαδικασία της εκμάθησης είναι επίσης διαφορετική.

4.5.3.1 Ο Αλγόριθμος του Δικτύου Hopfield

Ο αλγόριθμος που διέπει τη λειτουργία του δικτύου Hopfield παρουσιάζεται παρακάτω.

1. Βρίσκεται η κατάλληλη συνάρτηση ενέργειας E (Liarunov) του δικτύου.
2. Δίνονται οι τιμές των συνάψεων του δικτύου σύμφωνα με τη συνάρτηση ενέργειας E .
3. Εισάγεται ένα άγνωστο πρότυπο για ανάκληση στο δίκτυο.
4. Οι έξοδοι v_i , $i=1,2,\dots,n$ των νευρώνων του δικτύου δίνονται απ' τις σχέσεις - εξισώσεις κινήσεως

$$a) v_i(t+1) = f_h \left[\sum_{j=1}^n w_{ij} v_j(t) + I_i \right], \text{ για δυαδικές-διπολικές εξόδους}$$

(4.68)

$$b) C \frac{du_i}{dt} = f \left[\sum_{j=1}^n w_{ij} v_j(t) + I_i - \frac{u_i}{R} \right], \text{ για συνεχείς εξόδους}$$

όπου η συνάρτηση f_h είναι μη γραμμική και βηματική, I_i η εξωτερική διέγερση για το νευρώνα i , f η σιγμοειδής συνάρτηση και C, R παράμετροι.

5. Οι έξοδοι v_i , $i=1,2,\dots,n$ εισάγονται ξανά στο δίκτυο, μέχρι τη σύγκλιση (Οι έξοδοι να παραμένουν αμετάβλητες).

Όταν το δίκτυο λειτουργεί σαν συνειρμική μνήμη, τα βάρη μεταξύ των νευρώνων είναι τα πρότυπα για όλες τις κλάσεις. Αυτό είναι το εκπαιδευτικό στάδιο του αλγορίθμου και συνδέει κάθε πρότυπο με τον εαυτό του. Το στάδιο αναγνώρισης επέρχεται όταν η έξοδος του δικτύου εξαναγκάζεται να ταιριάζει με αυτήν ενός επιβαλλόμενου άγνωστου προτύπου στο χρόνο μηδέν. Τότε το δίκτυο αφήνεται να εκτελέσει επαναλήψεις ελεύθερα σε διακριτά χρονικά βήματα, μέχρι να φτάσει σε μια σταθερή κατάσταση όταν η έξοδος παραμένει αμετάβλητη. Το δίκτυο έτσι συγκλίνει στη λύση. Η αυτο-σύναψη των προτύπων σημαίνει ότι η παρουσία ενός εκφυλισμένου προτύπου εισόδου θα έχει σαν αποτέλεσμα την αναπαραγωγή του τέλει προτύπου σαν έξοδο.

4.5.3.2 Η Επιφάνεια της Ενέργειας

Μπορούμε να κατανοήσουμε καλύτερα το δίκτυο Hopfield με τη βοήθεια της επιφάνειας της ενέργειας που χρησιμοποιείται τελευταία πάρα πολύ. Παρουσιάζει επιτυχία στην περιγραφή των δικτύων perceptron, επειδή παρέχει μία οπτική αναλογία που μας επιτρέπει να σχηματίσουμε μία άποψη για το τι συμβαίνει. Το ίδιο ισχύει για το δίκτυο Hopfield. Η επιφάνεια της ενέργειας έχει κοιλότητες που παριστάνουν τα πρότυπα που αποθηκεύονται στο δίκτυο. Ένα άγνωστο πρότυπο εισόδου παριστάνει ένα ιδιαίτερο σημείο στην επιφάνεια της ενέργειας. Καθώς το δίκτυο εκτελεί επαναλήψεις κατευθυνόμενο προς μία λύση, το σημείο μετακινείται διαμέσου της επιφάνειας προς μία από τις κοιλότητες. Αυτές οι λεκάνες έλξης παριστάνουν τις σταθερές καταστάσεις του δικτύου. Η λύση επέρχεται όταν το σημείο μετακινείται μέσα στη χαμηλότερη περιοχή της λεκάνης. Οποιοδήποτε άλλο σημείο στην κοντινή γειτονιά βρίσκεται σε μεγαλύτερο ύψος και έτσι το αρχικό σημείο θα μείνει εκεί που είναι. Αυτό είναι ανάλογο με την τρισδιάστατη περίπτωση όπου μία μπάλα που τοποθετείται σε ένα τοπίο με κοιλάδες και λοφοπλαγιές θα μετακινηθεί προς τα κάτω προς την κοντινότερη κοιλάδα και θα καταλήξει σε μία σταθερή κατάσταση που δεν αλλάζει με το χρόνο όταν φτάσει στο χαμηλότερο σημείο.

Μπορούμε να προσδιορίσουμε μία κατάλληλη συνάρτηση ενέργειας για το δίκτυο Hopfield όπως

$$E = -\frac{1}{2} \sum_i \sum_{j \neq i} w_{ij} v_i v_j - \sum_i I_i v_i + \sum_i \theta_i v_i \quad (4.69)$$

όπου w_{ij} παριστάνει το βάρος μεταξύ του νευρώνα i και του νευρώνα j του δικτύου, v_i παριστάνει την έξοδο του νευρώνα i , θ_i την τιμή του κατωφλιού του νευρώνα i και με I_i την εξωτερική διέγερση του νευρώνα i . Καθώς η έξοδος τροφοδοτείται ξανά στο

δίκτυο, οι έξοδοι σε οποιοδήποτε χρόνο παριστάνουν το επόμενο σύνολο εισόδων. Τα βάρη στο δίκτυο περιέχουν την πληροφορία των προτύπων και έτσι όλα τα πρότυπα περιλαμβάνονται στη συνάρτηση ενέργειας. Οι νευρώνες δε συνδέονται άμεσα με τον εαυτό τους και έτσι οι όροι w_{ii} είναι μηδέν. Επειδή οι συνάψεις είναι συμμετρικές, $w_{ij} = w_{ji}$.

Αν αναγκάσουμε τα πρότυπά μας να καταλαμβάνουν τα χαμηλά σημεία στη συνάρτηση της ενέργειας, τότε μπορούμε να εκτελέσουμε κεκλιμένη κατάβαση στη συνάρτηση ενέργειας για να τερματίσουμε σε ένα από αυτά τα ελάχιστα, το οποίο θα παριστάνει τη λύση μας.

Υπάρχουν δύο ελαφρά διαφορετικές μέθοδοι για την πραγματική εκτέλεση της ενημέρωσης, οι οποίες παράγουν ελαφρά διαφορετική συμπεριφορά στο δίκτυο. Η ενημέρωση μπορεί να γίνει σε όλους τους νευρώνες ταυτόχρονα, όπου οι τιμές στο δίκτυο είναι προσωρινά παγωμένες και όλοι οι νευρώνες υπολογίζουν την επόμενη τους κατάσταση. Αυτή η νέα κατάσταση παριστάνει μία ενημέρωση κατά μήκος ολόκληρου του δικτύου, και μετά υπολογίζεται η επόμενη κατάσταση. Αυτή η λειτουργία είναι γνωστή σαν *σύγχρονη* ενημέρωση. Η εναλλακτική προσέγγιση, που καλείται *ασύγχρονη* ενημέρωση, συμβαίνει όταν ένας νευρώνας επιλέγεται τυχαία και ενημερώνει την έξοδό του σύμφωνα με την είσοδο που δέχεται. Μετά αυτή η διαδικασία επαναλαμβάνεται. Η κύρια διαφορά μεταξύ των μεθόδων είναι ότι στην περίπτωση της ασύγχρονης ενημέρωσης, η αλλαγή στην έξοδο ενός νευρώνα επηρεάζει την κατάσταση του συστήματος και έτσι μπορεί να επηρεάσει την αλλαγή του επόμενου νευρώνα. Αυτό σημαίνει ότι η σειρά με την οποία ενημερώνονται οι νευρώνες επηρεάζει τη συμπεριφορά του δικτύου σε κάποιο βαθμό. Τα αποτελέσματα είναι φανερά στο στάδιο της ανάκλησης, επειδή η τυχαία φύση της επιλογής του επόμενου νευρώνα που θα ενημερωθεί αλλάζει την ακολουθία των προτύπων που το δίκτυο αναπτύσσει. Με τη σύγχρονη ενημέρωση, όλοι οι νευρώνες ενημερώνονται μαζί και έτσι τα ενδιάμεσα πρότυπα δεν αλλάζουν. Η ασύγχρονη ενημέρωση προσθέτει κάποια αβεβαιότητα ή μη προκαθορισμό στο μονοπάτι που θα ακολουθηθεί από την είσοδο μέχρι την τελική σταθερή κατάσταση. Ωστόσο και οι δύο μέθοδοι έχουν τα ίδια γενικά χαρακτηριστικά του δικτύου και η χρήση της σύγχρονης ή της ασύγχρονης ενημέρωσης είναι σπάνια ένας σημαντικός παράγοντας.

4.5.3.3 Αποθήκευση Προτύπων - Συνειρμική Μνήμη

Για να αποθηκεύσουμε ένα πρότυπο, χρειάζεται να ελαχιστοποιήσουμε την τιμή της συνάρτησης ενέργειας γι' αυτό το συγκεκριμένο πρότυπο έτσι ώστε να καταλαμβάνει ένα σημείο ελαχίστου στην επιφάνεια της ενέργειας. Ωστόσο, θέλουμε επίσης να αφήσουμε τα προηγούμενως αποθηκευμένα πρότυπα στις κοιλότητες τους, έτσι ώστε η πρόσθεση νέων προτύπων να μην καταστρέφει όλη την προηγούμενη πληροφορία. Ο πίνακας των βαρών περιέχει την πληροφορία για τα αποθηκευμένα πρότυπα. Έτσι, θέλουμε να βρούμε μια έκφραση για τις τιμές των βαρών που θα παράγει ένα ελάχιστο στη συνάρτηση ενέργειας. Θεωρώντας τη συνάρτηση ενέργειας θέλουμε να

ελαχιστοποιήσουμε το $E = -\frac{1}{2} \sum_i \sum_{j \neq i} w_{ij} v_i v_j - \sum_i I_i v_i + \sum_i \theta_i v_i$ για ένα ιδιαίτερο πρότυπο s .

Το δίκτυο Hopfield δεν έχει κανέναν επαναληπτικό αλγόριθμο εκμάθησης. Τα πρότυπα απλά αποθηκεύονται μειώνοντας τις ενέργειές τους. Το δίκτυο δεν έχει κρυφές μονάδες, και έτσι δεν μπορεί να κωδικοποιήσει τα δεδομένα.

4.5.3.4 Ανάκληση Διπολικών Προτύπων

Μετά την αποθήκευση των προτύπων στο δίκτυο, η ανάκληση ενός προτύπου επιτυγχάνεται με κεκλιμένη κατάβαση στη συνάρτηση ενέργειας, και γι' αυτό χρειαζόμαστε μία μέθοδο. Θεωρώντας τη συνάρτηση ενέργειας στην (4.69), χρειάζεται να υπολογίσουμε τη συμβολή στην ενέργεια της τιμής ενός συγκεκριμένου νευρώνα, και μετά μπορούμε να το επαναλάβουμε για όλο το δίκτυο, μειώνοντας τη συμβολή του κάθε νευρώνα μέχρι η τιμή της ενέργειας να φτάσει σε ένα ελάχιστο. Μετά την ενεργοποίηση του νευρώνα k , η συνάρτηση ενέργειας γίνεται

$$E = -\frac{1}{2} \sum_{i \neq k} \sum_{j \neq k} w_{ij} v_i v_j + \sum_{i \neq k} \theta_i v_i - \frac{1}{2} v_k \sum_j w_{kj} v_j - \frac{1}{2} v_k \sum_i w_{ik} v_i + \theta_k v_k \quad (4.70)$$

Ο νευρώνας k αλλάζει την κατάσταση εξόδου από v_{k1} σε v_{k2} . Η διαφορά στην ενέργεια $\Delta E = E_2 - E_1$ που προκαλείται από την αλλαγή κατάστασης $\Delta v_k = v_{k2} - v_{k1}$ βρίσκεται εφαρμόζοντας την εξίσωση (4.68) για v_{k1} και v_{k2} και μετά αφαιρώντας. Έτσι, μπορεί να γραφεί ως εξής

$$\Delta E = -\frac{1}{2} \left[\Delta v_k \sum_j w_{kj} v_j + \Delta v_k \sum_i w_{ik} v_i \right] + \theta_k \Delta v_k \quad (4.71)$$

Οι πρώτοι δύο όροι στην (4.71) δεν επηρεάζονται από την αλλαγή του νευρώνα k , και έτσι παραμένουν αμετάβλητοι και απαλείφονται. Επειδή ο πίνακας W είναι συμμετρικός, μπορούμε να ανταλλάξουμε τους δείκτες και να απλοποιήσουμε την έκφραση ως εξής

$$\Delta E = -\Delta v_k \left[\sum_j w_{kj} v_j - \theta_k \right] \quad (4.72)$$

όπου $\sum_j w_{kj} v_j$ είναι το σταθμισμένο άθροισμα των εισόδων στο νευρώνα k και θ_k είναι το κατώφλι του νευρώνα k . Αλλά, το κατώφλι κάθε νευρώνα τέθηκε μηδέν στη φάση της αποθήκευσης, για να εξασφαλιστεί ότι τα πρότυπα καταλαμβάνουν τα ελάχιστα στη συνάρτηση ενέργειας. Αν θυμηθούμε ότι οι εξοδοί των νευρώνων είναι ή $+1$ ή -1 , μειώνοντας το ΔE_k , θα σημαίνει ότι η έξοδος θα είναι $+1$ αν το σταθμισμένο άθροισμα είναι μεγαλύτερο του μηδενός και -1 αν είναι μικρότερο του μηδενός, επειδή αυτό εξυπηρετεί πάντα για να μειωθεί η τιμή του ΔE_k . Αν συγκρίνουμε αυτό με τη

συνάρτηση ενημέρωσης των νευρώνων σε ένα δίκτυο Hopfield, που δίνεται από την (4.68), βλέπουμε ότι η συνάρτηση ενημέρωσης εκτελεί αυτή τη λειτουργία και έτσι εκτελεί κεκλιμένη κατάβαση στην E . Αυτό μας επιτρέπει να ανακαλέσουμε τα πρότυπά μας από το δίκτυο ανακυκλώνοντας διαμέσου διαδοχικών καταστάσεων, κάθε μία από τις οποίες έχει μια χαμηλότερη ενέργεια (ή, αν το σταθμισμένο άθροισμα είναι ίσο με το κατώφλι, μία ίση ενέργεια) από την προηγούμενη. Αυτή η χαλάρωση σε χαμηλότερες ενεργειακές καταστάσεις συνεχίζεται μέχρι να φτάσουμε σε μία σταθερή κατάσταση χαμηλής ενέργειας, όταν το δίκτυο έχει βρει το δρόμο του προς ένα ελάχιστο και έτσι έχει παράγει το πρότυπο.

4.5.3.5 Ανάκλιση Συνεχών Προτύπων

Στην περίπτωση των συνεχών προτύπων θα έχουμε

$$\frac{dE}{dt} = \nabla E(v)^T \dot{v}(t) = \sum_{j=1}^n \frac{\partial E}{\partial v_j} \frac{dv_j}{dt} = -C \sum_{j=1}^n \frac{du_j}{dt} \frac{dv_j}{dt} = -C \sum_{j=1}^n \frac{dv_j}{du_j} \left(\frac{du_j}{dt} \right)^2 \quad (4.73)$$

το οποίο είναι πάντα αρνητικό, αφού το $\frac{dv_j}{du_j}$ είναι πάντα θετικό, επειδή η συνάρτηση $v_j = f(u_j)$ είναι πάντα αύξουσα και μονότονη ή μηδέν.

5

Νευρωνικά Δίκτυα για την Επίλυση Γραμμικών Συστημάτων

5.1 Αρχιτεκτονικές και Αλγόριθμοι Δικτύων FeedForward

Έχουν προταθεί διάφορες αρχιτεκτονικές FeedForward δικτύων με γραμμικούς νευρώνες για την επίλυση συστημάτων γραμμικών εξισώσεων και γενικά προβλημάτων γραμμικής άλγεβρας [LuZQ91], [PoIo92], [WaMe91], [WaMe92]. Το αντίστοιχο πρόβλημα εκφράζεται με κάποια αρχιτεκτονική και έναν αλγόριθμο εκπαίδευσης, συνήθως Back-Propagation [RuHW86], ώστε το δίκτυο να αντιστοιχίσει τα επιθυμητά πρότυπα, ενώ η λύση του προβλήματος δίνεται απ' τις τιμές των συνάψεων, μετά την εκπαίδευση. Τα δίκτυα [LuZQ91], [PoIo92],[WaMe92] είναι δύο διαστάσεων (2-D), ενώ το [WaMe91] είναι τριών διαστάσεων (3-D), με αποτέλεσμα να υπάρχει μεγαλύτερου βαθμού παραλληλοποίηση. Όταν τα δίκτυα χρησιμοποιούνται για την επίλυση εξισώσεων, το δίκτυο της περίπτωσης [LuZQ91] χρησιμοποιεί μια απλή αρχιτεκτονική με n νευρώνες εισόδου και ένα νευρώνα εξόδου, ενώ το δίκτυο [PoIo92], (αποκαλούμενο Orthogonalized Back Propagation) χρησιμοποιεί μια αρχιτεκτονική με n νευρώνες εισόδου και m νευρώνες εξόδου. Τα δίκτυα [WaMe91], [WaMe92] εφαρμόζονται για την εύρεση του αντιστρόφου ενός πίνακα A χρησιμοποιώντας ένα δίκτυο με n νευρώνες εισόδου, ένα κρυφό στρώμα με n νευρώνες και n νευρώνες εξόδου [WaMe91], ενώ το [WaMe92] χρησιμοποιεί n δίκτυα που το καθένα έχει n νευρώνες εισόδου, ένα κρυφό στρώμα με n νευρώνες και n νευρώνες εξόδου. Οι γραμμές του πίνακα του αντίστοιχου προβλήματος της γραμμικής άλγεβρας εισάγονται στα δίκτυα [LuZQ91], [MAGE94], ενώ στα δίκτυα [PoIo92], [WaMe91] και [WaMe92], ο πίνακας αποθηκεύεται στις συνάψεις του δικτύου και εισάγεται ένα σύνολο γραμμικά ανεξάρτητων διανυσμάτων. Όλες οι παραπάνω αρχιτεκτονικές χρησιμοποιούν τη γραμμική συνάρτηση ενεργοποίησης, ενώ στην εκπαίδευση ο συντελεστής εκπαίδευσης μπορεί να είναι ευρεστικός [WaMe92] ή να υπολογίζεται κάθε φορά (adaptive), όπως στα [WaMe91], [PoIo92] και [LuZQ91].

Η αρχιτεκτονική που προτάθηκε απ' το Luo [LuZQ91] αποτελείται από ένα γραμμικό δίκτυο δύο στρωμάτων, ενός στρώματος εισόδου με r νευρώνες και ενός στρώματος εξόδου με s νευρώνες και χρησιμοποιείται για την επίλυση του συστήματος

$$WA = B \quad (5.1)$$

με $A \in \mathcal{R}^{m \times r}$, $B \in \mathcal{R}^{m \times s}$ και $W \in \mathcal{R}^{s \times r}$. Η επίλυση του συστήματος (5.1) ανάγεται στην ελαχιστοποίηση του μέσου τετραγωνικού σφάλματος (*mean-square-error*)

$$L(W) = \sum_{p=1}^m L_p(W) = \sum_{p=1}^m \|WA(p) - B(p)\|^2 \quad (5.2)$$

με $A(p), B(p)$ την p -στήλη των πινάκων A, B αντίστοιχα. Τα m πρότυπα $A(p)$, $p=1,2,\dots,m$ εισάγονται στο δίκτυο κυκλικά και τροποποιούνται οι συνάψεις [Koho74] σύμφωνα με τον τύπο

$$W^n(p+1) = W^n(p) - \alpha_n \nabla L_p = W^n(p) - \alpha_n [W^n(p)A(p) - B(p)]A^T(p) \quad (5.3)$$

με το συντελεστή εκπαίδευσης σε κάθε κύκλο να δίνεται απ' τον τύπο

$$\alpha_n = 1/n \quad (5.4)$$

όπου n ο αριθμός των επαναλήψεων.

Οι τιμές των συνάψεων W είναι η λύση ελαχίστων τετραγώνων του συστήματος (5.1) ή το ελάχιστο της συνάρτησης $L(W)$ στο σημείο $W = BA^+$, όπου A^+ είναι ο γενικευμένος ψευδο-αντίστροφος του A των Moore-Penrose [GoVL83].

Αν υπάρχει μόνο ένας νευρώνας εξόδου ($s=1$) το δίκτυο μπορεί να χρησιμοποιηθεί για την επίλυση του συστήματος $Ax = b$, ενώ ο αλγόριθμος εκπαίδευσης είναι ο μ -LMS [WiHo60].

Μια παρόμοια αρχιτεκτονική, με n νευρώνες εισόδου, ένα κρυφό στρώμα με n νευρώνες και n νευρώνες εξόδου, χρησιμοποιούν και οι Wang and Mendel [WaMe92] για την εύρεση του αντιστρόφου ενός πίνακα $A \in \mathcal{R}^{n \times n}$, ή την επίλυση του συστήματος

$$BA = I \quad (5.5)$$

με $A \in \mathcal{R}^{n \times n}$, $B \in \mathcal{R}^{n \times n}$ και $I \in \mathcal{R}^{n \times n}$. Τα στοιχεία του πίνακα A αποτελούν τις τιμές των συνάψεων μεταξύ των νευρώνων του στρώματος εισόδου με το κρυφό στρώμα, ενώ τα στοιχεία του προς εύρεση πίνακα B τις τιμές των συνάψεων μεταξύ των νευρώνων του κρυφού στρώματος με το στρώμα εξόδου. Ένα σύνολο γραμμικά ανεξάρτητων διανυσμάτων (τα μοναδιαία διανύσματα $e_i, i=1,2,\dots,n$) εισάγονται στο δίκτυο κυκλικά και χρησιμοποιείται ο αλγόριθμος *back-propagation* για την τροποποίηση των συνάψεων μεταξύ των νευρώνων του κρυφού στρώματος με το στρώμα εξόδου, οπότε οι νέες συνάψεις δίνονται απ' τον τύπο

$$b_{ij}(t+1) = b_{ij}(t) + \alpha [d_i - y_i] z_j \quad (5.6)$$

όπου d_i είναι η επιθυμητή έξοδος (ίση με το μοναδιαίο διάνυσμα εισόδου), y_i η τελική έξοδος και z_j η έξοδος του κρυφού στρώματος. Με την είσοδο του p -μοναδιαίου

διανύσματος, η έξοδος z_j η έξοδος του κρυφού στρώματος γίνεται $A^T(p)$, ενώ η τελική έξοδος y_i είναι $B^n(p)A(p)$, οπότε ο τύπος (5.6) δίνει

$$B^n(p+1) = B^n(p) - \alpha [B^n(p)A(p) - I(p)]A^T(p) \quad (5.7)$$

Ο συντελεστής α προσδιορίζεται με τον παρακάτω ευρεστικό (*heuristic*) τρόπο. Αρχίζει με την τιμή $\alpha = 0.15 \times 0.9^m$, με $m=0$, εκπαιδεύεται το δίκτυο και, αν αποκλίνει, το m παίρνει τις τιμές $m = 2, 3, \dots$, μέχρι να βρεθεί κάποια τιμή του m για την οποία συγκλίνει.

Μια παρόμοια αρχιτεκτονική με ένα δίκτυο τριών διαστάσεων (3-D), χρησιμοποιήθηκε απ' τους Wang και Mendel [WaMe91] για την εύρεση του αντιστρόφου ενός πίνακα $A \in \mathbb{R}^{n \times n}$ με τελικό στόχο την επίλυση του συστήματος

$$Ax = b \quad (5.8)$$

σύμφωνα με τον τύπο $x = A^{-1}b$. Η αρχιτεκτονική αυτή αποτελείται από n παράλληλα δίκτυα, στο καθένα απ' τα οποία, τα στοιχεία του προς εύρεση πίνακα B αποτελούν τις τιμές των συνάψεων μεταξύ των νευρώνων του στρώματος εισόδου με το κρυφό στρώμα, ενώ τα στοιχεία του πίνακα A τις τιμές των συνάψεων μεταξύ των νευρώνων του κρυφού στρώματος με το στρώμα εξόδου. Ένα σύνολο γραμμικά ανεξάρτητων διανυσμάτων (τα μοναδιαία διανύσματα $e_i, i=1, 2, \dots, n$) εισάγονται στο δίκτυο ταυτόχρονα και χρησιμοποιείται ο αλγόριθμος ASD, τύπου back-propagation, για την τροποποίηση των συνάψεων μεταξύ των νευρώνων του στρώματος εισόδου με το κρυφό στρώμα, οπότε οι νέες συνάψεις μόνο για κάθε υπο-δίκτυο $j, 1 \leq j \leq n$, δίνονται απ' τον τύπο

$$b_{ij}^j(t+1) = b_{ij}^j(t) + \alpha_j(t) [d_{ij} - y_{ij}(k)] \quad (5.9)$$

όπου $1 \leq i, j \leq n$, d_{ij} είναι η επιθυμητή έξοδος, ίση με το (i, j) στοιχείο του μοναδιαίου πίνακα I και y_{ij} η τελική έξοδος. Μετά την τροποποίηση των συνάψεων b_{ij}^j του κάθε υπο-δικτύου j , οι αντίστοιχες συνάψεις των υπολοίπων υπο-δικτύων παίρνουν τις ίδιες τιμές. Ο συντελεστής $\alpha_j(p)$ σε κάθε κύκλο εκπαίδευσης p δίνεται απ' τον τύπο

$$\alpha_j(p) = \frac{\sum_{i=1}^n (d_{ij} - y_{ij}(p))^2}{\sum_{i=1}^n \sum_{q=1}^n (d_{ij} - y_{ij}(p)) a_{iq} (d_{iq} - y_{iq}(p))} \quad (5.10)$$

Μια παρόμοια αρχιτεκτονική με αυτή του Luo χρησιμοποιείται απ' τους Polycargou και Ioannoy [PoIo92] σε έναν αλγόριθμο αποκαλούμενο Orthogonalized Back Propagation και χρησιμοποιείται για την επίλυση του συστήματος

$$WB = C \quad (5.11)$$

με $B \in \mathcal{R}^{m \times n}$, $C \in \mathcal{R}^{p \times n}$ και $W \in \mathcal{R}^{p \times m}$. Σαν είσοδοι στο παραπάνω δίκτυο είναι τα μοναδιαία διανύσματα $x(p)$ και έξοδοι το $y(p) = W(p-1)Bx(p)$. Η επιθυμητή έξοδος είναι $y^*(p) = W^*(p-1)Bx(p) = Cx(p)$, ενώ το σφάλμα του στρώματος εξόδου δίνεται απ' τον τύπο

$$\varepsilon(p) = y(p) - y^*(p) = W(p-1)z(p) - Cx(p) \quad (5.12)$$

με $z(p) = Bx(p)$. Οι διορθώσεις των συνάψεων έχουν τη μορφή

$$W(p) = W(p-1) - n(p)\varepsilon(p)\zeta^T(p) \quad (5.13)$$

όπου

$$\begin{aligned} \zeta(p) &= Q(p-1)z(p) \\ n(p) &= \begin{cases} 1/|\zeta(p)|^2 & \text{αν } \zeta(p) \neq 0 \\ 0, & \text{αν } \zeta(p) = 0 \end{cases} \\ Q(p) &= Q(p-1) - n(p)\zeta(p)\zeta^T(p) \end{aligned} \quad (5.14)$$

Οι Chu και Hsieh [ChHs92] για την επίλυση του συστήματος (5.1) προτείνουν τη συμπλήρωση του πίνακα A με την πρόσθεση γραμμών ή στηλών, οι οποίες δημιουργούνται απ' τις προηγούμενες, με τη χρήση λογικών AND και OR, ώστε να γίνει full rank (βαθμού n), πράγμα που υλοποιείται με functional link cells [PaoY89], ενώ οι Krishnamourthy [KrKr94] προτείνουν, για την εύρεση του γενικευμένου αντιστρόφου ενός πίνακα A , έναν αλγόριθμο που βασίζεται στον αλγόριθμο του Greville, με την παρατήρηση ότι, ο συγκεκριμένος αλγόριθμος μπορεί να υλοποιηθεί με κυκλώματα VLSI.

5.2 Αρχιτεκτονικές και Αλγόριθμοι Δικτύων Hopfield

Για την επίλυση του συστήματος $Ax=b$ με τη χρήση αναδρομικών δικτύων, το πρώτο βήμα είναι να βρεθεί η κατάλληλη *συνάρτηση ενέργειας* (Liarunon Function), τέτοια ώστε η κατάσταση ελάχιστης ενέργειας να αντιστοιχεί στη λύση του συστήματος x . Το επόμενο βήμα είναι να μετατραπεί το πρόβλημα της ελαχιστοποίησης σε ένα σύνολο διαφορικών εξισώσεων ή εξισώσεων διαφορών και να σχεδιασθεί το δίκτυο με τις απαραίτητες τιμές συνάψεων, εξωτερικών διεγέρσεων και συναρτήσεων ενεργοποίησης. Τα δίκτυα που έχουν προταθεί για την επίλυση συστημάτων γραμμικών εξισώσεων αποτελούνται από ένα ή περισσότερα στρώματα νευρώνων, χρησιμοποιούν τη γραμμική συνάρτηση ενεργοποίησης, ενώ τα στοιχεία των πινάκων A , b χρησιμοποιούνται σαν συνάψεις και εξωτερικές διεγέρσεις. Η ελαχιστοποίηση της συνάρτησης ενέργειας αποτελεί και τη λύση του συστήματος προς επίλυση. Στην πραγματικότητα βέβαια αντί του $Ax=b$ επιλύεται το σύστημα των κανονικών εξισώσεων $A^T Ax = A^T b$ ή $Cx = d$, με $C = A^T A$ και $d = A^T b$ [MALE92], όπου τα x_i , $i = 1, 2, \dots, n$ αποτελούν τις εξόδους του

δικτύου, ενώ οι συνάψεις μεταξύ δύο νευρώνων i, j , $1 \leq i, j \leq n$ και οι εξωτερικές διεγέρσεις δίνονται απ' τις σχέσεις

$$w_{ij} = \begin{cases} -c_{ij}/c_{ii}, & i \neq j \\ 0, & i = j \end{cases}, \quad I_i = d_i/c_{ii}, \quad (5.15)$$

Η συνάρτηση ενέργειας για το προηγούμενο δίκτυο ορίζεται σαν

$$\begin{aligned} E &= -\frac{1}{2} \sum_i \sum_{j \neq i} w_{ij} x_i x_j - \sum_i I_i x_i + \sum_i I/r_i \int_0^{x_i} g^{-1}(x) dx \\ &= \frac{1}{2} \sum_i \sum_{j \neq i} \frac{c_{ij}}{c_{ii}} x_i x_j - \sum_i \frac{d_i}{c_{ii}} x_i + \frac{1}{2} \sum_i (x_i)^2 = \frac{1}{2} \sum_i \sum_j \frac{c_{ij}}{c_{ii}} x_i x_j - \sum_i \frac{d_i}{c_{ii}} x_i \end{aligned} \quad (5.16)$$

όπου $r_i = 1$ και $g^{-1}(x) = x$. Ο τύπος (5.16) μπορεί να γραφεί σε μορφή πινάκων σαν

$E = \frac{1}{2} x^T C x - x^T d$ με παράγωγο την $\nabla E = Cx - d$, και το ελάχιστο είναι η λύση της $\nabla E = 0$. Με αρχικές εισόδους $x_i = 0$, $i = 1, 2, \dots, n$, οι έξοδοι του δικτύου $\dot{x}_i(t+1)$, $1 \leq i \leq n$, δίνονται απ' τη σχέση

$$\dot{x}_i(t+1) = \sum_{j=1}^n w_{ij} x_j(t) + I_i = \frac{1}{c_{ii}} \left(\sum_{j=1}^n c_{ij} x_j(t) + d_i \right) \quad (5.17)$$

Ο παραπάνω αλγόριθμος ανήκει στην κατηγορία των αλγορίθμων κλιμακωτής μείωσης και μπορεί να δοθεί με το γενικό τύπο

$$\dot{x}(t+1) = -\mu(t) \nabla E(x) \quad (5.18)$$

με $\mu(t) = \frac{1}{c_{ii}}$. Στην ίδια κατηγορία ανήκει και το δίκτυο ενός στρώματος n νευρώνων που προτάθηκε απ' τη J. Wang, [Wan92a], όπου οι συνάψεις μεταξύ δύο νευρώνων i, j , $1 \leq i, j \leq n$ και οι εξωτερικές διεγέρσεις ορίζονται σαν $W = -nA^T A$, $I = -nA^T b$. Με αρχικές εισόδους $x_i = 0$, $i = 1, 2, \dots, n$, οι έξοδοι του δικτύου δίνονται απ' τη σχέση

$$\dot{x}(t+1) = -nA^T Ax(t) + nA^T b \quad (5.19)$$

Ο παραπάνω αλγόριθμος ανήκει στην κατηγορία των αλγορίθμων κλιμακωτής μείωσης με σταθερό συντελεστή σύγκλισης $n > 0$ και μπορεί να δοθεί με το γενικό τύπο

$$\dot{x}(t+1) = -n \nabla E(x) \quad (5.20)$$

Το προηγούμενο δίκτυο μπορεί να απλοποιηθεί για ειδικά γραμμικά συστήματα εξισώσεων. Έτσι, αν η μήτρα A είναι συμμετρική ($A=A^T$), [WaLi94] η καταστατική εξίσωση κινήσεως δίνεται απ' τη σχέση

$$\dot{x}(t+1) = -\mu(A^2x(t) - Ab) \quad (5.21)$$

Αν η μήτρα A είναι θετικά ή αρνητικά ορισμένη, η καταστατική εξίσωση κινήσεως δίνεται απ' τη σχέση

$$\dot{x}(t+1) = -\mu \operatorname{sgn}(A)[Ax(t) - b] \quad (5.22)$$

με $\operatorname{sgn}(A)=I$, αν A είναι θετικά ορισμένη και $\operatorname{sgn}(A)=-I$, αν A είναι αρνητικά ορισμένη. Οι συνάψεις μεταξύ δύο νευρώνων i, j και οι εξωτερικές διεγέρσεις ορίζονται σαν $W = -\mu \operatorname{sgn}(A)A$ και $I = \mu \operatorname{sgn}(A)b$.

Αν οι τιμές των A, b είναι μιγαδικές, το πρόβλημα μπορεί να λυθεί με ένα δίκτυο δύο στρωμάτων [Wan92b] με $2n$ νευρώνες το καθένα. Οι νευρώνες του κάθε στρώματος συνδέονται με τους νευρώνες του άλλου στρώματος, αλλά δεν υπάρχουν συνδέσεις μεταξύ των νευρώνων του ίδιου στρώματος. Οι συνδέσεις των νευρώνων του πρώτου προς το δεύτερο στρώμα ορίζονται σαν $n_1 A^T$, ενώ οι συνδέσεις των νευρώνων του δεύτερου προς το πρώτο στρώμα ορίζονται σαν $n_2 A$ με n_1, n_2 τους συντελεστές σύγκλισης ενώ οι εξωτερικές διεγέρσεις μόνο του πρώτου στρώματος ορίζονται σαν $n_2 b$. Οι εξισώσεις κινήσεως για τους νευρώνες $u_i, v_i, i=1,2,\dots,2n$ δίνονται απ' τις σχέσεις

$$\dot{v}(t+1) = n_1 A^T u(t) \quad (5.23)$$

$$u(t+1) = -n_2 A v(t) + n_2 b \quad (5.24)$$

Στην περίπτωση που η μήτρα A είναι συμμετρική και θετικά ορισμένη [Wan92b], μπορεί να χρησιμοποιηθεί δίκτυο ενός μόνο στρώματος με συνάψεις μεταξύ των νευρώνων $-n_1 A$, εξωτερικές διεγέρσεις $n_1 b$ και εξισώσεις κινήσεως

$$\dot{v}(t+1) = n_1 A^T v(t) + n_1 b \quad (5.25)$$

Ένα άλλο δίκτυο δύο στρωμάτων που χρησιμοποιεί τον αλγόριθμο κλιμακωτής μείωσης (steepest descent) προτάθηκε απ' τους Carvalho και Barbosa [CaBa92]. Οι νευρώνες του κάθε στρώματος συνδέονται με τους νευρώνες του άλλου στρώματος, αλλά δεν υπάρχουν συνδέσεις μεταξύ των νευρώνων του ίδιου στρώματος. Οι συνδέσεις των νευρώνων του πρώτου προς το δεύτερο στρώμα ορίζονται σαν A^T , ενώ οι συνδέσεις των νευρώνων του δεύτερου προς το πρώτο στρώμα ορίζονται σαν A , ενώ οι εξωτερικές διεγέρσεις μόνο του πρώτου στρώματος ορίζονται σαν $-b$. Οι εξισώσεις κινήσεως για τους νευρώνες $u_i, i=1,2,\dots,m$ και $v_j, j=1,2,\dots,n$ δίνονται απ' τις σχέσεις

$$\begin{aligned}
u(t+1) &= -\sum_{j=1}^n w_{ij}v(t) + I_i = Av(t) - b, \\
C\dot{v}(t+1) &= \sum_{i=1}^m w_{ji}u_i - \frac{v_i}{R} = A^T u(t) - \frac{v_i}{R}
\end{aligned} \tag{5.26}$$

Για τιμές $C=10^{-6}$, $R=10^6$ και $\alpha=1/C$, ο αλγόριθμος ισοδυναμεί με τον αλγόριθμο κλιμακωτής μείωσης, όπου οι έξοδοι του δικτύου δίνονται απ' τη σχέση

$$\dot{v}(t+1) = -\alpha A^T Av(t) + \alpha A^T b \tag{5.27}$$

Το διάνυσμα εξόδου v είναι ένα σημείο στο χώρο των καταστάσεων του συστήματος, το οποίο αντιστοιχεί σε μια τιμή της συνάρτησης κόστους $E = \frac{1}{2} \|Ax - b\|^2$. Για ένα σύστημα εξισώσεων με $\text{rank}(A)=n$, κάθε σημείο ισορροπίας (*equilibrium point*) είναι ασυμπτωτικά σταθερό, ενώ η ταχύτητα σύγκλισης εξαρτάται απ' τις ιδιοτιμές του πίνακα A .

Μια πιο συστηματική μελέτη της επίλυσης συστημάτων γραμμικών εξισώσεων και γενικά της μεθόδου των Ελαχίστων Τετραγώνων έγινε απ' τους Cichocki και Unbehauen [CiUn92a], οι οποίοι χρησιμοποιούν διάφορα είδη συναρτήσεων ενέργειας. Για το κλασικό πρόβλημα Ελαχίστων Τετραγώνων, όπου η συνάρτηση ενέργειας δίνεται απ' τον τύπο $E_2 = \frac{1}{2} \|Ax - b\|_2^2$, με $\nabla E_2 = A^T (Ax - b)$, οι καταστατικές εξισώσεις δίνονται απ' τη σχέση

$$\dot{x}(t+1) = -\mu(t) \nabla E_2(x) = -\mu(t) A^T (Ax - b) \tag{5.28}$$

όπου $\mu(t) = [\mu_{ij}(t)]$ μια $n \times n$ θετικά ορισμένη μήτρα κατά προτίμηση διαγώνια. και οι τιμές της εξαρτώνται απ' το χρόνο και το x ή μπορεί να είναι σταθερές. Το δίκτυο που προτείνεται αποτελείται από τρία στρώματα συνδεδεμένων μεταξύ τους. Η μήτρα A αποτελεί τις συνάψεις του πρώτου (m νευρώνες) με το δεύτερο (n νευρώνες) στρώμα, η μήτρα μ αποτελεί τις συνάψεις του δεύτερου με το τρίτο (n νευρώνες) στρώμα. Το διάνυσμα b χρησιμοποιείται για τις εξωτερικές διεγέρσεις του πρώτου στρώματος, ενώ οι τελικές έξοδοι x του τρίτου στρώματος αποτελούν και τις νέες εισόδους στο πρώτο στρώμα νευρώνων. Η επίλυση του παραπάνω συστήματος διαφορικών εξισώσεων μπορεί να μετατραπεί στην επίλυση των εξισώσεων διαφορών

$$x_j^{(t+1)} = x_j^{(t)} - \mu_j^{(t)} \sum_{i=1}^m a_{ij} \left(\sum_{j=1}^n a_{ij} x_j^{(t)} - b_i \right) \tag{5.29}$$

Το δίκτυο και οι εξισώσεις κινήσεως απλοποιούνται για ειδικά συστήματα εξισώσεων. Έτσι, αν η μήτρα A είναι *συμμετρική και θετικά ορισμένη*, η εξίσωση (5.29) γίνεται

$$\dot{x}(t+1) = -\mu_0 (Ax - b) \quad (5.30)$$

Η συνάρτηση ενέργειας E σ' αυτή την περίπτωση είναι $E(x) = \frac{1}{2} x^T Ax - b^T x$ με $\nabla E(x) = Ax - b$ και $\nabla^2 E(x) = A$, οπότε η μοναδική λύση που ελαχιστοποιεί την $E(x)$ είναι η $x = A^{-1}b$, που είναι το μοναδικό καθολικό ελάχιστο της $E(x)$. Αντί της ελαχιστοποίησης της συνάρτησης ενέργειας $E(x)$ μπορούν να οριστούν n τοπικές συναρτήσεις ενέργειας

$$E_i(x) = |r_i| = \sum_{j=1}^m |a_{ij}x_j^{(t)} - b_i| \quad \text{ή} \quad E_i(x) = \frac{1}{2} r_i^2 = \frac{1}{2} \sum_{j=1}^n (a_{ij}x_j^{(t)} - b_i)^2$$

οι οποίες και ελαχιστοποιούνται με τη χρήση της εξίσωσης

$$\dot{x}(t+1) = -\mu_j a_{ji} \text{sign}(r_j) \quad \text{ή} \quad \dot{x}(t+1) = -\mu_j r_j \frac{dr_j}{dx_j} = -\mu_j r_j a_{ji} \quad (5.31)$$

Μια μέθοδος παρόμοια της Adaptive Steepest Descent χρησιμοποιεί σαν συντελεστή εκπαίδευσης $\mu_0(t)$ το $\mu_0(t) = \frac{\mu}{r^T A A^T r}$, όπου $r = Ax - b$ και $\mu = \frac{1}{\tau}$, με τ τη σταθερά ολοκλήρωσης. Σ' αυτή την περίπτωση το σύστημα βρίσκει το σταθερό σημείο ισορροπίας (stationary point) ανεξάρτητα απ' το μέγεθος του προβλήματος σε σταθερό χρόνο t_r που δίνεται απ' τον τύπο $t_r = \frac{\frac{1}{2} \sum_{i=0}^n b_i^2 - \delta}{\mu}$, όπου δ ένας μικρός τυχαίος αριθμός.

Οι Cichocki και Unbehauen [CiUn94a], [CiUn94b] επιχειρούν μια βελτίωση των προηγούμενων μεθόδων με εφαρμογές στην επίλυση του προβλήματος των ελαχίστων τετραγώνων (Least Squares) και Total Least Squares, όπου χρησιμοποιούν διακόπτες οι οποίοι ελέγχουν ποιες εισοδοί εισάγονται στο δίκτυο. Έτσι, οι διακόπτες μπορούν να ανοίγουν με τη σειρά ή ταυτόχρονα, οπότε επεξεργάζονται μια εξίσωση τη φορά ή όλες μαζί, ενώ η εξίσωση κινήσεως σε μορφή εξισώσεων διαφορών δίνεται απ' τον τύπο

$$x^{(t+1)} = x^{(t)} - n_i^{(t)} (a_i^T x - b_i) a_i \quad (5.32)$$

Στην πρώτη περίπτωση, χρησιμοποιώντας σαν $n_i^{(t)}$ το $n_i^{(t)} = \frac{\omega_i}{a_i^T a}$ ο τύπος (5.32) γίνεται

$$x^{(t+1)} = x^{(t)} - \omega_i \frac{(a_i^T x - b_i)}{a_i^T a} a_i \quad (5.33)$$

που είναι η μέθοδος SOR με ω_i το συντελεστή διαδοχικής υπερχαλάρωσης.

Παρόμοιες αρχιτεκτονικές δικτύων και αλγόριθμοι χρησιμοποιούνται απ' τους Cichocki και Unbehauen [CiU292b], όπου για τον ορισμό της συνάρτησης ενέργειας και

κατ' επέκταση των εξισώσεων κινήσεως χρησιμοποιούνται οι νόρμες l και ∞ (L_l και L_∞ norms) όπως και στο [CULW95] όπου οι παραπάνω τεχνικές χρησιμοποιούνται στη θεωρία σημάτων.

5.2.1. Δίκτυα τύπου Hopfield για την Εύρεση Αντιστρόφου Πίνακα

Τα δίκτυα που προτάθηκαν για την εύρεση του αντιστρόφου A^{-1} μιας $(n \times n)$ μη ιδιάζουσας μήτρας A αποτελούνται από n υπο-δίκτυα για την επίλυση n συστημάτων εξισώσεων. Η Jun Wang, [Wang93] χρησιμοποιεί ένα δίκτυο με n^2 νευρώνες τοποθετημένους σε μορφή πίνακα. Έτσι, οι συνάψεις μεταξύ των νευρώνων του κάθε υπο-δικτύου και οι εξωτερικές διεγέρσεις ορίζονται σαν $W = -nA^T A$, $I = -nA^T$, ενώ οι εξισώσεις κινήσεως δίνονται απ' τον τύπο

$$\dot{V}(t+1) = -nA^T AV(t) + nA^T \quad (5.34)$$

Ένα παρόμοιο δίκτυο [LuLi93] χρησιμοποιείται για την επίλυση του συστήματος $MX=Y$, όπου X , Y δεδομένες μήτρες $(n \times m)$ και $(p \times m)$ αντίστοιχα ή την εύρεση του ψευδο-αντιστρόφου $M=YX^+$. Το δίκτυο που προτείνεται χωρίζεται σε δύο τμήματα, ένα κύριο δίκτυο με $p \times n$ νευρώνες και ένα υπο-δίκτυο με $p \times m$ νευρώνες, στα οποία οι νευρώνες του κυρίως δικτύου και του υπο-δικτύου συνδέονται μεταξύ τους. Η μήτρα X αποτελεί τις συνάψεις των νευρώνων του κυρίως δικτύου με το υπο-δίκτυο, ενώ η μήτρα X^T τις συνάψεις των νευρώνων του υπο-δικτύου με το κυρίως δίκτυο. Οι εξωτερικές διεγέρσεις του υπο-δικτύου είναι τα στοιχεία της μήτρας Y , ενώ οι καταστατικές εξισώσεις για τους νευρώνες U_{ij} , $i=1,2,\dots,p$, $j=1,2,\dots,n$ του κυρίως δικτύου και ϕ_{il} , $i=1,2,\dots,p$, $l=1,2,\dots,m$ του υπο-δικτύου αντίστοιχα δίνονται απ' τις σχέσεις

$$\begin{aligned} C\dot{U}_{ij}(t+1) &= -\frac{U_{ij}(t)}{R} - \sum_{l=1}^m X_{jl}\phi_{il}(t) \\ \phi_{il}(t) &= k_2 \left(\sum_{j=1}^n X_{jl}U_{ij}(t) - Y_{il} \right) \end{aligned} \quad (5.35)$$

Το προηγούμενο δίκτυο χρησιμοποιείται απ' τους Luo και Bao [LuBa92], για την εύρεση του αντιστρόφου ενός $(n \times n)$ πίνακα Q με $B=I$.

Οι Wu, Wang και Hootman [WuWH94] χρησιμοποιούν τις ιδιότητες του ψευδο-αντιστρόφου δημιουργώντας μια πολύπλοκη αρχιτεκτονική, όπου η εξίσωση κινήσεως δίνεται απ' τον τύπο

$$\dot{X}(t+1) = -\mu_1 A^T U A^T - \mu_2 (V X^T A^T + A^T X^T U - V)$$

με $U = AXA - A$ και $V = XAX - X$ και μ_1, μ_2 θετικές σταθερές ελέγχου.

6

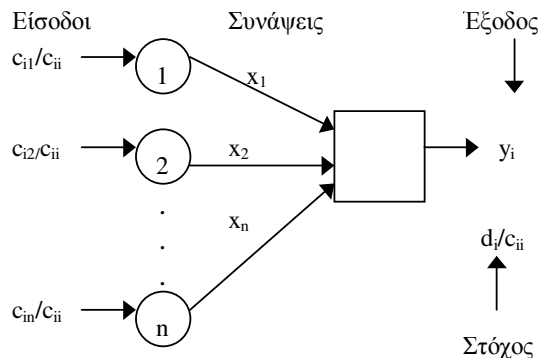
Νέες Αρχιτεκτονικές και Αλγόριθμοι Νευρωνικών Δικτύων για την Επίλυση Γραμμικών Συστημάτων με Επαναληπτικές Μεθόδους

6.1 Αρχιτεκτονικές και Αλγόριθμοι Δικτύων FeedForward

Η πρώτη προσπάθεια της σύνδεσης γνωστών επαναληπτικών μεθόδων της Αριθμητικής Ανάλυσης για την επίλυση συστημάτων γραμμικών εξισώσεων με αρχιτεκτονικές FeedForward Νευρωνικών Δικτύων που προσομοιώνουν αυτές τις μεθόδους έχει προταθεί από τους Μαργαρίτη, Αδαμόπουλο, Γουλιάνα [MAGE94]. Η αρχιτεκτονική που προτάθηκε αποτελείται από ένα γραμμικό δίκτυο δύο στρωμάτων, ενός στρώματος εισόδου με n νευρώνες και ενός στρώματος εξόδου με ένα νευρώνα, όπως φαίνεται στο σχήμα 6.1 και χρησιμοποιείται για την επίλυση του συστήματος

$$Ax = b \quad (6.1)$$

με $A \in \mathbb{R}^{n \times r}$, $b \in \mathbb{R}^n$ και $x \in \mathbb{R}^n$, αν η μήτρα A είναι συμμετρική και θετικά ορισμένη (SPD).



Σχήμα 6.1 Νευρωνικό Δίκτυο στην i επανάληψη

Αν η μήτρα A δεν είναι συμμετρική και θετικά ορισμένη, τότε επιλύεται το σύστημα

$$A^T Ax = A^T b \quad (6.2)$$

ή το ισοδύναμο

$$Cx = d \quad (6.3)$$

με $C \in \mathfrak{R}^{n \times r}$ και $b \in \mathfrak{R}^n$.

ΟΡΙΣΜΟΣ 6.1: Ορίζουμε σαν *κύκλο* ή *εποχή* k , $k=0, 1, 2, \dots$ το πέρασμα με τη σειρά όλων των γραμμών του πίνακα C .

ΟΡΙΣΜΟΣ 6.2: Ορίζουμε σαν *βήμα* i , $i = 1, 2, \dots, n$, την είσοδο στο δίκτυο (6.1) της i γραμμής του πίνακα C .

Οι αρχικές τιμές των συνάψεων του δικτύου x_1, x_2, \dots, x_n είναι τυχαίες ή μηδέν. Εισάγονται κυκλικά οι γραμμές του πίνακα C , αφού διαιρεθούν με το διαγώνιο στοιχείο. Έτσι, μετά την είσοδο της i γραμμής του πίνακα C , $i = 1, 2, \dots, n$, στον κύκλο $k+1$, υπολογίζεται η έξοδος $y_i = c_i^T x$ ή

$$y_i^{(k+1,i)} = \sum_{j=1}^n (c_{ij} / c_{ii}) x_j^{(k+1,i)} \quad (6.4)$$

όπου $(k+1, i)$ είναι το βήμα i στον κύκλο $k+1$.

6.1.1 Προσομοίωση της μεθόδου SOR

Για την υλοποίηση της SOR, εφαρμόζεται ο Κανόνας του Δέλτα μόνο για την i σύναψη και η διόρθωση της σύναψης x_i στον κύκλο $k+1$, μετά το βήμα i , θα έχει τη μορφή

$$x_i^{(k+1,i)} = x_i^{(k,i)} + \omega (d_i / c_{ii} - y_i^{(k+1,i)}) \quad (6.5)$$

με το $y_i^{(k+1,i)}$ να δίνεται απ' τη σχέση

$$y_i^{(k+1,i)} = \sum_{j=1}^{i-1} (c_{ij} / c_{ii}) x_j^{(k+1,i)} + \sum_{j=i}^n (c_{ij} / c_{ii}) x_j^{(k,i)} \quad (6.6)$$

αφού στον κύκλο $k+1$ έχουν αλλάξει τιμές μόνο οι συνάψεις x_j , $j=1, 2, \dots, i-1$, ενώ οι συνάψεις x_j , $j=i, i+1, \dots, n$, διατηρούν τις τιμές του κύκλου k .

Πρόταση 6.1: Με τη διόρθωση των συνάψεων του δικτύου του Σχήματος 6.1, όπως εκφράζεται στον τύπο (6.5) προσομοιώνεται η μέθοδος SOR.

Απόδειξη:

Η τιμή του x_i μετά την είσοδο της i γραμμής του πίνακα C , σύμφωνα με τις (6.5), (6.6) θα είναι

$$\begin{aligned}
 x_i^{(k+1,i)} &= x_i^{(k,i)} + \omega(d_i / c_{ii} - y_i^{(k+1,i)}) \\
 &= x_i^{(k,i)} + \omega(d_i / c_{ii} - \sum_{j=1}^{i-1} (c_{ij} / c_{ii}) x_j^{(k+1,i)} - \sum_{j=i}^n (c_{ij} / c_{ii}) x_j^{(k,i)}) \\
 &= x_i^{(k,i)} + \frac{\omega}{c_{ii}} (d_i - \sum_{j=1}^{i-1} c_{ij} x_j^{(k+1,i)} - \sum_{j=i}^n c_{ij} x_j^{(k,i)}) \\
 &= (1 - \omega) x_i^{(k,i)} + \frac{\omega}{c_{ii}} (d_i - \sum_{j=1}^{i-1} c_{ij} x_j^{(k+1,i)} - \sum_{j=i+1}^n c_{ij} x_j^{(k,i)}) \tag{6.7}
 \end{aligned}$$

που είναι ισοδύναμο με τον υπολογισμό του x_i στη μέθοδο SOR.

6.1.2 Προσομοίωση της μεθόδου JOR

Για την υλοποίηση της JOR, εφαρμόζεται ο Κανόνας του Δέλτα για όλες τις συνάψεις x_i , $i=1,2,\dots,n$ και η διόρθωση της σύνταξης x_i στον κύκλο $k+1$ θα έχει τη μορφή

$$x_i^{(k+1)} = x_i^{(k)} + \omega(d_i / c_{ii} - y_i^{(k+1)}) \tag{6.8}$$

με το $y_i^{(k+1)}$ να δίνεται απ' τη σχέση

$$y_i^{(k+1)} = \sum_{j=1}^n (c_{ij} / c_{ii}) x_j^{(k)} \tag{6.9}$$

αφού στον κύκλο $k+1$ για τον υπολογισμό των $y_i^{(k+1)}$ δεν έχουν αλλάξει τιμές οι συνάψεις x_j , $j=i,i+1,\dots,n$, οπότε διατηρούν τις τιμές του κύκλου k .

Πρόταση 6.2: Με τη διόρθωση των συνάψεων του δικτύου του Σχήματος 6.1, όπως εκφράζεται στον τύπο (6.7) προσομοιώνεται η μέθοδος JOR.

Απόδειξη:

Η τιμή του x_i μετά την είσοδο της i γραμμής του πίνακα C , σύμφωνα με τις (6.7), (6.8) θα είναι

$$\begin{aligned}
 x_i^{(k+1)} &= x_i^{(k)} + \omega(d_i / c_{ii} - y_i^{(k+1)}) \\
 &= x_i^{(k)} + \omega(d_i / c_{ii} - \sum_{j=1}^n (c_{ij} / c_{ii}) x_j^{(k)})
 \end{aligned}$$

$$\begin{aligned}
&= x_i^{(k)} + \frac{\omega}{c_{ii}} \left(d_i - \sum_{j=1}^n c_{ij} x_j^{(k)} \right) \\
&= (1 - \omega) x_i^{(k)} + \frac{\omega}{c_{ii}} \left(d_i - \sum_{\substack{j=1 \\ j \neq i}}^n c_{ij} x_j^{(k)} \right)
\end{aligned} \tag{6.10}$$

που είναι ισοδύναμο με τον υπολογισμό του x_i στη μέθοδο JOR.

6.1.3 Εξισώσεις της μορφής $AX=B$ και Αντιστροφή Πίνακα

Οι μέθοδοι που παρουσιάστηκαν προηγουμένως για την επίλυση εξισώσεων μπορούν να επεκταθούν για την επίλυση του συστήματος

$$AX = B \tag{6.11}$$

όπου A , X και B είναι $(m \times n)$, $(n \times p)$ και $(m \times p)$ πίνακες αντίστοιχα. Η επίλυση της εξίσωσης (6.11) ισοδυναμεί με την επίλυση του συστήματος των Κανονικών Εξισώσεων

$$A^T AX = A^T B \tag{6.12}$$

ή

$$CX = D \tag{6.13}$$

όπου C είναι ένας $(n \times n)$ πίνακας και D ένας $(n \times p)$ πίνακας. Η γενική λύση των παραπάνω συστημάτων (6.11) - (6.13) είναι

$$X = A^+ B \tag{6.14}$$

όπου ο πίνακας A^+ ικανοποιεί τις συνθήκες των Moore-Penrose για το γενικευμένο αντίστροφο [DaBi74], [Koho84].

Μια ειδική περίπτωση συστήματος είναι η εύρεση του αντιστρόφου πίνακα ή η επίλυση του συστήματος

$$AX = I \tag{6.15}$$

όπου A και X είναι $(n \times n)$ πίνακες. Η γενική λύση του συστήματος (6.15) δίνεται απ' τον τύπο

$$X = A^{-1} \tag{6.16}$$

Οι μέθοδοι που αναπτύχθηκαν μπορούν να χρησιμοποιηθούν για αυτο-συσχετιστικά ή ετερο-συσχετιστικά δίκτυα, όπου τα m πρότυπα εκπαίδευσης $a^{(k)} = a_1^{(k)}, a_2^{(k)}, \dots, a_n^{(k)}$,

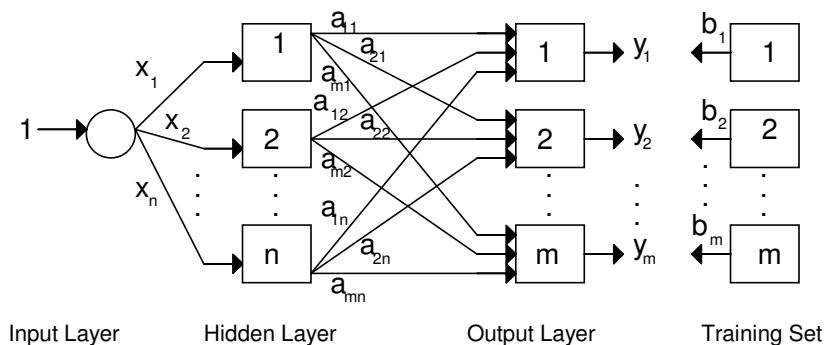
$k=1,2,\dots,m$ αποτελούν τις m γραμμές του πίνακα A , ενώ οι m στόχοι $b^{(k)} = b_1^{(k)}, b_2^{(k)}, \dots, b_p^{(k)}$ τις m στήλες του πίνακα B . Αφού δημιουργηθεί το σύστημα των κανονικών εξισώσεων (6.12), εφαρμόζονται οι μέθοδοι (6.5), (6.8) για τον υπολογισμό του γενικευμένου αντιστρόφου, οι οποίες πάντα συγκλίνουν, αφού η μήτρα A είναι Συμμετρική και Θετικά Ορισμένη [Hayo81], [Varg62], [Youn71], [Luen73].

6.2 Αρχιτεκτονικές και Αλγόριθμοι για επιτάχυνση της Batch-LMS

Το μειονέκτημα των μεθόδων που παρουσιάστηκαν στις παραγράφους 6.3 και 6.4 είναι η ανάγκη της δημιουργίας του συστήματος των Κανονικών Εξισώσεων, με εξωτερικό πολλαπλασιασμό πινάκων, πριν εισαχθούν τα πρότυπα (οι γραμμές του πίνακα A) στο δίκτυο. Αυτό μπορεί να αποφευχθεί με μια νέα αρχιτεκτονική τύπου Back-Propagation που προτείναμε [GoAM97], η οποία στηρίχθηκε σε μια ιδέα του κ. Αδαμόπουλου, στην οποία ο πίνακας των προτύπων αποθηκεύεται στις συνάψεις του δικτύου, ενώ για την εκπαίδευση του δικτύου χρησιμοποιούνται μορφές του Γενικευμένου Κανόνα του Δέλτα (Generalized Delta Rule), όπου ο συντελεστής εκπαίδευσης μπορεί να είναι *ευριστικός* (μέθοδος Heuristic Steepest Descent, ισοδύναμη με τη Batch-LMS) ή να υπολογίζεται σε κάθε κύκλο εκπαίδευσης (μέθοδος Adaptive Steepest Descent).

6.2.1 Αρχιτεκτονική του Δικτύου

Το δίκτυο που προτείνεται, όπως φαίνεται στο Σχήμα 6.2, αποτελείται από δύο στρώματα, ένα κρυφό στρώμα με n νευρώνες, πλήρως συνδεδεμένο με το μοναδικό νευρώνα του στρώματος εισόδου και το στρώμα εξόδου με m νευρώνες, πλήρως συνδεδεμένο με το κρυφό στρώμα. Το στρώμα εισόδου που δεν θεωρείται σαν ξεχωριστό στρώμα μπορεί να αγνοηθεί και, αντ' αυτού, να χρησιμοποιηθεί *εξωτερική διέγερση (bias)* πλήρως συνδεδεμένου με το κρυφό στρώμα.



Σχήμα 6.2 Νευρωνικό Δίκτυο για Επίλυση Συστημάτων Εξισώσεων

Ένα μειονέκτημα της παρούσης αρχιτεκτονικής είναι ο μεγάλος αριθμός νευρώνων και συνάψεων για πολλά πρότυπα ή όταν ο πίνακας A είναι μεγάλου μεγέθους. Σ' αυτή την

περίπτωση, μπορούμε να χρησιμοποιήσουμε *block-LMS*, με k ομάδες-blocks που το καθένα περιέχει $\left\lceil \frac{m}{k} \right\rceil$ νευρώνες.

6.2.2 Εκπαίδευση του Δικτύου

Το δίκτυο που προτείνεται χρησιμοποιεί τον αλγόριθμο κλιμακωτής μείωσης (gradient descent algorithm) [RuHW86] για την ελαχιστοποίηση του σφάλματος και λειτουργεί ως εξής: Αρχίζοντας με ένα τυχαίο σημείο $x^{(0)}$, δημιουργείται μια ακολουθία από βελτιωμένες προσεγγίσεις $x^{(1)}, x^{(2)}, \dots$, έτσι ώστε, για $t \geq 0$, $x^{(t+1)} = x^{(t)} + \alpha(-\nabla E(x^{(t)}))$, όπου $-\nabla E(x^{(t)})$ η κατεύθυνση κατάβασης, όπως ορίστηκε στην (3.24) και α το βήμα, το οποίο μπορεί να είναι σταθερό και να ορίζεται με ευρετικό τρόπο, οπότε έχουμε τη μέθοδο κλιμακωτής μείωσης με ευρετική αναζήτηση (steepest descent method with heuristic line search), ο κλασικός αλγόριθμος Back-propagation, ή να υπολογίζεται σε κάθε βήμα, οπότε έχουμε τη μέθοδο κλιμακωτής μείωσης adaptive steepest descent.

Ορίζουμε σαν $w_{0j} = x_j$, $j = 1, 2, \dots, n$ τις συνάψεις μεταξύ του νευρώνα j του κρυφού στρώματος και του νευρώνα εισόδου, έτσι ώστε η σύνδεση του νευρώνα j , ($1 \leq j \leq n$) του κρυφού στρώματος με το νευρώνα εισόδου να αντιστοιχεί στο j στοιχείο του διανύσματος x (τη λύση του συστήματος), με n στοιχεία και με $w_{ji} = a_{ij}$, $1 \leq j \leq n$, $1 \leq i \leq m$ τη σύναψη μεταξύ του νευρώνα εξόδου j και του νευρώνα του κρυφού στρώματος i , τα στοιχεία του πίνακα A .

6.2.2.1 Ο Αλγόριθμος Heuristic Steepest Descent (HSD)

Ο αλγόριθμος εκπαίδευσης HSD λειτουργεί ως εξής: Αρχικά οι συνάψεις, x_j , $j = 1, 2, \dots, n$ παίρνουν τυχαίες τιμές στο $(-1, 1)$. Εισάγεται στο στρώμα εισόδου η μονάδα και υπολογίζονται οι έξοδοι του κρυφού στρώματος $u_j^{(t)} = f(x_j^{(t)}) = x_j^{(t)}$, $j = 1, 2, \dots, n$, με $f(\cdot)$ την ταυτοτική συνάρτηση για το νευρώνα j , x_j τη σύναψη μεταξύ του νευρώνα εισόδου και του j νευρώνα του κρυφού στρώματος, και t το βήμα της επανάληψης ($t = 0, 1, \dots$). Για κάθε νευρώνα εξόδου i , $i = 1, 2, \dots, m$ στο βήμα t υπολογίζεται η τελική έξοδος y_i

$$y_i^{(t)} = f\left(\sum_{j=1}^n u_j^{(t)} w_{ji}\right) = f\left(\sum_{j=1}^n x_j^{(t)} a_{ij}\right) = \sum_{j=1}^n a_{ij} x_j^{(t)} \quad (6.17)$$

Σύμφωνα με τον Κανόνα του Δέλτα (Delta Rule) υπολογίζεται το σφάλμα μεταξύ της επιθυμητής (στόχος) και της πραγματικής εξόδου b_i και $y_i^{(t)}$, για $i = 1, 2, \dots, m$

$$d_i^{(t)} = b_i - y_i^{(t)} \quad (6.18)$$

Οι συνάψεις μεταξύ του κρυφού στρώματος και του στρώματος εξόδου (τα στοιχεία του πίνακα A) παραμένουν αμετάβλητες και σύμφωνα με τον αλγόριθμο back-propagation [RuHW86] ο υπολογισμός του δ_k , $k=1,2,\dots,n$ για το κρυφό στρώμα έχει τη μορφή

$$\delta_k^{(t)} = \sum_{i=1}^m w_{ki}^{(t)} d_i^{(t)} = \sum_{i=1}^m a_{ik} d_i^{(t)} \quad (6.19)$$

Η διόρθωση των συνάψεων του k , $k=1,2,\dots,n$ νευρώνα του κρυφού στρώματος με το νευρώνα εισόδου, επειδή η είσοδος είναι 1, θα έχει τη μορφή

$$x_k^{(t+1)} = x_k^{(t)} + \alpha \delta_k^{(t)} \quad (6.20)$$

όπου α ο συντελεστής εκπαίδευσης.

Ο παραπάνω αλγόριθμος επαναλαμβάνεται, μέχρις ότου το σφάλμα μεταξύ των εξόδων και των στόχων να είναι σε αποδεκτά πλαίσια. Η σύγκλιση του αλγορίθμου αποδεικνύεται με το θεώρημα που ακολουθεί.

ΘΕΩΡΗΜΑ 6.1 Το δίκτυο του Σχήματος 6.1 με τη χρήση του αλγορίθμου HSD συγκλίνει στη Λύση Ελαχίστων Τετραγώνων (Least Squares Solution) του συστήματος (3.24).

Απόδειξη.

Η τιμή του $\delta_k^{(t)}$, $k=1,2,\dots,n$ στο βήμα t , με τη χρήση των (6.17)-(6.19) θα είναι

$$\begin{aligned} \delta_k^{(t)} &= \sum_{i=1}^m a_{ik} d_i^{(t)} = \sum_{i=1}^m a_{ik} (b_i - y_i^{(t)}) = \sum_{i=1}^m a_{ik} (b_i - \sum_{j=1}^n a_{ij} x_j^{(t)}) \\ &= \sum_{i=1}^m a_{ik} b_i - \sum_{i=1}^m a_{ik} \sum_{j=1}^n a_{ij} x_j^{(t)} \end{aligned} \quad (6.21)$$

ή σε μορφή πινάκων

$$\delta^{(t)} = A^T (b - Ax^{(t)}) = A^T b - A^T Ax^{(t)} \quad (6.22)$$

ενώ, η τιμή του $x_k^{(t+1)}$, $k=1,2,\dots,n$ στο βήμα $t+1$, με τη χρήση των (6.20),(6.22) θα είναι

$$x_k^{(t+1)} = x_k^{(t)} + \alpha \delta_k^{(t)} = x_k^{(t)} + \alpha \left(\sum_{i=1}^m a_{ik} b_i - \sum_{i=1}^m a_{ik} \sum_{j=1}^n x_j^{(t)} a_{ij} \right), \text{ ή σε μορφή πινάκων}$$

$$x^{(t+1)} = x^{(t)} + \alpha A^T (b - Ax^{(t)}) = x^{(t)} - \alpha A^T (Ax^{(t)} - b) = x^{(t)} - \alpha \nabla E(x^{(t)}) \quad (6.23)$$

ισοδύναμης της εξίσωσης (4.47) του αλγορίθμου batch-LMS, οπότε οι τιμές x_k , $k=1,2,\dots,n$ του x συγκλίνουν στη λύση του συστήματος των Κανονικών Εξισώσεων (6.2), η οποία ελαχιστοποιεί το σφάλμα, δηλαδή $\lim_{t \rightarrow \infty} (E(x^{(t)}))=0$, αν $m \leq n$, ή το όριο $\lim_{t \rightarrow \infty} (E(x^{(t)}))$ γίνεται ελάχιστο, αν $m > n$.

6.2.2.2 Ο Αλγόριθμος Adaptive Steepest Descent (ASD)

Το μειονέκτημα του αλγορίθμου HSD είναι η αργή του σύγκλιση, συγκριτικά με τους αλγορίθμους με μεταβλητό συντελεστή εκπαίδευσης. Με την ίδια αρχιτεκτονική, τροποποιούμε τον αλγόριθμο HSD, μόνο στη διόρθωση των συνάψεων. Έτσι, ο νέος αλγόριθμος ASD διαφέρει απ' τον προηγούμενο στην εξίσωση (6.20), η οποία γίνεται $x_k^{(t+1)} = x_k^{(t)} + \alpha^{(t+1)} \delta_k^{(t)}$, $k=1,2,\dots,n$ ή σε μορφή πινάκων $x^{(t+1)} = x^{(t)} + \alpha^{(t+1)} \delta^{(t)}$, όπου

$$\alpha^{(t+1)} = \frac{\delta^{(t)} \delta^{(t)}}{\delta^{(t)} A^T A \delta^{(t)}} = \frac{\delta^{(t)} \delta^{(t)}}{\delta^{(t)} C \delta^{(t)}} \quad (6.24)$$

με το $\delta^{(t)}$ όπως ορίστηκε στην εξίσωση (6.21),(6.22). Η σύγκλιση του αλγορίθμου αποδεικνύεται με το θεώρημα που ακολουθεί.

ΘΕΩΡΗΜΑ 6.2 Το δίκτυο του Σχήματος 6.1 με τη χρήση του αλγορίθμου ASD και το $\alpha^{(t+1)}$ όπως ορίστηκε στη (6.24) προσομοιώνει τη μέθοδο κλιμακωτής μείωσης (Steepest Descent Method).

Απόδειξη.

Με το $E(x)$ όπως ορίστηκε στην εξίσωση (3.23) έχουμε

$$\nabla E(x) = Cx - d = A^T Ax - A^T b \quad (6.25)$$

Η (6.20) με τη χρήση της (6.25) γίνεται

$$\delta^{(t)} = A^T b - A^T Ax^{(t)} = d - Cx^{(t)} = -\nabla E(x^{(t)}) \quad (6.26)$$

Αν θέλουμε το $\alpha^{(t+1)}$ να ελαχιστοποιήσει την $E(x^{(t+1)}) = E(x^{(t)} + \alpha^{(t+1)} \delta^{(t)})$, τότε το $\alpha^{(t+1)}$ δίνεται απ' την εξίσωση $\nabla E(\alpha^{(t+1)}) = 0$. Με τη χρήση παραγώγων και τις (6.25), (6.26) έχουμε

$$\begin{aligned} \nabla E(x^{(t+1)}) &= \nabla E(x^{(t)} + \alpha^{(t+1)} \delta^{(t)}) \\ &= \nabla E(x^{(t)} + \alpha^{(t+1)} \delta^{(t)}) \delta^{(t)} \\ &= [C(x^{(t)} + \alpha^{(t+1)} \delta^{(t)}) - d] \delta^{(t)} \\ &= [Cx^{(t)} + \alpha^{(t+1)} C \delta^{(t)} - d] \delta^{(t)} \end{aligned}$$

$$= [\alpha^{(t+1)} C \delta^{(t)} - \delta^{(t)}] \delta^{(t)}$$

Αν θέσουμε $\nabla E(\alpha^{(t+1)}) = 0$, βλέπουμε ότι το $\alpha^{(t+1)}$ θα δίνεται απ' τη σχέση
$$\alpha^{(t+1)} = \frac{\delta^{(t)} \delta^{(t)}}{\delta^{(t)} C \delta^{(t)}}, \text{ όπως ορίστηκε στην εξίσωση (6.24), Ο.Ε.Δ..}$$

ΘΕΩΡΗΜΑ 6.3 Το δίκτυο του Σχήματος 6.1 με τη χρήση του αλγορίθμου ASD και το $\alpha^{(t+1)}$ όπως ορίστηκε στη (6.24) προσομοιώνει την άριστη μέθοδο κλιμακωτής μείωσης (Optimal Steepest Descent Method) [Tsyp71], με το $\alpha^{(t+1)}$ να ορίζεται σαν

$$\alpha_{opt}^{(t+1)} \approx \frac{\|\nabla E(x^{(t)})\|^2}{\nabla E(x^{(t)}) \nabla^2 E(x^{(t)}) \nabla E(x^{(t)})} \quad (6.27)$$

Απόδειξη.

Από την εξίσωση (6.25) έχουμε

$$\nabla^2 E(x^{(t)}) = \nabla(\nabla E(x^{(t)})) = \nabla(Cx^{(t)} - d) = C \quad (6.28)$$

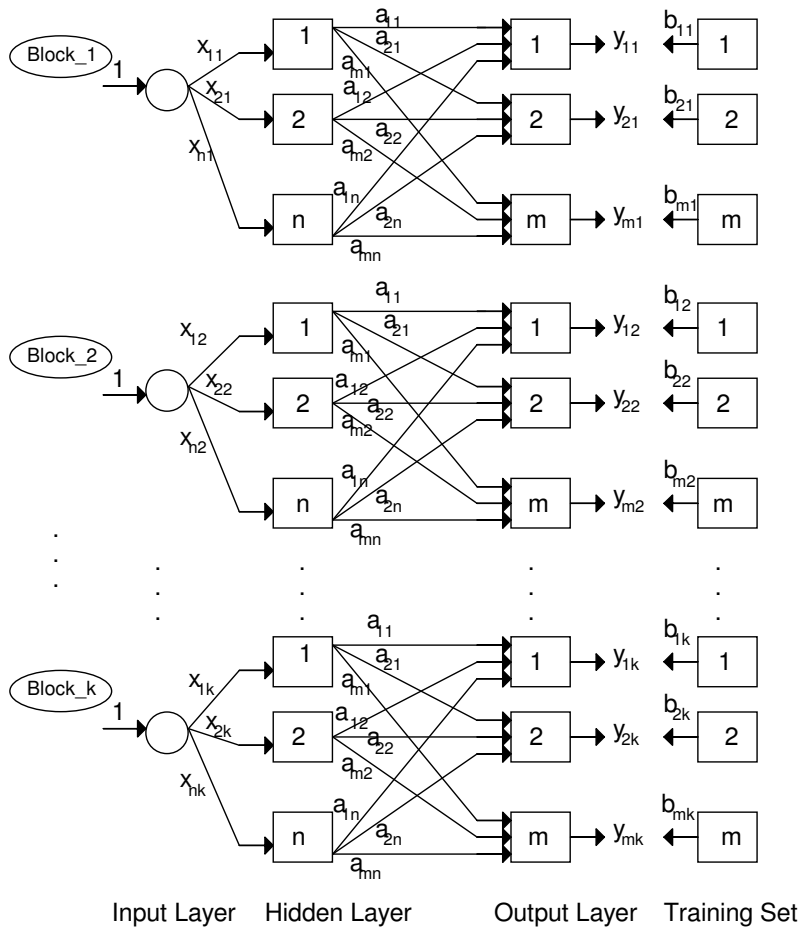
και από τις (6.25),(6.28), το $\alpha_{opt}^{(t+1)}$ γίνεται

$$\alpha_{opt}^{(t+1)} \approx \frac{\|\nabla E(x^{(t)})\|^2}{\nabla E(x^{(t)}) \nabla^2 E(x^{(t)}) \nabla E(x^{(t)})} = \frac{\|-\delta^{(t)}\|^2}{(-\delta^{(t)}) C (-\delta^{(t)})} = \frac{\delta^{(t)} \delta^{(t)}}{\delta^{(t)} C \delta^{(t)}} \quad (6.29)$$

όπως στην εξίσωση (6.24), Ο.Ε.Δ..

6.2.3 Εξισώσεις της μορφής $AX=B$ και Αντιστροφή Πίνακα

Οι μέθοδοι που παρουσιάστηκαν προηγουμένως για την επίλυση εξισώσεων μπορούν να επεκταθούν για την επίλυση του συστήματος (6.11), όπου A , X και B είναι $(m \times n)$, $(n \times p)$ και $(m \times p)$ πίνακες. Η επίλυση της εξίσωσης (6.11) ισοδυναμεί με την επίλυση k συστημάτων εξισώσεων με κοινό τον πίνακα των συντελεστών A . Έτσι, χρησιμοποιώντας k φορές το δίκτυο του Σχήματος 6.2, μπορούμε να δημιουργήσουμε τις k στήλες του πίνακα X . Μια άλλη δυνατότητα είναι η χρήση ενός τρισδιάστατου $(3-D)$ δικτύου, όπως φαίνεται στο Σχήμα 6.3, το οποίο αποτελείται από k δίκτυα του Σχήματος 6.2, τα οποία λειτουργούν παράλληλα.. Με τον ίδιο τρόπο μπορούμε να βρούμε τον αντίστροφο ενός πίνακα A με τη χρήση n δικτύων του Σχήματος 6.2. Οι αλγόριθμοι που εφαρμόζονται είναι οι ίδιοι, όπως και στο δισδιάστατο δίκτυο.



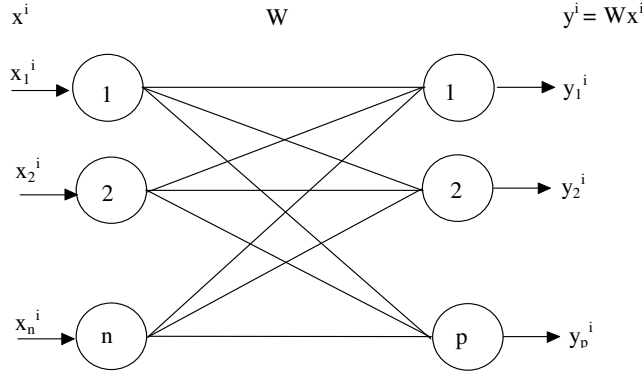
Σχήμα 6.3 Νευρωνικό Δίκτυο 3-D για Επίλυση Συστημάτων Εξισώσεων $AX=B$

6.3. Γενίκευση των Επαναληπτικών Μεθόδων

Τα πιο πολλά προβλήματα σε πολλούς τομείς επιστημονικών εφαρμογών, όπως η επεξεργασία σήματος και εικόνας, ρομποτική, συστήματα αυτομάτου ελέγχου, στατιστική, φυσική, ιατρική κ.λ.π., καταλήγουν σε ένα σύστημα γραμμικών εξισώσεων (συνήθως υπέρ-ορισμένο) του οποίου ζητείται η λύση ή η αντιστροφή ενός πίνακα σε πραγματικό χρόνο. Μια συλλογή από προβλήματα αυτού του τύπου είναι και η συλλογή *Proben1*, η οποία περιέχει πραγματικά δεδομένα (real world data) για 12 εφαρμογές που μπορούν να χρησιμοποιηθούν για την εκτίμηση αλγορίθμων νευρωνικών δικτύων και τη σύγκριση των αποτελεσμάτων. Οι παραπάνω εφαρμογές χωρίζονται σε *προβλήματα ταξινόμησης (pattern classification)* και *προβλήματα προσαρμογής συναρτήσεων (function approximation)*. Όλα τα προβλήματα απαιτούν εκπαίδευση με επίβλεψη (supervised learning) και μπορούν να χαρακτηρισθούν σαν στατικά, επειδή τα δεδομένα που πρέπει να μάθουν τα δίκτυα δεν αλλάζουν κατά τη διάρκεια της εκπαίδευσης και έχουν την παρακάτω μορφή:

Δίνεται ένα σύνολο προτύπων $x^i = [x_1^i, x_2^i, \dots, x_n^i]^T$ και στόχων $d^i = [d_1^i, d_2^i, \dots, d_p^i]^T$, $i=1,2,\dots,m$ και πρέπει να βρούμε ένα σύνολο συνάψεων $W \in \mathfrak{R}^{n \times p}$ που δίνει την καλύτερη απεικόνιση μεταξύ των ζευγών εισόδου/εξόδου (x^i, d^i) , $i=1,2,\dots,m$.

Μια απλή αρχιτεκτονική που υλοποιεί την παραπάνω απεικόνιση αποτελεί ένα ετερο-συσχετιστικό (hetero-associative) feedforward νευρωνικό δίκτυο δύο στρωμάτων, με n εισόδους και p εξόδους, όπως φαίνεται στο Σχήμα 6.4. Τα m πρότυπα εισάγονται στο στρώμα εισόδου κυκλικά δίνοντας μια έξοδο κάθε φορά $y^i = [y_1^i, y_2^i, \dots, y_p^i]^T$.



Σχήμα 6.3 Νευρωνικό Δίκτυο για την Επίλυση Συστημάτων Εξισώσεων $XW=D$

Η καλύτερη απεικόνιση θα ελαχιστοποιεί το Μέσο Τετραγωνικό Σφάλμα ή τη Συνάρτηση Κόστους

$$E(W) = \sum_{j=1}^p E(w^j) = \frac{1}{2} \sum_{j=1}^p \sum_{i=1}^m (d_j^i - y_j^i)^2 = \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^p (d_j^i - \sum_{k=1}^n x_k^i w_k^j)^2 = \frac{1}{2} \|X^T W - D\|^2 \quad (6.30)$$

με $X = [x^1, x^2, \dots, x^m]$, και $D = [d^1, d^2, \dots, d^m]$, w^j το διάνυσμα των συνάψεων του j νευρώνα εξόδου και $E(w^j)$ τη συνάρτηση κόστους αυτού του νευρώνα. Το ελάχιστο της συνάρτησης αυτής πρέπει να ικανοποιεί τη σχέση $\nabla E(W) = X(X^T W - D) = 0$, που μπορεί να γραφεί σαν

$$XX^T W = XD \quad (6.31)$$

ή το ισοδύναμο σύστημα των κανονικών εξισώσεων

$$BW = C \quad (6.32)$$

με $B = XX^T$, $B \in \mathfrak{R}^{n \times n}$ και $C = XD$, $C \in \mathfrak{R}^n$. Η εξίσωση (6.31) αποτελείται από p συστήματα κανονικών εξισώσεων, με $B = XX^T$ Συμμετρική και Θετικά Ορισμένη (SPD). Η λύση των παραπάνω συστημάτων για μια μη-ιδιάζουσα μήτρα X^T με $m \geq n$,

δίνει και μια μοναδική λύση ελαχίστων τετραγώνων $W = X^+ D$ με $X^+ = (XX^T)^{-1} X$ το γενικευμένο αντίστροφο του Moore-Penrose [GoVL83].

Οι συναρτήσεις κόστους $E(w^j)$ που ορίστηκαν στην (6.30) είναι τετραγωνικές μορφές ως προς τις συνάψεις w^j και για $m \geq n$ ορίζουν μια κλειστή υπέρ-παραβολοειδή επιφάνεια με ένα ελάχιστο, το καθολικό ελάχιστο που είναι η λύση των εξισώσεων (6.31), η οποία είναι μοναδική, αφού η Hessian μήτρα του $E(w^j)$ στο w^j , $\nabla^2 E(w^j) = XX^T$ είναι SPD.

Με το γραμμικό νευρωνικό δίκτυο του Σχήματος 6.4 μπορούν να προσομοιωθούν οι κλασσικές επαναληπτικές μέθοδοι, όπως SOR, JOR, Richardson και RGS, η βηματική-LMS (γνωστή σαν μέθοδος των Widrow-Hoff) [WiHo60], η ομαδική-LMS (γνωστή σαν batch-LMS [Hass95] ισοδύναμη της μεθόδου Richardson και της HSD, η μέθοδος ASD, καθώς και μέθοδοι συζυγούς κλίσης, όπως η μέθοδοι Fletcher-Reeves και Polak-Ribiere. Για τη σύγκριση βηματικών και ομαδικών μεθόδων δίνουμε τους παρακάτω ορισμούς.

ΟΡΙΣΜΟΣ 6.1. Ορίζεται σαν κύκλος εκπαίδευσης (*training cycle*) ή εποχή t το πέρασμα όλων των προτύπων εκπαίδευσης x^i , $i=1,2,\dots,m$.

ΟΡΙΣΜΟΣ 6.2. Ορίζεται σαν κύκλος επίσκεψης συνάψεων (*connection traversal cycle*) ct η διόρθωση όλων των συνάψεων. Να σημειωθεί ότι, στη μέθοδο LMS διορθώνονται οι συνάψεις με το πέρασμα κάθε προτύπου, (οπότε $ct=mt$), στις μεθόδους SOR και RGS διορθώνονται οι συνάψεις ενός νευρώνα σε κάθε κύκλο εκπαίδευσης, (οπότε $ct=div(t,m)$), ενώ στις υπόλοιπες μεθόδους RGS διορθώνονται όλες οι συνάψεις σε κάθε κύκλο εκπαίδευσης, (οπότε $ct=t$).

Οι αρχικές τιμές των συνάψεων του δικτύου w_k^j , $j=1,2,\dots,p$ και $k=1,2,\dots,n$ είναι τυχαίες ή μηδέν. Εισάγονται κυκλικά τα πρότυπα $x^i = [x_1^i, x_2^i, \dots, x_n^i]^T$, $i=1,2,\dots,m$. Έτσι, μετά την είσοδο του i προτύπου, στον κύκλο $t+1$, υπολογίζονται οι εξόδοι $y_j^{(t+1,i)}$ για κάθε νευρώνα εξόδου j , $j=1,2,\dots,p$, σύμφωνα με τον τύπο

$$y_j^{(t+1,i)} = w^{(t,j)} x^i = \sum_{k=1}^n w_k^{(t,j)} x_k^i \quad (6.33)$$

όπου $(t+1,i)$ είναι το βήμα i στον κύκλο $t+1$.

Σύμφωνα με τον Κανόνα του Δέλτα (Delta Rule) υπολογίζεται το σφάλμα μεταξύ της επιθυμητής εξόδου (στόχου) d_j^i και της πραγματικής εξόδου $y_j^{(t+1,i)}$, για κάθε πρότυπο i , $i=1,2,\dots,m$ και κάθε νευρώνα εξόδου j , $j=1,2,\dots,p$

$$\varepsilon_j^{(t+1,i)} = d_j^i - y_j^{(t+1,i)} \quad (6.34)$$

Για τις μεθόδους SOR και JOR, ο συντελεστής εκπαίδευσης ορίζεται σαν

$$\alpha_k = \frac{\omega}{(x^k)^T x^k} = \frac{\omega}{b_{kk}} \quad (6.35)$$

όπου ω ο συντελεστής υπερχαλαρώσεως, x^k , ($1 \leq k \leq n$), η k στήλη του πίνακα X , και b_{kk} το k διαγώνιο στοιχείο του πίνακα $B = XX^T$.

6.3.1 Προσομοίωση της μεθόδου LMS

Η διόρθωση των συνάψεων μετά την είσοδο του προτύπου x^i , $1 \leq i \leq m$ στον κύκλο $t+1$, θα έχει τη μορφή

$$\begin{aligned} w_k^{(t+1,j)} &= w_k^{(t,j)} + \omega \varepsilon_j^{(t+1,i)} = w_k^{(t,j)} + \omega (d_j^i - y_j^{(t+1,i)}) \\ &= w_k^{(t,j)} + \omega (d_j^i - \sum_{q=1}^n w_q^j x_q^i) \end{aligned} \quad (6.36)$$

όπου $k=1,2,\dots,n$, $j=1,2,\dots,p$ και ω ο συντελεστής εκπαίδευσης.

6.3.2 Προσομοίωση της μεθόδου Batch-LMS (B-LMS)

Η διόρθωση των συνάψεων μετά την είσοδο όλων των προτύπων x^i , $i=1,2,\dots,m$ στον κύκλο $t+1$, θα έχει τη μορφή

$$\begin{aligned} w_k^{(t+1,j)} &= w_k^{(t,j)} + \omega \sum_{i=1}^m (d_j^i - y_j^{(t+1,i)}) x_k^i \\ &= w_k^{(t,j)} + \omega \sum_{i=1}^m (d_j^i - \sum_{q=1}^n w_q^{(t,j)} x_q^i) x_k^i \\ &= w_k^{(t,j)} + \omega \sum_{i=1}^m (c_k^i - \sum_{q=1}^n b_q^k w_q^j) \end{aligned} \quad (6.37)$$

όπου $k=1,2,\dots,n$, $j=1,2,\dots,p$, ω ο συντελεστής εκπαίδευσης και b_q^k , c_k^i τα αντίστοιχα στοιχεία των πινάκων B και C , όπως ορίστηκαν στην (6.32).

6.3.3 Προσομοίωση της μεθόδου JOR

Η διόρθωση των συνάψεων μετά την είσοδο όλων των προτύπων x^i , $i=1,2,\dots,m$ στον κύκλο $t+1$, θα έχει τη μορφή

$$w_k^{(t+1,j)} = w_k^{(t,j)} + a_k \sum_{i=1}^m (d_j^i - y_j^{(t+1,i)}) x_k^i$$

$$\begin{aligned}
&= w_k^{(t,j)} + \frac{\omega}{b_{kk}} \sum_{i=1}^m (d_j^i - \sum_{q=1}^n w_q^{(t,j)} x_q^i) x_k^i \\
&= w_k^{(t,j)} + \frac{\omega}{b_{kk}} \sum_{i=1}^m (c_k^i - \sum_{q=1}^n b_q^k w_q^j)
\end{aligned} \tag{6.38}$$

όπου $k=1,2,\dots,n$, $j=1,2,\dots,p$, ω ο συντελεστής υπερχαλάρωσης, b_{kk} το k διαγώνιο στοιχείο του πίνακα $B = XX^T$ και b_q^k , c_k^i τα αντίστοιχα στοιχεία των πινάκων B και C , όπως ορίστηκαν στην (6.32). Να σημειωθεί ότι η παραπάνω μέθοδος είναι ισοδύναμη της Jacobi για $\omega = 1$.

6.3.4 Προσομοίωση της μεθόδου SOR

Η διόρθωση των συνάψεων του k νευρώνα εισόδου, όπου $k=\text{mod}(t,m)$, για $k=1,2,\dots,m-1$ και $k=m$, αν $\text{mod}(t,m)=0$ με όλους τους νευρώνες εξόδου, μετά την είσοδο όλων των προτύπων x^i , $i=1,2,\dots,m$ στον κύκλο εκπαίδευσης $t+1$ ή στον αντίστοιχο κύκλο επίσκεψης συνάψεων $ct+1$, θα έχει τη μορφή

$$\begin{aligned}
w_k^{(ct+1,j)} &= w_k^{(ct,j)} + a_k \sum_{i=1}^m (d_j^i - y_j^{(t+1,i)}) x_k^i \\
&= w_k^{(ct,j)} + \frac{\omega}{b_{kk}} \sum_{i=1}^m (d_j^i - \sum_{q=1}^{k-1} w_q^{(ct+1,j)} x_q^i - \sum_{q=k}^n w_q^{(ct,j)} x_q^i) x_k^i \\
&= (1-\omega) w_k^{(ct,j)} + \frac{\omega}{b_{kk}} \sum_{i=1}^m (d_j^i - \sum_{q=1}^{k-1} w_q^{(ct+1,j)} x_q^i - \sum_{q=k+1}^n w_q^{(ct,j)} x_q^i) x_k^i \\
&= (1-\omega) w_k^{(ct,j)} + \frac{\omega}{b_{kk}} (c_j^i - \sum_{q=1}^{k-1} w_q^{(ct+1,j)} b_q^i - \sum_{q=k+1}^n w_q^{(ct,j)} b_q^i)
\end{aligned} \tag{6.39}$$

όπου $k=1,2,\dots,n$, $j=1,2,\dots,p$, ω ο συντελεστής υπερχαλάρωσης, b_{kk} το k διαγώνιο στοιχείο του πίνακα $B = XX^T$, $w_q^{(ct+1,j)}$, $w_q^{(ct,j)}$ η σύναψη του j νευρώνα εξόδου με τον q νευρώνα εισόδου στον κύκλο $ct+1$ και ct και b_q^k , c_k^i τα αντίστοιχα στοιχεία των πινάκων B και C , όπως ορίστηκαν στην (6.32). Να σημειωθεί ότι η παραπάνω μέθοδος είναι ισοδύναμη της Gauss-Seidel για $\omega = 1$.

6.3.5 Προσομοίωση της μεθόδου Richardson-Gauss-Seidel (RGS)

Η διόρθωση των συνάψεων του k νευρώνα εισόδου, όπου $k=\text{mod}(t,m)$, για $k=1,2,\dots,m-1$ και $k=m$, αν $\text{mod}(t,m)=0$ με όλους τους νευρώνες εξόδου, μετά την είσοδο όλων των προτύπων x^i , $i=1,2,\dots,m$ στον κύκλο εκπαίδευσης $t+1$ ή στον αντίστοιχο κύκλο επίσκεψης συνάψεων $ct+1$, θα έχει τη μορφή

$$\begin{aligned}
w_k^{(ct+1,j)} &= w_k^{(ct,j)} + a_k \sum_{i=1}^m (d_j^i - y_j^{(t+1,i)}) x_k^i \\
&= w_k^{(ct,j)} + \omega \sum_{i=1}^m (d_j^i - \sum_{q=1}^{k-1} w_q^{(ct+1,j)} x_q^i - \sum_{q=k}^n w_q^{(ct,j)} x_q^i) x_k^i \\
&= w_k^{(ct,j)} + \omega (c_j^i - \sum_{q=1}^{k-1} w_q^{(ct+1,j)} b_q^i - \sum_{q=k}^n w_q^{(ct,j)} b_q^i)
\end{aligned} \tag{6.40}$$

όπου $k=1,2,\dots,n$, $j=1,2,\dots,p$, ω ο συντελεστής υπερχαλάρωσης, b_{kk} το k διαγώνιο στοιχείο του πίνακα $B = XX^T$, $w_q^{(ct+1,j)}$, $w_q^{(ct,j)}$ η σύναψη του j νευρώνα εξόδου με τον q νευρώνα εισόδου στον κύκλο $ct+1$ και ct και b_q^k , c_k^i τα αντίστοιχα στοιχεία των πινάκων B και C , όπως ορίστηκαν στην (6.32).

6.3.6 Προσομοίωση της μεθόδου Adaptive Steepest Descent (ASD)

Ορίζουμε για κάθε νευρώνα εισόδου k , $k=1,2,\dots,n$ το ομαδικό σφάλμα $\delta_k^{(t+1)}$ σαν

$$\delta_k^{(t+1)} = \sum_{i=1}^m \varepsilon_j^{(t+1,i)} x_k^i = \sum_{i=1}^m (d_j^i - y_j^{(t+1,i)}) x_k^i \tag{6.41}$$

ή σε μορφή διανύσματος

$$\delta^{(t+1)} = \sum_{i=1}^m \varepsilon_j^{(t+1,i)} x^i = \sum_{i=1}^m (d_j^i - y_j^{(t+1,i)}) x^i \tag{6.42}$$

και το συντελεστή εκπαίδευσης $\alpha_k^{(t+1)}$ σαν

$$\alpha_k^{(t+1)} = \frac{\delta_k^{(t+1)} \delta_k^{(t+1)}}{\delta_k^{(t+1)} X^T X \delta_k^{(t+1)}} \tag{6.43}$$

οπότε, η διόρθωση των συνάψεων μετά την είσοδο όλων των προτύπων x^i , $i=1,2,\dots,m$ στον κύκλο $t+1$, θα έχει τη μορφή

$$\begin{aligned}
w_k^{(t+1,j)} &= w_k^{(t,j)} + \alpha_k^{(t+1)} \sum_{i=1}^m (d_j^i - y_j^{(t+1,i)}) x_k^i \\
&= w_k^{(t,j)} + \frac{\delta^{(t+1)} \delta^{(t+1)}}{\delta^{(t+1)} X^T X \delta^{(t+1)}} \sum_{i=1}^m (d_j^i - \sum_{q=1}^n w_q^{(t,j)} x_q^i) x_k^i \\
&= w_k^{(t,j)} + \frac{\delta^{(t+1)} \delta^{(t+1)}}{\delta^{(t+1)} C \delta^{(t+1)}} \sum_{i=1}^m (c_k^i - \sum_{q=1}^n w_q^{(t,j)} b_q^i)
\end{aligned} \tag{6.44}$$

όπου $k=1,2,\dots,n$ και b_q^k, c_k^i τα αντίστοιχα στοιχεία των πινάκων B και C , όπως ορίστηκαν στην (6.32).

6.3.7 Προσομοίωση της μεθόδου Polak-Ribiere (PR)

Με το $\delta^{(t+1)}$ όπως ορίστηκε στην (6.42), ορίζουμε το συντελεστή εκπαίδευσης $\beta_k^{(t+1)}$ σαν

$$\beta_k^{(t+1)} = \frac{(\delta^{(t+1)} - \delta^{(t)})\delta_k^{(t+1)}}{\delta^{(t)}\delta^{(t)}} \quad (6.45)$$

οπότε, η διόρθωση των συνάψεων μετά την είσοδο όλων των προτύπων $x^i, i=1,2,\dots,m$ στον κύκλο $t+1$, θα έχει τη μορφή

$$\begin{aligned} w_k^{(t+1,j)} &= w_k^{(t,j)} - \omega \sum_{i=1}^m (d_j^i - y_j^{(t+1,i)})x_k^i + \beta^{(t+1)}\Delta w_k^{(t,j)} \\ &= w_k^{(t,j)} - \omega \sum_{i=1}^m (d_j^i - \sum_{q=1}^n w_q^{(t,j)}x_q^i)x_k^i + \frac{(\delta^{(t+1)} - \delta^{(t)})\delta^{(t+1)}}{\delta^{(t)}\delta^{(t)}}\Delta w_k^{(t,j)} \\ &= w_k^{(t,j)} - \omega \sum_{i=1}^m (c_j^i - \sum_{q=1}^n w_q^{(t,j)}b_q^i) + \frac{(\delta^{(t+1)} - \delta^{(t)})\delta^{(t+1)}}{\delta^{(t)}\delta^{(t)}}\Delta w_k^{(t,j)} \end{aligned} \quad (6.46)$$

όπου $k=1,2,\dots,n$, ω ο συντελεστής εκπαίδευσης και b_q^k, c_k^i τα αντίστοιχα στοιχεία των πινάκων B και C , όπως ορίστηκαν στην (6.32).

6.3.8 Προσομοίωση της μεθόδου Fletcher-Reeves (FR)

Με το $\delta^{(t+1)}$ όπως ορίστηκε στην (6.42), ορίζουμε το συντελεστή εκπαίδευσης $\beta_k^{(t+1)}$ σαν

$$\beta_k^{(t+1)} = \frac{\delta^{(t+1)}\delta_k^{(t+1)}}{\delta^{(t)}\delta^{(t)}} \quad (6.47)$$

οπότε, η διόρθωση των συνάψεων μετά την είσοδο όλων των προτύπων $x^i, i=1,2,\dots,m$ στον κύκλο $t+1$, θα έχει τη μορφή

$$w_k^{(t+1,j)} = w_k^{(t,j)} - \omega \sum_{i=1}^m (d_j^i - y_j^{(t+1,i)})x_k^i + \beta^{(t+1)}\Delta w_k^{(t,j)}$$

$$\begin{aligned}
&= w_k^{(t,j)} - \omega \sum_{i=1}^m (d_j^i - \sum_{q=1}^n w_q^{(t,j)} x_q^i) x^i + \frac{\delta^{(t+1)} \delta^{(t+1)}}{\delta^{(t)} \delta^{(t)}} \Delta w_k^{(t,j)} \\
&= w_k^{(t,j)} - \omega \sum_{i=1}^m (c_j^i - \sum_{q=1}^n w_q^{(t,j)} b_q^i) + \frac{\delta^{(t+1)} \delta^{(t+1)}}{\delta^{(t)} \delta^{(t)}} \Delta w_k^{(t,j)} \quad (6.48)
\end{aligned}$$

όπου $k=1,2,\dots,n$, ω ο συντελεστής εκπαίδευσης και b_q^k , c_k^i τα αντίστοιχα στοιχεία των πινάκων B και C , όπως ορίστηκαν στην (6.32).

6.3.9 Προσομοίωση της μεθόδου Optimal Fletcher-Reeves (OFR)

Η διόρθωση των συνάψεων μετά την είσοδο όλων των προτύπων x^i , $i=1,2,\dots,m$ στον κύκλο $t+1$, θα έχει τη μορφή

$$\begin{aligned}
w_k^{(t+1,j)} &= w_k^{(t,j)} + \alpha_k^{(t+1)} \sum_{i=1}^m (d_j^i - y_j^{(t+1,i)}) x_k^i + \beta^{(t+1)} \Delta w_k^{(t,j)} \\
&= w_k^{(t,j)} + \frac{\delta^{(t+1)} \delta^{(t+1)}}{\delta^{(t+1)} C \delta^{(t+1)}} \sum_{i=1}^m (c_k^i - \sum_{q=1}^n w_q^{(t,j)} b_q^i) + \frac{\delta^{(t+1)} \delta^{(t+1)}}{\delta^{(t)} \delta^{(t)}} \Delta w_k^{(t,j)} \quad (6.44)
\end{aligned}$$

όπου $k=1,2,\dots,n$ και b_q^k , c_k^i τα αντίστοιχα στοιχεία των πινάκων B και C , όπως ορίστηκαν στην (6.32).

6.3.10 Προσομοίωση της μεθόδου RPROP

Η διόρθωση των συντελεστών $\Delta_k^{(t+1,j)}$ μετά την είσοδο όλων των προτύπων x^i , $i=1,2,\dots,m$ στον κύκλο $t+1$, θα έχει τη μορφή

$$\Delta_k^{(t+1,j)} = \begin{cases} \Delta_k^{(t,j)} * u, & \alpha v \frac{\partial \mathcal{E}(w^{(t)})}{\partial w_{kj}} * \frac{\partial \mathcal{E}(w^{(t-1)})}{\partial w_{kj}} > 0 \\ \Delta_k^{(t,j)} * d, & \alpha v \frac{\partial \mathcal{E}(w^{(t)})}{\partial w_{kj}} * \frac{\partial \mathcal{E}(w^{(t-1)})}{\partial w_{kj}} < 0 \\ \Delta_{max}, & \alpha v \Delta_k^{(t,j)} > \Delta_{max} \\ \Delta_{min}, & \alpha v \Delta_k^{(t,j)} \leq \Delta_{min} \end{cases} \quad (6.45)$$

όπου $\Delta_{max}=50$, $\Delta_{min}=0.000001$, $u=1.2$ και $d=0.5$. Στη συνέχεια, αν ισχύει $\frac{\partial \mathcal{E}(w^{(t)})}{\partial w_{kj}} * \frac{\partial \mathcal{E}(w^{(t-1)})}{\partial w_{kj}} \geq 0$, τροποποιούνται οι συνάψεις σύμφωνα με τον τύπο

$$\Delta w_k^{(t+1,j)} = \begin{cases} -\Delta_k^{(t,j)} & \alpha\nu \frac{\partial E(w^{(t)})}{\partial w_{kj}} > 0 \\ +\Delta_k^{(t,j)} & \alpha\nu \frac{\partial E(w^{(t)})}{\partial w_{kj}} < 0 \end{cases} \quad (6.46)$$

ενώ, αν ισχύει $\frac{\partial E(w^{(t)})}{\partial w_{kj}} * \frac{\partial E(w^{(t-1)})}{\partial w_{kj}} < 0$, οι συνάψεις παίρνουν τις παλιές τιμές, δηλαδή

$$w_k^{(t+1,j)} = w_k^{(t-1,j)} \quad (6.47)$$

6.3.11 Σύγκλιση των μεθόδων 6.3.1-6.3.10

Το πλεονέκτημα των ομαδικών μεθόδων (Richardson ή batch-LMS, JOR), είναι η παραλληλοποίησή τους, αφού σε κάθε κύκλο εκπαίδευσης παίρνουμε τις εξόδους για κάθε πρότυπο (τη γραμμή του Πίνακα X^T). Στις βηματικές μεθόδους (SOR, και Richardson-Gauss-Seidel) δεν υπάρχει παραλληλοποίηση, αφού τροποποιούνται μόνο οι συνάψεις του νευρώνα εξόδου με ένα νευρώνα εισόδου σε κάθε κύκλο εκπαίδευσης, αλλά οι κύκλοι διόρθωσης συνάψεων είναι πολύ λιγότεροι.

Όλοι οι αλγόριθμοι είναι αλγόριθμοι κλιμακωτής μείωσης, δηλαδή για κάθε διάνυσμα w^j που ικανοποιεί τη σχέση $\nabla E(w^j) \neq 0$ τροποποιούν το w^j κατά μήκος μιας φθίνουσας κατεύθυνσης.

Με τη χρήση του Λήμματος Μείωσης (Descent Lemma) του Θεωρήματος 2.6 και του Θεωρήματος Συνέχειας του Lipschitz (Lipschitz Continuity) [BeTs89] αποδεικνύεται η σύγκλιση της μεθόδου Richardson-Gauss-Seidel.

Με την προϋπόθεση ότι $\nabla_{kk}^2 E(w^j) > 0, \forall w^j \in R^n$, ο αλγόριθμος JOR μπορεί να θεωρηθεί σαν ένας κλιμακωτός αλγόριθμος μείωσης [BeTs89] αφού το k στοιχείο της μεταβολής $-\omega \nabla E((w^j)^{(t)})$ διαιρείται με $\frac{1}{\nabla_{kk}^2 E((w^j)^{(t)})}$.

Ο αλγόριθμος SOR μπορεί να θεωρηθεί σαν ένας κλιμακωτός αλγόριθμος της μεθόδου RGS, και συγκλίνει για $0 < \omega < 2$ [BeTs89]. Η μέθοδος Gauss-Seidel είναι ειδική περίπτωση της SOR.

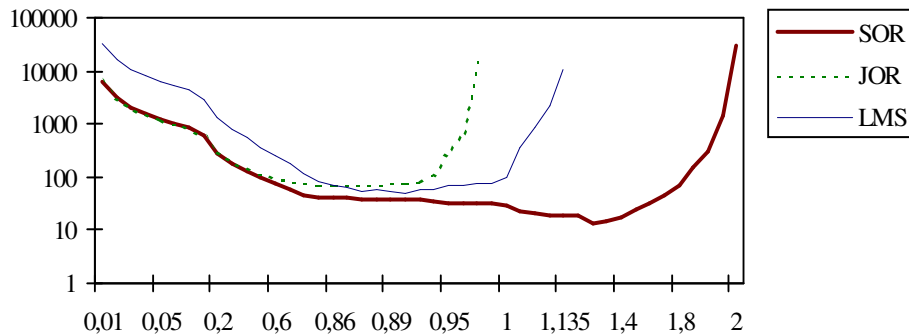
Η μέθοδος Jacobi μπορεί να θεωρηθεί σαν μια προσέγγιση της μεθόδου του Newton, όπου αγνοούνται τα μη διαγώνια στοιχεία του πίνακα $B = XX^T$ [BeLC88]. Είναι μέθοδος κλιμακωτής μείωσης [BeTs89], αλλά δεν συγκλίνει πάντοτε (η σύγκλιση εξαρτάται απ' τις ιδιοτιμές του πίνακα XX^T). Για να αποφύγουμε την αναζήτηση σε λάθος κατευθύνσεις, και τη μη σύγκλιση του αλγορίθμου, το $\nabla_{kk}^2 E((w^j)^{(t)})$ μπορεί να αντικατασταθεί με το $|\nabla_{kk}^2 E((w^j)^{(t)})| + \mu$, όπου μ μια σταθερά με μικρή τιμή.

7

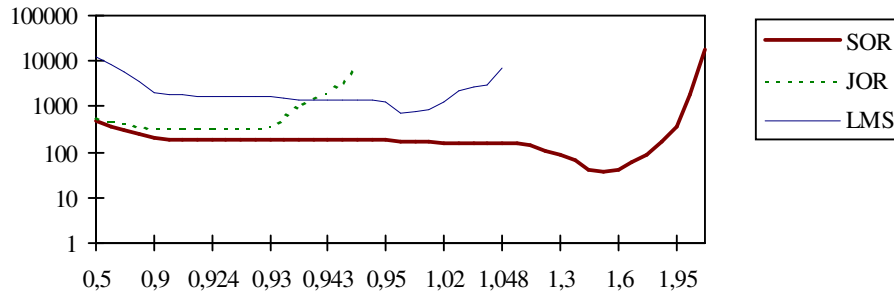
Απόδοση των Προτεινόμενων Αλγορίθμων στην Επίλυση Γραμμικών Συστημάτων και Προβλημάτων Ταξινόμησης και Προσαρμογής

7.1 Αρχιτεκτονική του Σχήματος 6.1 με Συμμετρική και Θετικά Ορισμένη Μήτρα A

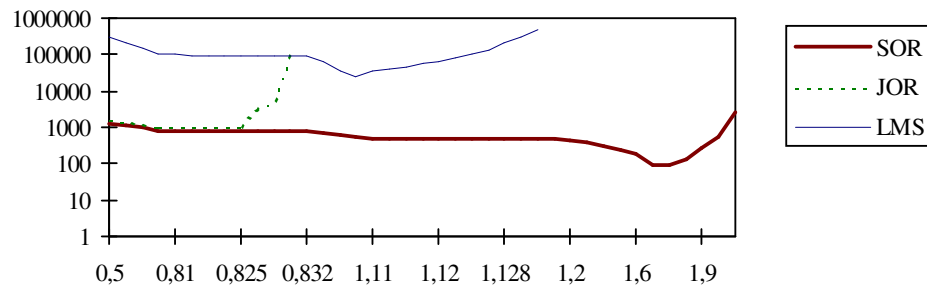
Στην αρχιτεκτονική που προτάθηκε για την επίλυση του συστήματος $Ax = b$ ή του $Cx = d$ της εξίσωσης (6.3), αν η μήτρα A δεν είναι Συμμετρική και Θετικά Ορισμένη συγκρίνονται οι μέθοδοι SOR και JOR με τη μέθοδο LMS. Για το σκοπό αυτό χρησιμοποιήθηκαν διάφορα συστήματα με τυχαίες τιμές, με διάσταση $n=3$, $n=5$ και $n=10$. Το πρώτο πείραμα εξέταζε την επίδραση του ω ($0 < \omega < 2$) στην ταχύτητα της σύγκλισης, τον αριθμό επαναλήψεων που απαιτούνταν για τη σύγκλιση της κάθε μεθόδου, όπως φαίνεται στα διαγράμματα 7.1, 7.2 και 7.3.



Διάγραμμα 7.1. Μεταβολή των Επαναλήψεων ως προς το ω για Σύστημα (5×5) $Ax=b$



Διάγραμμα 7.2. Μεταβολή των Επαναλήψεων ως προς το ω για Σύστημα $(10 \times 10) Ax=b$

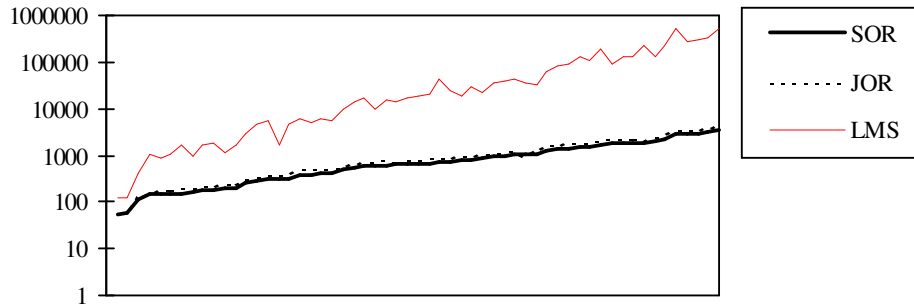


Διάγραμμα 7.3. Μεταβολή των Επαναλήψεων ως προς το ω για Σύστημα $(20 \times 20) Ax=b$

Για τα προηγούμενα παραδείγματα υπάρχει μια βέλτιστη τιμή του ω για κάθε μέθοδο, όπου το κάθε δίκτυο απαιτεί τον ελάχιστο αριθμό επαναλήψεων για να συγκλίνει. Η μέθοδος SOR συγκλίνει για οποιαδήποτε τιμή του ω ($0 < \omega < 2$), ενώ οι μέθοδοι JOR και LMS δεν συγκλίνουν, μετά από κάποια μέγιστη τιμή του ω , ω_{max} . Καθώς το μέγεθος του προβλήματος αυξάνεται, το ω_{max} μειώνεται για τη JOR, ενώ αυξάνεται για την LMS. Καθώς το ω πλησιάζει το ανώτατο ή κατώτατο όριο, ο αριθμός των επαναλήψεων αυξάνεται δραματικά.

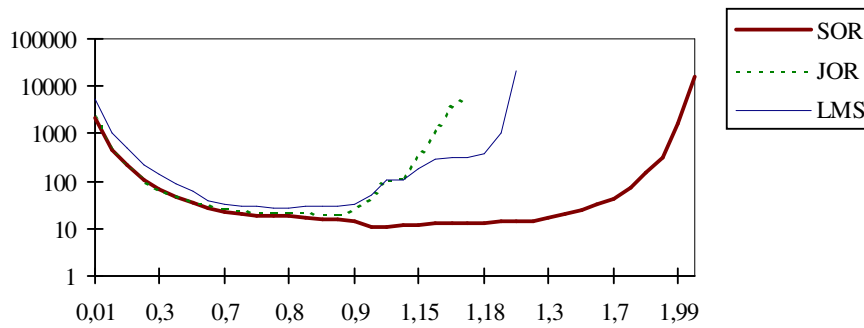
Για όλα τα συστήματα που ελέγχθηκαν, η μέθοδος SOR συγκλίνει πιο γρήγορα απ' τις υπόλοιπες. Η JOR έχει καλύτερη συμπεριφορά απ' την LMS, όταν το ω δεν βρίσκεται κοντά στο ω_{max} .

Σε ένα άλλο πείραμα με σταθερό ω ($\omega=0,4$), ένας αριθμός συστημάτων επιλύθηκε με τις τρεις παραπάνω μεθόδους και η μεταβολή του αριθμού των επαναλήψεων φαίνεται στο Διάγραμμα 7.4. Και σ' αυτό το πείραμα, η SOR υπερτερεί ως προς της υπόλοιπες στην ταχύτητα σύγκλισης.

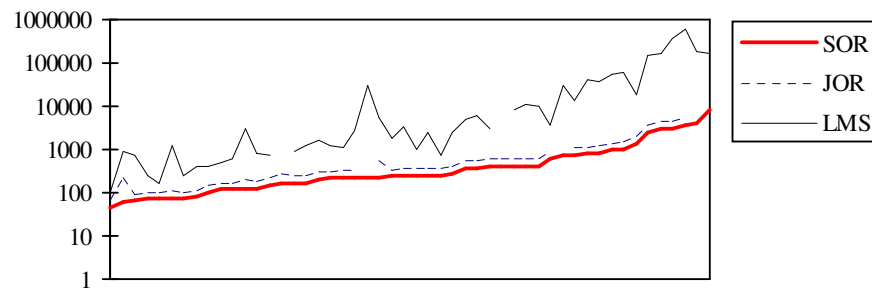


Διάγραμμα 7.4. Μεταβολή των Επαναλήψεων για 77 Συστήματα (5x5) $Ax=b$

Σε ένα άλλο πείραμα, οι μέθοδοι SOR και JOR συγκρίνονται με την LMS για τον υπολογισμό του αντιστρόφου ενός πίνακα 3×3 για διάφορες τιμές του ω ($0 < \omega < 2$) και για 50 συστήματα 3×3 με σταθερή τιμή του ω ($\omega=0.7$). Τα αποτελέσματα, όπως φαίνεται στα διαγράμματα 7.5 και 7.6, είναι τα ίδια με τα προηγούμενα πειράματα.



Διάγραμμα 7.5. Μεταβολή των Επαναλήψεων ως προς το ω για Σύστημα (5x5) $AX=I$



Διάγραμμα 7.6. Μεταβολή των Επαναλήψεων για 50 Συστήματα (5x5) $AX=I$

7.2 Αρχιτεκτονική του Σχήματος 6.2 - Αλγόριθμοι HSD και ASD

Η απόδοση των αλγορίθμων στην επίλυση συστημάτων της μορφής $Ax = b$ ελέγχεται με συγκεκριμένα παραδείγματα και συγκρίνεται η λύση τους με τη *Λύση Ελαχίστων Τετραγώνων*, *Least Square Solution*, LS. Παρουσιάζονται επίσης και τα αντίστοιχα διαγράμματα που αφορούν τη σύγκλιση των αλγορίθμων, ως προς την ελαχιστοποίηση του *Μέσου Τετραγωνικού Σφάλματος* (*Mean Square Error*).

Παράδειγμα 1: Το Τετραγωνικό Σύστημα προς επίλυση είναι το

$$\begin{bmatrix} 1.2 & 0.8 & 0.7 & 0.5 \\ 0.4 & 1.5 & 0.3 & 0.1 \\ 0.1 & 0.5 & 1.7 & 0.9 \\ 0.1 & 0.5 & 0.6 & 1.2 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} \quad (7.1)$$

Με μηδενικές αρχικές τιμές στις συνάψεις και

$$E(x) = \|Ax - b\|^2 \quad (7.2)$$

το τελικό σφάλμα, όπως χρησιμοποιήθηκε απ' τους Wang L.X. και Mendel J. [WaMe91], η σύγκλιση των τριών αλγορίθμων φαίνεται στο Διάγραμμα 7.7, όπου ο οριζόντιος άξονας απεικονίζει τον κύκλο εκπαίδευσης t και ο κάθετος άξονας το Μέσο Τετραγωνικό Σφάλμα, όπως δίνεται στην (7.2). Με $\varepsilon = 10^{-5}$, ο αλγόριθμος ASD συγκλίνει σε 50 κύκλους στη λύση

$$x_{ASD} = [x_1, x_2, x_3, x_4]^T = [-1.30014, 1.51324, -0.12329, 2.87085]^T$$

με

$$r_{ASD} = [r_1, r_2, r_3, r_4]^T = [0.00046, 0.00010, -0.00078, 0.00235]^T$$

μια λύση πολύ κοντά στη λύση Ελαχίστων Τετραγώνων του συστήματος (7.1)

$$x_{LS} = [-1.29991, 1.51346, -0.12535, 2.87372]^T \quad (7.3)$$

την οποία παίρνουμε με τον αλγόριθμο του Greville και καλύτερη απ' την προσέγγιση των Wang L.X. και Mendel J. [WaMe91]

$$x = [-1.21653, 1.47053, 0.16055, 2.35300]^T$$

$$r = [0.00548, 0.00265, 0.00465, 0.00215]^T$$

Με μηδενικές αρχικές τιμές στις συνάψεις και $\alpha = 0.01$ σαν συντελεστή εκπαίδευσης στις μεθόδους LMS και HSD, ο αλγόριθμος HSD (ισοδύναμος της Batch-LMS) συγκλίνει σε 1351 κύκλους, με $\varepsilon = 10^{-5}$, στη λύση

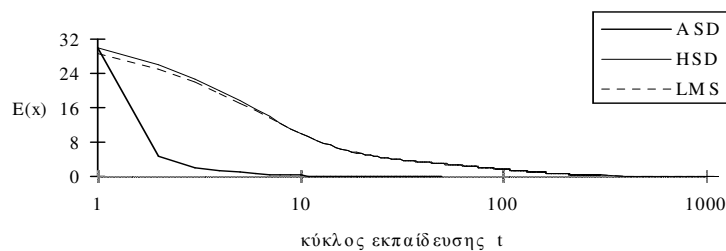
$$x_{HSD} = [-1.29998, 1.51329, -0.12241, 2.87013]^T$$

$$r_{HSD} = [-0.00003, -0.00023, -0.00167, 0.00265]^T$$

ενώ ο αλγόριθμος LMS συγκλίνει σε 1341 κύκλους, με $\varepsilon = 10^{-5}$, στη λύση

$$x_{LMS} = [-1.30002, 1.51330, -0.12242, 2.87016]^T$$

$$r_{LMS} = [-0.00001, -0.00024, -0.00168, 0.00261]^T$$



Διάγραμμα 7.7 Σύγκλιση LMS, HSD, και ASD για το σύστημα (7.1) του Παραδείγματος 1

Όπως φαίνεται στο Διάγραμμα 7.8, η συμπεριφορά των LMS και HSD είναι παρόμοια, όπως έχει δειχθεί [WiHo60], [Batt92], ενώ με τη μέθοδο ASD το σφάλμα μειώνεται δραματικά και ο αριθμός των επαναλήψεων που απαιτούνται είναι πολύ μικρός συγκριτικά με τις άλλες δύο μεθόδους. Αν χρησιμοποιήσουμε σαν $\varepsilon = 10^{-9}$ οι προηγούμενοι αλγόριθμοι συγκλίνουν στη Λύση Ελαχίστων Τετραγώνων (7.3) του συστήματος (7.1), αλλά αυξάνεται ο αριθμός των επαναλήψεων που απαιτούνται για τη σύγκλιση.

Παράδειγμα 2 : Το Αόριστο Σύστημα προς επίλυση είναι το

$$\begin{bmatrix} 2 & -1 & 4 & 0 & 3 & 1 \\ 5 & 1 & -3 & 1 & 2 & 0 \\ 1 & -2 & 1 & -5 & -1 & 4 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ -4 \end{bmatrix} \quad (7.4)$$

Με μηδενικές αρχικές τιμές στις συνάψεις και

$$E(x) = \frac{1}{2} \|Ax - b\|^2 \quad (7.5)$$

το τελικό σφάλμα, όπως ορίστηκε στην (3.23), η σύγκλιση των τριών αλγορίθμων φαίνεται στο Διάγραμμα 7.8, όπου ο οριζόντιος άξονας απεικονίζει τον κύκλο εκπαίδευσης t και ο κάθετος άξονας το Μέσο Τετραγωνικό Σφάλμα, όπως δίνεται στην (7.5). Με σφάλμα $\varepsilon = 10^{-5}$, ο αλγόριθμος ASD συγκλίνει σε 8 κύκλους στη λύση

$$x_{ASD} = [0.08826, 0.10826, 0.27321, 0.50457, 0.38275, -0.30965]^T$$

$$r_{ASD} = [0.00029, 0.00001, -0.00076]^T$$

μια λύση πολύ κοντά στη λύση Ελαχίστων Τετραγώνων του συστήματος (7.4) και κοντά στην προσέγγιση των Cichocki και Unbehauen [CiUn94a]

$$x = [0.0882, 0.1083, 0.2733, 0.5047, 0.3828, -0.3097]^T$$

Με μηδενικές αρχικές τιμές στις συνάψεις και $\alpha = 0.01$ σαν συντελεστή εκπαίδευσης στις μεθόδους LMS και HSD, ο αλγόριθμος HSD συγκλίνει σε 24 κύκλους, με $\varepsilon = 10^{-5}$, στη λύση

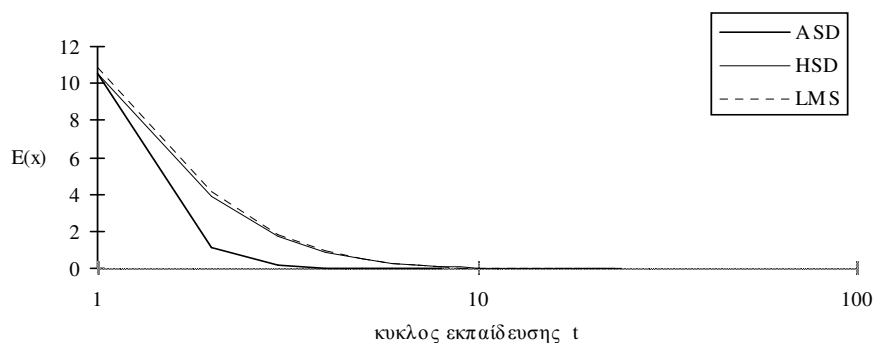
$$x_{HSD} = [0.08832, 0.10832, 0.27278, 0.50444, 0.38254, -0.30960]^T$$

$$r_{HSD} = [0.00254, -0.00112, -0.00133]^T$$

ενώ ο αλγόριθμος LMS συγκλίνει σε 23 κύκλους, $\varepsilon = 10^{-5}$, στη λύση

$$x_{LMS} = [0.08827, 0.10834, 0.27281, 0.50451, 0.38255, -0.30965]^T$$

$$r_{LMS} = [0.00256, -0.00085, -0.00071]^T$$



Διάγραμμα 7.8 Σύγκλιση LMS, HSD, και ASD για το Αόριστο σύστημα (7.4) του Παραδείγματος 2

Παράδειγμα 3 : Το Υπερορισμένο Σύστημα προς επίλυση είναι το

$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \\ 3 \end{bmatrix} \quad (7.6)$$

Αφού το σύστημα (7.6) είναι υπερορισμένο, οι αλγόριθμοι θα συγκλίνουν στη λύση ελαχίστων τετραγώνων, αλλά το σφάλμα, όπως ορίστηκε στην (7.5), θα ελαχιστοποιηθεί, αλλά δεν θα μηδενισθεί. Έτσι, ορίζουμε το κριτήριο σύγκλισης

$$\Delta x = \|x^{(t+1)} - x^{(t)}\|_1 = \sum_{j=1}^n |x_j^{(t+1)} - x_j^{(t)}| \leq 10^{-5} \quad (7.7)$$

για να τερματίσουμε τους αλγόριθμους, με $t=0,1,\dots$ τον κύκλο εκπαίδευσης. Η σύγκλιση των τριών αλγορίθμων φαίνεται στο Διάγραμμα 7.9, όπου ο οριζόντιος άξονας απεικονίζει τον κύκλο εκπαίδευσης t και ο κάθετος άξονας το Μέσο Τετραγωνικό Σφάλμα, όπως δίνεται στην (7.5). Με $\varepsilon = 10^{-5}$, ο αλγόριθμος ASD συγκλίνει σε 4 κύκλους εκπαίδευσης, με $E(x) = 0.20000$, στη λύση

$$x_{ASD} = [x_1, x_2]^T = [0.20000, 0.60000]^T$$

$$r_{ASD} = [r_1, r_2, r_3, r_4, r_5]^T = [-0.20000, 0.40000, 0.00000, -0.40000, 0.20000]^T$$

που είναι η λύση Ελαχίστων Τετραγώνων του συστήματος (7.6), όπως υπολογίζεται με τον αλγόριθμο του Greville

$$x_{LS} = [x_1, x_2]^T = [0.20000, 0.60000]^T \quad (7.8)$$

και καλύτερη απ' τη λύση των Cichocki και Unbehauen [CiUn92a]

$$x = [0.206, 0.598]^T$$

$$r = [-0.196, 0.402, 0, -0.402, 0.196]^T$$

Με μηδενικές αρχικές τιμές στις συνάψεις και $\alpha = 0.01$ σαν συντελεστή εκπαίδευσης στις μεθόδους LMS και HSD, ο αλγόριθμος HSD συγκλίνει σε 417 κύκλους, με $E(x) = 0.20000$, στη λύση

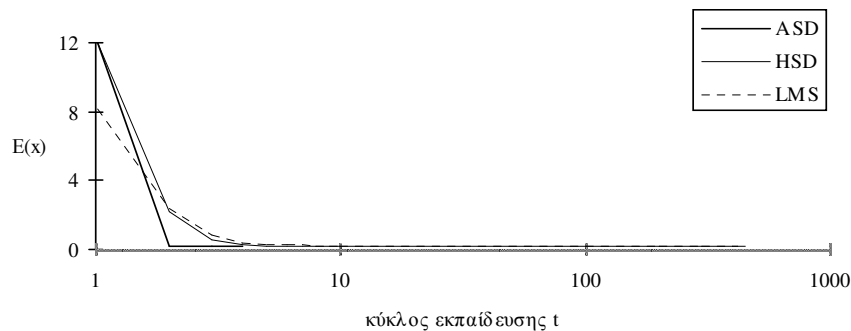
$$x_{HSD} = [0.19909, 0.60025]^T$$

$$r_{HSD} = [-0.20066, 0.39959, -0.00015, -0.39990, 0.20035]^T$$

ενώ ο αλγόριθμος LMS συγκλίνει σε 456 κύκλους, με $E(x) = 0.22634$, στη λύση

$$x_{LMS} = [0.22298, 0.59080]^T$$

$$r_{LMS} = [-0.18622, 0.40459, -0.00461, -0.41381, 0.17699]^T$$



Διάγραμμα 7.8 Σύγκλιση LMS, HSD, και ASD για το υπεριορισμένο σύστημα (7.6) του Παραδείγματος 3

7.2.1 Συμπεράσματα

Για Τετραγωνικά και Αόριστα Συστήματα Εξισώσεων, όπως στις εξισώσεις (7.1), (7.4), οι τρεις αλγόριθμοι, για μικρές τιμές του συντελεστή εκπαίδευσης οι HSD και LMS συγκλίνουν στη μοναδική Λύση Ελαχίστων Τετραγώνων (Least Square Solution) $A^{-1}b$ για Τετραγωνικά Συστήματα ή $A^{+}b$ για Αόριστα Συστήματα, ή μια προσέγγιση της Λύσης Ελαχίστων Τετραγώνων για Αόριστα Συστήματα. Η απόδοση της μεθόδου ASD γι' αυτά τα συστήματα είναι καλύτερη απ' την απόδοση των μεθόδων HSD και LMS (ο αριθμός των κύκλων εκπαίδευσης που απαιτούνται για τη σύγκλιση είναι μικρότερος για την ASD).

Για Υπέρ-ορισμένα Συστήματα Εξισώσεων, όπως στην εξίσωση (7.6), μόνο η μέθοδος ASD συγκλίνει στη Λύση Ελαχίστων Τετραγώνων $A^{+}b$. Οι μέθοδοι HSD και LMS συγκλίνουν σε μια προσέγγιση του $A^{+}b$, με καλύτερη προσέγγιση αυτή που δίνει η μέθοδος HSD. Και γι' αυτά τα συστήματα η απόδοση της μεθόδου ASD είναι καλύτερη απ' την απόδοση των μεθόδων HSD και LMS.

Παράδειγμα 4: Να βρεθεί ο ψευδο-αντίστροφος του πίνακα

$$A = \begin{bmatrix} 1.2 & 0.8 & 0.7 & 0.5 \\ 0.4 & 1.5 & 0.3 & 0.1 \\ 0.1 & 0.5 & 1.7 & 0.9 \\ 0.1 & 0.5 & 0.6 & 1.2 \end{bmatrix} \quad (7.9)$$

Αν χρησιμοποιήσουμε το δίκτυο του Σχήματος (6.3) ή τέσσερις φορές το δίκτυο του Σχήματος (6.2) και τα μοναδιαία διανύσματα $[1, 0, 0, 0]^T$, $[0, 1, 0, 0]^T$, $[0, 0, 1, 0]^T$, $[0, 0, 0, 1]^T$ σαν στοιχεία του πίνακα B , βρίσκουμε τις τέσσερις στήλες του πίνακα $X = A^{+}$.

Με μηδενικές αρχικές τιμές στις συνάψεις, και $\alpha = 0.01$ σαν συντελεστή εκπαίδευσης στις μεθόδους HSD και LMS, ο αλγόριθμος ASD συγκλίνει με $\varepsilon = 10^{-5}$, σε 30 κύκλους για τη στήλη 1, σε 37 κύκλους για τη στήλη 2, σε 36 κύκλους για τη στήλη 3 και σε 40 κύκλους για τη στήλη 4, στη λύση

$$X_{ASD} = A^+ = \begin{bmatrix} 0.99493 & -0.37772 & -0.28277 & -0.17232 \\ -0.26761 & 0.79883 & -0.06353 & 0.09261 \\ 0.00769 & -0.07447 & 0.80851 & -0.60289 \\ 0.02377 & -0.26244 & -0.35240 & 1.10766 \end{bmatrix}$$

μια καλή προσέγγιση της Λύσης Ελαχίστων Τετραγώνων που παίρνουμε με τον αλγόριθμο του Greville

$$X_{LS} = A^{-1} = \begin{bmatrix} 0.99814 & -0.38069 & -0.28319 & -0.17177 \\ -0.26927 & 0.80037 & -0.06314 & 0.09285 \\ 0.00696 & -0.07242 & 0.81198 & -0.60585 \\ 0.02553 & -0.26555 & -0.35608 & 1.11188 \end{bmatrix} \quad (7.10)$$

ο αλγόριθμος HSD συγκλίνει με $\varepsilon = 10^{-5}$, σε 843 κύκλους για τη στήλη 1, σε 841 κύκλους για τη στήλη 2, σε 1059 κύκλους για τη στήλη 3 και σε 1149 κύκλους για τη στήλη 4, στη λύση

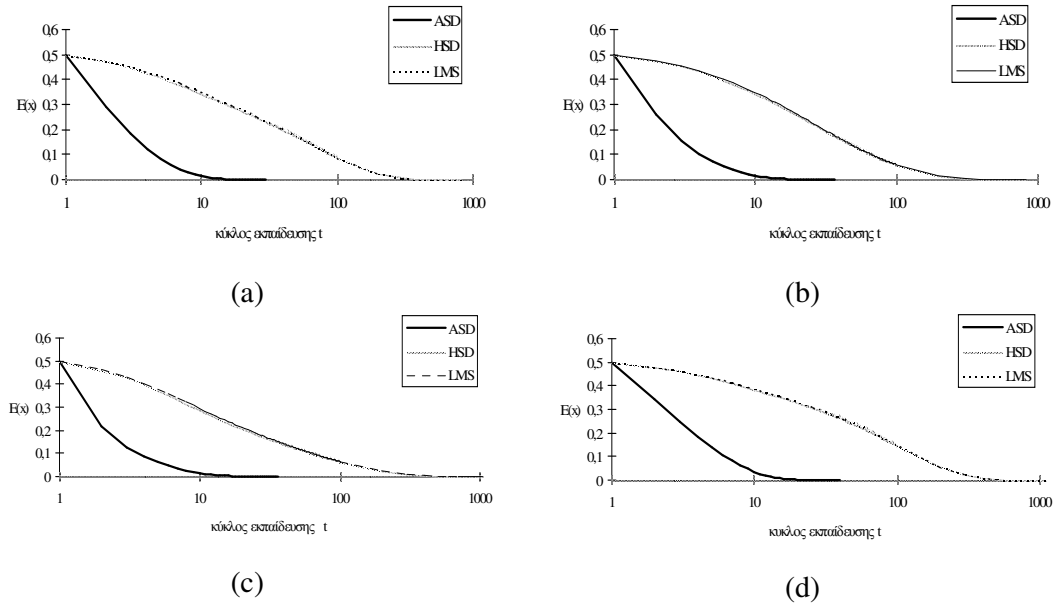
$$X_{HSD} = A^+ = \begin{bmatrix} 0.99357 & -0.37687 & -0.28241 & -0.17223 \\ -0.26638 & 0.79808 & -0.06333 & 0.09283 \\ 0.00809 & -0.07493 & 0.80786 & -0.60170 \\ 0.02376 & -0.26215 & -0.35099 & 1.10678 \end{bmatrix}$$

και ο αλγόριθμος LMS συγκλίνει με $\varepsilon = 10^{-5}$, σε 834 κύκλους για τη στήλη 1, σε 834 κύκλους για τη στήλη 2, σε 1049 κύκλους για τη στήλη 3 και σε 1140 κύκλους για τη στήλη 4, στη λύση

$$X_{LMS} = A^+ = \begin{bmatrix} 0.99357 & -0.37687 & -0.28238 & -0.17227 \\ -0.26641 & 0.79811 & -0.06334 & 0.09284 \\ 0.00794 & -0.07485 & 0.80790 & -0.60174 \\ 0.02394 & -0.26227 & -0.35107 & 1.10686 \end{bmatrix}$$

Η συμπεριφορά των τριών μεθόδων στην εύρεση του αντιστρόφου του πίνακα του Παραδείγματος 4 φαίνεται στο Διάγραμμα (7.9). Όπως βλέπουμε, η συμπεριφορά των HSD και LMS είναι σχεδόν ίδια, ενώ απαιτείται σχεδόν ο ίδιος αριθμός επαναλήψεων για να συγκλίνουν. Οι επαναλήψεις που απαιτούνται για τη μέθοδο ASD είναι πολύ

λίγες συγκριτικά με τις άλλες μεθόδους, ενώ το σφάλμα μειώνεται δραματικά. Αν χρησιμοποιήσουμε $\varepsilon = 10^{-9}$, οι παραπάνω αλγόριθμοι συγκλίνουν στη Λύση Ελαχίστων Τετραγώνων (7.10), αλλά αυξάνεται υπερβολικά ο αριθμός των επαναλήψεων που απαιτούνται.



Διάγραμμα 7.9 Σύγκλιση LMS, HSD, και ASD για τις στήλες 1 (a), 2 (b), 3 (c) και 4 (d) του πίνακα A , για την εύρεση του αντιστρόφου του Παραδείγματος 4.

7.3 Εφαρμογή της Γενικευμένης Αρχιτεκτονικής 6.3 σε Προβλήματα Ταξινόμησης και Προσαρμογής

Πολλά προβλήματα σε διάφορους τομείς επιστημονικών εφαρμογών καταλήγουν σε ένα σύστημα γραμμικών εξισώσεων (συνήθως υπέρ-ορισμένο) του οποίου ζητείται η λύση ή η αντιστροφή ενός πίνακα σε πραγματικό χρόνο. Μια συλλογή από προβλήματα αυτού του τύπου είναι και η συλλογή Proben1, η οποία περιέχει πραγματικά δεδομένα (real world data) για 14 εφαρμογές που μπορούν να χρησιμοποιηθούν για την εκτίμηση αλγορίθμων νευρωνικών δικτύων και τη σύγκριση των αποτελεσμάτων. Οι παραπάνω εφαρμογές μπορούν να χωριστούν σε *προβλήματα ταξινόμησης (pattern classification)* και *προβλήματα προσαρμογής συναρτήσεων (function approximation)*. Όλα τα προβλήματα απαιτούν εκπαίδευση με επίβλεψη (supervised learning) και μπορούν να χαρακτηρισθούν σαν στατικά, επειδή τα δεδομένα που πρέπει να μάθουν τα δίκτυα δεν αλλάζουν κατά τη διάρκεια της εκπαίδευσης. Απ' την παραπάνω συλλογή δεν εξετάζουμε το πρόβλημα με τα δηλητηριώδημανιτάρια, το οποίο δεν έχει πραγματικά δεδομένα.

Τα πιο πολλά προβλήματα που χρησιμοποιήθηκαν για την εκπαίδευση νευρωνικών δικτύων, όπως η πύλη XOR, η γενικευμένη πύλη XOR ή n-parity, ο κωδικοποιητής n-bit encoder, το πρόβλημα της συμμετρίας, το πρόβλημα με τα 2

σπιδράλ (two spirals) ή τους 3 δίσκους [Fahl88], [RuML86], αποτελούνταν από τεχνητά δεδομένα και τα πιο πολλά απ' αυτά μπορούσαν να λυθούν σωστά κατά 100%, αλλά δε μπορούμε να βγάλουμε κανένα συμπέρασμα απ' τα αποτελέσματα για τη συμπεριφορά τους σε πραγματικά προβλήματα.

7.3.1 Περιγραφή των Προβλημάτων Ταξινόμησης

7.3.1.1 Διάγνωση Καρκίνου Μαστού

Το δίκτυο ταξινομεί έναν όγκο σε καλοήγη ή κακοήγη, μετά από μικροσκοπική ανάλυση κυττάρων. Τις εισόδους αποτελούν η πυκνότητα του όγκου, η ομοιομορφία του μεγέθους και του σχήματος του κυττάρου, κ.λ.π.. Τα δεδομένα αποτελούνται από 9 εισόδους και 2 εξόδους και 699 πρότυπα και βασίστηκαν στο σύνολο δεδομένων του καρκίνου στήθους του Wisconsin, “breast cancer Wisconsin”, που βρίσκεται στη συλλογή δεδομένων για την εκμάθηση μηχανών (repository of machine datasets) UCI, ενώ τα πρωτότυπα δεδομένα οφείλονται στο Dr. William H. Wolberg, University of Wisconsin Hospitals, Madison.

7.3.1.2 Έγκριση/Απόρριψη Πιστωτικής Κάρτας

Το δίκτυο ταξινομεί τα στοιχεία της αίτησης ενός πελάτη σε μια τράπεζα για πιστωτική κάρτα σε έγκριση ή απόρριψη. Τις εισόδους αποτελούν στοιχεία που αφορούν εισόδημα, ηλικία, υγεία κ.λ.π.. Τα δεδομένα αποτελούνται από 51 εισόδους και 2 εξόδους και 690 πρότυπα και βασίστηκαν στο σύνολο δεδομένων “credit screening”, που βρίσκεται στη συλλογή δεδομένων για την εκμάθηση μηχανών (repository of machine datasets) UCI.

7.3.1.3 Διάγνωση Διαβήτη

Το δίκτυο ταξινομεί έναν Ινδιάνο της φυλής Pima σε διαβητικό ή μη διαβητικό. Τις εισόδους αποτελούν προσωπικά δεδομένα όπως ηλικία, αριθμός φορών εγκυμοσύνης γυναικών και αποτελέσματα εργαστηριακών αναλύσεων, όπως πίεση αίματος, αντοχή σε γλυκόζη κ.λ.π.. Τα δεδομένα αποτελούνται από 8 εισόδους και 2 εξόδους και 768 πρότυπα και βασίστηκαν στο σύνολο δεδομένων “Pima indians diabetes”, που βρίσκεται στη συλλογή δεδομένων για την εκμάθηση μηχανών (repository of machine datasets) UCI.

7.3.1.4 Κατάταξη Γονιδίων

Το δίκτυο ταξινομεί το εσωτερικό μιας αλληλουχίας στοιχείων DNA (νουκλεϊνοειδή) σε εσωτερικά, εξωτερικά ή κανένα απ' τα παραπάνω. Τις εισόδους αποτελούν τα νουκλεϊνοειδή. Τα δεδομένα αποτελούνται από 120 εισόδους και 3 εξόδους και 3175 πρότυπα και βασίστηκαν στο σύνολο δεδομένων "Splice Junction", που βρίσκεται στη συλλογή δεδομένων για την εκμάθηση μηχανών (repository of machine datasets) UCI.

7.3.1.5 Τύποι Γυαλιού

Το δίκτυο ταξινομεί ένα γυαλί σε επίπεδο ή ανώμαλο τζάμι κτιρίου, τζάμι αυτοκινήτου, μπουκάλι, δοχείο φαγητού ή φανάρι αυτοκινήτου. Τις εισόδους αποτελούν αποτελέσματα χημικής ανάλυσης σε κομμάτια από γυαλιά και χρησιμοποιήθηκαν σε ανάλυση ευρημάτων σε υποθέσεις εγκλημάτων. Τα δεδομένα αποτελούνται από 9 εισόδους και 6 εξόδους και 214 πρότυπα και βασίστηκαν στο σύνολο δεδομένων "Glass", που βρίσκεται στη συλλογή δεδομένων για την εκμάθηση μηχανών (repository of machine datasets) UCI.

7.3.1.6 Διάγνωση Καρδιοπάθειας

Το δίκτυο ταξινομεί έναν ασθενή σε καρδιοπαθή ή μη, ανάλογα με το αν έχει απόφραξη πάνω από 50% στις 4 βασικές αρτηρίες. Τις εισόδους αποτελούν προσωπικά δεδομένα όπως ηλικία, φύλο, βάρος, κάπνισμα και αποτελέσματα εργαστηριακών αναλύσεων, όπως πίεση αίματος, ηλεκτροκαρδιογραφήματα κ.λ.π.. Τα δεδομένα αποτελούνται από 35 εισόδους και 2 εξόδους και 920 πρότυπα και αποτελούνται απ' την ένωση τεσσάρων συνόλων δεδομένων των νοσοκομείων Cleveland Clinic Foundation, Robert Detrano M.D. Ph. D., Hungarian Institute of Cardiology, Budapest, Andras Janosi M.D., V.A. Medical Center Long Beach, Robert Detrano M.D. Ph. D., University Hospital Basel, Switzerland, Matthias Pfisterer, M.D. και University Hospital Zurich, Switzerland, William Steinbrunn, M.D.. Επειδή λείπουν κάποιες τιμές απ' τα παραπάνω δεδομένα, χρησιμοποιήθηκε το σύνολο προτύπων *heartc* που περιέχει 303 πρότυπα απ' το νοσοκομείο Cleveland Clinic Foundation, απ' το οποίο λείπουν μόνο δύο τιμές.

7.3.1.7 Διάγνωση Κολικού Αλόγου

Το δίκτυο προβλέπει αν ένα άλογο με κολικό θα ζήσει, θα πεθάνει ή θα υποστεί ευθανασία.. Τις εισόδους αποτελούν τα αποτελέσματα κτηνιατρικών εξετάσεων. Τα δεδομένα αποτελούνται από 58 εισόδους και 3 εξόδους και 364 πρότυπα και βασίστηκαν στο σύνολο δεδομένων "horse colic", που βρίσκεται στη συλλογή δεδομένων για την εκμάθηση μηχανών (repository of machine datasets) UCI.

7.3.1.8 Αρρώστιες Φασολιών

Το δίκτυο αναγνωρίζει 19 διαφορετικές Αρρώστιες φασολιών. Τις εισόδους αποτελούν η εμφάνιση του φασολιού (μέγεθος, χρώμα) και του φυτού (σημάδια στα φύλλα, ανάπτυξη, ρίζες κ.λ.π.) καθώς και η ιστορία του (αλλαγή σπόρων, θερμοκρασία εδάφους κ.λ.π.. Τα δεδομένα αποτελούνται από 35 εισόδους και 19 εξόδους και 683 πρότυπα και βασίστηκαν στο σύνολο δεδομένων “soybean large”, που βρίσκεται στη συλλογή δεδομένων για την εκμάθηση μηχανών (repository of machine datasets) UCI.

7.3.1.9 Διάγνωση υπέρ-θυρεοειδή ή Υπο-θυρεοειδή

Το δίκτυο ταξινομεί έναν ασθενή σε τρεις κατηγορίες, ανάλογα με το αν στο θυρεοειδή του γίνεται υπερέκκριση, υποέκκριση ή λειτουργεί κανονικά.. Τις εισόδους αποτελούν αποτελέσματα εργαστηριακών εξετάσεων και απαντήσεις του ασθενούς ή μη σε ερωτηματολόγιο. Τα δεδομένα αποτελούνται από 21 εισόδους και 3 εξόδους και 7200 πρότυπα και βασίστηκαν στο σύνολο δεδομένων “ann”, που βρίσκεται στη συλλογή δεδομένων για την εκμάθηση μηχανών (repository of machine datasets) UCI.

7.3.2 Περιγραφή των Προβλημάτων Προσαρμογής

7.3.2.1 Οικονομία Ενέργειας Κτιρίων

Το δίκτυο προβλέπει την ωριαία οικονομία ενός κτιρίου σε ηλεκτρικό, ζεστό νερό και κρύο νερό. Τις εισόδους αποτελούν περιβαλλοντολογικές μετρήσεις, όπως η μέρα, η ώρα, θερμοκρασία εξωτερικού περιβάλλοντος, υγρασία, ηλιακή ακτινοβολία, αέρας κ.λ.π.. Τα δεδομένα αποτελούνται από 14 εισόδους και 3 εξόδους και 4208 πρότυπα, που αποτελούν τις ωριαίες μετρήσεις τεσσάρων συνεχόμενων μηνών και πρόβλεψη για τους επόμενους δύο μήνες και βασίστηκαν στο σύνολο δεδομένων “The Great Energy Predictor Shootout-the first data analysis and prediction problem”, που οργανώθηκε το 1993 για την ASHRAE, στο Denver, Colorado.

7.3.2.2 Ηλιακές Κηλίδες

Το δίκτυο προβλέπει την εμφάνιση μικρών, μεσαίων και μεγάλων ηλιακών κηλίδων στις επόμενες 24 ώρες σε μια συγκεκριμένη περιοχή της επιφάνειας του ηλίου που βρίσκεται σε δραστηριότητα.. Τις εισόδους αποτελούν η προηγούμενη δραστηριότητα της περιοχής και η ιστορία της. Τα δεδομένα αποτελούνται από 24 εισόδους και 3 εξόδους και 1066 πρότυπα και βασίστηκαν στο σύνολο δεδομένων “solar flare”, που βρίσκεται στη συλλογή δεδομένων για την εκμάθηση μηχανών (repository of machine datasets) UCI.

7.3.2.3 Διάγνωση Καρδιοπάθειας

Το δίκτυο ταξινομεί έναν ασθενή σε καρδιοπαθή ή μη, ανάλογα με το αν έχει απόφραξη πάνω από 50% στις 4 βασικές αρτηρίες.. Τις εισόδους αποτελούν προσωπικά δεδομένα όπως ηλικία, φύλο, βάρος, κάπνισμα και αποτελέσματα εργαστηριακών αναλύσεων, όπως πίεση αίματος, ηλεκτροκαρδιογραφήματα κ.λ.π.. Τα δεδομένα αποτελούνται από 35 εισόδους και 2 εξόδους και 303 πρότυπα και βασίζονται στα δεδομένα *heartc* του νοσοκομείου Clinical Clinic Foundation, απ' το οποίο λείπουν μόνο δύο τιμές.

7.3.3 Δεδομένα Εκπαίδευσης - Εγκυρότητας - Ανάκλησης

Τα δεδομένα που χρησιμοποιούνται για τον έλεγχο της απόδοσης των μεθόδων εκπαίδευσης χωρίζονται σε δύο κατηγορίες, τα *δεδομένα εκπαίδευσης (training data)* στα *δεδομένα ανάκλησης (test set)*, ελέγχου της εγκυρότητας της εκπαίδευσης σε πραγματικό περιβάλλον. Τα δεδομένα εκπαίδευσης χωρίζονται στο *σύνολο δεδομένων εκπαίδευσης (training set)* και στο *σύνολο δεδομένων ελέγχου της εγκυρότητας της εκπαίδευσης (validation set)*, τα οποία χρησιμοποιούνται σαν ψευδο-δεδομένα ανάκλησης με σκοπό να ελέγχουν την ποιότητα της εκπαίδευσης, μια μέθοδος που ονομάζεται *διανομοποίηση (cross validation)* και είναι απαραίτητη για την αποφυγή της *υπέρ-εκπαίδευσης (overtraining ή overfitting)*, ένα φαινόμενο που έχει σαν αποτέλεσμα το δίκτυο να επικεντρώνεται στις παραξενιές του συνόλου των δεδομένων εκμάθησης και να χάνει τις κανονικότητες που απαιτούνται για μια καλή γενίκευση [GeBD92]. Μια μέθοδος εφαρμογής της αποφυγής της υπέρ-εκπαίδευσης είναι και το *έγκαιρο σταμάτημα της εκπαίδευσης (early stopping)* [AMMFY95], [FiHZ93], [NeII91], [SarI95], [SjLj92], [Wang94], [WaVJ94], [Weig94], [MoBo90], όπου το δίκτυο εκπαιδεύεται μέχρι να ελαχιστοποιηθεί το σφάλμα στα δεδομένα ελέγχου της εγκυρότητας της εκπαίδευσης (validation set), αντί του σφάλματος στα δεδομένα εκπαίδευσης (training set) και ορίζεται ως εξής. Αν $E_{tr}(t)$, $E_{va}(t)$ και $E_{te}(t)$ το σφάλμα των προτύπων εκπαίδευσης, ελέγχου εγκυρότητας και ανάκλησης στον κύκλο t και $E_{opt}(t)$ το ελάχιστο σφάλμα των προτύπων ελέγχου εγκυρότητας μέχρι στον κύκλο t , δηλαδή

$$E_{opt}(t) = \min_{t' \leq t} E_{va}(t') \quad (7.11)$$

η *απώλεια της γενίκευσης (generalization loss)* ορίζεται σαν

$$GL_{\alpha}(t) = 100 \cdot \left(\frac{E_{va}(t)}{E_{opt}(t)} - 1 \right) \quad (7.12)$$

ενώ η εκπαίδευση σταματά, όταν η απώλεια της γενίκευσης $GL_{\alpha}(t)$ στον κύκλο t , ξεπεράσει την τιμή α .

Η *πρόοδος της εκπαίδευσης (training progress)* ορίζεται σαν

$$P_k(t) = 1000 \cdot \left(\frac{\sum_{t' \in t-k+1..t} E_{tr}(t')}{k \cdot \min_{t' \in t-k+1..t} E_{tr}(t')} - 1 \right) \quad (7.13)$$

Στις προσομοιώσεις των μεθόδων εκπαίδευσης χρησιμοποιήθηκε η συλλογή προβλημάτων Proben1, όπου τα δεδομένα για κάθε πρόβλημα χωρίζονται στις παραπάνω τρεις κατηγορίες, όπου το 50% αποτελούν τα δεδομένα εκπαίδευσης, 25% τα δεδομένα ελέγχου της εγκυρότητας της εκπαίδευσης και 25% τα δεδομένα ανάκλησης. Για την εξαγωγή συμπερασμάτων χρησιμοποιούνται για το κάθε πρόβλημα τρία σύνολα δεδομένων, τα οποία διαφέρουν στη σειρά των προτύπων και προκύπτουν από διαφορετικούς συνδυασμούς των προτύπων εκπαίδευσης, εγκυρότητας και ανάκλησης. Για παράδειγμα, στο πρόβλημα της ταξινόμησης των διαφόρων τύπων γυαλιού, το πρόβλημα GLASS, υπάρχουν τρεις συνδυασμοί προτύπων με τα ονόματα GLASS1, GLASS2 και GLASS3, τα οποία χρησιμοποιούνται για τον έλεγχο της απόδοσης των αλγορίθμων, με σκοπό την εξαγωγή γενικότερων συμπερασμάτων.

7.3.4. Απόδοση των Αλγορίθμων

Για την υλοποίηση των αλγορίθμων χρησιμοποιείται ένα απλό δίκτυο δύο στρωμάτων με πλήρεις συνδέσεις, όπου ο αριθμός των εισόδων είναι οι εισοδοί για κάθε πρόβλημα και ο αριθμός των εξόδων οι αντίστοιχοι στόχοι. Σαν συνάρτηση ενεργοποίησης χρησιμοποιείται η ταυτοτική συνάρτηση $y=x$.

Για τον έλεγχο της απόδοσης των αλγορίθμων χρησιμοποιείται το μέσο τετραγωνικό σφάλμα (*mean squared error*), όπως περιγράφεται στην (3.23), το οποίο είναι ανεξάρτητο απ' το μέγεθος του συνόλου των δεδομένων, αλλά εξαρτάται απ' τον αριθμό των εξόδων και την κλίμακα των τιμών τους, γι' αυτό και χρησιμοποιείται, όπως και στο Proben1, ένα κανονικοποιημένο μέσο τετραγωνικό σφάλμα, το Ποσοστό Τετραγωνικού Σφάλματος (*Squared Error Percentage*)

$$E = 100 \cdot \frac{o_{max} - o_{min}}{N \cdot P} \sum_{p=1}^P \sum_{i=1}^N (o_{pi} - t_{pi})^2 \quad (7.13)$$

όπου o_{max} , o_{min} είναι η μέγιστη και ελάχιστη τιμή των εξόδων-στόχων, N ο αριθμός των εξόδων του δικτύου και P ο αριθμός των προτύπων.

Ένα άλλο κριτήριο της απόδοσης των αλγορίθμων είναι το Σφάλμα της Ταξινόμησης (*Classification Error*), το ποσοστό δηλαδή των προτύπων που ταξινομούνται σε λάθος κατηγορία. Επειδή όλες οι μέθοδοι εφαρμόζονται σε γραμμικά δίκτυα, επιλέχθηκε η μέθοδος του νικητή (*winner-takes-all*) για τον προσδιορισμό της κλάσης, δηλαδή η κατηγορία στην οποία ανήκει το πρότυπο είναι εκείνη για την οποία ο αντίστοιχος νευρώνας εξόδου έχει τη μέγιστη τιμή για τα προβλήματα ταξινόμησης, ενώ για τα προβλήματα προσέγγισης χρησιμοποιείται κατώφλι 0.3, όπου αν η έξοδος διαφέρει κατά 0.3 απ' το στόχο θεωρείται επιτυχής κατάταξη.

Εκτός απ' τις μεθόδους αριθμητικής ανάλυσης που παρουσιάστηκαν στην (6.3) χρησιμοποιήθηκε και η μέθοδος RPROP [RiBr93], για να συγκριθούν τα αποτελέσματα με αυτά του Prechelt [Prec94], ο οποίος χρησιμοποίησε αυτή τη μέθοδο, την οποία αποκαλούμε P-RPROP, για τα παραπάνω προβλήματα με παραμέτρους $n^+ = 1.2$, $n^- = 0.5$, $\Delta_0 \in [0.005, 0.02]$, $\Delta_{max} = 50$, $\Delta_{min} = 0$ και αρχικές τιμές συνάψεων στο διάστημα $[-0.01, 0.01]$ και το GL_5 κριτήριο.

Τα κριτήρια τερματισμού των αλγορίθμων είναι τα παρακάτω τρία και λειτουργούν διαζευκτικά, δηλαδή ο αλγόριθμος τερματίζει, αν ικανοποιείται *τουλάχιστον ένα* απ' αυτά.

1. Ικανοποιείται το GL_5 κριτήριο.
2. Η πρόοδος της εκπαίδευσης (training progress) $P_5(t)$ έχει γίνει μικρότερη του 0.1 .
3. Έχουν γίνει 3000 επαναλήψεις.

Για την κάθε μέθοδο έγιναν 20 τρεξίματα με διαφορετικές τυχαίες αρχικές τιμές των συνάψεων για τις μεθόδους με σταθερό συντελεστή εκπαίδευσης (ASD, RPROP), ενώ για τις μεθόδους SOR, JOR, RGS, LMS, BATCH-LMS, Polak-Ribiere (PR), Fletcher-Reeves (FR), χρησιμοποιήθηκαν συντελεστές εκπαίδευσης $0.0001, 0.001, 0.01, 0.1, 0.2, \dots, 1.9$, για τις οποίες συγκλίνει η κάθε μέθοδος με σκοπό την εξαγωγή συμπερασμάτων για τη συμπεριφορά της κάθε μεθόδου για διάφορες τιμές των συντελεστών εκπαίδευσης. Η πρώτη δοκιμή έγινε με το συντελεστή εκπαίδευσης ή την παράμετρο υπερχαλαρώσεως ω να αρχίζει με την τιμή 1.9 και μειώνονταν κάθε φορά κατά 0.1 μέχρι και την τιμή 0.1 . Επειδή όμως η μέθοδος Batch-LMS δε συγκλίνει γι' αυτές τις τιμές, ο συντελεστής εκπαίδευσης, μετά την τιμή 0.1 , διαιρούνταν δια του 10 , παίρνοντας τις τιμές $0.01, 0.001$, κ.λ.π. μέχρι να συγκλίνει η μέθοδος. Επειδή, εκτός της μεθόδου SOR, η οποία συγκλίνει για όλες τις παραπάνω τιμές, οι υπόλοιπες μέθοδοι απαιτούν μικρό συντελεστή εκπαίδευσης ή παράμετρο υπερχαλαρώσεως ω , για να πάρουμε τη μέση συμπεριφορά της κάθε μεθόδου, τα πειράματα χρησιμοποιούσαν, όσες φορές απαιτούνταν, την κλίμακα των τιμών του μικρό συντελεστή εκπαίδευσης για τις οποίες συγκλίνει η κάθε μέθοδος. Για παράδειγμα, για κάποια μέθοδο που συγκλίνει μόνο για τις τιμές 0.001 και 0.01 , 10 πειράματα έγιναν (με διαφορετικές τυχαίες αρχικές τιμές) με την πρώτη και 10 με τη δεύτερη τιμή.

Για το κάθε πείραμα το σφάλμα ταξινόμησης είναι διαφορετικό, οπότε υπολογίστηκε η μέση τιμή (*mean*) και η απόκλιση (*standard deviation*) στο σημείο του αλγορίθμου, όπου το σφάλμα των προτύπων ελέγχου εγκυρότητας πήρε την ελάχιστη τιμή, για τα παρακάτω στοιχεία: τις *εποχές* ή *κύκλους εκπαίδευσης* που απαιτήθηκαν, μέχρι να ικανοποιηθεί κάποιο απ' τα τρία κριτήρια τερματισμού, τις *σχετικές εποχές* που απαιτήθηκαν, μέχρι να γίνει ελάχιστο το σφάλμα των προτύπων ελέγχου εγκυρότητας, τους *κύκλους* που απαιτήθηκαν για την *επίσκεψη* (και *διόρθωση*) όλων των *συνάψεων*, το *ποσοστιαίο σφάλμα των προτύπων εκπαίδευσης* (*Training Set Error*), το *ποσοστιαίο σφάλμα των προτύπων ελέγχου εγκυρότητας*, (*Validation Set Error*), το *ποσοστιαίο σφάλμα των προτύπων ανάκλησης*, (*Test Set Error*), το *σφάλμα ταξινόμησης των προτύπων ανάκλησης*, (*Test Set Classification*), καθώς και την *απώλεια γενίκευσης*.

Τα αποτελέσματα αυτά φαίνονται στους πίνακες (7.1)-(7.10) για τα προβλήματα ταξινόμησης.

Σε όλους τους πίνακες, η πρώτη γραμμή περιέχει τα αποτελέσματα των πειραμάτων του Prechelt [Prec94] για τα ίδια προβλήματα, ο οποίος έκανε 10 πειράματα για το κάθε πρόβλημα, ενώ η δεύτερη γραμμή περιέχει τα αποτελέσματα των πειραμάτων της μεθόδου linear-RPROP με τις ίδιες παραμέτρους που χρησιμοποίησε ο Prechelt και για 20 πειράματα.

Πίνακας 7.1 Μέση Τιμή και Απόκλιση για το Πρόβλημα Διάγνωσης Καρκίνου

| CANCER1 | | | | | | | | | | | | | | | | |
|----------------|--------|-------|-----------------|-------|---------------------|-------|--------------------|-------|----------------------|-------|----------------|-------|-------------------------|-------|--------------------|--------|
| Μέθοδος | Εποχές | | Σχετικές Εποχές | | Διορθώσεις Συνάψεων | | Training Set Error | | Validation Set Error | | Test Set Error | | Test Set Classification | | Απόλεια Γενίκευσης | |
| | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev |
| P-RPROP | 129 | 13 | 104 | 31 | 129 | 13 | 4,25 | 0,00 | 2,91 | 0,01 | 3,52 | 0,04 | 2,93 | 0,18 | 0,55 | 0,59 |
| RPROP | 85 | 14 | 73 | 18 | 85 | 14 | 20,75 | 0,01 | 20,21 | 0,12 | 18,61 | 0,13 | 15,15 | 0,50 | 0,18 | 0,28 |
| ASD | 40 | 0 | 23 | 1 | 40 | 0 | 20,74 | 0,00 | 20,20 | 0,00 | 18,63 | 0,00 | 16,09 | 0,00 | 0,16 | 0,01 |
| OFR | 20 | 1 | 9 | 0 | 20 | 1 | 20,74 | 0,00 | 20,07 | 0,04 | 18,61 | 0,05 | 15,79 | 0,39 | 0,83 | 0,28 |
| LMS | 38 | 82 | 33 | 83 | 38 | 82 | 47,86 | 25,84 | 45,79 | 22,68 | 47,63 | 26,91 | 41,08 | 19,71 | 9,97 | 6,20 |
| B-LMS | 275 | 0 | 275 | 0 | 275 | 0 | 20,81 | 0,00 | 20,21 | 0,00 | 18,64 | 0,00 | 16,09 | 0,00 | 0,00 | 0,00 |
| PR | 275 | 0 | 275 | 0 | 275 | 0 | 20,81 | 0,00 | 20,21 | 0,00 | 18,64 | 0,00 | 16,09 | 0,00 | 0,00 | 0,00 |
| FR | 274 | 2 | 274 | 2 | 274 | 2 | 20,81 | 0,00 | 20,21 | 0,00 | 18,64 | 0,00 | 16,09 | 0,00 | 0,00 | 0,00 |
| RGS | 510 | 337 | 498 | 349 | 57 | 37 | 21,34 | 0,50 | 20,61 | 0,80 | 19,16 | 0,52 | 15,76 | 0,86 | 0,08 | 0,07 |
| JOR | 757 | 1042 | 756 | 1043 | 757 | 1042 | 29,38 | 15,16 | 28,39 | 13,96 | 28,60 | 17,48 | 25,11 | 19,42 | 43,17 | 83,61 |
| SOR | 234 | 153 | 181 | 172 | 26 | 17 | 22,45 | 4,77 | 21,04 | 4,49 | 20,48 | 4,98 | 17,60 | 10,66 | 2,86 | 2,55 |
| CANCER2 | | | | | | | | | | | | | | | | |
| Μέθοδος | Εποχές | | Σχετικές Εποχές | | Διορθώσεις Συνάψεων | | Training Set Error | | Validation Set Error | | Test Set Error | | Test Set Classification | | Απόλεια Γενίκευσης | |
| | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev |
| P-RPROP | 87 | 51 | 79 | 51 | 87 | 51 | 3,95 | 0,52 | 3,77 | 0,47 | 4,77 | 0,39 | 5,00 | 0,61 | 5,36 | 10,21 |
| RPROP | 84 | 23 | 63 | 19 | 84 | 23 | 18,71 | 0,18 | 20,56 | 0,10 | 22,03 | 0,19 | 19,03 | 0,45 | 1,12 | 2,68 |
| ASD | 50 | 0 | 30 | 1 | 50 | 0 | 18,65 | 0,00 | 20,51 | 0,00 | 22,02 | 0,01 | 18,39 | 0,00 | 0,13 | 0,01 |
| OFR | 24 | 4 | 11 | 0 | 24 | 4 | 18,64 | 0,00 | 20,47 | 0,02 | 22,14 | 0,03 | 18,94 | 0,13 | 0,37 | 0,08 |
| LMS | 48 | 90 | 40 | 92 | 48 | 90 | 35,11 | 26,60 | 32,49 | 17,44 | 37,50 | 23,97 | 24,74 | 10,08 | 0,48 | 0,98 |
| B-LMS | 305 | 0 | 305 | 0 | 305 | 0 | 18,73 | 0,00 | 20,51 | 0,00 | 21,90 | 0,00 | 18,39 | 0,00 | 0,00 | 0,00 |
| PR | 305 | 0 | 305 | 0 | 305 | 0 | 18,73 | 0,00 | 20,51 | 0,00 | 21,90 | 0,00 | 18,39 | 0,00 | 0,00 | 0,00 |
| FR | 305 | 0 | 305 | 0 | 305 | 0 | 18,73 | 0,00 | 20,51 | 0,00 | 21,90 | 0,00 | 18,39 | 0,00 | 0,00 | 0,00 |
| RGS | 574 | 404 | 571 | 407 | 64 | 45 | 19,22 | 0,48 | 20,81 | 0,34 | 21,82 | 0,01 | 17,54 | 0,29 | 0,04 | 0,04 |
| JOR | 777 | 1044 | 770 | 1048 | 777 | 1044 | 27,64 | 15,65 | 31,02 | 19,02 | 31,71 | 18,19 | 28,37 | 19,49 | 43,81 | 84,82 |
| SOR | 283 | 167 | 268 | 172 | 31 | 19 | 20,65 | 5,22 | 22,55 | 5,21 | 23,88 | 4,59 | 21,05 | 5,11 | 0,85 | 2,02 |
| CANCER3 | | | | | | | | | | | | | | | | |
| Μέθοδος | Εποχές | | Σχετικές Εποχές | | Διορθώσεις Συνάψεων | | Training Set Error | | Validation Set Error | | Test Set Error | | Test Set Classification | | Απόλεια Γενίκευσης | |
| | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev |
| P-RPROP | 115 | 18 | 92 | 29 | 115 | 18 | 3,30 | 0,00 | 4,23 | 0,04 | 4,11 | 0,03 | 5,17 | 0,00 | 0,35 | 0,64 |
| RPROP | 85 | 14 | 73 | 18 | 85 | 14 | 20,75 | 0,01 | 20,21 | 0,12 | 18,61 | 0,13 | 15,15 | 0,50 | 0,18 | 0,28 |
| ASD | 48 | 2 | 21 | 0 | 48 | 2 | 20,43 | 0,00 | 18,70 | 0,00 | 20,34 | 0,01 | 17,66 | 0,25 | 0,46 | 0,03 |
| OFR | 23 | 5 | 13 | 5 | 23 | 5 | 20,43 | 0,01 | 18,75 | 0,03 | 20,28 | 0,07 | 16,88 | 0,28 | 0,33 | 0,16 |
| LMS | 48 | 92 | 40 | 83 | 48 | 92 | 27,59 | 8,50 | 26,21 | 9,81 | 30,55 | 11,46 | 23,05 | 5,42 | 10,53 | 21,32 |
| B-LMS | 310 | 0 | 279 | 2 | 310 | 0 | 20,52 | 0,00 | 18,71 | 0,00 | 20,33 | 0,00 | 17,82 | 0,00 | 0,02 | 0,00 |
| PR | 310 | 0 | 279 | 1 | 310 | 0 | 20,52 | 0,00 | 18,71 | 0,00 | 20,33 | 0,00 | 17,82 | 0,00 | 0,02 | 0,00 |
| FR | 310 | 0 | 279 | 1 | 310 | 0 | 20,52 | 0,00 | 18,71 | 0,00 | 20,33 | 0,00 | 17,82 | 0,00 | 0,02 | 0,00 |
| RGS | 619 | 404 | 595 | 425 | 69 | 45 | 20,94 | 0,46 | 18,77 | 0,43 | 20,73 | 0,27 | 19,81 | 0,29 | 0,11 | 0,09 |
| JOR | 761 | 1038 | 760 | 1039 | 761 | 1038 | 29,01 | 14,96 | 29,75 | 19,63 | 29,36 | 15,99 | 27,04 | 18,11 | 54,74 | 106,01 |
| SOR | 222 | 158 | 187 | 169 | 25 | 18 | 21,88 | 3,96 | 19,76 | 4,58 | 22,52 | 4,67 | 19,48 | 6,58 | 1,90 | 2,53 |

Πίνακας 7.2 Μέση Τιμή και Απόκλιση για το Πρόβλημα της Πιστωτικής Κάρτας

| CARD1 | | | | | | | | | | | | | | | | |
|---------|--------|-------|-----------------|-------|---------------------|-------|--------------------|-------|----------------------|-------|----------------|-------|-------------------------|-------|--------------------|-------|
| Μέθοδος | Εποχές | | Σχετικές Εποχές | | Διορθώσεις Συνάψεων | | Training Set Error | | Validation Set Error | | Test Set Error | | Test Set Classification | | Απόλεια Γενίκευσης | |
| | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev |
| P-RPROP | 62 | 9 | 26 | 3 | 62 | 9 | 9,82 | 0,01 | 8,89 | 0,11 | 10,61 | 0,11 | 13,37 | 0,67 | 4,57 | 1,05 |
| RPROP | 64 | 9 | 27 | 5 | 64 | 9 | 9,83 | 0,01 | 8,93 | 0,14 | 10,67 | 0,20 | 13,71 | 0,74 | 4,19 | 1,23 |
| ASD | 88 | 3 | 13 | 0 | 88 | 3 | 10,04 | 0,01 | 8,21 | 0,01 | 10,47 | 0,01 | 13,95 | 0,00 | 5,08 | 0,06 |
| OFR | 49 | 13 | 13 | 0 | 49 | 13 | 10,05 | 0,072 | 8,20 | 0,01 | 10,46 | 0,01 | 13,95 | 0,00 | 5,90 | 0,86 |
| LMS | 73 | 123 | 11 | 18 | 73 | 123 | 14,21 | 5,75 | 10,96 | 4,28 | 15,25 | 7,02 | 17,01 | 5,83 | 8,86 | 3,66 |
| B-LMS | 324 | 11 | 45 | 1 | 324 | 11 | 10,03 | 0,01 | 8,22 | 0,01 | 10,47 | 0,01 | 13,95 | 0,00 | 5,02 | 0,01 |
| PR | 324 | 13 | 45 | 1 | 324 | 13 | 10,03 | 0,01 | 8,22 | 0,01 | 10,48 | 0,01 | 13,95 | 0,00 | 5,02 | 0,01 |
| FR | 324 | 13 | 46 | 1 | 324 | 12 | 10,03 | 0,01 | 8,22 | 0,01 | 10,47 | 0,01 | 13,95 | 0,00 | 5,03 | 0,01 |
| RGS | 98 | 87 | 93 | 91 | 1 | 2 | 31,83 | 12,42 | 32,35 | 14,10 | 31,35 | 12,10 | 31,43 | 14,99 | 41,94 | 39,80 |
| JOR | 871 | 1129 | 849 | 1141 | 871 | 1129 | 20,43 | 17,19 | 19,40 | 16,88 | 21,22 | 17,39 | 15,48 | 2,34 | 22,25 | 36,52 |
| SOR | 147 | 104 | 139 | 106 | 2 | 2 | 17,08 | 8,10 | 17,34 | 9,43 | 19,07 | 9,45 | 21,66 | 9,76 | 2,13 | 3,94 |
| CARD2 | | | | | | | | | | | | | | | | |
| Μέθοδος | Εποχές | | Σχετικές Εποχές | | Διορθώσεις Συνάψεων | | Training Set Error | | Validation Set Error | | Test Set Error | | Test Set Classification | | Απόλεια Γενίκευσης | |
| | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev |
| P-RPROP | 65 | 10 | 23 | 5 | 65 | 10 | 8,24 | 0,01 | 10,80 | 0,16 | 14,91 | 0,55 | 19,24 | 0,43 | 4,22 | 1,08 |
| RPROP | 55 | 19 | 24 | 6 | 55 | 19 | 8,38 | 0,35 | 10,76 | 0,24 | 14,93 | 0,26 | 19,37 | 0,38 | 4,38 | 1,49 |
| ASD | 231 | 4 | 21 | 1 | 231 | 4 | 8,32 | 0,00 | 9,71 | 0,01 | 13,67 | 0,03 | 19,77 | 0,00 | 5,02 | 0,04 |
| OFR | 85 | 30 | 19 | 0 | 85 | 30 | 8,34 | 0,03 | 9,71 | 0,01 | 13,72 | 0,02 | 19,77 | 0,00 | 5,59 | 0,77 |
| LMS | 92 | 112 | 16 | 24 | 92 | 112 | 16,56 | 11,62 | 16,00 | 9,07 | 21,61 | 11,33 | 23,99 | 7,35 | 5,40 | 2,80 |
| B-LMS | 303 | 3 | 60 | 1 | 303 | 3 | 8,42 | 0,00 | 9,72 | 0,01 | 13,62 | 0,03 | 19,77 | 0,00 | 2,47 | 0,07 |
| PR | 302 | 3 | 60 | 1 | 302 | 3 | 8,43 | 0,00 | 9,72 | 0,01 | 13,60 | 0,02 | 19,77 | 0,00 | 2,48 | 0,08 |
| FR | 302 | 3 | 60 | 1 | 302 | 3 | 8,43 | 0,00 | 9,72 | 0,01 | 13,61 | 0,02 | 19,77 | 0,00 | 2,45 | 0,09 |
| RGS | 108 | 119 | 97 | 112 | 2 | 2 | 33,13 | 16,03 | 34,01 | 15,80 | 33,44 | 14,43 | 30,75 | 12,01 | 13,17 | 12,39 |
| JOR | 873 | 1139 | 864 | 1146 | 873 | 1139 | 18,74 | 17,03 | 21,26 | 16,52 | 23,56 | 14,68 | 19,65 | 1,11 | 16,23 | 27,10 |
| SOR | 128 | 106 | 109 | 100 | 2 | 2 | 17,36 | 9,36 | 18,57 | 7,53 | 20,14 | 6,94 | 26,38 | 9,85 | 5,59 | 6,47 |
| CARD3 | | | | | | | | | | | | | | | | |
| Μέθοδος | Εποχές | | Σχετικές Εποχές | | Διορθώσεις Συνάψεων | | Training Set Error | | Validation Set Error | | Test Set Error | | Test Set Classification | | Απόλεια Γενίκευσης | |
| | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev |
| P-RPROP | 102 | 9 | 44 | 12 | 102 | 9 | 9,47 | 0,00 | 8,39 | 0,07 | 12,67 | 0,17 | 14,42 | 0,46 | 1,52 | 0,69 |
| RPROP | 104 | 11 | 45 | 13 | 104 | 11 | 9,47 | 0,00 | 8,39 | 0,10 | 12,60 | 0,21 | 14,60 | 0,65 | 1,75 | 1,19 |
| ASD | 158 | 4 | 26 | 1 | 158 | 4 | 9,64 | 0,00 | 7,68 | 0,01 | 12,34 | 0,01 | 15,45 | 0,29 | 5,05 | 0,04 |
| OFR | 76 | 35 | 20 | 2 | 76 | 35 | 9,65 | 0,04 | 7,67 | 0,01 | 12,34 | 0,03 | 15,51 | 0,42 | 5,69 | 0,81 |
| LMS | 133 | 144 | 51 | 27 | 133 | 144 | 16,53 | 9,04 | 15,31 | 9,26 | 21,84 | 12,11 | 22,34 | 8,26 | 2,30 | 2,44 |
| B-LMS | 413 | 3 | 76 | 1 | 413 | 3 | 9,67 | 0,00 | 7,67 | 0,01 | 12,33 | 0,01 | 15,18 | 0,18 | 4,49 | 0,08 |
| PR | 413 | 2 | 76 | 1 | 413 | 2 | 9,67 | 0,00 | 7,68 | 0,01 | 12,33 | 0,01 | 15,18 | 0,18 | 4,46 | 0,09 |
| FR | 413 | 2 | 75 | 1 | 413 | 3 | 9,67 | 0,00 | 7,68 | 0,01 | 12,33 | 0,01 | 15,15 | 0,13 | 4,49 | 0,07 |
| RGS | 96 | 90 | 91 | 89 | 1 | 2 | 30,57 | 11,82 | 30,91 | 12,87 | 31,73 | 11,39 | 34,70 | 11,97 | 6,73 | 16,50 |
| JOR | 803 | 1025 | 802 | 1026 | 803 | 1025 | 19,83 | 16,71 | 19,46 | 17,26 | 24,25 | 18,01 | 18,94 | 6,12 | 21,01 | 35,17 |
| SOR | 96 | 111 | 88 | 111 | 1 | 2 | 19,93 | 7,72 | 21,32 | 7,14 | 23,54 | 6,98 | 34,76 | 9,27 | 4,99 | 4,21 |

Πίνακας 7.3 Μέση Τιμή και Απόκλιση για το Πρόβλημα της Διάγνωσης Διαβήτη

| DIABETES1 | | | | | | | | | | | | | | | | |
|------------------|--------|-------|-----------------|-------|---------------------|-------|--------------------|-------|----------------------|-------|----------------|-------|-------------------------|-------|--------------------|--------|
| Μέθοδος | Εποχές | | Σχετικές Εποχές | | Διορθώσεις Συνάψεων | | Training Set Error | | Validation Set Error | | Test Set Error | | Test Set Classification | | Απόλεια Γενίκευσης | |
| | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev |
| P-RPROP | 209 | 50 | 203 | 47 | 209 | 50 | 15,39 | 0,01 | 16,30 | 0,04 | 17,22 | 0,06 | 25,83 | 0,56 | 0,05 | 0,07 |
| RPROP | 103 | 15 | 85 | 19 | 103 | 15 | 20,34 | 0,01 | 22,58 | 0,03 | 24,06 | 0,09 | 38,65 | 0,60 | 0,19 | 0,16 |
| ASD | 30 | 0 | 10 | 0 | 30 | 0 | 20,32 | 0,00 | 22,54 | 0,00 | 24,02 | 0,00 | 38,54 | 0,00 | 0,54 | 0,01 |
| OFR | 20 | 1 | 8 | 0 | 20 | 1 | 20,31 | 0,00 | 22,51 | 0,00 | 23,89 | 0,01 | 37,50 | 0,00 | 0,88 | 0,04 |
| LMS | 23 | 51 | 14 | 43 | 23 | 51 | 25,39 | 7,02 | 26,02 | 5,59 | 29,59 | 9,70 | 38,71 | 1,80 | 0,63 | 0,79 |
| B-LMS | 235 | 0 | 195 | 1 | 235 | 0 | 20,38 | 0,00 | 22,55 | 0,00 | 24,07 | 0,00 | 39,58 | 0,00 | 0,05 | 0,00 |
| PR | 235 | 0 | 194 | 1 | 235 | 0 | 20,38 | 0,00 | 22,55 | 0,00 | 24,07 | 0,00 | 39,58 | 0,00 | 0,05 | 0,00 |
| FR | 235 | 0 | 195 | 1 | 235 | 0 | 20,38 | 0,00 | 22,55 | 0,00 | 24,07 | 0,00 | 39,58 | 0,00 | 0,05 | 0,00 |
| RGS | 311 | 154 | 273 | 185 | 38 | 19 | 21,02 | 0,37 | 22,77 | 0,26 | 23,98 | 0,47 | 39,91 | 0,87 | 0,31 | 0,26 |
| JOR | 393 | 402 | 392 | 402 | 393 | 402 | 31,17 | 19,45 | 34,31 | 21,39 | 35,84 | 21,59 | 39,39 | 1,04 | 121,10 | 234,51 |
| SOR | 147 | 93 | 140 | 93 | 18 | 12 | 22,60 | 3,97 | 24,53 | 3,73 | 25,66 | 3,99 | 38,65 | 1,34 | 1,36 | 2,10 |
| DIABETES2 | | | | | | | | | | | | | | | | |
| Μέθοδος | Εποχές | | Σχετικές Εποχές | | Διορθώσεις Συνάψεων | | Training Set Error | | Validation Set Error | | Test Set Error | | Test Set Classification | | Απόλεια Γενίκευσης | |
| | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev |
| P-RPROP | 209 | 32 | 204 | 34 | 209 | 32 | 14,93 | 0,01 | 17,47 | 0,02 | 17,69 | 0,04 | 24,69 | 0,61 | 0,02 | 0,02 |
| RPROP | 100 | 16 | 99 | 16 | 100 | 16 | 21,05 | 0,01 | 20,87 | 0,06 | 24,29 | 0,08 | 37,31 | 0,78 | 0,02 | 0,03 |
| ASD | 30 | 0 | 15 | 0 | 30 | 0 | 21,03 | 0,00 | 20,71 | 0,00 | 23,98 | 0,00 | 35,94 | 0,00 | 0,16 | 0,01 |
| OFR | 20 | 0 | 11 | 1 | 20 | 0 | 21,02 | 0,00 | 20,72 | 0,00 | 24,16 | 0,09 | 36,38 | 0,72 | 0,22 | 0,02 |
| LMS | 25 | 54 | 16 | 56 | 25 | 54 | 49,53 | 22,84 | 51,42 | 24,33 | 55,26 | 24,20 | 37,99 | 0,92 | 0,58 | 0,60 |
| B-LMS | 250 | 0 | 250 | 0 | 250 | 0 | 21,10 | 0,00 | 20,78 | 0,00 | 23,87 | 0,00 | 35,42 | 0,00 | 0,00 | 0,00 |
| PR | 250 | 1 | 250 | 1 | 250 | 1 | 21,10 | 0,00 | 20,78 | 0,00 | 23,87 | 0,00 | 35,42 | 0,00 | 0,00 | 0,00 |
| FR | 250 | 1 | 250 | 1 | 250 | 1 | 21,10 | 0,00 | 20,78 | 0,00 | 23,87 | 0,00 | 35,42 | 0,00 | 0,00 | 0,00 |
| RGS | 342 | 145 | 341 | 146 | 42 | 18 | 21,66 | 0,45 | 21,56 | 0,59 | 24,04 | 0,23 | 37,50 | 0,00 | 0,03 | 0,03 |
| JOR | 360 | 349 | 359 | 350 | 360 | 349 | 32,19 | 19,95 | 32,55 | 20,43 | 33,93 | 18,46 | 37,53 | 0,73 | 151,66 | 293,70 |
| SOR | 169 | 88 | 166 | 86 | 21 | 11 | 22,88 | 4,67 | 23,14 | 5,087 | 25,87 | 4,71 | 37,42 | 1,56 | 0,46 | 1,66 |
| DIABETES3 | | | | | | | | | | | | | | | | |
| Μέθοδος | Εποχές | | Σχετικές Εποχές | | Διορθώσεις Συνάψεων | | Training Set Error | | Validation Set Error | | Test Set Error | | Test Set Classification | | Απόλεια Γενίκευσης | |
| | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev |
| P-RPROP | 214 | 22 | 185 | 46 | 214 | 22 | 14,78 | 0,02 | 18,21 | 0,04 | 16,50 | 0,05 | 22,92 | 0,35 | 0,12 | 0,17 |
| RPROP | 106 | 11 | 35 | 8 | 106 | 11 | 20,40 | 0,01 | 23,03 | 0,14 | 22,64 | 0,17 | 37,06 | 1,24 | 2,55 | 0,57 |
| ASD | 25 | 0 | 3 | 0 | 25 | 0 | 20,38 | 0,00 | 22,87 | 0,01 | 22,04 | 0,01 | 33,74 | 0,47 | 3,70 | 0,03 |
| OFR | 27 | 8 | 3 | 0 | 27 | 8 | 20,39 | 0,01 | 22,89 | 0,01 | 22,03 | 0,01 | 33,28 | 0,53 | 3,65 | 0,22 |
| LMS | 26 | 58 | 7 | 24 | 26 | 58 | 40,03 | 21,38 | 43,13 | 21,31 | 42,64 | 20,43 | 53,02 | 12,48 | 1,72 | 1,30 |
| B-LMS | 266 | 2 | 110 | 1 | 266 | 2 | 20,46 | 0,00 | 22,77 | 0,01 | 22,64 | 0,01 | 35,94 | 0,00 | 2,06 | 0,03 |
| PR | 267 | 3 | 110 | 1 | 267 | 3 | 20,46 | 0,00 | 22,77 | 0,01 | 22,63 | 0,00 | 35,94 | 0,00 | 2,08 | 0,03 |
| FR | 267 | 2 | 110 | 1 | 267 | 2 | 20,46 | 0,00 | 22,77 | 0,01 | 22,63 | 0,00 | 35,94 | 0,00 | 2,07 | 0,03 |
| RGS | 174 | 34 | 101 | 38 | 21 | 4 | 21,88 | 1,43 | 23,94 | 0,08 | 24,04 | 0,31 | 35,86 | 0,94 | 1,50 | 1,43 |
| JOR | 413 | 390 | 341 | 428 | 413 | 390 | 31,79 | 20,11 | 34,01 | 20,68 | 34,30 | 21,62 | 36,68 | 1,06 | 150,96 | 289,93 |
| SOR | 157 | 60 | 135 | 71 | 19 | 7 | 21,84 | 1,40 | 24,82 | 2,71 | 24,38 | 2,46 | 37,86 | 2,03 | 1,38 | 2,03 |

Πίνακας 7.4 Μέση Τιμή και Απόκλιση για το Πρόβλημα Κατάταξης Γονιδίων

| GENE1 | | | | | | | | | | | | | | | | |
|---------|--------|-------|-----------------|-------|---------------------|-------|--------------------|-------|----------------------|-------|----------------|-------|-------------------------|-------|--------------------|--------|
| Μέθοδος | Εποχές | | Σχετικές Εποχές | | Διορθώσεις Συνάψεων | | Training Set Error | | Validation Set Error | | Test Set Error | | Test Set Classification | | Απόλεια Γενίκευσης | |
| | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev |
| P-RPROP | 47 | 6 | 43 | 10 | 47 | 6 | 8,42 | 0,00 | 9,58 | 0,01 | 9,92 | 0,01 | 13,64 | 0,10 | 0,03 | 0,07 |
| RPROP | 30 | 0 | 18 | 6 | 30 | 0 | 21,48 | 0,00 | 25,55 | 0,05 | 25,03 | 0,04 | 39,06 | 0,60 | 0,24 | 0,20 |
| ASD | 19 | 2 | 2 | 0 | 19 | 2 | 21,48 | 0,00 | 24,73 | 0,10 | 24,46 | 0,06 | 40,70 | 0,68 | 3,52 | 0,43 |
| OFR | 20 | 2 | 2 | 0 | 20 | 2 | 21,48 | 0,00 | 24,74 | 0,08 | 24,53 | 0,07 | 40,66 | 0,66 | 3,46 | 0,34 |
| LMS | 16 | 13 | 3 | 2 | 16 | 13 | 35,30 | 17,77 | 39,26 | 19,13 | 38,54 | 18,37 | 45,71 | 5,94 | 2,85 | 2,18 |
| B-LMS | 35 | 0 | 5 | 0 | 35 | 0 | 21,48 | 0,00 | 24,33 | 0,05 | 24,22 | 0,07 | 41,52 | 0,55 | 4,80 | 0,19 |
| PR | 35 | 0 | 5 | 0 | 35 | 0 | 21,48 | 0,00 | 24,33 | 0,06 | 24,21 | 0,06 | 41,22 | 0,51 | 4,81 | 0,25 |
| FR | 35 | 2 | 5 | 0 | 35 | 2 | 21,48 | 0,00 | 24,33 | 0,08 | 24,23 | 0,06 | 41,43 | 0,53 | 4,80 | 0,26 |
| RGS | 130 | 162 | 109 | 159 | 1 | 1 | 32,09 | 6,71 | 33,17 | 6,36 | 33,48 | 6,75 | 62,45 | 14,89 | 197,82 | 290,31 |
| JOR | 448 | 876 | 413 | 883 | 448 | 876 | 23,03 | 2,56 | 25,52 | 1,88 | 25,32 | 1,79 | 40,82 | 0,78 | 21,05 | 42,84 |
| SOR | 232 | 139 | 178 | 136 | 2 | 1 | 25,36 | 4,88 | 27,71 | 3,89 | 27,60 | 4,02 | 52,50 | 16,95 | 2,00 | 2,31 |
| GENE2 | | | | | | | | | | | | | | | | |
| Μέθοδος | Εποχές | | Σχετικές Εποχές | | Διορθώσεις Συνάψεων | | Training Set Error | | Validation Set Error | | Test Set Error | | Test Set Classification | | Απόλεια Γενίκευσης | |
| | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev |
| P-RPROP | 46 | 4 | 40 | 6 | 46 | 4 | 8,39 | 0,00 | 9,90 | 0,00 | 9,51 | 0,00 | 12,30 | 0,14 | 0,02 | 0,03 |
| RPROP | 30 | 0 | 19 | 5 | 30 | 0 | 21,62 | 0,00 | 25,20 | 0,03 | 24,99 | 0,04 | 39,74 | 0,50 | 0,15 | 0,09 |
| ASD | 15 | 0 | 2 | 0 | 15 | 0 | 21,62 | 0,00 | 24,54 | 0,10 | 24,54 | 0,09 | 41,14 | 0,58 | 2,75 | 0,39 |
| OFR | 15 | 0 | 2 | 0 | 15 | 0 | 21,62 | 0,00 | 24,54 | 0,11 | 24,57 | 0,11 | 41,05 | 0,77 | 2,75 | 0,45 |
| LMS | 17 | 12 | 3 | 2 | 17 | 12 | 34,59 | 16,71 | 37,06 | 16,43 | 37,50 | 17,02 | 43,51 | 3,91 | 2,21 | 1,68 |
| B-LMS | 311 | 2 | 5 | 0 | 31 | 2 | 21,62 | 0,00 | 24,24 | 0,05 | 24,29 | 0,05 | 42,03 | 0,54 | 3,70 | 0,22 |
| PR | 31 | 2 | 5 | 0 | 31 | 2 | 21,62 | 0,00 | 24,25 | 0,06 | 24,31 | 0,07 | 42,22 | 0,59 | 3,69 | 0,25 |
| FR | 31 | 2 | 5 | 0 | 31 | 2 | 21,62 | 0,00 | 24,26 | 0,05 | 24,32 | 0,05 | 42,06 | 0,72 | 3,66 | 0,21 |
| RGS | 100 | 122 | 78 | 115 | 1 | 1 | 31,50 | 5,19 | 32,20 | 4,88 | 32,31 | 4,87 | 62,62 | 14,29 | 142,36 | 194,27 |
| JOR | 444 | 876 | 408 | 881 | 444 | 876 | 23,08 | 2,40 | 25,33 | 1,79 | 25,41 | 1,89 | 41,59 | 0,78 | 14,29 | 28,18 |
| SOR | 170 | 100 | 142 | 105 | 1 | 1 | 25,92 | 4,64 | 27,98 | 3,84 | 28,01 | 3,85 | 52,29 | 16,25 | 2,31 | 2,69 |
| GENE3 | | | | | | | | | | | | | | | | |
| Μέθοδος | Εποχές | | Σχετικές Εποχές | | Διορθώσεις Συνάψεων | | Training Set Error | | Validation Set Error | | Test Set Error | | Test Set Classification | | Απόλεια Γενίκευσης | |
| | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev |
| P-RPROP | 42 | 4 | 39 | 6 | 42 | 4 | 8,21 | 0,00 | 9,36 | 0,01 | 10,61 | 0,01 | 15,41 | 0,13 | 0,03 | 0,06 |
| RPROP | 30 | 0 | 18 | 5 | 30 | 0 | 21,88 | 0,00 | 24,33 | 0,05 | 25,37 | 0,07 | 41,93 | 0,64 | 0,21 | 0,20 |
| ASD | 15 | 0 | 3 | 0 | 15 | 0 | 21,88 | 0,00 | 24,01 | 0,08 | 24,88 | 0,08 | 42,37 | 0,41 | 1,45 | 0,32 |
| OFR | 15 | 0 | 3 | 0 | 15 | 0 | 21,88 | 0,00 | 24,00 | 0,06 | 24,86 | 0,06 | 42,50 | 0,34 | 1,50 | 0,26 |
| LMS | 16 | 12 | 3 | 3 | 16 | 12 | 33,87 | 15,45 | 35,46 | 15,08 | 36,36 | 15,25 | 46,72 | 4,87 | 1,78 | 1,38 |
| B-LMS | 30 | 0 | 6 | 0 | 30 | 0 | 21,88 | 0,00 | 23,66 | 0,06 | 24,40 | 0,05 | 43,54 | 0,41 | 2,73 | 0,23 |
| PR | 30 | 0 | 6 | 0 | 30 | 0 | 21,88 | 0,00 | 23,67 | 0,06 | 24,42 | 0,06 | 43,47 | 0,30 | 2,69 | 0,25 |
| FR | 30 | 0 | 6 | 0 | 30 | 0 | 21,88 | 0,00 | 23,67 | 0,06 | 24,41 | 0,07 | 43,47 | 0,39 | 2,67 | 0,26 |
| RGS | 148 | 146 | 127 | 136 | 1 | 1 | 30,28 | 5,60 | 31,27 | 5,46 | 31,89 | 5,29 | 53,30 | 10,40 | 302,59 | 402,96 |
| JOR | 432 | 845 | 408 | 854 | 432 | 845 | 23,30 | 2,31 | 24,71 | 1,69 | 25,41 | 1,60 | 43,07 | 0,97 | 16,73 | 35,27 |
| SOR | 219 | 86 | 184 | 96 | 1 | 1 | 24,80 | 3,62 | 26,49 | 3,50 | 27,31 | 3,43 | 47,22 | 9,24 | 1,67 | 2,23 |

Πίνακας 7.5 Μέση Τιμή και Απόκλιση για το Πρόβλημα Ταξινόμησης Τύπων Γυαλιού

| GLASS1 | | | | | | | | | | | | | | | | |
|---------|--------|-------|-----------------|-------|---------------------|-------|--------------------|-------|----------------------|-------|----------------|-------|-------------------------|-------|--------------------|--------|
| Μέθοδος | Εποχές | | Σχετικές Εποχές | | Διορθώσεις Συνάψεων | | Training Set Error | | Validation Set Error | | Test Set Error | | Test Set Classification | | Απόλεια Γενίκευσης | |
| | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev |
| P-RPROP | 129 | 13 | 23 | 5 | 129 | 13 | 8,83 | 0,01 | 9,70 | 0,04 | 9,98 | 0,10 | 46,04 | 2,21 | 3,81 | 0,42 |
| RPROP | 124 | 14 | 24 | 5 | 124 | 14 | 8,84 | 0,01 | 9,71 | 0,07 | 10,13 | 0,12 | 47,67 | 3,05 | 3,75 | 0,69 |
| ASD | 157 | 3 | 44 | 1 | 157 | 3 | 8,80 | 0,00 | 9,95 | 0,01 | 9,77 | 0,01 | 47,17 | 0,00 | 1,65 | 0,06 |
| OFR | 75 | 14 | 29 | 8 | 75 | 14 | 8,79 | 0,02 | 9,86 | 0,08 | 9,80 | 0,11 | 43,30 | 2,07 | 3,68 | 1,95 |
| LMS | 58 | 135 | 38 | 130 | 58 | 135 | 18,72 | 8,18 | 19,99 | 8,18 | 22,96 | 10,24 | 66,24 | 9,30 | 8,38 | 4,20 |
| B-LMS | 401 | 170 | 331 | 236 | 401 | 170 | 8,98 | 0,12 | 9,83 | 0,02 | 9,69 | 0,01 | 47,07 | 1,88 | 0,51 | 0,48 |
| PR | 401 | 169 | 331 | 236 | 401 | 169 | 8,98 | 0,12 | 9,83 | 0,02 | 9,69 | 0,01 | 47,07 | 1,88 | 0,5080 | 0,48 |
| FR | 401 | 170 | 330 | 237 | 401 | 169 | 8,98 | 0,12 | 9,83 | 0,02 | 9,69 | 0,01 | 47,07 | 1,88 | 0,52 | 0,50 |
| RGS | 414 | 367 | 412 | 368 | 46 | 41 | 10,78 | 1,62 | 11,21 | 1,41 | 11,09 | 1,37 | 56,01 | 5,20 | 42,87 | 56,11 |
| JOR | 317 | 468 | 257 | 491 | 317 | 468 | 13,25 | 6,56 | 14,00 | 6,59 | 13,87 | 5,79 | 47,86 | 5,17 | 1656,2 | 3677,0 |
| SOR | 180 | 96 | 146 | 103 | 20 | 11 | 9,50 | 1,39 | 10,46 | 1,09 | 10,67 | 0,89 | 42,60 | 5,17 | 3,31 | 8,20 |
| GLASS2 | | | | | | | | | | | | | | | | |
| Μέθοδος | Εποχές | | Σχετικές Εποχές | | Διορθώσεις Συνάψεων | | Training Set Error | | Validation Set Error | | Test Set Error | | Test Set Classification | | Απόλεια Γενίκευσης | |
| | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev |
| P-RPROP | 34 | 6 | 14 | 2 | 34 | 6 | 8,71 | 0,09 | 10,28 | 0,19 | 10,34 | 0,15 | 55,28 | 1,27 | 5,74 | 0,67 |
| RPROP | 33 | 9 | 16 | 3 | 33 | 9 | 8,75 | 0,16 | 10,27 | 0,16 | 10,30 | 0,14 | 55,91 | 2,39 | 6,19 | 1,09 |
| ASD | 25 | 0 | 7 | 0 | 25 | 0 | 8,62 | 0,01 | 10,54 | 0,01 | 10,37 | 0,01 | 54,72 | 0,00 | 5,40 | 0,14 |
| OFR | 13 | 3 | 5 | 0 | 13 | 3 | 8,70 | 0,11 | 10,59 | 0,02 | 10,48 | 0,01 | 52,83 | 0,00 | 6,16 | 0,97 |
| LMS | 53 | 141 | 11 | 33 | 53 | 141 | 16,51 | 9,23 | 18,56 | 8,88 | 16,59 | 8,13 | 53,53 | 2,88 | 6,19 | 1,34 |
| B-LMS | 340 | 289 | 78 | 68 | 340 | 289 | 8,66 | 0,01 | 10,34 | 0,01 | 10,17 | 0,00 | 52,83 | 0,00 | 5,11 | 0,13 |
| PR | 342 | 292 | 78 | 68 | 342 | 292 | 8,66 | 0,00 | 10,34 | 0,01 | 10,17 | 0,01 | 52,83 | 0,00 | 5,08 | 0,07 |
| FR | 343 | 292 | 79 | 68 | 343 | 292 | 8,66 | 0,01 | 10,34 | 0,01 | 10,17 | 0,00 | 52,83 | 0,00 | 5,12 | 0,11 |
| RGS | 577 | 542 | 450 | 586 | 64 | 60 | 10,62 | 2,03 | 11,17 | 1,08 | 11,11 | 1,23 | 56,21 | 4,56 | 82,35 | 104,69 |
| JOR | 433 | 620 | 305 | 607 | 433 | 620 | 10,88 | 3,91 | 11,94 | 2,91 | 11,79 | 2,65 | 58,29 | 2,66 | 98,45 | 183,07 |
| SOR | 78 | 88 | 30 | 34 | 8 | 10 | 9,60 | 1,48 | 10,84 | 1,27 | 10,76 | 0,76 | 53,53 | 3,98 | 5,91 | 2,35 |
| GLASS3 | | | | | | | | | | | | | | | | |
| Μέθοδος | Εποχές | | Σχετικές Εποχές | | Διορθώσεις Συνάψεων | | Training Set Error | | Validation Set Error | | Test Set Error | | Test Set Classification | | Απόλεια Γενίκευσης | |
| | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev |
| P-RPROP | 135 | 30 | 27 | 11 | 135 | 30 | 8,71 | 0,02 | 9,37 | 0,06 | 11,07 | 0,15 | 60,57 | 3,82 | 1,76 | 0,57 |
| RPROP | 124 | 22 | 26 | 12 | 124 | 22 | 8,72 | 0,02 | 9,36 | 0,05 | 11,18 | 0,20 | 60,77 | 4,13 | 2,03 | 0,55 |
| ASD | 182 | 3 | 46 | 2 | 182 | 3 | 8,67 | 0,00 | 9,41 | 0,00 | 10,94 | 0,01 | 56,60 | 0,00 | 1,36 | 0,04 |
| OFR | 66 | 22 | 28 | 7 | 66 | 22 | 8,68 | 0,06 | 9,41 | 0,03 | 11,01 | 0,07 | 56,31 | 1,26 | 4,07 | 4,40 |
| LMS | 139 | 113 | 110 | 119 | 139 | 113 | 14,77 | 3,19 | 15,15 | 2,86 | 21,96 | 5,23 | 65,74 | 2,31 | 0,14 | 0,26 |
| B-LMS | 380 | 211 | 329 | 259 | 380 | 211 | 8,84 | 0,08 | 9,40 | 0,01 | 10,86 | 0,02 | 58,49 | 0,00 | 0,25 | 0,24 |
| PR | 380 | 210 | 329 | 259 | 380 | 210 | 8,84 | 0,08 | 9,40 | 0,01 | 10,86 | 0,02 | 58,49 | 0,00 | 0,25 | 0,24 |
| FR | 380 | 211 | 329 | 259 | 380 | 211 | 8,84 | 0,08 | 9,40 | 0,01 | 10,86 | 0,02 | 58,59 | 0,42 | 0,25 | 0,24 |
| RGS | 390 | 336 | 388 | 337 | 43 | 38 | 10,62 | 1,57 | 10,73 | 1,28 | 12,16 | 1,50 | 65,64 | 4,92 | 65,82 | 86,17 |
| JOR | 318 | 472 | 233 | 486 | 318 | 472 | 13,16 | 6,60 | 13,50 | 6,43 | 15,78 | 7,04 | 68,72 | 5,98 | 1853,9 | 4113,4 |
| SOR | 208 | 108 | 188 | 113 | 23 | 12 | 9,36 | 1,40 | 9,94 | 0,86 | 11,79 | 1,22 | 60,68 | 3,48 | 1,46 | 3,41 |

Πίνακας 7.6 Μέση Τιμή και Απόκλιση για το Πρόβλημα Διάγνωσης Καρδιοπάθειας (heart)

| HEART1 | | | | | | | | | | | | | | | | |
|---------------|--------|-------|-----------------|-------|---------------------|-------|--------------------|-------|----------------------|-------|----------------|-------|-------------------------|-------|--------------------|-------|
| Μέθοδος | Εποχές | | Σχετικές Εποχές | | Διορθώσεις Συνάψεων | | Training Set Error | | Validation Set Error | | Test Set Error | | Test Set Classification | | Απόλεια Γενίκευσης | |
| | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev |
| P-RPROP | 134 | 15 | 41 | 5 | 134 | 15 | 11,19 | 0,01 | 13,28 | 0,06 | 14,29 | 0,05 | 20,65 | 0,31 | 1,14 | 0,45 |
| RPROP | 138 | 15 | 49 | 9 | 138 | 15 | 11,19 | 0,01 | 13,25 | 0,06 | 14,31 | 0,05 | 21,08 | 0,41 | 1,22 | 0,55 |
| ASD | 161 | 2 | 160 | 2 | 161 | 2 | 11,21 | 0,00 | 13,13 | 0,00 | 14,13 | 0,00 | 20,43 | 0,00 | 0,01 | 0,01 |
| OFR | 75 | 17 | 69 | 15 | 75 | 17 | 11,19 | 0,02 | 13,11 | 0,02 | 14,08 | 0,02 | 20,48 | 0,13 | 0,06 | 0,06 |
| LMS | 76 | 75 | 65 | 82 | 76 | 75 | 31,12 | 29,23 | 33,67 | 30,53 | 33,20 | 27,26 | 31,17 | 13,96 | 1,14 | 1,62 |
| B-LMS | 216 | 2 | 216 | 2 | 216 | 2 | 11,25 | 0,00 | 13,13 | 0,00 | 14,18 | 0,00 | 20,87 | 0,00 | 0,00 | 0,00 |
| PR | 216 | 2 | 216 | 2 | 216 | 2 | 11,25 | 0,00 | 13,13 | 0,00 | 14,18 | 0,00 | 20,87 | 0,00 | 0,00 | 0,00 |
| FR | 216 | 2 | 216 | 2 | 216 | 2 | 11,25 | 0,00 | 13,13 | 0,00 | 14,18 | 0,00 | 20,87 | 0,00 | 0,00 | 0,00 |
| RGS | 200 | 195 | 192 | 197 | 5 | 5 | 18,18 | 5,56 | 19,59 | 5,53 | 20,84 | 5,37 | 26,29 | 3,50 | 5,37 | 5,08 |
| JOR | 488 | 474 | 472 | 484 | 488 | 474 | 22,46 | 18,16 | 24,19 | 18,02 | 25,08 | 17,83 | 24,35 | 5,60 | 52,62 | 88,03 |
| SOR | 198 | 82 | 192 | 81 | 5 | 2 | 15,43 | 8,35 | 17,12 | 8,01 | 18,43 | 7,95 | 24,32 | 7,41 | 1,15 | 2,64 |
| HEART2 | | | | | | | | | | | | | | | | |
| Μέθοδος | Εποχές | | Σχετικές Εποχές | | Διορθώσεις Συνάψεων | | Training Set Error | | Validation Set Error | | Test Set Error | | Test Set Classification | | Απόλεια Γενίκευσης | |
| | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev |
| P-RPROP | 184 | 14 | 146 | 48 | 184 | 14 | 11,66 | 0,01 | 12,22 | 0,02 | 13,52 | 0,06 | 16,43 | 0,40 | 0,13 | 0,09 |
| RPROP | 189 | 24 | 168 | 26 | 189 | 24 | 11,67 | 0,02 | 12,21 | 0,02 | 13,53 | 0,03 | 16,50 | 0,22 | 0,15 | 0,11 |
| ASD | 216 | 3 | 215 | 3 | 216 | 3 | 11,66 | 0,00 | 12,22 | 0,00 | 13,60 | 0,00 | 16,52 | 0,00 | 0,00 | 0,00 |
| OFR | 81 | 23 | 72 | 23 | 81 | 23 | 11,64 | 0,02 | 12,22 | 0,01 | 13,63 | 0,04 | 16,54 | 0,10 | 0,07 | 0,08 |
| LMS | 100 | 108 | 86 | 115 | 100 | 108 | 20,75 | 13,88 | 20,77 | 12,89 | 23,20 | 14,94 | 20,78 | 4,92 | 0,43 | 0,44 |
| B-LMS | 298 | 2 | 298 | 2 | 298 | 2 | 11,72 | 0,00 | 12,26 | 0,00 | 13,55 | 0,00 | 16,52 | 0,00 | 0,00 | 0,00 |
| PR | 298 | 2 | 298 | 2 | 298 | 2 | 11,72 | 0,00 | 12,26 | 0,00 | 13,55 | 0,00 | 16,52 | 0,00 | 0,00 | 0,00 |
| FR | 299 | 2 | 299 | 2 | 299 | 2 | 11,72 | 0,00 | 12,26 | 0,00 | 13,55 | 0,00 | 16,52 | 0,00 | 0,00 | 0,00 |
| RGS | 203 | 190 | 200 | 189 | 5 | 5 | 19,54 | 6,15 | 19,12 | 5,94 | 20,67 | 6,36 | 20,14 | 2,94 | 3,42 | 3,43 |
| JOR | 536 | 469 | 535 | 470 | 536 | 469 | 23,13 | 18,24 | 23,56 | 18,47 | 24,15 | 17,78 | 20,46 | 6,61 | 47,69 | 79,80 |
| SOR | 227 | 104 | 219 | 101 | 6 | 3 | 16,43 | 8,54 | 17,11 | 9,17 | 18,22 | 9,24 | 18,44 | 5,78 | 1,31 | 2,54 |
| HEART3 | | | | | | | | | | | | | | | | |
| Μέθοδος | Εποχές | | Σχετικές Εποχές | | Διορθώσεις Συνάψεων | | Training Set Error | | Validation Set Error | | Test Set Error | | Test Set Classification | | Απόλεια Γενίκευσης | |
| | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev |
| P-RPROP | 142 | 15 | 113 | 53 | 142 | 15 | 11,11 | 0,01 | 10,77 | 0,02 | 16,39 | 0,18 | 22,65 | 0,69 | 0,14 | 0,23 |
| RPROP | 141 | 13 | 85 | 48 | 141 | 13 | 11,11 | 0,00 | 10,73 | 0,06 | 16,40 | 0,11 | 23,18 | 0,90 | 0,42 | 0,48 |
| ASD | 187 | 2 | 137 | 2 | 187 | 2 | 11,11 | 0,00 | 10,61 | 0,00 | 16,27 | 0,00 | 22,61 | 0,00 | 0,09 | 0,02 |
| OFR | 87 | 21 | 57 | 10 | 87 | 21 | 11,09 | 0,01 | 10,56 | 0,02 | 16,25 | 0,04 | 22,68 | 0,32 | 0,58 | 0,32 |
| LMS | 81 | 95 | 62 | 103 | 81 | 95 | 17,08 | 8,66 | 16,02 | 7,99 | 22,56 | 9,46 | 25,47 | 3,06 | 1,32 | 1,23 |
| B-LMS | 263 | 3 | 263 | 3 | 263 | 2 | 11,15 | 0,00 | 10,60 | 0,00 | 16,27 | 0,00 | 22,61 | 0,00 | 0,00 | 0,00 |
| PR | 262 | 2 | 262 | 2 | 262 | 2 | 11,15 | 0,00 | 10,59 | 0,00 | 16,27 | 0,00 | 22,61 | 0,00 | 0,00 | 0,00 |
| FR | 262 | 2 | 262 | 2 | 262 | 2 | 11,15 | 0,00 | 10,60 | 0,00 | 16,27 | 0,00 | 22,61 | 0,00 | 0,00 | 0,00 |
| RGS | 204 | 174 | 202 | 176 | 6 | 5 | 18,44 | 5,82 | 17,67 | 5,42 | 25,30 | 8,00 | 27,37 | 2,69 | 21,95 | 20,84 |
| JOR | 490 | 468 | 487 | 470 | 490 | 468 | 22,34 | 18,11 | 22,55 | 19,26 | 26,81 | 17,34 | 27,55 | 7,31 | 46,27 | 77,43 |
| SOR | 226 | 99 | 218 | 101 | 6 | 3 | 14,85 | 7,71 | 14,15 | 6,78 | 20,42 | 8,66 | 24,99 | 1,92 | 0,75 | 1,76 |

Πίνακας 7.7 Μέση Τιμή και Απόκλιση για το Πρόβλημα Διάγνωσης Καρδιοπάθειας (heartc)

| HEARTC1 | | | | | | | | | | | | | | | | |
|---------|--------|-------|-----------------|-------|---------------------|-------|--------------------|-------|----------------------|-------|----------------|-------|-------------------------|-------|--------------------|--------|
| Μέθοδος | Εποχές | | Σχετικές Εποχές | | Διορθώσεις Συνάψεων | | Training Set Error | | Validation Set Error | | Test Set Error | | Test Set Classification | | Απόλεια Γενίκευσης | |
| | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev |
| P-RPROP | 128 | 10 | 124 | 23 | 128 | 10 | 10,17 | 0,01 | 9,65 | 0,03 | 16,12 | 0,04 | 19,73 | 0,56 | 0,15 | 0,11 |
| RPROP | 120 | 27 | 82 | 40 | 120 | 27 | 10,20 | 0,15 | 9,62 | 0,12 | 16,47 | 0,64 | 20,00 | 0,97 | 0,82 | 1,40 |
| ASD | 155 | 0 | 155 | 0 | 155 | 0 | 10,18 | 0,00 | 9,61 | 0,00 | 16,00 | 0,00 | 20,00 | 0,00 | 0,00 | 0,00 |
| OFR | 66 | 13 | 44 | 18 | 66 | 13 | 10,16 | 0,01 | 9,57 | 0,03 | 16,04 | 0,07 | 19,44 | 0,66 | 0,40 | 0,25 |
| LMS | 86 | 108 | 85 | 109 | 86 | 108 | 33,88 | 41,79 | 31,85 | 38,65 | 31,66 | 23,66 | 23,79 | 8,57 | 137,77 | 254,11 |
| B-LMS | 319 | 2 | 319 | 2 | 319 | 2 | 10,30 | 0,00 | 9,68 | 0,01 | 16,03 | 0,00 | 18,67 | 0,00 | 0,00 | 0,00 |
| PR | 155 | 156 | 153 | 158 | 155 | 156 | 17,44 | 7,63 | 17,14 | 7,90 | 24,25 | 9,04 | 24,91 | 8,34 | 2074,3 | 5869,7 |
| FR | 319 | 4 | 319 | 4 | 319 | 4 | 10,30 | 0,00 | 9,68 | 0,01 | 16,03 | 0,01 | 18,67 | 0,00 | 0,00 | 0,00 |
| RGS | 303 | 47 | 302 | 46 | 9 | 1 | 12,54 | 1,83 | 11,87 | 1,79 | 17,31 | 0,44 | 21,82 | 3,33 | 0,03 | 0,06 |
| JOR | 490 | 536 | 489 | 537 | 490 | 536 | 18,93 | 14,09 | 18,32 | 14,24 | 24,19 | 12,93 | 19,86 | 0,41 | 3,47 | 5,81 |
| SOR | 92 | 74 | 85 | 73 | 2 | 2 | 18,43 | 9,19 | 16,25 | 7,77 | 24,67 | 10,14 | 23,58 | 3,87 | 3,96 | 4,42 |

| HEARTC2 | | | | | | | | | | | | | | | | |
|---------|--------|-------|-----------------|-------|---------------------|-------|--------------------|--------|----------------------|--------|----------------|--------|-------------------------|-------|--------------------|--------|
| Μέθοδος | Εποχές | | Σχετικές Εποχές | | Διορθώσεις Συνάψεων | | Training Set Error | | Validation Set Error | | Test Set Error | | Test Set Classification | | Απόλεια Γενίκευσης | |
| | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev |
| P-RPROP | 136 | 22 | 25 | 10 | 136 | 22 | 11,23 | 0,03 | 16,51 | 0,08 | 6,34 | 0,25 | 3,20 | 1,56 | 3,98 | 0,56 |
| RPROP | 96 | 61 | 26 | 16 | 96 | 61 | 11,78 | 0,93 | 16,58 | 0,43 | 6,66 | 0,98 | 3,79 | 1,90 | 5,38 | 2,25 |
| ASD | 170 | 0 | 28 | 1 | 170 | 0 | 11,23 | 0,00 | 16,54 | 0,01 | 6,15 | 0,01 | 3,23 | 0,66 | 2,87 | 0,04 |
| OFR | 72 | 19 | 22 | 2 | 72 | 19 | 11,22 | 0,01 | 16,53 | 0,04 | 6,22 | 0,10 | 2,46 | 0,49 | 3,41 | 0,41 |
| LMS | 101 | 162 | 31 | 59 | 101 | 162 | 82,05 | 132,18 | 78,11 | 113,16 | 61,37 | 100,76 | 8,35 | 4,74 | 9,68 | 7,49 |
| B-LMS | 459 | 2 | 160 | 4 | 459 | 2 | 11,36 | 0,00 | 16,55 | 0,01 | 6,16 | 0,01 | 2,88 | 0,49 | 1,11 | 0,03 |
| PR | 244 | 226 | 86 | 79 | 244 | 226 | 27,18 | 19,35 | 28,04 | 14,58 | 21,69 | 19,05 | 20,49 | 20,63 | 668,04 | 2333,5 |
| FR | 460 | 2 | 161 | 3 | 460 | 2 | 11,36 | 0,00 | 16,55 | 0,01 | 6,16 | 0,00 | 2,74 | 0,30 | 1,10 | 0,02 |
| RGS | 305 | 47 | 296 | 43 | 9 | 2 | 14,09 | 1,92 | 17,50 | 0,90 | 8,90 | 2,74 | 5,26 | 1,33 | 0,45 | 0,22 |
| JOR | 605 | 549 | 394 | 584 | 605 | 549 | 20,25 | 14,11 | 24,55 | 13,21 | 15,79 | 15,59 | 5,89 | 1,00 | 4,01 | 5,56 |
| SOR | 106 | 106 | 98 | 103 | 3 | 3 | 20,46 | 9,04 | 24,64 | 8,59 | 16,38 | 10,54 | 11,86 | 10,47 | 7,03 | 6,93 |

| HEARTC3 | | | | | | | | | | | | | | | | |
|---------|--------|-------|-----------------|-------|---------------------|-------|--------------------|-------|----------------------|-------|----------------|-------|-------------------------|-------|--------------------|--------|
| Μέθοδος | Εποχές | | Σχετικές Εποχές | | Διορθώσεις Συνάψεων | | Training Set Error | | Validation Set Error | | Test Set Error | | Test Set Classification | | Απόλεια Γενίκευσης | |
| | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev |
| P-RPROP | 26 | 9 | 12 | 3 | 26 | 9 | 10,48 | 0,31 | 13,88 | 0,33 | 12,53 | 0,44 | 14,27 | 1,67 | 6,23 | 1,15 |
| RPROP | 20 | 6 | 13 | 2 | 20 | 6 | 11,12 | 0,62 | 13,99 | 0,35 | 13,21 | 0,41 | 13,75 | 1,44 | 8,31 | 3,33 |
| ASD | 24 | 2 | 5 | 0 | 24 | 2 | 10,44 | 0,06 | 13,12 | 0,03 | 12,08 | 0,02 | 15,44 | 0,66 | 6,06 | 0,53 |
| OFR | 20 | 0 | 5 | 0 | 20 | 0 | 10,33 | 0,02 | 13,11 | 0,03 | 12,09 | 0,03 | 15,72 | 0,54 | 8,75 | 0,50 |
| LMS | 25 | 41 | 6 | 10 | 25 | 41 | 30,94 | 32,72 | 40,45 | 46,42 | 42,08 | 53,35 | 21,75 | 8,14 | 23,71 | 20,20 |
| B-LMS | 125 | 3 | 29 | 1 | 125 | 3 | 10,51 | 0,01 | 13,10 | 0,02 | 12,05 | 0,02 | 15,16 | 0,64 | 5,15 | 0,07 |
| PR | 68 | 59 | 17 | 13 | 68 | 59 | 23,48 | 18,00 | 28,78 | 20,69 | 25,45 | 18,55 | 22,95 | 8,87 | 73,32 | 124,70 |
| FR | 123 | 4 | 29 | 1 | 123 | 4 | 10,51 | 0,02 | 13,10 | 0,02 | 12,05 | 0,02 | 15,37 | 0,67 | 5,15 | 0,07 |
| RGS | 303 | 40 | 197 | 59 | 8 | 1 | 13,12 | 2,32 | 14,58 | 1,03 | 13,67 | 0,88 | 15,86 | 1,29 | 1,77 | 1,64 |
| JOR | 478 | 716 | 291 | 480 | 478 | 716 | 19,13 | 14,11 | 21,48 | 12,94 | 20,18 | 13,33 | 13,82 | 0,64 | 6,14 | 3,57 |
| SOR | 85 | 100 | 60 | 88 | 2 | 3 | 18,99 | 9,90 | 20,06 | 7,51 | 21,48 | 10,93 | 24,98 | 11,90 | 7,87 | 4,82 |

Πίνακας 7.8 Μέση Τιμή και Απόκλιση για το Πρόβλημα Διάγνωσης Κολικού Αλόγου

| HORSE1 | | | | | | | | | | | | | | | | |
|----------------|--------|-------|-----------------|-------|---------------------|-------|--------------------|-------|----------------------|-------|----------------|-------|-------------------------|-------|--------------------|--------|
| Μέθοδος | Εποχές | | Σχετικές Εποχές | | Διορθώσεις Συνάψεων | | Training Set Error | | Validation Set Error | | Test Set Error | | Test Set Classification | | Απόλεια Γενίκευσης | |
| | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev |
| P-RPROP | 27 | 7 | 9 | 2 | 27 | 7 | 11,31 | 0,16 | 15,53 | 0,29 | 12,93 | 0,38 | 26,70 | 1,87 | 6,22 | 0,57 |
| RPROP | 30 | 6 | 13 | 4 | 30 | 6 | 11,30 | 0,20 | 15,55 | 0,24 | 12,87 | 0,33 | 26,20 | 2,60 | 6,25 | 0,64 |
| ASD | 32 | 2 | 5 | 0 | 32 | 2 | 11,25 | 0,05 | 15,21 | 0,04 | 12,76 | 0,05 | 26,49 | 1,23 | 5,59 | 0,36 |
| OFR | 29 | 2 | 5 | 0 | 29 | 2 | 11,10 | 0,07 | 15,20 | 0,05 | 12,75 | 0,04 | 26,95 | 1,09 | 6,84 | 0,82 |
| LMS | 60 | 48 | 25 | 16 | 60 | 48 | 19,10 | 9,88 | 24,56 | 12,38 | 20,77 | 10,66 | 30,19 | 5,50 | 3,60 | 2,78 |
| B-LMS | 128 | 4 | 29 | 1 | 128 | 4 | 11,25 | 0,02 | 15,18 | 0,05 | 12,64 | 0,03 | 25,45 | 0,89 | 5,14 | 0,08 |
| PR | 127 | 3 | 29 | 1 | 127 | 3 | 11,26 | 0,02 | 15,18 | 0,03 | 12,63 | 0,03 | 25,51 | 0,67 | 5,14 | 0,08 |
| FR | 126 | 4 | 29 | 1 | 126 | 4 | 11,26 | 0,02 | 15,16 | 0,05 | 12,63 | 0,03 | 25,51 | 0,57 | 5,15 | 0,08 |
| RGS | 188 | 194 | 117 | 139 | 3 | 3 | 24,24 | 10,44 | 25,23 | 9,13 | 24,11 | 10,21 | 44,42 | 21,64 | 1341,4 | 2158,4 |
| JOR | 628 | 697 | 293 | 370 | 628 | 697 | 11,50 | 0,27 | 15,01 | 0,05 | 12,62 | 0,04 | 24,87 | 0,53 | 4,20 | 1,97 |
| SOR | 141 | 163 | 72 | 104 | 2 | 3 | 16,20 | 5,01 | 17,47 | 2,75 | 14,59 | 2,26 | 27,18 | 3,44 | 6,48 | 3,39 |
| HORSE 2 | | | | | | | | | | | | | | | | |
| Μέθοδος | Εποχές | | Σχετικές Εποχές | | Διορθώσεις Συνάψεων | | Training Set Error | | Validation Set Error | | Test Set Error | | Test Set Classification | | Απόλεια Γενίκευσης | |
| | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev |
| P-RPROP | 42 | 16 | 13 | 3 | 42 | 16 | 8,62 | 0,28 | 15,99 | 0,21 | 17,43 | 0,45 | 34,84 | 1,38 | 5,54 | 0,47 |
| RPROP | 32 | 7 | 14 | 3 | 32 | 7 | 8,90 | 0,21 | 15,73 | 0,27 | 17,34 | 0,50 | 36,50 | 1,31 | 5,60 | 0,96 |
| ASD | 60 | 0 | 15 | 1 | 60 | 0 | 8,36 | 0,01 | 15,68 | 0,02 | 16,62 | 0,06 | 35,80 | 0,54 | 5,22 | 0,13 |
| OFR | 27 | 3 | 13 | 1 | 27 | 3 | 8,44 | 0,06 | 15,69 | 0,03 | 16,66 | 0,08 | 35,51 | 0,51 | 6,82 | 1,31 |
| LMS | 86 | 109 | 18 | 23 | 86 | 109 | 13,36 | 6,08 | 21,35 | 7,54 | 22,94 | 8,41 | 42,86 | 9,25 | 6,32 | 0,97 |
| B-LMS | 254 | 5 | 52 | 2 | 254 | 5 | 8,42 | 0,01 | 15,61 | 0,02 | 16,27 | 0,05 | 36,32 | 0,25 | 5,09 | 0,04 |
| PR | 253 | 6 | 52 | 2 | 253 | 6 | 8,42 | 0,01 | 15,61 | 0,03 | 16,26 | 0,07 | 36,26 | 0,00 | 5,07 | 0,04 |
| FR | 253 | 5 | 52 | 3 | 253 | 5 | 8,42 | 0,01 | 15,61 | 0,23 | 16,26 | 0,06 | 36,26 | 0,00 | 5,08 | 0,04 |
| RGS | 249 | 279 | 237 | 276 | 4 | 4 | 22,66 | 11,33 | 24,88 | 8,47 | 25,53 | 9,18 | 51,82 | 17,00 | 1847,3 | 2990,8 |
| JOR | 787 | 881 | 342 | 432 | 787 | 881 | 9,03 | 0,24 | 15,33 | 0,03 | 15,37 | 0,08 | 33,49 | 0,75 | 4,09 | 1,67 |
| SOR | 179 | 199 | 83 | 129 | 3 | 3 | 14,48 | 5,87 | 19,04 | 3,89 | 19,47 | 4,31 | 37,65 | 4,05 | 4,67 | 1,63 |
| HORSE 3 | | | | | | | | | | | | | | | | |
| Μέθοδος | Εποχές | | Σχετικές Εποχές | | Διορθώσεις Συνάψεων | | Training Set Error | | Validation Set Error | | Test Set Error | | Test Set Classification | | Απόλεια Γενίκευσης | |
| | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev |
| P-RPROP | 26 | 6 | 8 | 3 | 26 | 6 | 10,43 | 0,27 | 15,59 | 0,30 | 15,50 | 0,45 | 32,42 | 2,65 | 6,34 | 1,07 |
| RPROP | 22 | 7 | 9 | 2 | 22 | 7 | 10,73 | 0,42 | 15,43 | 0,28 | 15,24 | 0,34 | 31,87 | 1,82 | 5,89 | 0,90 |
| ASD | 25 | 1 | 7 | 1 | 25 | 1 | 10,26 | 0,03 | 15,33 | 0,04 | 15,13 | 0,05 | 33,20 | 0,67 | 5,30 | 0,29 |
| OFR | 20 | 0 | 7 | 1 | 20 | 0 | 10,24 | 10,03 | 15,31 | 0,04 | 15,13 | 0,05 | 33,49 | 0,75 | 6,06 | 0,35 |
| LMS | 35 | 40 | 8 | 10 | 35 | 40 | 22,65 | 16,11 | 25,35 | 13,14 | 25,10 | 13,21 | 45,98 | 17,77 | 19,24 | 17,71 |
| B-LMS | 97 | 4 | 22 | 1 | 97 | 4 | 10,30 | 0,03 | 15,30 | 0,05 | 15,03 | 0,04 | 33,31 | 0,51 | 5,14 | 0,08 |
| PR | 98 | 3 | 22 | 1 | 98 | 3 | 10,29 | 0,03 | 15,33 | 0,03 | 14,99 | 0,05 | 33,02 | 0,25 | 5,12 | 0,07 |
| FR | 98 | 3 | 22 | 1 | 98 | 3 | 10,29 | 0,025 | 15,32 | 0,035 | 15,00 | 0,05 | 33,26 | 0,60 | 5,16 | 0,08 |
| RGS | 325 | 438 | 257 | 435 | 5 | 7 | 24,60 | 12,70 | 26,07 | 10,69 | 25,36 | 9,72 | 48,70 | 18,20 | 1370,9 | 2190,2 |
| JOR | 673 | 817 | 225 | 284 | 673 | 817 | 10,63 | 0,12 | 14,98 | 0,04 | 14,79 | 0,06 | 31,46 | 0,53 | 4,72 | 0,75 |
| SOR | 129 | 156 | 60 | 73 | 2 | 3 | 14,86 | 4,68 | 16,99 | 3,11 | 18,35 | 3,80 | 33,95 | 3,63 | 6,61 | 2,91 |

Πίνακας 7.9 Μέση Τιμή και Απόκλιση για το Πρόβλημα της Αρρώστιας των Φασολιών

| SOYBEAN1 | | | | | | | | | | | | | | | | |
|----------|--------|------|-----------------|------|---------------------|------|--------------------|------|----------------------|------|----------------|------|-------------------------|-------|--------------------|--------|
| Μέθοδος | Εποχές | | Σχετικές Εποχές | | Διορθώσεις Συνάψεων | | Training Set Error | | Validation Set Error | | Test Set Error | | Test Set Classification | | Απόλεια Γενίκευσης | |
| P-RPROP | 553 | 11 | 418 | 41 | 553 | 11 | 0,65 | 0,00 | 0,98 | 0,00 | 1,16 | 0,00 | 9,47 | 0,51 | 0,28 | 0,18 |
| RPROP | 543 | 15 | 420 | 51 | 543 | 15 | 0,65 | 0,00 | 0,98 | 0,00 | 1,16 | 0,00 | 9,57 | 0,32 | 0,26 | 0,11 |
| ASD | 5 | 1 | 2 | 1 | 5 | 1 | 3,73 | 0,15 | 3,88 | 0,14 | 3,94 | 0,13 | 58,08 | 6,45 | 2507,9 | 2011,2 |
| OFR | 5 | 0 | 2 | 0 | 5 | 0 | 3,78 | 0,08 | 3,93 | 0,08 | 3,98 | 0,08 | 60,62 | 4,94 | 2409,8 | 2052,4 |
| LMS | 1324 | 1106 | 1271 | 1160 | 1324 | 1106 | 0,71 | 0,03 | 1,00 | 0,03 | 1,20 | 0,03 | 9,44 | 0,49 | 0,63 | 0,82 |
| B-LMS | 2879 | 7 | 2879 | 7 | 2879 | 7 | 0,75 | 0,00 | 1,01 | 0,00 | 1,19 | 0,00 | 8,82 | 0,00 | 0,00 | 0,00 |
| PR | 1366 | 1435 | 1364 | 1437 | 1366 | 1435 | 3,42 | 2,53 | 3,67 | 2,52 | 3,79 | 2,47 | 43,22 | 32,74 | 1396,4 | 1329,3 |
| FR | 2880 | 5 | 2880 | 5 | 2880 | 5 | 0,75 | 0,00 | 1,01 | 0,00 | 1,19 | 0,00 | 8,82 | 0,00 | 0,00 | 0,00 |
| RGS | 791 | 562 | 790 | 564 | 9 | 7 | 2,94 | 0,78 | 3,17 | 0,79 | 3,28 | 0,77 | 39,23 | 15,70 | 3,83 | 5,03 |
| JOR | 2049 | 1248 | 2048 | 1250 | 2049 | 1248 | 2,12 | 1,44 | 2,31 | 1,36 | 2,44 | 1,29 | 27,59 | 19,67 | 136,39 | 229,57 |
| SOR | 722 | 384 | 680 | 345 | 8 | 5 | 1,25 | 0,87 | 1,55 | 0,84 | 1,70 | 0,82 | 16,47 | 16,13 | 0,66 | 1,21 |
| SOYBEAN2 | | | | | | | | | | | | | | | | |
| Μέθοδος | Εποχές | | Σχετικές Εποχές | | Διορθώσεις Συνάψεων | | Training Set Error | | Validation Set Error | | Test Set Error | | Test Set Classification | | Απόλεια Γενίκευσης | |
| P-RPROP | 509 | 19 | 504 | 18 | 509 | 19 | 0,80 | 0,00 | 0,81 | 0,00 | 1,05 | 0,00 | 4,24 | 0,25 | 0,02 | 0,02 |
| RPROP | 491 | 18 | 474 | 28 | 491 | 18 | 0,80 | 0,00 | 0,81 | 0,00 | 1,05 | 0,00 | 4,12 | 0,00 | 0,05 | 0,04 |
| ASD | 5 | 1 | 2 | 1 | 5 | 1 | 3,89 | 0,11 | 3,75 | 0,12 | 4,15 | 0,09 | 55,26 | 4,22 | 2668,2 | 2138,2 |
| OFR | 5 | 1 | 2 | 1 | 5 | 1 | 3,85 | 0,12 | 3,71 | 0,13 | 4,12 | 0,10 | 55,91 | 4,61 | 2275,0 | 2047,1 |
| LMS | 1147 | 936 | 1120 | 962 | 1147 | 936 | 0,86 | 0,03 | 0,85 | 0,02 | 1,12 | 0,05 | 4,98 | 1,14 | 0,07 | 0,09 |
| B-LMS | 2468 | 5 | 2468 | 5 | 2468 | 5 | 0,90 | 0,00 | 0,87 | 0,00 | 1,18 | 0,00 | 4,12 | 0,00 | 0,00 | 0,00 |
| PR | 1172 | 1231 | 1171 | 1233 | 1172 | 1231 | 3,66 | 2,62 | 3,51 | 2,51 | 3,99 | 2,66 | 37,24 | 31,53 | 2128,8 | 2022,7 |
| FR | 2469 | 4 | 2469 | 4 | 2469 | 4 | 0,90 | 0,00 | 0,87 | 0,00 | 1,18 | 0,00 | 4,12 | 0,00 | 0,00 | 0,00 |
| RGS | 443 | 515 | 442 | 516 | 5 | 6 | 3,40 | 1,13 | 3,29 | 1,17 | 3,85 | 1,02 | 46,69 | 19,62 | 61,07 | 7,95 |
| JOR | 1971 | 1234 | 1970 | 1235 | 1971 | 1234 | 2,17 | 1,31 | 2,08 | 1,35 | 2,45 | 1,27 | 24,86 | 15,51 | 117,35 | 197,36 |
| SOR | 692 | 318 | 683 | 315 | 8 | 4 | 1,20 | 0,09 | 1,08 | 0,07 | 1,53 | 0,10 | 12,29 | 5,50 | 0,32 | 0,33 |
| SOYBEAN3 | | | | | | | | | | | | | | | | |
| Μέθοδος | Εποχές | | Σχετικές Εποχές | | Διορθώσεις Συνάψεων | | Training Set Error | | Validation Set Error | | Test Set Error | | Test Set Classification | | Απόλεια Γενίκευσης | |
| P-RPROP | 533 | 27 | 522 | 28 | 533 | 27 | 0,78 | 0,00 | 0,96 | 0,00 | 1,03 | 0,00 | 7,00 | 0,19 | 0,03 | 0,04 |
| RPROP | 518 | 20 | 498 | 34 | 518 | 20 | 0,78 | 0,00 | 0,96 | 0,00 | 1,04 | 0,00 | 7,09 | 0,13 | 0,04 | 0,03 |
| ASD | 5 | 0 | 2 | 0 | 5 | 0 | 3,85 | 0,07 | 3,87 | 0,07 | 3,96 | 0,07 | 50,06 | 4,53 | 2130,8 | 2171,0 |
| OFR | 5 | 0 | 2 | 0 | 5 | 0 | 3,88 | 0,08 | 3,90 | 0,09 | 3,99 | 0,08 | 50,34 | 5,43 | 2976,4 | 2237,4 |
| LMS | 1211 | 955 | 1176 | 990 | 1211 | 955 | 0,84 | 0,04 | 1,00 | 0,03 | 1,07 | 0,03 | 6,22 | 0,52 | 0,10 | 0,14 |
| B-LMS | 2541 | 5 | 2541 | 5 | 2541 | 5 | 0,89 | 0,00 | 1,04 | 0,00 | 1,10 | 0,00 | 7,12 | 0,38 | 0,00 | 0,00 |
| PR | 1206 | 1265 | 1206 | 1265 | 1206 | 1265 | 3,56 | 2,54 | 3,63 | 2,46 | 3,70 | 2,47 | 39,44 | 30,81 | 1756,8 | 1672,8 |
| FR | 2541 | 5 | 2541 | 5 | 2541 | 5 | 0,89 | 0,00 | 1,04 | 0,00 | 1,10 | 0,00 | 7,18 | 0,36 | 0,00 | 0,00 |
| RGS | 858 | 630 | 857 | 631 | 10 | 8 | 2,96 | 0,76 | 3,07 | 0,75 | 3,25 | 0,77 | 35,11 | 11,38 | 5,62 | 7,36 |
| JOR | 1980 | 1234 | 1978 | 1236 | 1980 | 1234 | 2,15 | 1,32 | 2,26 | 1,27 | 2,31 | 1,24 | 24,27 | 16,27 | 126,01 | 211,19 |
| SOR | 805 | 377 | 794 | 360 | 9 | 5 | 1,14 | 0,10 | 1,30 | 0,08 | 1,43 | 0,11 | 10,99 | 2,16 | 0,29 | 0,62 |

Πίνακας 7.10 Μέση Τιμή και Απόκλιση για το Πρόβλημα Διάγνωσης Θυρεοειδούς

| THYROID1 | | | | | | | | | | | | | | | | |
|----------|--------|-------|-----------------|-------|---------------------|-------|--------------------|-------|----------------------|-------|----------------|-------|-------------------------|-------|--------------------|-------|
| Μέθοδος | Εποχές | | Σχετικές Εποχές | | Διορθώσεις Συνάψεων | | Training Set Error | | Validation Set Error | | Test Set Error | | Test Set Classification | | Απόλεια Γενίκευσης | |
| | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev |
| P-RPROP | 104 | 16 | 99 | 22 | 104 | 16 | 3,76 | 0,00 | 3,78 | 0,01 | 3,84 | 0,012 | 6,56 | 0,00 | 0,01 | 0,03 |
| RPROP | 246 | 50 | 243 | 51 | 246 | 50 | 3,91 | 0,01 | 3,95 | 0,02 | 4,08 | 0,02 | 6,53 | 0,03 | 0,03 | 0,03 |
| ASD | 810 | 11 | 810 | 11 | 810 | 11 | 4,03 | 0,00 | 4,17 | 0,00 | 4,22 | 0,00 | 6,56 | 0,00 | 0,01 | 0,01 |
| OFR | 346 | 121 | 344 | 120 | 346 | 121 | 3,97 | 0,04 | 4,05 | 0,09 | 4,13 | 0,08 | 6,55 | 0,02 | 0,01 | 0,01 |
| LMS | 90 | 197 | 88 | 197 | 90 | 196 | 10,72 | 7,68 | 10,28 | 7,37 | 10,81 | 7,87 | 16,45 | 14,72 | 0,21 | 0,60 |
| B-LMS | 874 | 2 | 874 | 2 | 874 | 2 | 4,05 | 0,00 | 4,21 | 0,00 | 4,25 | 0,00 | 6,56 | 0,00 | 0,00 | 0,00 |
| PR | 874 | 2 | 874 | 2 | 874 | 2 | 4,05 | 0,00 | 4,21 | 0,00 | 4,25 | 0,00 | 6,56 | 0,00 | 0,00 | 0,00 |
| FR | 873 | 2 | 873 | 2 | 873 | 2 | 4,05 | 0,00 | 4,21 | 0,00 | 4,25 | 0,00 | 6,56 | 0,00 | 0,00 | 0,00 |
| RGS | 163 | 51 | 163 | 51 | 7 | 2 | 6,98 | 0,45 | 7,10 | 0,43 | 6,87 | 0,55 | 7,33 | 0,00 | 0,00 | 0,00 |
| JOR | 572 | 347 | 572 | 347 | 572 | 347 | 4,05 | 0,16 | 4,12 | 0,19 | 4,24 | 0,17 | 6,55 | 0,06 | 0,00 | 0,00 |
| SOR | 123 | 88 | 119 | 89 | 5 | 4 | 6,87 | 6,28 | 7,07 | 6,24 | 7,06 | 6,23 | 6,70 | 0,21 | 0,06 | 0,06 |

| THYROID2 | | | | | | | | | | | | | | | | |
|----------|--------|-------|-----------------|-------|---------------------|-------|--------------------|-------|----------------------|-------|----------------|-------|-------------------------|-------|--------------------|-------|
| Μέθοδος | Εποχές | | Σχετικές Εποχές | | Διορθώσεις Συνάψεων | | Training Set Error | | Validation Set Error | | Test Set Error | | Test Set Classification | | Απόλεια Γενίκευσης | |
| | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev |
| P-RPROP | 98 | 16 | 96 | 16 | 98 | 16 | 3,93 | 0,00 | 3,55 | 0,01 | 3,71 | 0,01 | 6,56 | 0,00 | 0,01 | 0,02 |
| RPROP | 251 | 40 | 248 | 39 | 251 | 40 | 4,10 | 0,01 | 3,68 | 0,02 | 3,86 | 0,02 | 6,38 | 0,00 | 0,05 | 0,05 |
| ASD | 914 | 10 | 913 | 10 | 914 | 10 | 4,23 | 0,00 | 3,81 | 0,00 | 3,99 | 0,00 | 6,38 | 0,00 | 0,01 | 0,01 |
| OFR | 453 | 110 | 452 | 110 | 453 | 110 | 4,17 | 0,04 | 3,77 | 0,04 | 3,95 | 0,04 | 6,38 | 0,00 | 0,01 | 0,01 |
| LMS | 119 | 219 | 117 | 219 | 119 | 219 | 12,08 | 8,76 | 11,72 | 8,75 | 12,31 | 9,24 | 19,64 | 12,55 | 0,06 | 0,19 |
| B-LMS | 965 | 0 | 965 | 0 | 965 | 0 | 4,24 | 0,00 | 3,82 | 0,00 | 4,01 | 0,00 | 6,38 | 0,00 | 0,00 | 0,00 |
| PR | 965 | 1 | 965 | 1 | 965 | 1 | 4,24 | 0,00 | 3,82 | 0,00 | 4,01 | 0,00 | 6,38 | 0,00 | 0,00 | 0,00 |
| FR | 965 | 1 | 965 | 1 | 965 | 1 | 4,24 | 0,00 | 3,82 | 0,00 | 4,01 | 0,00 | 6,38 | 0,00 | 0,00 | 0,00 |
| RGS | 180 | 0 | 174 | 6 | 8 | 0 | 7,00 | 0,01 | 6,56 | 0,01 | 6,65 | 0,01 | 7,13 | 0,00 | 0,00 | 0,00 |
| JOR | 590 | 329 | 590 | 329 | 590 | 329 | 4,27 | 0,19 | 3,83 | 0,15 | 4,01 | 0,15 | 6,39 | 0,05 | 0,00 | 0,00 |
| SOR | 138 | 106 | 131 | 107 | 6 | 5 | 6,99 | 6,23 | 6,52 | 6,34 | 6,68 | 6,30 | 6,48 | 0,20 | 0,37 | 0,72 |

| THYROID3 | | | | | | | | | | | | | | | | |
|----------|--------|-------|-----------------|-------|---------------------|-------|--------------------|-------|----------------------|-------|----------------|-------|-------------------------|-------|--------------------|-------|
| Μέθοδος | Εποχές | | Σχετικές Εποχές | | Διορθώσεις Συνάψεων | | Training Set Error | | Validation Set Error | | Test Set Error | | Test Set Classification | | Απόλεια Γενίκευσης | |
| | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev |
| P-RPROP | 114 | 22 | 109 | 21 | 114 | 22 | 3,85 | 0,00 | 3,39 | 0,00 | 4,02 | 0,00 | 7,23 | 0,02 | 0,02 | 0,02 |
| RPROP | 280 | 65 | 275 | 67 | 280 | 65 | 4,01 | 0,01 | 3,48 | 0,01 | 4,19 | 0,01 | 7,23 | 0,02 | 0,05 | 0,03 |
| ASD | 904 | 7 | 903 | 7 | 904 | 7 | 4,16 | 0,00 | 3,65 | 0,00 | 4,33 | 0,00 | 7,17 | 0,00 | 0,02 | 0,01 |
| OFR | 424 | 126 | 422 | 125 | 424 | 126 | 4,08 | 0,06 | 3,56 | 0,06 | 4,25 | 0,05 | 7,22 | 0,06 | 0,02 | 0,02 |
| LMS | 96 | 214 | 94 | 215 | 96 | 214 | 5,37 | 2,50 | 4,81 | 2,56 | 5,47 | 2,25 | 7,80 | 1,33 | 0,02 | 0,06 |
| B-LMS | 944 | 2 | 944 | 2 | 944 | 2 | 4,19 | 0,00 | 3,67 | 0,00 | 4,36 | 0,00 | 7,22 | 0,00 | 0,00 | 0,00 |
| PR | 943 | 2 | 943 | 2 | 943 | 2 | 4,19 | 0,00 | 3,67 | 0,00 | 4,36 | 0,00 | 7,22 | 0,00 | 0,00 | 0,00 |
| FR | 942 | 2 | 942 | 2 | 942 | 2 | 4,19 | 0,00 | 3,67 | 0,00 | 4,36 | 0,00 | 7,22 | 0,00 | 0,00 | 0,00 |
| RGS | 180 | 0 | 180 | 0 | 8 | 0 | 6,98 | 0,01 | 6,30 | 0,01 | 6,90 | 0,01 | 7,94 | 0,00 | 0,00 | 0,00 |
| JOR | 628 | 311 | 628 | 311 | 628 | 311 | 4,19 | 0,20 | 3,65 | 0,18 | 4,35 | 0,19 | 7,22 | 0,07 | 0,00 | 0,00 |
| SOR | 137 | 106 | 135 | 107 | 6 | 5 | 6,91 | 6,25 | 6,40 | 6,37 | 7,00 | 6,22 | 7,23 | 0,29 | 0,02 | 0,02 |

7.3.4.1 Παρατηρήσεις στα Αποτελέσματα των Πινάκων (7.1)-(7.10)

Στους πίνακες (7.1)-(7.10), μπορούν να γίνουν οι παρακάτω παρατηρήσεις:

1. Εκτός της μεθόδου SOR, η οποία, σε όλα τα προβλήματα, συγκλίνει για όλες τις τιμές του συντελεστή εκπαίδευσης ω στο διάστημα $[0.01, 1.9]$, η μέγιστη τιμή του συντελεστή εκπαίδευσης ω , ω_{max} , για τον οποίο συγκλίνει η JOR παίρνει τιμές στο διάστημα $[0.1, 0.5]$. Το ω_{max} για την RGS είναι 0.01 για τα περισσότερα προβλήματα, εκτός του προβλήματος HORSE, όπου $\omega_{max}=0.2$ και του προβλήματος THYROID, που $\omega_{max}=0.001$. Το ω_{max} για την LMS παίρνει τιμές στο διάστημα $[0.01, 1]$, όπου $\omega_{max}=1$ για τα προβλήματα DIABETES, GLASS και THYROID και 0.01 για τα προβλήματα, GENE και SOYBEAN. Το ω_{max} για τις μεθόδους BATCH-LMS, PR και FR είναι 0.001 για τα περισσότερα προβλήματα, εκτός των προβλημάτων GENE και SOYBEAN, όπου $\omega_{max}=0.0001$ και του προβλήματος GLASS, όπου $\omega_{max}=0.01$
2. Σε μερικά απ' τα προβλήματα, στα αποτελέσματα της μεθόδου RPROP (αν και χρησιμοποιήσαμε την ίδια μέθοδο και τις ίδιες παραμέτρους της μεθόδου P-RPROP που δημοσίευσε ο Prechelt [Prec94]), υπήρχε απόκλιση απ' τα αποτελέσματα της μεθόδου P-RPROP, όπου η απόκλιση είναι μεγάλη για το πρόβλημα της Διάγνωσης Καρκίνου CANCER και των Γονιδίων GENE και μικρή για το πρόβλημα της Διάγνωσης Διαβήτη DIABETES.
3. Η Μέση Τιμή και Απόκλιση στα 20 πειράματα που έγιναν για κάθε μέθοδο δεν επηρεάζονται απ' το συντελεστή εκπαίδευσης ω , ο οποίος υπολογίζεται κάθε φορά, για τις μεθόδους RPROP, ASD και OFR. Για τις υπόλοιπες μεθόδους, όμως, ιδιαίτερα τις SOR, JOR και RGS οι διαφορετικές τιμές του ω σε κάθε πείραμα στο διάστημα $[0.0001, 1.9]$ δίνουν εντελώς διαφορετικά αποτελέσματα, οπότε η Μέση Τιμή και η Απόκλιση, γι' αυτές τις μεθόδους, παίρνουν πολύ μεγάλες τιμές.
4. Στα προβλήματα CARD και HORSE έχουν μεγάλη απώλεια γενίκευσης (πάνω από 5) όλες σχεδόν οι μέθοδοι. Επί πλέον, η μέθοδος JOR έχει μεγάλη απώλεια γενίκευσης σε όλα σχεδόν τα προβλήματα, εκτός του THYROID, η μέθοδος RGS στα προβλήματα GENE, GLASS, HEART και SOYBEAN, και η μέθοδος LMS στα προβλήματα CANCER, GLASS και HEARTC.
5. Μεγάλα σφάλματα παρατηρήθηκαν στο ποσοστιαίο σφάλμα των προτύπων εκπαίδευσης (*Training Set Error*), το ποσοστιαίο σφάλμα των προτύπων ελέγχου εγκυρότητας, (*Validation Set Error*), και το ποσοστιαίο σφάλμα των προτύπων ανάκλησης, (*Test Set Error*) για τη μέθοδο LMS σε όλα τα προβλήματα, εκτός του SOYBEAN, για τη μέθοδο JOR στα πιο πολλά προβλήματα, εκτός των GENE, HORSE, THYROID και SOYBEAN, για τη μέθοδο SOR στα προβλήματα, CARD και HEARTC, για τη μέθοδο RGS στα προβλήματα CARD, GENE, HORSE, HEART και SOYBEAN, για τη μέθοδο PR στα προβλήματα HEARTC και SOYBEAN, και για τις μεθόδους ASD και OFR στο πρόβλημα SOYBEAN.
6. Μεγάλα σφάλματα παρατηρήθηκαν στο σφάλμα ταξινόμησης των προτύπων ανάκλησης, (*Test Set Classification Error*) για τη μέθοδο LMS σε όλα τα

προβλήματα, εκτός των DIABETES και SOYBEAN, για τη μέθοδο JOR στα προβλήματα, CANCER και SOYBEAN, για τη μέθοδο SOR στα προβλήματα, CARD, GENE, HEARTC και SOYBEAN, για τη μέθοδο RGS στα προβλήματα CARD, GENE, HORSE και SOYBEAN και για τις μεθόδους ASD και OFR στο πρόβλημα SOYBEAN.

7. Στα πιο πολλά προβλήματα, οι μέθοδοι OFR και ASD απαιτούν το μικρότερο αριθμό εποχών και σχετικών εποχών που απαιτούνται για τη σύγκλιση τους και το σφάλμα ταξινόμησης των προτύπων ανάκλησης είναι το μικρότερο, ή ισοδύναμο με το σφάλμα ταξινόμησης των προτύπων ανάκλησης της μεθόδου RPROP.
8. Οι μέθοδοι SOR, JOR και RGS εμφανίζουν μεγάλες τιμές στις εποχές και στο σφάλμα ταξινόμησης των προτύπων ανάκλησης επειδή, τα αποτελέσματα που παίρνουμε για πολύ μικρές τιμές του συντελεστή εκπαίδευσης ω δεν είναι καλά. Από όλες τις μεθόδους, η μέθοδος JOR εμφανίζει τις μεγαλύτερες τιμές στις εποχές που απαιτούνται, τις σχετικές εποχές και τις επισκέψεις-διορθώσεις των συνάψεων, σε όλα τα προβλήματα, εκτός του προβλήματος GLASS, όπου πιο αργή μέθοδος είναι η RGS, του προβλήματος SOYBEAN, όπου πιο αργή μέθοδος είναι η FR και του προβλήματος THYROID, όπου πιο αργή μέθοδος είναι η B-LMS.
9. Οι μέθοδοι SOR και RGS εμφανίζουν τις μικρότερες τιμές στις επισκέψεις των συνάψεων, επειδή σε κάθε κύκλο εκπαίδευσης διορθώνονται μόνο οι συνάψεις ενός νευρώνα εισόδου, ενώ στις υπόλοιπες μεθόδους διορθώνονται όλες οι συνάψεις.
10. Οι μέθοδοι B-LMS, PR και FR εμφανίζουν παρόμοια συμπεριφορά, όσον αφορά τον αριθμό επαναλήψεων, τα σφάλματα των προτύπων και το σφάλμα ταξινόμησης, καθώς και τις τιμές του συντελεστή εκπαίδευσης, για τις οποίες συγκλίνουν.
11. Σε πολλά προβλήματα, υπάρχουν διαφορές στη συμπεριφορά των αλγορίθμων στους τρεις διαφορετικούς συνδυασμούς δεδομένων για το κάθε πρόβλημα. Έτσι, ο αριθμός επαναλήψεων που απαιτείται για τη σύγκλιση της κάθε μεθόδου στο πρόβλημα GLASS (GLASS1, GLASS2 και GLASS3) έχει μεγάλη απόκλιση για όλες τις μεθόδους, όπως και για το πρόβλημα HORSE. Απόκλιση εμφανίζει και το πρόβλημα HEARTC στο ποσοστιαίο σφάλμα των προτύπων εκπαίδευσης και ελέγχου εγκυρότητας, τα προβλήματα HEARTC και HORSE στο ποσοστιαίο σφάλμα των προτύπων ανάκλησης και τα προβλήματα CARD, GLASS, HEART, HEARTC, HORSE και SOYBEAN στο σφάλμα ταξινόμησης των προτύπων ανάκλησης (Test Set Classification Error).
12. Σε κάποια προβλήματα, τα σφάλματα έχουν μια ανώμαλη συμπεριφορά. Π.χ., στα προβλήματα CARD1, CARD3, DIABETES2, HEART3, HEARTC1 και THYROID3 το ποσοστιαίο σφάλμα των προτύπων ελέγχου εγκυρότητας είναι μικρότερο απ' το ποσοστιαίο σφάλμα των προτύπων εκπαίδευσης, ενώ στα προβλήματα CANCER1, CANCER3, HEARTC2 και THYROID2 το ποσοστιαίο σφάλμα των προτύπων ανάκλησης είναι μικρότερο απ' το ποσοστιαίο σφάλμα των προτύπων εκπαίδευσης.

Στους πίνακες (7.11)-(7.14) που ακολουθούν παρουσιάζονται τα αποτελέσματα των προσομοιώσεων για τα προβλήματα προσαρμογής.

Πίνακας 7.11 Μέση Τιμή και Απόκλιση για το Πρόβλημα Εξοικονόμησης Ενέργειας Κτιρίου

| BUILDING1 | | | | | | | | | | | | | | | | |
|------------------|--------|-------|-----------------|-------|---------------------|-------|--------------------|-------|----------------------|-------|----------------|-------|-------------------------|-------|--------------------|--------|
| Μέθοδος | Εποχές | | Σχετικές Εποχές | | Διορθώσεις Συνάψεων | | Training Set Error | | Validation Set Error | | Test Set Error | | Test Set Classification | | Απόλεια Γενίκευσης | |
| | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev |
| P-RPROP | 407 | 138 | 401 | 142 | 407 | 148 | 0,21 | 0,01 | 0,92 | 0,06 | 0,78 | 0,02 | | | 2,15 | 4,64 |
| RPROP | 348 | 35 | 344 | 35 | 348 | 35 | 0,34 | 0,00 | 0,37 | 0,00 | 0,35 | 0,00 | 0,29 | 0,00 | 0,03 | 0,03 |
| ASD | 544 | 24 | 544 | 24 | 544 | 24 | 0,34 | 0,00 | 0,37 | 0,00 | 0,35 | 0,00 | 0,29 | 0,00 | 0,00 | 0,00 |
| OFR | 343 | 130 | 342 | 129 | 343 | 130 | 0,33 | 0,00 | 0,37 | 0,00 | 0,34 | 0,00 | 0,29 | 0,00 | 0,00 | 0,01 |
| LMS | 64 | 135 | 60 | 137 | 64 | 135 | 0,70 | 0,47 | 0,73 | 0,47 | 0,72 | 0,50 | 2,09 | 4,12 | 0,19 | 0,33 |
| B-LMS | 5 | 0 | 2 | 0 | 5 | 0 | 19,22 | 0,27 | 19,84 | 0,34 | 19,75 | 0,50 | 96,21 | 6,35 | 120,76 | 34,95 |
| PR | 5 | 0 | 2 | 0 | 5 | 0 | 19,20 | 0,27 | 18,88 | 0,26 | 19,15 | 0,26 | 99,98 | 0,05 | 96,62 | 22,65 |
| FR | 5 | 0 | 2 | 0 | 5 | 0 | 19,33 | 0,26 | 19,96 | 0,35 | 19,94 | 0,50 | 97,01 | 5,76 | 128,89 | 33,76 |
| RGS | 704 | 736 | 699 | 735 | 50 | 52 | 14,10 | 12,99 | 14,38 | 13,21 | 14,11 | 12,98 | 52,77 | 49,79 | 59,43 | 56,37 |
| JOR | 1141 | 1040 | 1140 | 1040 | 1141 | 1040 | 5,14 | 10,85 | 5,12 | 10,68 | 5,02 | 10,51 | 15,89 | 35,42 | 108,20 | 249,89 |
| SOR | 967 | 750 | 963 | 750 | 68 | 54 | 0,93 | 2,33 | 0,97 | 2,33 | 0,96 | 2,40 | 3,07 | 11,85 | 0,45 | 1,49 |
| BUILDING2 | | | | | | | | | | | | | | | | |
| Μέθοδος | Εποχές | | Σχετικές Εποχές | | Διορθώσεις Συνάψεων | | Training Set Error | | Validation Set Error | | Test Set Error | | Test Set Classification | | Απόλεια Γενίκευσης | |
| | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev |
| P-RPROP | 298 | 23 | 297 | 23 | 298 | 23 | 0,34 | 0,00 | 0,37 | 0,00 | 0,35 | 0,00 | | | 0,00 | 0,01 |
| RPROP | 340 | 26 | 335 | 27 | 340 | 26 | 0,34 | 0,00 | 0,37 | 0,00 | 0,35 | 0,00 | 0,29 | 0,00 | 0,04 | 0,02 |
| ASD | 525 | 31 | 525 | 31 | 525 | 31 | 0,34 | 0,00 | 0,37 | 0,00 | 0,35 | 0,00 | 0,29 | 0,00 | 0,00 | 0,00 |
| OFR | 367 | 118 | 366 | 118 | 367 | 118 | 0,33 | 0,00 | 0,37 | 0,00 | 0,34 | 0,00 | 0,29 | 0,00 | 0,00 | 0,00 |
| LMS | 64 | 135 | 60 | 137 | 64 | 135 | 0,70 | 0,47 | 0,73 | 0,47 | 0,72 | 0,50 | 2,09 | 4,12 | 0,19 | 0,32 |
| B-LMS | 5 | 0 | 2 | 0 | 5 | 0 | 19,36 | 0,24 | 20,04 | 0,28 | 19,75 | 0,60 | 94,01 | 7,03 | 132,52 | 29,42 |
| PR | 5 | 0 | 2 | 0 | 5 | 0 | 19,23 | 0,31 | 18,92 | 0,30 | 19,19 | 0,30 | 99,97 | 0,07 | 94,23 | 27,52 |
| FR | 5 | 0 | 2 | 0 | 5 | 0 | 19,33 | 0,26 | 19,90 | 0,30 | 20,04 | 0,37 | 98,50 | 4,35 | 116,56 | 30,18 |
| RGS | 704 | 736 | 699 | 735 | 50 | 52 | 14,08 | 12,97 | 14,36 | 13,19 | 14,10 | 12,96 | 52,77 | 49,79 | 59,34 | 56,28 |
| JOR | 1141 | 1039 | 1141 | 1040 | 1141 | 1039 | 5,13 | 10,81 | 5,10 | 10,63 | 5,00 | 10,46 | 15,88 | 35,39 | 108,18 | 249,82 |
| SOR | 967 | 750 | 963 | 750 | 68 | 54 | 0,93 | 2,33 | 0,97 | 2,33 | 0,96 | 2,40 | 3,07 | 11,85 | 0,45 | 1,49 |
| BUILDING3 | | | | | | | | | | | | | | | | |
| Μέθοδος | Εποχές | | Σχετικές Εποχές | | Διορθώσεις Συνάψεων | | Training Set Error | | Validation Set Error | | Test Set Error | | Test Set Classification | | Απόλεια Γενίκευσης | |
| | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev |
| P-RPROP | 229 | 107 | 217 | 102 | 229 | 107 | 0,37 | 0,04 | 0,38 | 0,07 | 0,38 | 0,08 | | | 1,99 | 4,45 |
| RPROP | 351 | 23 | 341 | 22 | 351 | 23 | 0,35 | 0,00 | 0,35 | 0,00 | 0,34 | 0,00 | 0,29 | 0,00 | 0,02 | 0,01 |
| ASD | 462 | 44 | 462 | 44 | 462 | 44 | 0,35 | 0,00 | 0,35 | 0,00 | 0,35 | 0,00 | 0,29 | 0,00 | 0,01 | 0,01 |
| OFR | 304 | 59 | 298 | 59 | 304 | 59 | 0,35 | 0,00 | 0,35 | 0,00 | 0,35 | 0,00 | 0,25 | 0,05 | 0,04 | 0,07 |
| LMS | 63 | 135 | 54 | 138 | 63 | 135 | 0,83 | 0,74 | 0,83 | 0,74 | 0,84 | 0,76 | 3,05 | 6,21 | 0,70 | 0,58 |
| B-LMS | 5 | 0 | 1 | 0 | 5 | 0 | 18,88 | 0,27 | 19,07 | 0,26 | 19,41 | 0,28 | 85,22 | 0,55 | 167,34 | 31,35 |
| PR | 5 | 0 | 2 | 0 | 5 | 0 | 18,62 | 0,30 | 18,35 | 0,32 | 19,05 | 0,36 | 99,61 | 0,25 | 136,95 | 33,24 |
| FR | 5 | 0 | 1 | 0 | 5 | 0 | 18,88 | 0,24 | 19,08 | 0,25 | 19,41 | 0,24 | 85,70 | 0,66 | 145,65 | 24,01 |
| RGS | 704 | 736 | 702 | 738 | 50 | 52 | 14,58 | 13,43 | 14,56 | 13,43 | 14,60 | 13,46 | 52,72 | 49,84 | 54,38 | 51,59 |
| JOR | 1148 | 1038 | 1147 | 1039 | 1148 | 1038 | 5,13 | 10,78 | 5,17 | 10,88 | 5,17 | 10,89 | 15,83 | 35,37 | 109,26 | 252,32 |
| SOR | 968 | 749 | 967 | 750 | 68 | 54 | 0,76 | 1,58 | 0,75 | 1,54 | 0,77 | 1,63 | 2,83 | 10,87 | 0,34 | 1,30 |

Πίνακας 7.12 Μέση Τιμή και Απόκλιση για το Πρόβλημα των Ηλιακών Κηλίδων

| FLARE1 | | | | | | | | | | | | | | | | |
|---------------|--------|-------|-----------------|-------|---------------------|-------|--------------------|-------|----------------------|-------|----------------|-------|-------------------------|-------|--------------------|----------|
| Μέθοδος | Εποχές | | Σχετικές Εποχές | | Διορθώσεις Συνάψεων | | Training Set Error | | Validation Set Error | | Test Set Error | | Test Set Classification | | Απόλεια Γενίκευσης | |
| | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev |
| P-RPROP | 41 | 5 | 12 | 4 | 41 | 5 | 0,37 | 0,00 | 0,34 | 0,01 | 0,52 | 0,01 | | | 2,17 | 1,61 |
| RPROP | 24 | 20 | 8 | 6 | 24 | 20 | 0,39 | 0,02 | 0,34 | 0,01 | 0,55 | 0,03 | 3,84 | 0,45 | 6,94 | 4,82 |
| ASD | 5 | 0 | 1 | 0 | 5 | 0 | 0,51 | 0,03 | 0,43 | 0,03 | 0,70 | 0,04 | 5,26 | 0,00 | 0,00 | 0,00 |
| OFR | 5 | 0 | 1 | 0 | 5 | 0 | 0,52 | 0,03 | 0,43 | 0,02 | 0,70 | 0,03 | 5,26 | 0,00 | 0,00 | 0,00 |
| LMS | 40 | 51 | 9 | 16 | 40 | 51 | 0,47 | 0,11 | 0,40 | 0,08 | 0,66 | 0,12 | 4,55 | 0,45 | 3,99 | 3,90 |
| B-LMS | 151 | 7 | 42 | 6 | 151 | 7 | 0,37 | 0,00 | 0,33 | 0,00 | 0,53 | 0,00 | 3,76 | 0,00 | 1,05 | 0,22 |
| PR | 150 | 4 | 43 | 7 | 150 | 4 | 0,37 | 0,00 | 0,33 | 0,00 | 0,53 | 0,00 | 3,76 | 0,00 | 1,05 | 0,32 |
| FR | 152 | 6 | 41 | 6 | 152 | 6 | 0,37 | 0,00 | 0,33 | 0,00 | 0,53 | 0,00 | 3,76 | 0,00 | 1,06 | 0,24 |
| RGS | 17 | 41 | 12 | 40 | 1 | 2 | 0,49 | 0,04 | 0,41 | 0,03 | 0,68 | 0,05 | 5,10 | 0,46 | 61152,68 | 52715,43 |
| JOR | 106 | 206 | 64 | 166 | 106 | 206 | 0,51 | 0,17 | 0,48 | 0,20 | 0,65 | 0,14 | 3,42 | 0,79 | 9162,84 | 18985,49 |
| SOR | 37 | 29 | 24 | 25 | 1 | 1 | 0,40 | 0,03 | 0,33 | 0,03 | 0,53 | 0,05 | 3,94 | 0,49 | 4,42 | 2,65 |
| FLARE2 | | | | | | | | | | | | | | | | |
| Μέθοδος | Εποχές | | Σχετικές Εποχές | | Διορθώσεις Συνάψεων | | Training Set Error | | Validation Set Error | | Test Set Error | | Test Set Classification | | Απόλεια Γενίκευσης | |
| | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev |
| P-RPROP | 37 | 3 | 16 | 10 | 37 | 3 | 0,42 | 0,00 | 0,46 | 0,00 | 0,31 | 0,02 | | | 0,72 | 0,90 |
| RPROP | 19 | 20 | 7 | 6 | 19 | 20 | 0,46 | 0,03 | 0,48 | 0,02 | 0,31 | 0,01 | 2,59 | 0,24 | 6,29 | 4,67 |
| ASD | 5 | 0 | 1 | 0 | 5 | 0 | 0,61 | 0,03 | 0,59 | 0,03 | 0,33 | 0,02 | 3,01 | 0,00 | 0,00 | 0,00 |
| OFR | 5 | 0 | 1 | 0 | 5 | 0 | 0,61 | 0,03 | 0,59 | 0,03 | 0,33 | 0,02 | 3,01 | 0,00 | 0,00 | 0,00 |
| LMS | 46 | 55 | 10 | 10 | 46 | 55 | 1,47 | 1,36 | 1,69 | 1,54 | 1,49 | 1,54 | 3,52 | 2,45 | 1,99 | 2,02 |
| B-LMS | 164 | 6 | 20 | 2 | 164 | 6 | 0,42 | 0,00 | 0,45 | 0,00 | 0,27 | 0,00 | 2,41 | 0,19 | 4,35 | 0,26 |
| PR | 164 | 6 | 20 | 1 | 164 | 6 | 0,42 | 0,00 | 0,45 | 0,00 | 0,27 | 0,00 | 2,33 | 0,15 | 4,42 | 0,32 |
| FR | 165 | 5 | 21 | 2 | 165 | 5 | 0,42 | 0,00 | 0,45 | 0,00 | 0,28 | 0,00 | 2,43 | 0,19 | 4,16 | 0,31 |
| RGS | 18 | 41 | 12 | 40 | 1 | 2 | 0,59 | 0,05 | 0,56 | 0,04 | 0,31 | 0,02 | 2,95 | 0,18 | 45494 | 38658 |
| JOR | 136 | 258 | 62 | 159 | 136 | 258 | 0,58 | 0,19 | 0,64 | 0,23 | 0,44 | 0,21 | 1,86 | 0,83 | 8005,10 | 16644,0 |
| SOR | 43 | 37 | 24 | 25 | 1 | 2 | 0,46 | 0,04 | 0,46 | 0,02 | 0,28 | 0,03 | 2,65 | 0,19 | 6,42 | 3,63 |
| FLARE3 | | | | | | | | | | | | | | | | |
| Μέθοδος | Εποχές | | Σχετικές Εποχές | | Διορθώσεις Συνάψεων | | Training Set Error | | Validation Set Error | | Test Set Error | | Test Set Classification | | Απόλεια Γενίκευσης | |
| | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev |
| P-RPROP | 35 | 6 | 18 | 12 | 35 | 6 | 0,39 | 0,00 | 0,46 | 0,00 | 0,35 | 0,00 | | | 0,57 | 0,73 |
| RPROP | 34 | 19 | 11 | 7 | 34 | 19 | 0,41 | 0,03 | 0,47 | 0,02 | 0,36 | 0,02 | 3,01 | 0,12 | 4,63 | 6,01 |
| ASD | 5 | 0 | 1 | 0 | 5 | 0 | 0,58 | 0,02 | 0,57 | 0,02 | 0,43 | 0,02 | 3,76 | 0,00 | 0,00 | 0,00 |
| OFR | 5 | 0 | 1 | 0 | 5 | 0 | 0,58 | 0,03 | 0,58 | 0,03 | 0,44 | 0,03 | 3,76 | 0,00 | 0,00 | 0,00 |
| LMS | 37 | 51 | 32 | 52 | 37 | 51 | 0,61 | 0,28 | 0,65 | 0,24 | 0,54 | 0,25 | 3,96 | 1,11 | 0,05 | 0,09 |
| B-LMS | 154 | 5 | 154 | 5 | 154 | 5 | 0,39 | 0,00 | 0,46 | 0,00 | 0,35 | 0,00 | 3,01 | 0,00 | 0,00 | 0,00 |
| PR | 157 | 5 | 156 | 5 | 157 | 5 | 0,39 | 0,00 | 0,46 | 0,00 | 0,35 | 0,00 | 3,01 | 0,00 | 0,00 | 0,00 |
| FR | 155 | 4 | 155 | 4 | 155 | 4 | 0,39 | 0,00 | 0,46 | 0,00 | 0,35 | 0,00 | 3,01 | 0,00 | 0,00 | 0,00 |
| RGS | 15 | 39 | 11 | 40 | 0 | 2 | 0,56 | 0,05 | 0,56 | 0,03 | 0,42 | 0,03 | 3,64 | 0,35 | 57767,00 | 52627,47 |
| JOR | 120 | 220 | 100 | 221 | 120 | 220 | 0,57 | 0,22 | 0,66 | 0,26 | 0,55 | 0,26 | 2,97 | 0,67 | 8539,64 | 18007,73 |
| SOR | 57 | 34 | 31 | 39 | 2 | 1 | 0,41 | 0,04 | 0,46 | 0,03 | 0,37 | 0,03 | 2,97 | 0,32 | 3,72 | 3,27 |

Πίνακας 7.13 Μέση Τιμή και Απόκλιση για το Πρόβλημα Διάγνωσης Καρδιοπάθειας (hearta)

| HEARTA1 | | | | | | | | | | | | | | | | |
|----------------|--------|-------|-----------------|-------|---------------------|-------|--------------------|-------|----------------------|-------|----------------|-------|-------------------------|-------|--------------------|--------|
| Μέθοδος | Εποχές | | Σχετικές Εποχές | | Διορθώσεις Συνάψεων | | Training Set Error | | Validation Set Error | | Test Set Error | | Test Set Classification | | Απόλεια Γενίκευσης | |
| | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev |
| P-RPROP | 118 | 12 | 27 | 10 | 118 | 12 | 3,82 | 0,00 | 4,42 | 0,03 | 4,47 | 0,06 | | | 1,68 | 0,68 |
| RPROP | 63 | 57 | 17 | 13 | 63 | 57 | 4,85 | 1,33 | 5,07 | 1,06 | 5,25 | 1,18 | 14,07 | 4,72 | 8,47 | 6,09 |
| ASD | 225 | 4 | 8 | 1 | 225 | 4 | 3,84 | 0,00 | 4,36 | 0,01 | 4,65 | 0,02 | 10,92 | 0,44 | 2,59 | 0,26 |
| OFR | 87 | 23 | 12 | 10 | 87 | 23 | 3,84 | 0,01 | 4,35 | 0,01 | 4,62 | 0,08 | 10,85 | 0,52 | 3,17 | 0,65 |
| LMS | 94 | 110 | 23 | 34 | 94 | 110 | 6,96 | 4,40 | 6,68 | 2,73 | 7,72 | 4,30 | 15,84 | 8,09 | 3,70 | 3,37 |
| B-LMS | 5 | 0 | 4 | 0 | 5 | 0 | 5,42 | 0,04 | 5,37 | 0,05 | 6,58 | 0,07 | 21,62 | 0,73 | 14,21 | 0,35 |
| PR | 5 | 0 | 4 | 0 | 5 | 0 | 5,08 | 0,02 | 5,02 | 0,03 | 6,05 | 0,05 | 19,15 | 0,54 | 11,73 | 0,26 |
| FR | 5 | 0 | 4 | 0 | 5 | 0 | 5,43 | 0,04 | 5,39 | 0,05 | 6,58 | 0,07 | 21,58 | 0,81 | 13,85 | 0,60 |
| RGS | 207 | 213 | 200 | 210 | 6 | 6 | 7,64 | 3,11 | 7,98 | 3,34 | 8,06 | 3,085 | 21,60 | 8,77 | 9,15 | 8,61 |
| JOR | 495 | 462 | 450 | 492 | 495 | 462 | 11,49 | 12,59 | 13,10 | 14,59 | 11,98 | 12,71 | 26,98 | 28,22 | 50,94 | 84,68 |
| SOR | 143 | 95 | 129 | 93 | 4 | 3 | 6,07 | 3,25 | 6,37 | 3,33 | 7,41 | 4,62 | 17,35 | 8,48 | 3,01 | 3,17 |
| HEARTA2 | | | | | | | | | | | | | | | | |
| Μέθοδος | Εποχές | | Σχετικές Εποχές | | Διορθώσεις Συνάψεων | | Training Set Error | | Validation Set Error | | Test Set Error | | Test Set Classification | | Απόλεια Γενίκευσης | |
| | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev |
| P-RPROP | 112 | 10 | 107 | 15 | 112 | 10 | 4,17 | 0,00 | 4,28 | 0,02 | 4,19 | 0,01 | | | 0,06 | 0,13 |
| RPROP | 98 | 12 | 90 | 11 | 98 | 12 | 4,17 | 0,01 | 4,27 | 0,03 | 4,19 | 0,01 | 10,69 | 0,57 | 0,10 | 0,12 |
| ASD | 176 | 3 | 176 | 3 | 176 | 3 | 4,18 | 0,00 | 4,25 | 0,00 | 4,19 | 0,00 | 10,87 | 0,00 | 0,01 | 0,01 |
| OFR | 75 | 19 | 73 | 19 | 75 | 19 | 4,17 | 0,01 | 4,25 | 0,01 | 4,18 | 0,01 | 10,59 | 0,21 | 0,01 | 0,02 |
| LMS | 77 | 80 | 68 | 87 | 77 | 80 | 9,60 | 7,12 | 10,12 | 7,72 | 9,51 | 7,24 | 17,92 | 9,00 | 1,60 | 2,15 |
| B-LMS | 227 | 3 | 227 | 3 | 227 | 3 | 4,20 | 0,00 | 4,27 | 0,00 | 4,21 | 0,00 | 10,87 | 0,00 | 0,00 | 0,00 |
| PR | 224 | 3 | 224 | 3 | 224 | 3 | 4,20 | 0,00 | 4,27 | 0,00 | 4,21 | 0,00 | 10,87 | 0,00 | 0,00 | 0,00 |
| FR | 224 | 4 | 224 | 4 | 224 | 4 | 4,20 | 0,00 | 4,27 | 0,00 | 4,21 | 0,00 | 10,87 | 0,00 | 0,00 | 0,00 |
| RGS | 231 | 239 | 227 | 238 | 6 | 7 | 7,81 | 3,14 | 7,63 | 2,94 | 7,61 | 2,87 | 18,56 | 9,45 | 18,28 | 17,37 |
| JOR | 350 | 409 | 328 | 416 | 350 | 409 | 11,91 | 12,81 | 11,70 | 12,36 | 11,49 | 12,09 | 26,41 | 24,97 | 68,31 | 114,22 |
| SOR | 102 | 67 | 86 | 69 | 3 | 2 | 7,65 | 4,62 | 7,17 | 4,10 | 7,47 | 4,22 | 17,07 | 8,78 | 2,44 | 3,10 |
| HEARTA3 | | | | | | | | | | | | | | | | |
| Μέθοδος | Εποχές | | Σχετικές Εποχές | | Διορθώσεις Συνάψεων | | Training Set Error | | Validation Set Error | | Test Set Error | | Test Set Classification | | Απόλεια Γενίκευσης | |
| | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev |
| P-RPROP | 116 | 8 | 110 | 10 | 116 | 8 | 4,06 | 0,00 | 4,14 | 0,02 | 4,54 | 0,01 | | | 0,05 | 0,05 |
| RPROP | 95 | 31 | 81 | 34 | 95 | 31 | 4,09 | 0,08 | 4,15 | 0,06 | 4,59 | 0,12 | 12,22 | 1,10 | 0,92 | 2,21 |
| ASD | 143 | 3 | 143 | 3 | 143 | 3 | 4,08 | 0,00 | 4,12 | 0,00 | 4,58 | 0,00 | 12,11 | 0,16 | 0,02 | 0,02 |
| OFR | 58 | 7 | 53 | 9 | 58 | 7 | 4,06 | 0,01 | 4,10 | 0,02 | 4,56 | 0,02 | 11,69 | 0,31 | 0,07 | 0,11 |
| LMS | 72 | 64 | 68 | 67 | 72 | 64 | 11,23 | 7,45 | 10,02 | 6,12 | 11,87 | 7,64 | 32,63 | 20,15 | 0,01 | 0,01 |
| B-LMS | 191 | 3 | 191 | 3 | 191 | 3 | 4,09 | 0,00 | 4,12 | 0,00 | 4,59 | 0,00 | 11,74 | 0,00 | 0,00 | 0,00 |
| PR | 191 | 3 | 191 | 3 | 191 | 3 | 4,09 | 0,00 | 4,12 | 0,00 | 4,59 | 0,00 | 11,69 | 0,13 | 0,00 | 0,00 |
| FR | 192 | 3 | 192 | 3 | 192 | 3 | 4,09 | 0,00 | 4,12 | 0,00 | 4,59 | 0,00 | 11,74 | 0,00 | 0,00 | 0,00 |
| RGS | 253 | 232 | 234 | 237 | 7 | 7 | 7,44 | 2,86 | 7,41 | 3,00 | 8,65 | 3,74 | 20,11 | 6,26 | 10,96 | 16,99 |
| JOR | 375 | 436 | 311 | 459 | 375 | 436 | 11,70 | 12,63 | 11,72 | 12,70 | 11,78 | 12,09 | 27,78 | 25,82 | 63,92 | 106,17 |
| SOR | 102 | 80 | 76 | 77 | 2 | 3 | 7,64 | 4,66 | 7,39 | 4,36 | 8,35 | 4,91 | 18,81 | 8,72 | 2,79 | 2,77 |

Πίνακας 7.14 Μέση Τιμή και Απόκλιση για το Πρόβλημα Διάγνωσης Καρδιοπάθειας (heartac)

| HEARTAC1 | | | | | | | | | | | | | | | | |
|----------|--------|-------|-----------------|-------|---------------------|-------|--------------------|-------|----------------------|-------|----------------|-------|-------------------------|-------|--------------------|--------|
| Μέθοδος | Εποχές | | Σχετικές Εποχές | | Διορθώσεις Συνάψεων | | Training Set Error | | Validation Set Error | | Test Set Error | | Test Set Classification | | Απόλεια Γενίκευσης | |
| | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev |
| P-RPROP | 98 | 10 | 96 | 11 | 98 | 10 | 4,05 | 0,00 | 4,70 | 0,02 | 2,69 | 0,02 | | | 0,01 | 0,02 |
| RPROP | 75 | 41 | 69 | 41 | 75 | 41 | 4,10 | 0,09 | 4,71 | 0,04 | 2,73 | 0,14 | 5,05 | 0,82 | 2,74 | 4,35 |
| ASD | 129 | 3 | 128 | 2 | 129 | 3 | 4,05 | 0,00 | 4,67 | 0,00 | 2,65 | 0,01 | 5,33 | 0,00 | 0,03 | 0,02 |
| OFR | 74 | 24 | 71 | 25 | 74 | 24 | 4,05 | 0,01 | 4,66 | 0,02 | 2,68 | 0,03 | 4,91 | 0,62 | 0,04 | 0,05 |
| LMS | 84 | 113 | 70 | 119 | 84 | 113 | 34,83 | 55,62 | 33,91 | 51,45 | 30,36 | 50,87 | 18,46 | 20,57 | 35,44 | 64,19 |
| B-LMS | 332 | 6 | 332 | 6 | 332 | 6 | 4,09 | 0,00 | 4,75 | 0,01 | 2,61 | 0,00 | 5,19 | 0,41 | 0,00 | 0,00 |
| PR | 160 | 164 | 159 | 165 | 160 | 164 | 37,80 | 32,02 | 39,04 | 32,58 | 43,64 | 38,98 | 31,86 | 25,17 | 4252,3 | 4162,5 |
| FR | 332 | 5 | 332 | 5 | 332 | 5 | 4,09 | 0,00 | 4,75 | 0,01 | 2,61 | 0,01 | 5,26 | 0,30 | 0,00 | 0,00 |
| RGS | 136 | 105 | 134 | 108 | 4 | 3 | 5,46 | 0,53 | 5,98 | 0,55 | 4,76 | 0,75 | 10,04 | 2,42 | 0,13 | 0,18 |
| JOR | 382 | 460 | 353 | 469 | 382 | 460 | 8,76 | 7,75 | 8,91 | 6,80 | 7,66 | 8,36 | 20,35 | 26,89 | 7,47 | 12,21 |
| SOR | 29 | 35 | 23 | 35 | 1 | 1 | 8,13 | 4,09 | 9,54 | 4,06 | 8,60 | 5,51 | 19,58 | 12,58 | 6,63 | 7,37 |

| HEARTAC2 | | | | | | | | | | | | | | | | |
|----------|--------|-------|-----------------|-------|---------------------|-------|--------------------|-------|----------------------|-------|----------------|-------|-------------------------|-------|--------------------|--------|
| Μέθοδος | Εποχές | | Σχετικές Εποχές | | Διορθώσεις Συνάψεων | | Training Set Error | | Validation Set Error | | Test Set Error | | Test Set Classification | | Απόλεια Γενίκευσης | |
| | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev |
| P-RPROP | 19 | 4 | 13 | 4 | 19 | 4 | 3,37 | 0,11 | 5,21 | 0,21 | 3,87 | 0,16 | | | 6,99 | 2,27 |
| RPROP | 13 | 6 | 9 | 4 | 13 | 6 | 4,23 | 1,18 | 5,44 | 0,93 | 4,81 | 1,10 | 11,23 | 3,25 | 13,97 | 10,63 |
| ASD | 5 | 0 | 2 | 0 | 5 | 0 | 3,89 | 0,02 | 4,67 | 0,04 | 4,07 | 0,04 | 9,47 | 0,41 | 15,25 | 0,59 |
| OFR | 5 | 0 | 2 | 0 | 5 | 0 | 3,88 | 0,02 | 4,67 | 0,02 | 4,07 | 0,03 | 9,47 | 0,41 | 15,10 | 0,34 |
| LMS | 34 | 35 | 5 | 6 | 34 | 35 | 15,56 | 20,12 | 17,99 | 20,31 | 17,84 | 21,19 | 21,05 | 15,70 | 73,30 | 133,39 |
| B-LMS | 118 | 9 | 19 | 1 | 118 | 9 | 3,26 | 0,02 | 5,21 | 0,03 | 3,71 | 0,02 | 7,93 | 0,30 | 5,13 | 0,07 |
| PR | 64 | 56 | 11 | 8 | 64 | 56 | 49,33 | 48,59 | 59,99 | 57,80 | 54,84 | 53,95 | 26,53 | 19,62 | 3435,4 | 3654,2 |
| FR | 115 | 8 | 19 | 0 | 115 | 8 | 3,27 | 0,017 | 5,20 | 0,02 | 3,71 | 0,02 | 7,93 | 0,30 | 5,12 | 0,07 |
| RGS | 120 | 95 | 112 | 91 | 3 | 3 | 5,06 | 0,12 | 6,12 | 0,53 | 5,43 | 0,67 | 11,16 | 3,03 | 4,08 | 3,19 |
| JOR | 452 | 718 | 231 | 379 | 452 | 718 | 8,90 | 9,36 | 8,82 | 6,47 | 9,48 | 9,47 | 19,79 | 18,48 | 31,05 | 43,43 |
| SOR | 80 | 2 | 45 | 80 | 2 | 4 | 7,83 | 4,98 | 8,20 | 3,46 | 9,79 | 5,23 | 20,49 | 10,80 | 9,52 | 3,40 |

| HEARTAC3 | | | | | | | | | | | | | | | | |
|----------|--------|-------|-----------------|-------|---------------------|-------|--------------------|-------|----------------------|-------|----------------|-------|-------------------------|-------|--------------------|--------|
| Μέθοδος | Εποχές | | Σχετικές Εποχές | | Διορθώσεις Συνάψεων | | Training Set Error | | Validation Set Error | | Test Set Error | | Test Set Classification | | Απόλεια Γενίκευσης | |
| | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev | mean | stdev |
| P-RPROP | 29 | 9 | 14 | 3 | 29 | 9 | 2,85 | 0,09 | 5,66 | 0,16 | 5,43 | 0,23 | | | 6,06 | 0,99 |
| RPROP | 13 | 10 | 8 | 5 | 13 | 10 | 4,04 | 1,28 | 5,89 | 0,79 | 5,99 | 0,82 | 16,42 | 3,77 | 12,44 | 6,71 |
| ASD | 5 | 0 | 2 | 0 | 5 | 0 | 3,24 | 0,01 | 5,10 | 0,04 | 5,45 | 0,02 | 14,95 | 0,69 | 6,62 | 0,21 |
| OFR | 5 | 0 | 2 | 0 | 5 | 0 | 3,24 | 0,01 | 5,08 | 0,04 | 5,45 | 0,03 | 15,09 | 0,97 | 6,81 | 0,25 |
| LMS | 27 | 17 | 5 | 4 | 27 | 17 | 21,49 | 33,28 | 19,35 | 26,55 | 21,43 | 29,70 | 23,72 | 10,94 | 15,68 | 21,63 |
| B-LMS | 37 | 3 | 12 | 0 | 37 | 3 | 3,11 | 0,03 | 5,18 | 0,05 | 5,40 | 0,04 | 13,12 | 1,08 | 5,40 | 0,24 |
| PR | 20 | 16 | 7 | 5 | 20 | 16 | 59,76 | 53,87 | 55,18 | 47,55 | 54,40 | 46,57 | 32,49 | 17,91 | 3368,2 | 3439,7 |
| FR | 37 | 2 | 12 | 0 | 37 | 2 | 3,11 | 0,02 | 5,19 | 0,04 | 5,39 | 0,02 | 12,91 | 0,97 | 5,27 | 0,19 |
| RGS | 147 | 99 | 128 | 107 | 4 | 3 | 4,08 | 0,36 | 5,96 | 0,52 | 5,95 | 0,36 | 14,04 | 2,26 | 2,90 | 2,63 |
| JOR | 360 | 556 | 170 | 278 | 360 | 556 | 8,37 | 9,22 | 9,84 | 7,47 | 9,90 | 7,91 | 21,47 | 12,70 | 21,37 | 27,75 |
| SOR | 58 | 67 | 30 | 34 | 1 | 2 | 7,60 | 5,24 | 8,29 | 3,43 | 10,20 | 5,71 | 18,53 | 8,21 | 8,40 | 5,40 |

7.3.4.2 Παρατηρήσεις στα Αποτελέσματα των Πινάκων (7.11)-(7.14)

Όπως φαίνεται στους πίνακες (7.11)-(7.14), μπορούν να γίνουν οι παρακάτω παρατηρήσεις στα αποτελέσματα:

1. Η μέθοδος SOR, σε όλα τα προβλήματα συγκλίνει για όλες τις τιμές του συντελεστή εκπαίδευσης ω στο διάστημα $[0.01, 1.9]$, εκτός του προβλήματος BUILDING, για το οποίο $\omega_{max} = 1.5$. Η μέγιστη τιμή του συντελεστή εκπαίδευσης ω , ω_{max} , για τον οποίο συγκλίνει η JOR παίρνει τιμές στο διάστημα $[0.2, 0.6]$. Το ω_{max} για την RGS είναι 0.01 για τα περισσότερα προβλήματα, εκτός του προβλήματος FLARE, όπου $\omega_{max} = 1$. Το ω_{max} για την LMS παίρνει τιμές στο διάστημα $[0.2, 0.5]$, ενώ το ω_{max} για τις μεθόδους BATCH-LMS, PR και FR είναι 0.001 για όλα τα προβλήματα προσαρμογής.
2. Σε όλα τα προβλήματα, τα αποτελέσματα της μεθόδου RPROP είναι σχεδόν τα ίδια με τα αποτελέσματα της μεθόδου P-RPROP που δημοσίευσε ο Prechelt [Prec94].
3. Στο πρόβλημα HEARTAC έχουν μεγάλη απώλεια γενίκευσης (πάνω από 5) όλες οι μέθοδοι, εκτός της RGS. Επί πλέον, οι μέθοδοι JOR και RGS έχουν μεγάλη απώλεια γενίκευσης σε όλα τα υπόλοιπα προβλήματα, η μέθοδος RPROP στα προβλήματα FLARE, και HEARTA και οι μέθοδοι B-LMS, PR και FR στα προβλήματα BUILDING και HEARTA.
4. Δεν παρατηρήθηκαν μεγάλα σφάλματα στο ποσοστιαίο σφάλμα των προτύπων εκπαίδευσης (*Training Set Error*), το ποσοστιαίο σφάλμα των προτύπων ελέγχου εγκυρότητας, (*Validation Set Error*) και το ποσοστιαίο σφάλμα των προτύπων ανάκλησης, (*Test Set Error*) για καμιά μέθοδο στο πρόβλημα FLARE. Μεγάλα σφάλματα παρατηρήθηκαν για τη μέθοδο JOR στα υπόλοιπα τρία προβλήματα, για τη μέθοδο LMS στα προβλήματα HEARTA και HEARTAC, για τη μέθοδο RGS στο πρόβλημα BUILDING, για τη μέθοδο SOR στο πρόβλημα HEARTAC, για τις μεθόδους B-LMS και FR στο πρόβλημα BUILDING και για τη μέθοδο PR στα προβλήματα BUILDING και HEARTAC.
5. Δεν παρατηρήθηκαν μεγάλα σφάλματα στο σφάλμα ταξινόμησης των προτύπων ανάκλησης, (*Test Set Classification Error*) για καμιά μέθοδο στο πρόβλημα FLARE. Μεγάλα σφάλματα παρατηρήθηκαν για τη μέθοδο JOR στα υπόλοιπα τρία προβλήματα, για τη μέθοδο LMS στα προβλήματα HEARTA και HEARTAC, για τη μέθοδο RGS στο πρόβλημα BUILDING, για τη μέθοδο SOR στο πρόβλημα HEARTAC, για τις μεθόδους B-LMS και FR στο πρόβλημα BUILDING και για τη μέθοδο PR PR στα προβλήματα BUILDING και HEARTAC.
6. Στα πιο πολλά προβλήματα, οι μέθοδοι OFR και ASD απαιτούν το μικρότερο αριθμό εποχών και σχετικών εποχών που απαιτούνται για τη σύγκλισή τους και το σφάλμα ταξινόμησης των προτύπων ανάκλησης είναι το μικρότερο, ή ισοδύναμο με το σφάλμα ταξινόμησης των προτύπων ανάκλησης της μεθόδου RPROP. Η μέθοδος LMS απαιτεί το μικρότερο αριθμό εποχών και σχετικών εποχών για το πρόβλημα BUILDING και η μέθοδος SOR για το πρόβλημα HEARTAC, αλλά έχουν μεγάλα

σφάλματα ταξινόμησης των προτύπων ανάκλησης, και στα ποσοστιαία σφάλματα των προτύπων εκπαίδευσης, ελέγχου εγκυρότητας και ανάκλησης.

7. Από όλες τις μεθόδους, η μέθοδος JOR εμφανίζει τις μεγαλύτερες τιμές στις εποχές που απαιτούνται, τις σχετικές εποχές και τις επισκέψεις-διορθώσεις των συνάψεων, σε όλα τα προβλήματα, εκτός του προβλήματος FLARE, όπου πιο αργή μέθοδος είναι η B-LMS.
8. Οι μέθοδοι SOR και RGS εμφανίζουν τις μικρότερες τιμές στις επισκέψεις των συνάψεων, επειδή σε κάθε κύκλο εκπαίδευσης διορθώνονται μόνο οι συνάψεις ενός νευρώνα εισόδου, ενώ στις υπόλοιπες μεθόδους διορθώνονται όλες οι συνάψεις.
9. Οι μέθοδοι B-LMS, PR και FR εμφανίζουν παρόμοια συμπεριφορά, όσον αφορά τον αριθμό επαναλήψεων, τα σφάλματα των προτύπων και το σφάλμα ταξινόμησης, καθώς και τις τιμές του συντελεστή εκπαίδευσης, για τις οποίες συγκλίνουν.
10. Σε πολλά προβλήματα, υπάρχουν διαφορές στη συμπεριφορά των αλγορίθμων στους τρεις διαφορετικούς συνδυασμούς δεδομένων για το κάθε πρόβλημα. Έτσι, ο αριθμός επαναλήψεων που απαιτείται για τη σύγκλιση της κάθε μεθόδου στα προβλήματα HEARTA και HEARTAC έχει μεγάλη απόκλιση για όλες τις μεθόδους. Απόκλιση εμφανίζει και το πρόβλημα BUILDING στο ποσοστιαίο σφάλμα των προτύπων ελέγχου εγκυρότητας, και το πρόβλημα HEARTAC στο ποσοστιαίο σφάλμα των προτύπων ανάκλησης και στο σφάλμα ταξινόμησης των προτύπων ανάκλησης (Test Set Classification Error).
11. Σε κάποια προβλήματα, τα σφάλματα έχουν μια ανώμαλη συμπεριφορά. Π.χ., στο πρόβλημα FLARE1 το ποσοστιαίο σφάλμα των προτύπων ελέγχου εγκυρότητας είναι μικρότερο απ' το ποσοστιαίο σφάλμα των προτύπων εκπαίδευσης, ενώ στα προβλήματα FLARE2, FLARE3 και HEARTAC1 το ποσοστιαίο σφάλμα των προτύπων ανάκλησης είναι μικρότερο απ' το ποσοστιαίο σφάλμα των προτύπων εκπαίδευσης.

Στους πίνακες (7.15)-(7.24) που ακολουθούν παρουσιάζονται τα αποτελέσματα του καλύτερου πειράματος της κάθε μεθόδου του κάθε προβλήματος για τα προβλήματα ταξινόμησης, στο σημείο του αλγορίθμου, όπου το σφάλμα των προτύπων ελέγχου εγκυρότητας πήρε την ελάχιστη τιμή. Οι πίνακες περιέχουν την τιμή του συντελεστή εκπαίδευσης του καλύτερου πειράματος, τις εποχές που απαιτούνται, τις σχετικές εποχές, τις επισκέψεις-διορθώσεις των συνάψεων, το ποσοστιαίο σφάλμα των προτύπων εκπαίδευσης (Training Set Error), το ποσοστιαίο σφάλμα των προτύπων ελέγχου εγκυρότητας, (Validation Set Error), το ποσοστιαίο σφάλμα των προτύπων ανάκλησης, (Test Set Error), το σφάλμα ταξινόμησης των προτύπων ανάκλησης και την απώλεια γενίκευσης.

Πίνακας 7.15 Καλύτερο Πείραμα για το Πρόβλημα Διάγνωσης Καρκίνου CANCER

| CANCER1 | | | | | | | | | |
|----------------|---------------------------|--------|-----------------|---------------------|--------------------|----------------------|----------------|-------------------------|--------------------|
| Μέθοδος | Συντελεστής Εκπ/σης-Ωμέγα | Εποχές | Σχετικές Εποχές | Διορθώσεις Συνάψεων | Training Set Error | Validation Set Error | Test Set Error | Test Set Classification | Απόλεια Γενίκευσης |
| RPROP | | 105 | 79 | 105 | 20,75 | 20,03 | 18,45 | 14,37 | 0,38 |
| ASD | | 40 | 23 | 40 | 20,74 | 20,19 | 18,63 | 16,09 | 0,17 |
| OFR | | 20 | 9 | 20 | 20,75 | 20,01 | 18,54 | 16,67 | 1,30 |
| LMS | 0,001 | 275 | 275 | 275 | 20,81 | 20,22 | 18,64 | 16,09 | 0,00 |
| B-LMS | 0,001 | 275 | 275 | 275 | 20,81 | 20,21 | 18,64 | 16,09 | 0,00 |
| PR | 0,001 | 275 | 275 | 275 | 20,81 | 20,21 | 18,63 | 16,09 | 0,00 |
| FR | 0,001 | 275 | 275 | 275 | 20,81 | 20,21 | 18,64 | 16,09 | 0,00 |
| RGS | 0,01 | 190 | 167 | 21 | 20,87 | 19,85 | 18,67 | 14,94 | 0,15 |
| JOR | 0,2 | 110 | 110 | 110 | 20,76 | 20,27 | 18,68 | 15,52 | 0,00 |
| SOR | 1,4 | 160 | 86 | 17 | 21,59 | 19,76 | 19,66 | 16,09 | 5,01 |
| CANCER2 | | | | | | | | | |
| Μέθοδος | Συντελεστής Εκπ/σης-Ωμέγα | Εποχές | Σχετικές Εποχές | Διορθώσεις Συνάψεων | Training Set Error | Validation Set Error | Test Set Error | Test Set Classification | Απόλεια Γενίκευσης |
| RPROP | | 95 | 43 | 95 | 18,66 | 20,38 | 21,95 | 17,82 | 0,72 |
| ASD | | 50 | 31 | 50 | 18,65 | 20,51 | 22,02 | 18,39 | 0,13 |
| OFR | | 30 | 10 | 30 | 18,66 | 20,44 | 22,19 | 18,97 | 0,32 |
| LMS | 0,001 | 305 | 305 | 305 | 18,73 | 20,52 | 21,90 | 18,39 | 0,00 |
| B-LMS | 0,001 | 305 | 305 | 305 | 18,73 | 20,51 | 21,90 | 18,39 | 0,00 |
| PR | 0,001 | 305 | 305 | 305 | 18,73 | 20,51 | 21,90 | 18,39 | 0,00 |
| FR | 0,001 | 305 | 305 | 305 | 18,73 | 20,51 | 21,90 | 18,39 | 0,00 |
| RGS | 0,01 | 190 | 185 | 21 | 18,76 | 20,48 | 21,82 | 17,82 | 0,07 |
| JOR | 0,1 | 205 | 196 | 205 | 18,69 | 20,49 | 21,97 | 18,97 | 0,00 |
| SOR | 0,5 | 240 | 230 | 26 | 18,70 | 20,43 | 21,84 | 18,39 | 0,03 |
| CANCER3 | | | | | | | | | |
| Μέθοδος | Συντελεστής Εκπ/σης-Ωμέγα | Εποχές | Σχετικές Εποχές | Διορθώσεις Συνάψεων | Training Set Error | Validation Set Error | Test Set Error | Test Set Classification | Απόλεια Γενίκευσης |
| RPROP | | 100 | 49 | 100 | 20,44 | 18,48 | 20,39 | 16,09 | 1,23 |
| ASD | | 50 | 21 | 50 | 20,42 | 18,70 | 20,34 | 17,82 | 0,50 |
| OFR | | 35 | 11 | 35 | 20,42 | 18,71 | 20,29 | 17,24 | 0,54 |
| LMS | 0,001 | 315 | 279 | 315 | 20,52 | 18,70 | 20,32 | 17,24 | 0,03 |
| B-LMS | 0,001 | 310 | 276 | 310 | 20,52 | 18,70 | 20,33 | 17,82 | 0,02 |
| PR | 0,001 | 310 | 276 | 310 | 20,52 | 18,70 | 20,34 | 17,82 | 0,02 |
| FR | 0,001 | 310 | 277 | 310 | 20,52 | 18,70 | 20,34 | 17,82 | 0,02 |
| RGS | 0,01 | 235 | 196 | 26 | 20,50 | 18,35 | 20,47 | 19,54 | 0,20 |
| JOR | 0,2 | 125 | 125 | 125 | 20,46 | 18,86 | 20,36 | 17,24 | 0,00 |
| SOR | 1,1 | 225 | 119 | 25 | 20,58 | 18,39 | 21,11 | 17,82 | 3,31 |

Πίνακας 7.16 Καλύτερο Πείραμα για το Πρόβλημα της Πιστωτικής Κάρτας CARD

| CARD1 | | | | | | | | | |
|---------|---------------------------|--------|-----------------|---------------------|--------------------|----------------------|----------------|-------------------------|--------------------|
| Μέθοδος | Συντελεστής Εκπ/σης-Ωμέγα | Εποχές | Σχετικές Εποχές | Διορθώσεις Συνάψεων | Training Set Error | Validation Set Error | Test Set Error | Test Set Classification | Απόλεια Γενίκευσης |
| RPROP | | 45 | 28 | 45 | 9,86 | 8,64 | 10,67 | 13,95 | 6,24 |
| ASD | | 80 | 13 | 80 | 10,06 | 8,19 | 10,46 | 13,95 | 5,03 |
| OFR | | 55 | 13 | 55 | 10,05 | 8,19 | 10,44 | 13,95 | 5,21 |
| LMS | 0,01 | 20 | 4 | 20 | 10,22 | 8,08 | 10,38 | 13,95 | 5,08 |
| B-LMS | 0,001 | 300 | 44 | 300 | 10,05 | 8,21 | 10,48 | 13,95 | 5,03 |
| PR | 0,001 | 300 | 45 | 300 | 10,05 | 8,20 | 10,46 | 13,95 | 5,03 |
| FR | 0,001 | 295 | 45 | 295 | 10,05 | 8,20 | 10,48 | 13,95 | 5,02 |
| RGS | 0,001 | 190 | 189 | 3 | 18,67 | 17,37 | 18,56 | 15,12 | 0,00 |
| JOR | 0,1 | 135 | 76 | 135 | 9,83 | 9,12 | 10,62 | 13,95 | 1,23 |
| SOR | 1,4 | 280 | 275 | 5 | 10,57 | 10,27 | 11,99 | 13,95 | 0,35 |
| CARD2 | | | | | | | | | |
| Μέθοδος | Συντελεστής Εκπ/σης-Ωμέγα | Εποχές | Σχετικές Εποχές | Διορθώσεις Συνάψεων | Training Set Error | Validation Set Error | Test Set Error | Test Set Classification | Απόλεια Γενίκευσης |
| RPROP | | 20 | 15 | 20 | 8,48 | 10,07 | 14,85 | 19,77 | 6,56 |
| ASD | | 225 | 21 | 225 | 8,32 | 9,70 | 13,65 | 19,77 | 5,01 |
| OFR | | 65 | 19 | 65 | 8,35 | 9,69 | 13,70 | 19,77 | 5,10 |
| LMS | 0,001 | 300 | 61 | 300 | 8,42 | 9,71 | 13,59 | 19,77 | 2,48 |
| B-LMS | 0,001 | 305 | 59 | 305 | 8,42 | 9,68 | 13,57 | 19,77 | 2,64 |
| PR | 0,001 | 305 | 60 | 305 | 8,42 | 9,69 | 13,59 | 19,77 | 2,61 |
| FR | 0,001 | 300 | 59 | 300 | 8,42 | 9,69 | 13,57 | 19,77 | 2,58 |
| RGS | 0,001 | 380 | 357 | 7 | 12,67 | 13,70 | 15,33 | 17,44 | 0,26 |
| JOR | 0,1 | 130 | 97 | 130 | 8,24 | 11,17 | 15,13 | 19,77 | 0,11 |
| SOR | 0,7 | 220 | 198 | 4 | 9,31 | 11,90 | 14,30 | 19,19 | 1,23 |
| CARD3 | | | | | | | | | |
| Μέθοδος | Συντελεστής Εκπ/σης-Ωμέγα | Εποχές | Σχετικές Εποχές | Διορθώσεις Συνάψεων | Training Set Error | Validation Set Error | Test Set Error | Test Set Classification | Απόλεια Γενίκευσης |
| RPROP | | 115 | 23 | 115 | 9,48 | 8,16 | 12,23 | 13,95 | 4,45 |
| ASD | | 155 | 27 | 155 | 9,64 | 7,67 | 12,35 | 15,70 | 5,00 |
| OFR | | 75 | 24 | 75 | 9,68 | 7,68 | 12,27 | 16,28 | 5,06 |
| LMS | 0,001 | 415 | 74 | 415 | 9,67 | 7,70 | 12,37 | 14,53 | 4,68 |
| B-LMS | 0,001 | 415 | 73 | 415 | 9,66 | 7,66 | 12,32 | 15,12 | 4,68 |
| PR | 0,001 | 415 | 75 | 415 | 9,66 | 7,67 | 12,33 | 15,12 | 4,59 |
| FR | 0,001 | 410 | 73 | 410 | 9,67 | 7,67 | 12,32 | 15,12 | 4,61 |
| RGS | 0,001 | 190 | 185 | 3 | 18,01 | 17,19 | 19,54 | 22,09 | 0,03 |
| JOR | 0,1 | 145 | 145 | 145 | 9,53 | 8,47 | 12,82 | 14,53 | 0,00 |
| SOR | 0,6 | 215 | 212 | 4 | 10,71 | 11,54 | 14,84 | 19,19 | 0,00 |

Πίνακας 7.17 Καλύτερο Πείραμα για το Πρόβλημα της Διάγνωσης Διαβήτη DIABETES

| DIABETES1 | | | | | | | | | |
|------------------|---------------------------|--------|-----------------|---------------------|--------------------|----------------------|----------------|-------------------------|--------------------|
| Μέθοδος | Συντελεστής Εκπ/σης-Ωμέγα | Εποχές | Σχετικές Εποχές | Διορθώσεις Συνάψεων | Training Set Error | Validation Set Error | Test Set Error | Test Set Classification | Απόλεια Γενίκευσης |
| RPROP | | 110 | 87 | 110 | 20,33 | 22,51 | 23,96 | 38,54 | 0,27 |
| ASD | | 30 | 10 | 30 | 20,32 | 22,54 | 24,01 | 38,54 | 0,55 |
| OFR | | 20 | 8 | 20 | 20,31 | 22,50 | 23,87 | 37,50 | 0,91 |
| LMS | 0,001 | 235 | 196 | 235 | 20,38 | 22,55 | 24,06 | 39,58 | 0,04 |
| B-LMS | 0,001 | 235 | 193 | 235 | 20,38 | 22,54 | 24,07 | 39,58 | 0,05 |
| PR | 0,001 | 235 | 193 | 235 | 20,38 | 22,54 | 24,06 | 39,58 | 0,05 |
| FR | 0,001 | 235 | 193 | 235 | 20,38 | 22,55 | 24,07 | 39,58 | 0,05 |
| RGS | 0,01 | 165 | 98 | 20 | 20,66 | 22,51 | 23,52 | 39,58 | 0,59 |
| JOR | 0,3 | 130 | 130 | 130 | 20,34 | 22,67 | 24,18 | 38,54 | 0,00 |
| SOR | 1,4 | 250 | 228 | 31 | 20,46 | 22,93 | 24,27 | 37,50 | 0,45 |
| DIABETES2 | | | | | | | | | |
| Μέθοδος | Συντελεστής Εκπ/σης-Ωμέγα | Εποχές | Σχετικές Εποχές | Διορθώσεις Συνάψεων | Training Set Error | Validation Set Error | Test Set Error | Test Set Classification | Απόλεια Γενίκευσης |
| RPROP | | 100 | 97 | 100 | 21,03 | 20,75 | 24,34 | 38,02 | 0,03 |
| ASD | | 30 | 15 | 30 | 21,03 | 20,71 | 23,99 | 35,94 | 0,17 |
| OFR | | 20 | 10 | 20 | 21,02 | 20,71 | 24,20 | 35,94 | 0,24 |
| LMS | 0,01 | 45 | 35 | 45 | 21,04 | 20,78 | 24,07 | 37,50 | 0,04 |
| B-LMS | 0,001 | 250 | 250 | 250 | 21,10 | 20,77 | 23,87 | 35,42 | 0,00 |
| PR | 0,001 | 255 | 255 | 255 | 21,10 | 20,77 | 23,88 | 35,42 | 0,00 |
| FR | 0,001 | 250 | 250 | 250 | 21,10 | 20,77 | 23,87 | 35,42 | 0,00 |
| RGS | 0,01 | 205 | 202 | 25 | 21,23 | 21,00 | 24,25 | 37,50 | 0,06 |
| JOR | 0,3 | 125 | 125 | 125 | 21,05 | 20,94 | 24,29 | 38,02 | 0,00 |
| SOR | 1,7 | 295 | 280 | 36 | 21,74 | 20,32 | 25,61 | 34,38 | 0,37 |
| DIABETES3 | | | | | | | | | |
| Μέθοδος | Συντελεστής Εκπ/σης-Ωμέγα | Εποχές | Σχετικές Εποχές | Διορθώσεις Συνάψεων | Training Set Error | Validation Set Error | Test Set Error | Test Set Classification | Απόλεια Γενίκευσης |
| RPROP | | 105 | 33 | 105 | 20,40 | 22,83 | 22,50 | 37,50 | 3,22 |
| ASD | | 25 | 3 | 25 | 20,38 | 22,86 | 22,05 | 34,38 | 3,73 |
| OFR | | 30 | 3 | 30 | 20,39 | 22,88 | 22,03 | 33,33 | 3,70 |
| LMS | 0,001 | 270 | 110 | 270 | 20,46 | 22,77 | 22,63 | 35,94 | 2,10 |
| B-LMS | 0,001 | 265 | 109 | 265 | 20,46 | 22,77 | 22,64 | 35,94 | 2,07 |
| PR | 0,001 | 265 | 109 | 265 | 20,46 | 22,76 | 22,63 | 35,94 | 2,08 |
| FR | 0,001 | 265 | 109 | 265 | 20,46 | 22,77 | 22,63 | 35,94 | 2,06 |
| RGS | 0,001 | 175 | 175 | 21 | 23,17 | 23,84 | 24,14 | 38,54 | 0,00 |
| JOR | 0,1 | 325 | 120 | 325 | 20,48 | 22,93 | 22,59 | 36,46 | 1,55 |
| SOR | 0,4 | 160 | 154 | 20 | 21,45 | 23,07 | 22,79 | 38,54 | 0,04 |

Πίνακας 7.18 Καλύτερο Πείραμα για το Πρόβλημα Κατάταξης Γονιδίων GENE

| GENE1 | | | | | | | | | |
|--------------|---------------------------|--------|-----------------|---------------------|--------------------|----------------------|----------------|-------------------------|--------------------|
| Μέθοδος | Συντελεστής Εκπ/σης-Ωμέγα | Εποχές | Σχετικές Εποχές | Διορθώσεις Συνάψεων | Training Set Error | Validation Set Error | Test Set Error | Test Set Classification | Απόλεια Γενίκευσης |
| RPROP | | 30 | 13 | 30 | 21,48 | 25,45 | 25,01 | 39,34 | 0,65 |
| ASD | | 20 | 2 | 20 | 21,48 | 24,52 | 24,36 | 41,74 | 4,38 |
| OFR | | 15 | 2 | 15 | 21,48 | 24,63 | 24,48 | 40,48 | 3,82 |
| LMS | 0,0001 | 35 | 6 | 35 | 21,49 | 24,31 | 24,18 | 41,24 | 4,81 |
| B-LMS | 0,0001 | 35 | 5 | 35 | 21,48 | 24,24 | 24,17 | 42,12 | 5,19 |
| PR | 0,0001 | 35 | 5 | 35 | 21,48 | 24,25 | 24,24 | 40,86 | 5,12 |
| FR | 0,0001 | 30 | 5 | 30 | 21,48 | 24,15 | 24,30 | 41,11 | 5,25 |
| RGS | 0,0001 | 375 | 339 | 3 | 23,20 | 24,70 | 24,65 | 41,36 | 0,19 |
| JOR | 0,2 | 25 | 4 | 25 | 21,48 | 24,25 | 24,13 | 41,74 | 5,02 |
| SOR | 0,4 | 315 | 191 | 2 | 21,92 | 24,31 | 24,22 | 42,37 | 0,48 |
| GENE2 | | | | | | | | | |
| Μέθοδος | Συντελεστής Εκπ/σης-Ωμέγα | Εποχές | Σχετικές Εποχές | Διορθώσεις Συνάψεων | Training Set Error | Validation Set Error | Test Set Error | Test Set Classification | Απόλεια Γενίκευσης |
| RPROP | | 30 | 16 | 30 | 21,62 | 25,15 | 24,91 | 38,59 | 0,32 |
| ASD | | 15 | 2 | 15 | 21,62 | 24,34 | 24,50 | 40,98 | 3,58 |
| OFR | | 15 | 2 | 15 | 21,62 | 24,38 | 24,49 | 42,12 | 3,38 |
| LMS | 0,0001 | 35 | 6 | 35 | 21,62 | 24,20 | 24,25 | 41,61 | 3,91 |
| B-LMS | 0,0001 | 35 | 5 | 35 | 21,62 | 24,17 | 24,41 | 41,87 | 4,14 |
| PR | 0,0001 | 30 | 5 | 30 | 21,62 | 24,16 | 24,32 | 42,62 | 4,00 |
| FR | 0,0001 | 30 | 5 | 30 | 21,62 | 24,17 | 24,24 | 42,37 | 3,96 |
| RGS | 0,0001 | 375 | 316 | 3 | 23,37 | 24,56 | 24,75 | 42,37 | 0,22 |
| JOR | 0,2 | 25 | 4 | 25 | 21,62 | 24,17 | 24,24 | 42,24 | 4,06 |
| SOR | 0,3 | 315 | 314 | 2 | 22,25 | 24,17 | 24,28 | 41,61 | 0,00 |
| GENE3 | | | | | | | | | |
| Μέθοδος | Συντελεστής Εκπ/σης-Ωμέγα | Εποχές | Σχετικές Εποχές | Διορθώσεις Συνάψεων | Training Set Error | Validation Set Error | Test Set Error | Test Set Classification | Απόλεια Γενίκευσης |
| RPROP | | 30 | 9 | 30 | 21,88 | 24,20 | 25,28 | 40,23 | 0,77 |
| ASD | | 15 | 3 | 15 | 21,88 | 23,91 | 24,85 | 42,62 | 1,89 |
| OFR | | 15 | 3 | 15 | 21,88 | 23,90 | 24,75 | 42,62 | 1,96 |
| LMS | 0,0001 | 35 | 7 | 35 | 21,88 | 23,64 | 24,41 | 43,25 | 2,94 |
| B-LMS | 0,0001 | 30 | 6 | 30 | 21,88 | 23,58 | 24,34 | 44,14 | 3,04 |
| PR | 0,0001 | 30 | 6 | 30 | 21,88 | 23,60 | 24,53 | 43,38 | 2,94 |
| FR | 0,0001 | 30 | 6 | 30 | 21,88 | 23,58 | 24,50 | 42,88 | 3,03 |
| RGS | 0,0001 | 495 | 431 | 4 | 23,06 | 23,73 | 24,51 | 44,39 | 0,29 |
| JOR | 0,1 | 45 | 10 | 45 | 21,88 | 23,58 | 24,33 | 43,51 | 2,95 |
| SOR | 0,4 | 255 | 191 | 2 | 22,51 | 23,58 | 24,31 | 44,39 | 0,86 |

Πίνακας 7.19 Καλύτερο Πείραμα για το Πρόβλημα Ταξινόμησης Τύπων Γυαλιού GLASS

| GLASS1 | | | | | | | | | |
|---------|---------------------------|--------|-----------------|---------------------|--------------------|----------------------|----------------|-------------------------|--------------------|
| Μέθοδος | Συντελεστής Εκπ/σης-Ωμέγα | Εποχές | Σχετικές Εποχές | Διορθώσεις Συνάψεων | Training Set Error | Validation Set Error | Test Set Error | Test Set Classification | Απόλεια Γενίκευσης |
| RPROP | | 140 | 18 | 140 | 8,84 | 9,59 | 10,32 | 49,06 | 4,87 |
| ASD | | 160 | 45 | 160 | 8,80 | 9,94 | 9,75 | 47,17 | 1,79 |
| OFR | | 90 | 34 | 90 | 8,78 | 9,73 | 9,63 | 43,40 | 3,88 |
| LMS | 0,01 | 240 | 107 | 240 | 8,87 | 9,84 | 9,72 | 45,28 | 0,99 |
| B-LMS | 0,01 | 240 | 106 | 240 | 8,86 | 9,81 | 9,68 | 45,28 | 0,98 |
| PR | 0,01 | 240 | 107 | 240 | 8,86 | 9,81 | 9,68 | 45,28 | 0,97 |
| FR | 0,01 | 240 | 106 | 240 | 8,86 | 9,81 | 9,68 | 45,28 | 0,99 |
| RGS | 0,01 | 410 | 408 | 45 | 9,16 | 9,89 | 9,73 | 47,17 | 0,01 |
| JOR | 0,3 | 130 | 22 | 130 | 8,85 | 9,74 | 10,21 | 45,28 | 3,21 |
| SOR | 0,2 | 270 | 264 | 30 | 9,11 | 9,76 | 10,20 | 45,28 | 0,02 |
| GLASS2 | | | | | | | | | |
| Μέθοδος | Συντελεστής Εκπ/σης-Ωμέγα | Εποχές | Σχετικές Εποχές | Διορθώσεις Συνάψεων | Training Set Error | Validation Set Error | Test Set Error | Test Set Classification | Απόλεια Γενίκευσης |
| RPROP | | 20 | 14 | 20 | 8,95 | 9,93 | 10,33 | 54,72 | 5,45 |
| ASD | | 25 | 7 | 25 | 8,62 | 10,51 | 10,36 | 54,72 | 5,61 |
| OFR | | 15 | 5 | 15 | 8,58 | 10,52 | 10,46 | 52,83 | 8,71 |
| LMS | 0,001 | 650 | 150 | 650 | 8,67 | 10,35 | 10,18 | 52,83 | 5,03 |
| B-LMS | 0,001 | 640 | 148 | 640 | 8,67 | 10,32 | 10,17 | 52,83 | 5,03 |
| PR | 0,01 | 65 | 14 | 65 | 8,66 | 10,32 | 10,18 | 52,83 | 5,23 |
| FR | 0,01 | 65 | 14 | 65 | 8,66 | 10,32 | 10,16 | 52,83 | 5,28 |
| RGS | 0,001 | 1315 | 1309 | 146 | 9,48 | 10,33 | 10,17 | 52,83 | 0,03 |
| JOR | 0,1 | 100 | 19 | 100 | 8,67 | 10,41 | 10,41 | 56,60 | 5,21 |
| SOR | 0,8 | 45 | 13 | 5 | 8,99 | 10,10 | 10,37 | 58,49 | 5,70 |
| GLASS3 | | | | | | | | | |
| Μέθοδος | Συντελεστής Εκπ/σης-Ωμέγα | Εποχές | Σχετικές Εποχές | Διορθώσεις Συνάψεων | Training Set Error | Validation Set Error | Test Set Error | Test Set Classification | Απόλεια Γενίκευσης |
| RPROP | | 140 | 18 | 140 | 8,71 | 9,24 | 11,02 | 60,38 | 3,33 |
| ASD | | 185 | 43 | 185 | 8,67 | 9,40 | 10,95 | 56,60 | 1,44 |
| OFR | | 100 | 28 | 100 | 8,61 | 9,36 | 10,95 | 56,60 | 3,38 |
| LMS | 0,01 | 180 | 90 | 180 | 8,77 | 9,38 | 11,04 | 60,38 | 0,38 |
| B-LMS | 0,01 | 180 | 82 | 180 | 8,76 | 9,39 | 10,88 | 58,49 | 0,48 |
| PR | 0,01 | 180 | 83 | 180 | 8,76 | 9,39 | 10,89 | 58,49 | 0,48 |
| FR | 0,01 | 180 | 82 | 180 | 8,76 | 9,40 | 10,88 | 58,49 | 0,49 |
| RGS | 0,01 | 410 | 409 | 45 | 9,01 | 9,45 | 10,82 | 58,49 | 0,01 |
| JOR | 0,3 | 140 | 16 | 140 | 8,73 | 9,37 | 11,25 | 64,15 | 1,62 |
| SOR | 0,2 | 270 | 203 | 30 | 8,97 | 9,40 | 11,27 | 64,15 | 0,20 |

Πίνακας 7.20 Καλύτερο Πείραμα για το Πρόβλημα Διάγνωσης Καρδιοπάθειας HEART

| HEART1 | | | | | | | | | |
|---------------|---------------------------|--------|-----------------|---------------------|--------------------|----------------------|----------------|-------------------------|--------------------|
| Μέθοδος | Συντελεστής Εκπ/σης-Ωμέγα | Εποχές | Σχετικές Εποχές | Διορθώσεις Συνάψεων | Training Set Error | Validation Set Error | Test Set Error | Test Set Classification | Απόλεια Γενίκευσης |
| RPROP | | 135 | 39 | 135 | 11,18 | 13,15 | 14,36 | 22,17 | 1,82 |
| ASD | | 160 | 159 | 160 | 11,21 | 13,12 | 14,13 | 20,43 | 0,02 |
| OFR | | 60 | 52 | 60 | 11,18 | 13,07 | 14,06 | 20,43 | 0,11 |
| LMS | 0,001 | 215 | 215 | 215 | 11,25 | 13,11 | 14,14 | 20,87 | 0,00 |
| B-LMS | 0,001 | 215 | 215 | 215 | 11,25 | 13,12 | 14,17 | 20,87 | 0,00 |
| PR | 0,001 | 220 | 220 | 220 | 11,25 | 13,13 | 14,18 | 20,87 | 0,00 |
| FR | 0,001 | 215 | 215 | 215 | 11,25 | 13,12 | 14,17 | 20,87 | 0,00 |
| RGS | 0,001 | 410 | 404 | 11 | 12,30 | 13,73 | 15,16 | 22,61 | 0,02 |
| JOR | 0,1 | 240 | 179 | 240 | 11,23 | 13,21 | 14,25 | 20,87 | 0,06 |
| SOR | 0,5 | 270 | 270 | 7 | 11,95 | 13,70 | 14,95 | 21,30 | 0,00 |
| HEART2 | | | | | | | | | |
| Μέθοδος | Συντελεστής Εκπ/σης-Ωμέγα | Εποχές | Σχετικές Εποχές | Διορθώσεις Συνάψεων | Training Set Error | Validation Set Error | Test Set Error | Test Set Classification | Απόλεια Γενίκευσης |
| RPROP | | 160 | 137 | 160 | 11,67 | 12,18 | 13,51 | 16,52 | 0,17 |
| ASD | | 220 | 220 | 220 | 11,66 | 12,22 | 13,60 | 16,52 | 0,00 |
| OFR | | 55 | 44 | 55 | 11,64 | 12,20 | 13,59 | 16,52 | 0,33 |
| LMS | 0,001 | 300 | 300 | 300 | 11,72 | 12,26 | 13,54 | 16,09 | 0,00 |
| B-LMS | 0,001 | 300 | 300 | 300 | 11,72 | 12,25 | 13,55 | 16,52 | 0,00 |
| PR | 0,001 | 300 | 300 | 300 | 11,72 | 12,25 | 13,56 | 16,52 | 0,00 |
| FR | 0,001 | 300 | 300 | 300 | 11,72 | 12,25 | 13,55 | 16,52 | 0,00 |
| RGS | 0,001 | 410 | 400 | 11 | 13,04 | 12,86 | 13,96 | 16,96 | 0,00 |
| JOR | 0,1 | 320 | 320 | 320 | 11,71 | 12,24 | 13,49 | 16,09 | 0,00 |
| SOR | 0,4 | 345 | 330 | 9 | 12,54 | 12,82 | 13,72 | 16,96 | 0,06 |
| HEART3 | | | | | | | | | |
| Μέθοδος | Συντελεστής Εκπ/σης-Ωμέγα | Εποχές | Σχετικές Εποχές | Διορθώσεις Συνάψεων | Training Set Error | Validation Set Error | Test Set Error | Test Set Classification | Απόλεια Γενίκευσης |
| RPROP | | 130 | 34 | 130 | 11,10 | 10,59 | 16,56 | 23,48 | 1,42 |
| ASD | | 190 | 135 | 190 | 11,11 | 10,61 | 16,27 | 22,61 | 0,12 |
| OFR | | 65 | 52 | 65 | 11,09 | 10,53 | 16,21 | 22,17 | 0,70 |
| LMS | 0,001 | 260 | 260 | 260 | 11,15 | 10,57 | 16,23 | 22,61 | 0,00 |
| B-LMS | 0,001 | 265 | 265 | 265 | 11,15 | 10,59 | 16,26 | 22,61 | 0,00 |
| PR | 0,001 | 260 | 260 | 260 | 11,15 | 10,59 | 16,26 | 22,61 | 0,00 |
| FR | 0,001 | 265 | 265 | 265 | 11,15 | 10,59 | 16,27 | 22,61 | 0,00 |
| RGS | 0,001 | 395 | 395 | 11 | 12,30 | 11,93 | 16,88 | 24,78 | 0,00 |
| JOR | 0,1 | 235 | 211 | 235 | 11,14 | 10,70 | 16,32 | 22,61 | 0,00 |
| SOR | 0,4 | 275 | 275 | 7 | 11,86 | 11,25 | 17,01 | 25,65 | 0,00 |

Πίνακας 7.21 Καλύτερο Πείραμα για το Πρόβλημα Διάγνωσης Καρδιοπάθειας HEARTC

| HEARTC1 | | | | | | | | | |
|----------------|---------------------------|--------|-----------------|---------------------|--------------------|----------------------|----------------|-------------------------|--------------------|
| Μέθοδος | Συντελεστής Εκπ/σης-Ωμέγα | Εποχές | Σχετικές Εποχές | Διορθώσεις Συνάψεων | Training Set Error | Validation Set Error | Test Set Error | Test Set Classification | Απώλεια Γενίκευσης |
| RPROP | | 125 | 17 | 125 | 10,16 | 9,44 | 17,45 | 22,67 | 2,19 |
| ASD | | 155 | 155 | 155 | 10,18 | 9,60 | 16,00 | 20,00 | 0,00 |
| OFR | | 75 | 24 | 75 | 10,14 | 9,51 | 16,04 | 18,67 | 0,74 |
| LMS | 0,01 | 110 | 110 | 110 | 10,20 | 9,62 | 16,10 | 20,00 | 0,00 |
| B-LMS | 0,001 | 320 | 320 | 320 | 10,30 | 9,67 | 16,03 | 18,67 | 0,00 |
| PR | 0,001 | 320 | 320 | 320 | 10,30 | 9,68 | 16,02 | 18,67 | 0,00 |
| FR | 0,001 | 315 | 315 | 315 | 10,30 | 9,67 | 16,04 | 18,67 | 0,00 |
| RGS | 0,01 | 360 | 355 | 10 | 10,77 | 10,14 | 16,87 | 18,67 | 0,15 |
| JOR | 0,1 | 215 | 215 | 215 | 10,20 | 9,67 | 16,23 | 20,00 | 0,00 |
| SOR | 0,5 | 210 | 194 | 6 | 10,93 | 9,98 | 17,19 | 21,33 | 0,19 |
| HEARTC2 | | | | | | | | | |
| Μέθοδος | Συντελεστής Εκπ/σης-Ωμέγα | Εποχές | Σχετικές Εποχές | Διορθώσεις Συνάψεων | Training Set Error | Validation Set Error | Test Set Error | Test Set Classification | Απώλεια Γενίκευσης |
| RPROP | | 30 | 15 | 30 | 11,87 | 15,80 | 6,08 | 2,67 | 5,31 |
| ASD | | 170 | 27 | 170 | 11,23 | 16,52 | 6,14 | 4,00 | 2,99 |
| OFR | | 95 | 26 | 95 | 11,23 | 16,39 | 6,27 | 1,33 | 4,01 |
| LMS | 0,001 | 460 | 161 | 460 | 11,36 | 16,63 | 6,23 | 4,00 | 1,11 |
| B-LMS | 0,001 | 460 | 154 | 460 | 11,36 | 16,54 | 6,15 | 4,00 | 1,17 |
| PR | 0,001 | 455 | 155 | 455 | 11,36 | 16,54 | 6,15 | 2,67 | 1,14 |
| FR | 0,001 | 460 | 161 | 460 | 11,36 | 16,54 | 6,16 | 2,67 | 1,12 |
| RGS | 0,001 | 350 | 337 | 10 | 12,27 | 16,62 | 6,29 | 4,00 | 0,66 |
| JOR | 0,1 | 290 | 23 | 290 | 11,27 | 16,63 | 6,42 | 5,33 | 1,80 |
| SOR | 0,2 | 305 | 302 | 8 | 12,62 | 16,79 | 6,59 | 5,33 | 0,14 |
| HEARTC3 | | | | | | | | | |
| Μέθοδος | Συντελεστής Εκπ/σης-Ωμέγα | Εποχές | Σχετικές Εποχές | Διορθώσεις Συνάψεων | Training Set Error | Validation Set Error | Test Set Error | Test Set Classification | Απώλεια Γενίκευσης |
| RPROP | | 20 | 11 | 20 | 10,66 | 13,24 | 12,64 | 16,00 | 12,72 |
| ASD | | 20 | 5 | 20 | 10,55 | 13,06 | 12,08 | 16,00 | 5,19 |
| OFR | | 20 | 5 | 20 | 10,32 | 13,05 | 12,06 | 16,00 | 9,17 |
| LMS | 0,01 | 15 | 3 | 15 | 10,53 | 12,57 | 13,11 | 13,33 | 7,74 |
| B-LMS | 0,001 | 120 | 29 | 120 | 10,53 | 13,06 | 12,00 | 14,67 | 5,14 |
| PR | 0,001 | 125 | 29 | 125 | 10,50 | 13,10 | 12,08 | 16,00 | 5,27 |
| FR | 0,001 | 115 | 28 | 115 | 10,55 | 13,06 | 12,00 | 14,67 | 5,00 |
| RGS | 0,001 | 340 | 142 | 9 | 10,92 | 13,60 | 12,82 | 14,67 | 3,34 |
| JOR | 0,1 | 30 | 11 | 30 | 10,61 | 13,72 | 12,17 | 13,33 | 5,60 |
| SOR | 0,2 | 235 | 177 | 6 | 11,45 | 13,74 | 12,34 | 14,67 | 2,04 |

Πίνακας 7.22 Καλύτερο Πείραμα για το Πρόβλημα Διάγνωσης Κολικού Αλόγου HORSE

| HORSE1 | | | | | | | | | |
|---------------|---------------------------|--------|-----------------|---------------------|--------------------|----------------------|----------------|-------------------------|--------------------|
| Μέθοδος | Συντελεστής Εκπ/σης-Ωμέγα | Εποχές | Σχετικές Εποχές | Διορθώσεις Συνάψεων | Training Set Error | Validation Set Error | Test Set Error | Test Set Classification | Απόλεια Γενίκευσης |
| RPROP | | 25 | 14 | 25 | 11,43 | 15,18 | 12,91 | 27,47 | 5,60 |
| ASD | | 30 | 5 | 30 | 11,29 | 15,13 | 12,70 | 26,37 | 5,45 |
| OFR | | 25 | 5 | 25 | 11,24 | 15,10 | 12,73 | 26,37 | 5,45 |
| LMS | 0,001 | 125 | 29 | 125 | 11,28 | 15,06 | 12,57 | 25,27 | 5,11 |
| B-LMS | 0,001 | 120 | 27 | 120 | 11,29 | 15,08 | 12,64 | 25,27 | 5,22 |
| PR | 0,001 | 120 | 28 | 120 | 11,30 | 15,10 | 12,58 | 25,27 | 5,12 |
| FR | 0,001 | 120 | 28 | 120 | 11,30 | 15,06 | 12,63 | 26,37 | 5,11 |
| RGS | 0,01 | 420 | 161 | 7 | 11,67 | 15,24 | 12,63 | 20,88 | 3,69 |
| JOR | 0,01 | 290 | 78 | 290 | 11,40 | 14,89 | 12,67 | 25,27 | 5,04 |
| SOR | 0,1 | 535 | 457 | 9 | 12,48 | 14,98 | 12,61 | 25,27 | 0,34 |
| HORSE2 | | | | | | | | | |
| Μέθοδος | Συντελεστής Εκπ/σης-Ωμέγα | Εποχές | Σχετικές Εποχές | Διορθώσεις Συνάψεων | Training Set Error | Validation Set Error | Test Set Error | Test Set Classification | Απόλεια Γενίκευσης |
| RPROP | | 20 | 12 | 20 | 9,36 | 15,18 | 17,48 | 35,16 | 9,65 |
| ASD | | 60 | 15 | 60 | 8,38 | 15,63 | 16,61 | 35,16 | 5,14 |
| OFR | | 30 | 13 | 30 | 8,41 | 15,65 | 16,68 | 35,16 | 6,45 |
| LMS | 0,001 | 240 | 48 | 240 | 8,45 | 15,53 | 16,20 | 36,26 | 5,13 |
| B-LMS | 0,001 | 245 | 49 | 245 | 8,43 | 15,57 | 16,21 | 36,26 | 5,05 |
| PR | 0,001 | 240 | 47 | 240 | 8,45 | 15,56 | 16,13 | 36,26 | 5,05 |
| FR | 0,001 | 245 | 50 | 245 | 8,44 | 15,57 | 16,21 | 36,26 | 5,04 |
| RGS | 0,01 | 380 | 296 | 6 | 9,58 | 15,94 | 16,39 | 36,26 | 0,47 |
| JOR | 0,01 | 380 | 100 | 380 | 8,89 | 15,29 | 15,39 | 32,97 | 5,01 |
| SOR | 0,1 | 695 | 547 | 11 | 9,97 | 15,33 | 15,40 | 32,97 | 0,27 |
| HORSE3 | | | | | | | | | |
| Μέθοδος | Συντελεστής Εκπ/σης-Ωμέγα | Εποχές | Σχετικές Εποχές | Διορθώσεις Συνάψεων | Training Set Error | Validation Set Error | Test Set Error | Test Set Classification | Απόλεια Γενίκευσης |
| RPROP | | 15 | 9 | 15 | 11,01 | 14,71 | 15,33 | 30,77 | 9,08 |
| ASD | | 25 | 5 | 25 | 10,24 | 15,22 | 15,00 | 34,07 | 5,88 |
| OFR | | 20 | 7 | 20 | 10,23 | 15,25 | 15,12 | 32,97 | 6,37 |
| LMS | 0,001 | 90 | 22 | 90 | 10,36 | 15,23 | 14,95 | 32,97 | 5,14 |
| B-LMS | 0,001 | 90 | 22 | 90 | 10,36 | 15,22 | 14,99 | 32,97 | 5,02 |
| PR | 0,001 | 95 | 22 | 95 | 10,32 | 15,25 | 14,99 | 32,97 | 5,27 |
| FR | 0,001 | 95 | 22 | 95 | 10,32 | 15,26 | 14,96 | 32,97 | 5,27 |
| RGS | 0,01 | 300 | 98 | 5 | 10,94 | 14,81 | 15,35 | 32,97 | 5,00 |
| JOR | 0,01 | 235 | 64 | 235 | 10,59 | 14,92 | 14,73 | 31,87 | 5,10 |
| SOR | 0,6 | 120 | 59 | 2 | 11,23 | 14,64 | 15,17 | 30,77 | 5,50 |

Πίνακας 7.23 Καλύτερο Πείραμα για το Πρόβλημα Διάγνωσης Ταξινόμησης Φασολιών SOYBEAN

| SOYBEAN1 | | | | | | | | | |
|----------|---------------------------|--------|-----------------|---------------------|--------------------|----------------------|----------------|-------------------------|--------------------|
| Μέθοδος | Συντελεστής Εκπ/σης-Ωμέγα | Εποχές | Σχετικές Εποχές | Διορθώσεις Συνάψεων | Training Set Error | Validation Set Error | Test Set Error | Test Set Classification | Απώλεια Γενίκευσης |
| RPROP | | 525 | 486 | 525 | 0,65 | 0,98 | 1,16 | 8,82 | 0,22 |
| ASD | | 5 | 2 | 5 | 3,72 | 3,90 | 3,97 | 61,76 | 320,96 |
| OFR | | 5 | 2 | 5 | 3,78 | 3,90 | 3,97 | 57,65 | 549,93 |
| LMS | 0,001 | 1035 | 1035 | 1035 | 0,67 | 0,97 | 1,16 | 9,41 | 0,00 |
| B-LMS | 0,0001 | 2880 | 2880 | 2880 | 0,75 | 1,01 | 1,19 | 8,82 | 0,00 |
| PR | 0,0001 | 2870 | 2870 | 2870 | 0,75 | 1,01 | 1,19 | 8,82 | 0,00 |
| FR | 0,0001 | 2880 | 2880 | 2880 | 0,75 | 1,01 | 1,19 | 8,82 | 0,00 |
| RGS | 0,001 | 1200 | 1200 | 4 | 1,79 | 2,01 | 2,15 | 17,06 | 0,00 |
| JOR | 0,01 | 2390 | 2390 | 2390 | 0,74 | 1,01 | 1,22 | 10,00 | 0,00 |
| SOR | 1,2 | 685 | 659 | 8 | 0,96 | 1,27 | 1,43 | 11,76 | 1,02 |
| SOYBEAN2 | | | | | | | | | |
| Μέθοδος | Συντελεστής Εκπ/σης-Ωμέγα | Εποχές | Σχετικές Εποχές | Διορθώσεις Συνάψεων | Training Set Error | Validation Set Error | Test Set Error | Test Set Classification | Απώλεια Γενίκευσης |
| RPROP | | 520 | 484 | 520 | 0,80 | 0,81 | 1,05 | 4,12 | 0,11 |
| ASD | | 5 | 2 | 5 | 3,91 | 3,78 | 4,17 | 56,47 | 354,39 |
| OFR | | 5 | 2 | 5 | 3,86 | 3,72 | 4,11 | 56,47 | 323,37 |
| LMS | 0,001 | 890 | 890 | 890 | 0,82 | 0,81 | 1,06 | 4,12 | 0,00 |
| B-LMS | 0,0001 | 2460 | 2460 | 2460 | 0,90 | 0,87 | 1,18 | 4,12 | 0,00 |
| PR | 0,0001 | 2470 | 2470 | 2470 | 0,90 | 0,87 | 1,18 | 4,12 | 0,00 |
| FR | 0,0001 | 2475 | 2475 | 2475 | 0,90 | 0,87 | 1,18 | 4,12 | 0,00 |
| RGS | 0,001 | 1200 | 1200 | 4 | 1,89 | 1,76 | 2,36 | 25,29 | 0,00 |
| JOR | 0,01 | 2095 | 2095 | 2095 | 0,89 | 0,85 | 1,15 | 5,29 | 0,00 |
| SOR | 0,9 | 545 | 534 | 6 | 1,14 | 0,99 | 1,43 | 10,59 | 0,36 |
| SOYBEAN3 | | | | | | | | | |
| Μέθοδος | Συντελεστής Εκπ/σης-Ωμέγα | Εποχές | Σχετικές Εποχές | Διορθώσεις Συνάψεων | Training Set Error | Validation Set Error | Test Set Error | Test Set Classification | Απώλεια Γενίκευσης |
| RPROP | | 555 | 548 | 555 | 0,78 | 0,96 | 1,03 | 7,06 | 0,04 |
| ASD | | 5 | 2 | 5 | 3,87 | 3,90 | 3,94 | 54,71 | 782,47 |
| OFR | | 5 | 2 | 5 | 3,88 | 3,87 | 3,99 | 56,47 | 328,17 |
| LMS | 0,001 | 990 | 990 | 990 | 0,80 | 0,95 | 1,02 | 5,88 | 0,00 |
| B-LMS | 0,0001 | 2540 | 2540 | 2540 | 0,89 | 1,04 | 1,10 | 6,47 | 0,00 |
| PR | 0,0001 | 2545 | 2545 | 2545 | 0,89 | 1,04 | 1,10 | 7,06 | 0,00 |
| FR | 0,0001 | 2545 | 2545 | 2545 | 0,89 | 1,04 | 1,10 | 6,47 | 0,00 |
| RGS | 0,001 | 1200 | 1200 | 14 | 1,84 | 1,97 | 2,13 | 18,24 | 0,00 |
| JOR | 0,01 | 2115 | 2115 | 2115 | 0,87 | 1,04 | 1,09 | 7,65 | 0,00 |
| SOR | 1,3 | 790 | 787 | 9 | 1,05 | 1,21 | 1,32 | 8,24 | 0,02 |

Πίνακας 7.24 Καλύτερο Πείραμα για το Πρόβλημα Διάγνωσης Θυρεοειδούς THYROID

| THYROID1 | | | | | | | | | |
|----------|---------------------------|--------|-----------------|---------------------|--------------------|----------------------|----------------|-------------------------|--------------------|
| Μέθοδος | Συντελεστής Εκπ/σης-Ωμέγα | Εποχές | Σχετικές Εποχές | Διορθώσεις Συνάψεων | Training Set Error | Validation Set Error | Test Set Error | Test Set Classification | Απώλεια Γενίκευσης |
| RPROP | | 290 | 287 | 290 | 3,91 | 3,96 | 4,08 | 6,56 | 0,01 |
| ASD | | 815 | 814 | 815 | 4,03 | 4,17 | 4,22 | 6,56 | 0,02 |
| OFR | | 555 | 548 | 555 | 3,88 | 3,87 | 3,99 | 6,56 | 0,03 |
| LMS | 0,01 | 325 | 325 | 325 | 3,92 | 3,95 | 4,05 | 6,50 | 0,00 |
| B-LMS | 0,001 | 875 | 875 | 875 | 4,05 | 4,21 | 4,25 | 6,56 | 0,00 |
| PR | 0,001 | 875 | 875 | 875 | 4,05 | 4,21 | 4,25 | 6,56 | 0,00 |
| FR | 0,001 | 875 | 875 | 875 | 4,05 | 4,21 | 4,25 | 6,56 | 0,00 |
| RGS | 0,001 | 180 | 180 | 8 | 6,81 | 6,95 | 6,67 | 7,33 | 0,00 |
| JOR | 0,3 | 375 | 375 | 375 | 3,90 | 3,95 | 4,07 | 6,56 | 0,00 |
| SOR | 0,3 | 195 | 194 | 9 | 4,32 | 4,40 | 4,53 | 6,56 | 0,00 |
| THYROID2 | | | | | | | | | |
| Μέθοδος | Συντελεστής Εκπ/σης-Ωμέγα | Εποχές | Σχετικές Εποχές | Διορθώσεις Συνάψεων | Training Set Error | Validation Set Error | Test Set Error | Test Set Classification | Απώλεια Γενίκευσης |
| RPROP | | 310 | 297 | 310 | 4,08 | 3,65 | 3,82 | 6,38 | 0,14 |
| ASD | | 930 | 929 | 930 | 4,22 | 3,81 | 3,99 | 6,38 | 0,02 |
| OFR | | 370 | 370 | 370 | 4,10 | 3,67 | 3,83 | 6,38 | 0,00 |
| LMS | 0,01 | 325 | 325 | 325 | 4,09 | 3,68 | 3,84 | 6,38 | 0,00 |
| B-LMS | 0,001 | 965 | 965 | 965 | 4,24 | 3,82 | 4,01 | 6,38 | 0,00 |
| PR | 0,001 | 965 | 965 | 965 | 4,24 | 3,82 | 4,01 | 6,38 | 0,00 |
| FR | 0,001 | 965 | 965 | 965 | 4,24 | 3,82 | 4,01 | 6,38 | 0,00 |
| RGS | 0,001 | 180 | 168 | 8 | 6,98 | 6,55 | 6,64 | 7,13 | 0,00 |
| JOR | 0,3 | 385 | 385 | 385 | 4,09 | 3,68 | 3,86 | 6,38 | 0,00 |
| SOR | 0,2 | 225 | 223 | 10 | 4,56 | 4,07 | 4,27 | 6,38 | 0,01 |
| THYROID3 | | | | | | | | | |
| Μέθοδος | Συντελεστής Εκπ/σης-Ωμέγα | Εποχές | Σχετικές Εποχές | Διορθώσεις Συνάψεων | Training Set Error | Validation Set Error | Test Set Error | Test Set Classification | Απώλεια Γενίκευσης |
| RPROP | | 340 | 337 | 340 | 4,02 | 3,48 | 4,19 | 7,22 | 0,05 |
| ASD | | 915 | 915 | 915 | 4,16 | 3,64 | 4,32 | 7,17 | 0,00 |
| OFR | | 510 | 503 | 510 | 3,98 | 3,46 | 4,16 | 7,28 | 0,07 |
| LMS | 0,01 | 370 | 370 | 370 | 4,00 | 3,47 | 4,17 | 7,28 | 0,00 |
| B-LMS | 0,001 | 945 | 945 | 945 | 4,19 | 3,67 | 4,35 | 7,22 | 0,00 |
| PR | 0,001 | 945 | 945 | 945 | 4,19 | 3,67 | 4,35 | 7,22 | 0,00 |
| FR | 0,001 | 945 | 945 | 945 | 4,19 | 3,67 | 4,35 | 7,22 | 0,00 |
| RGS | 0,001 | 180 | 180 | 8 | 6,97 | 6,28 | 6,89 | 7,94 | 0,00 |
| JOR | 0,3 | 450 | 450 | 450 | 4,00 | 3,48 | 4,18 | 7,22 | 0,00 |
| SOR | 0,2 | 225 | 224 | 10 | 4,49 | 3,92 | 4,63 | 7,11 | 0,00 |

7.3.4.3 Παρατηρήσεις στα Αποτελέσματα των Πινάκων (7.15)-(7.24)

1. Ο συντελεστής εκπαίδευσης του καλύτερου πειράματος, ω_{best} , για τη μέθοδο SOR παίρνει τιμές στο διάστημα $[0.1, 1.7]$, όπου $\omega_{best} = 0.1$ για τα προβλήματα HORSE1 και HORSE2 και $\omega_{best} = 1.7$ για το πρόβλημα DIABETES2. Για τη μέθοδο JOR, το ω_{best} παίρνει τιμές στο διάστημα $[0.01, 0.3]$, όπου $\omega_{best} = 0.01$ για τα προβλήματα HORSE και SOYBEAN και $\omega_{best} = 0.3$ για τα προβλήματα DIABETES1, DIABETES2, GLASS1, GLASS3 και THYROID. Για τη μέθοδο RGS, το $\omega_{best} = 0.001$ για τα περισσότερα προβλήματα, εκτός του προβλήματος GENE, όπου $\omega_{best} = 0.0001$ και των προβλημάτων CANCER, DIABETES1, DIABETES2, GLASS1, GLASS3, HEARTC1 και HORSE, όπου $\omega_{best} = 0.01$. Για τη μέθοδο LMS το $\omega_{best} = 0.001$ ή $\omega_{best} = 0.01$, εκτός του προβλήματος GENE, όπου $\omega_{best} = 0.0001$. Για τη μέθοδο LMS το $\omega_{best} = 0.001$ ή $\omega_{best} = 0.01$, εκτός του προβλήματος GENE, όπου $\omega_{best} = 0.0001$. Για τις μεθόδους BATCH-LMS, PR και FR το $\omega_{best} = 0.001$, εκτός των προβλημάτων GENE και SOYBEAN, όπου $\omega_{best} = 0.0001$ και του προβλήματος GLASS, όπου $\omega_{best} = 0.01$.
2. Δεν υπήρχε μεγάλη απώλεια γενίκευσης (πάνω από 5) σε καμιά μέθοδο για τα προβλήματα CANCER2, CANCER3, DIABETES1, DIABETES2, DIABETES3, GENE2, GENE3, GLASS1, GLASS3, HEART, HEARTC1 και THYROID. Οι καλύτερες μέθοδοι με απώλεια γενίκευσης στα λιγότερα προβλήματα, ήταν οι RGS, SOR, JOR και LMS. Μεγάλη απώλεια γενίκευσης για όλες σχεδόν τις μεθόδους είχαν τα προβλήματα CARD1, GLASS2, HEARTC3 και HORSE.
3. Μεγάλα σφάλματα παρατηρήθηκαν στο ποσοστιαίο σφάλμα των προτύπων εκπαίδευσης (Training Set Error), το ποσοστιαίο σφάλμα των προτύπων ελέγχου εγκυρότητας, (Validation Set Error), και το ποσοστιαίο σφάλμα των προτύπων ανάκλησης, (Test Set Error) για τη μέθοδο RGS στα προβλήματα CARD και SOYBEAN και για τις μεθόδους ASD και OFR στο πρόβλημα SOYBEAN.
4. Μεγάλα σφάλματα παρατηρήθηκαν στο σφάλμα ταξινόμησης των προτύπων ανάκλησης, (Test Set Classification Error) για τις μεθόδους RGS, ASD και OFR στο πρόβλημα SOYBEAN. Στα πιο πολλά προβλήματα δε, στο σφάλμα ταξινόμησης των προτύπων ανάκλησης, είναι το ίδιο, για όλες σχεδόν τις μεθόδους.
5. Στα πιο πολλά προβλήματα, οι μέθοδοι OFR και ASD απαιτούν το μικρότερο αριθμό εποχών και σχετικών εποχών που απαιτούνται για τη σύγκλιση τους και το σφάλμα ταξινόμησης των προτύπων ανάκλησης είναι το μικρότερο, ή ισοδύναμο με το σφάλμα ταξινόμησης των προτύπων ανάκλησης της μεθόδου RPROP.
6. Οι χειρότερες μέθοδοι ως προς τις εποχές και τις σχετικές εποχές που απαιτούνται για σύγκλιση ήταν οι μέθοδοι BATCH-LMS, PR και FR και σε λιγότερα οι μέθοδοι SOR και RGS, στις περιπτώσεις όπου ο συντελεστής εκπαίδευσης του καλύτερου πειράματος, ω_{best} έχει πολύ μικρή τιμή.

7. Οι μέθοδοι SOR και RGS εμφανίζουν τις μικρότερες τιμές στις επισκέψεις των συνάψεων, επειδή σε κάθε κύκλο εκπαίδευσης διορθώνονται μόνο οι συνάψεις ενός νευρώνα εισόδου, ενώ στις υπόλοιπες μεθόδους διορθώνονται όλες οι συνάψεις.
8. Οι μέθοδοι LMS, PR και FR εμφανίζουν παρόμοια συμπεριφορά με τη μέθοδο B-LMS, όσον αφορά τον αριθμό επαναλήψεων, τα σφάλματα των προτύπων και το σφάλμα ταξινόμησης, στα περισσότερα παραδείγματα, όπου ο συντελεστής εκπαίδευσης, για τον οποίο συγκλίνουν είναι ο ίδιος. Τα αποτελέσματα της LMS είναι διαφορετικά απ' αυτά των υπολοίπων, για τα παραδείγματα CARD1, DIABETES2, HEARTC1, HEARTC3, SOYBEAN1, SOYBEAN2, SOYBEAN3, THYROID1, THYROID2 και THYROID3, όπου ο συντελεστής εκπαίδευσης, για τον οποίο συγκλίνει η LMS έχει μεγαλύτερη τιμή απ' αυτόν της BATCH-LMS, ενώ οι μέθοδοι PR και FR έχουν διαφορετικά αποτελέσματα για το πρόβλημα GLASS2.
9. Σε πολλά προβλήματα, υπάρχουν διαφορές στη συμπεριφορά των αλγορίθμων στους τρεις διαφορετικούς συνδυασμούς δεδομένων για το κάθε πρόβλημα. Έτσι, ο αριθμός επαναλήψεων που απαιτείται για τη σύγκλιση της κάθε μεθόδου στο πρόβλημα GLASS (GLASS1, GLASS2 και GLASS3) έχει μεγάλη απόκλιση για όλες τις μεθόδους, εκτός της μεθόδου JOR, όπως και για το πρόβλημα HEARTC και HORSE. Απόκλιση εμφανίζει και το πρόβλημα HEARTC στο ποσοστιαίο σφάλμα των προτύπων ελέγχου εγκυρότητας και ανάκλησης, το πρόβλημα HORSE στο ποσοστιαίο σφάλμα των προτύπων ανάκλησης και τα προβλήματα CARD, GLASS, HEART, HEARTC, HORSE και SOYBEAN στο σφάλμα ταξινόμησης των προτύπων ανάκλησης (Test Set Classification Error).
10. Σε κάποια προβλήματα, τα σφάλματα έχουν μια ανώμαλη συμπεριφορά. Π.χ., στα προβλήματα CANCER3, CARD1, CARD3, DIABETES2, HEART3, HEARTC1 και THYROID3 το ποσοστιαίο σφάλμα των προτύπων ελέγχου εγκυρότητας είναι μικρότερο απ' το ποσοστιαίο σφάλμα των προτύπων εκπαίδευσης, ενώ στα προβλήματα CANCER1, HEARTC2 και THYROID2 το ποσοστιαίο σφάλμα των προτύπων ανάκλησης είναι μικρότερο απ' το ποσοστιαίο σφάλμα των προτύπων εκπαίδευσης.

Στους πίνακες (7.25)-(7.28) που ακολουθούν παρουσιάζονται τα αποτελέσματα του καλύτερου πειράματος, της κάθε μεθόδου, του κάθε προβλήματος, όσον αφορά το ελάχιστο σφάλμα των προτύπων ελέγχου εγκυρότητας για τα προβλήματα προσαρμογής, όπως επίσης και η αντίστοιχη τιμή του συντελεστή εκπαίδευσης ω , η οποία χρησιμοποιήθηκε στο καλύτερο πείραμα.

Πίνακας 7.25 Καλύτερο Πείραμα για το Πρόβλημα Ενέργειας Κτιρίου BUILDING

| BUILDING1 | | | | | | | | | |
|------------------|---------------------------|--------|-----------------|---------------------|--------------------|----------------------|----------------|-------------------------|--------------------|
| Μέθοδος | Συντελεστής Εκπ/σης-Ωμέγα | Εποχές | Σχετικές Εποχές | Διορθώσεις Συνάψεων | Training Set Error | Validation Set Error | Test Set Error | Test Set Classification | Απόλεια Γενίκευσης |
| RPROP | | 390 | 390 | 390 | 0,33 | 0,37 | 0,35 | 0,29 | 0,00 |
| ASD | | 470 | 470 | 470 | 0,34 | 0,37 | 0,35 | 0,29 | 0,00 |
| OFR | | 395 | 394 | 395 | 0,33 | 0,37 | 0,34 | 0,29 | 0,00 |
| LMS | 0,001 | 455 | 455 | 455 | 0,34 | 0,37 | 0,35 | 0,29 | 0,00 |
| B-LMS | 0,001 | 5 | 2 | 5 | 18,74 | 19,21 | 19,45 | 100,00 | 85,26 |
| PR | 0,001 | 5 | 2 | 5 | 18,84 | 18,50 | 18,71 | 100,00 | 107,65 |
| FR | 0,001 | 5 | 2 | 5 | 18,62 | 19,05 | 19,40 | 100,00 | 87,32 |
| RGS | 0,001 | 1480 | 1474 | 105 | 0,41 | 0,46 | 0,44 | 0,29 | 0,02 |
| JOR | 0,4 | 460 | 460 | 460 | 0,34 | 0,37 | 0,35 | 0,29 | 0,00 |
| SOR | 1,5 | 500 | 500 | 35 | 0,34 | 0,37 | 0,34 | 0,29 | 0,00 |
| BUILDING2 | | | | | | | | | |
| Μέθοδος | Συντελεστής Εκπ/σης-Ωμέγα | Εποχές | Σχετικές Εποχές | Διορθώσεις Συνάψεων | Training Set Error | Validation Set Error | Test Set Error | Test Set Classification | Απόλεια Γενίκευσης |
| RPROP | | 360 | 357 | 360 | 0,33 | 0,37 | 0,35 | 0,29 | 0,06 |
| ASD | | 475 | 475 | 475 | 0,34 | 0,37 | 0,35 | 0,29 | 0,00 |
| OFR | | 480 | 475 | 480 | 0,33 | 0,37 | 0,34 | 0,29 | 0,00 |
| LMS | 0,001 | 455 | 455 | 455 | 0,34 | 0,37 | 0,35 | 0,29 | 0,00 |
| B-LMS | 0,001 | 5 | 2 | 5 | 18,89 | 19,41 | 19,65 | 100,00 | 94,33 |
| PR | 0,001 | 5 | 2 | 5 | 18,44 | 18,14 | 18,44 | 100,00 | 49,77 |
| FR | 0,001 | 5 | 2 | 5 | 18,93 | 19,33 | 19,61 | 100,00 | 77,18 |
| RGS | 0,001 | 1480 | 1474 | 105 | 0,42 | 0,46 | 0,44 | 0,29 | 0,02 |
| JOR | 0,4 | 460 | 460 | 460 | 0,34 | 0,37 | 0,35 | 0,29 | 0,00 |
| SOR | 1,5 | 500 | 500 | 35 | 0,34 | 0,37 | 0,34 | 0,29 | 0,00 |
| BUILDING3 | | | | | | | | | |
| Μέθοδος | Συντελεστής Εκπ/σης-Ωμέγα | Εποχές | Σχετικές Εποχές | Διορθώσεις Συνάψεων | Training Set Error | Validation Set Error | Test Set Error | Test Set Classification | Απόλεια Γενίκευσης |
| RPROP | | 305 | 295 | 305 | 0,35 | 0,34 | 0,34 | 0,29 | 0,03 |
| ASD | | 525 | 525 | 525 | 0,35 | 0,34 | 0,34 | 0,29 | 0,00 |
| OFR | | 285 | 278 | 285 | 0,35 | 0,34 | 0,34 | 0,29 | 0,03 |
| LMS | 0,01 | 455 | 454 | 455 | 0,35 | 0,35 | 0,34 | 0,29 | 0,00 |
| B-LMS | 0,001 | 5 | 1 | 5 | 18,38 | 18,63 | 18,89 | 85,46 | 148,21 |
| PR | 0,001 | 5 | 2 | 5 | 18,14 | 17,87 | 18,54 | 99,62 | 121,68 |
| FR | 0,001 | 5 | 1 | 5 | 18,40 | 18,59 | 18,92 | 85,08 | 163,08 |
| RGS | 0,001 | 1480 | 1480 | 105 | 0,43 | 0,41 | 0,42 | 0,19 | 0,00 |
| JOR | 0,4 | 465 | 465 | 465 | 0,35 | 0,34 | 0,34 | 0,29 | 0,00 |
| SOR | 1,3 | 430 | 428 | 30 | 0,35 | 0,35 | 0,35 | 0,29 | 0,01 |

Πίνακας 7.26 Καλύτερο Πείραμα για το Πρόβλημα των Ηλιακών Κηλίδων FLARE

| FLARE1 | | | | | | | | | |
|---------|---------------------------|--------|-----------------|---------------------|--------------------|----------------------|----------------|-------------------------|--------------------|
| Μέθοδος | Συντελεστής Εκπ/σης-Ωμέγα | Εποχές | Σχετικές Εποχές | Διορθώσεις Συνάψεων | Training Set Error | Validation Set Error | Test Set Error | Test Set Classification | Απόλεια Γενίκευσης |
| RPROP | | 20 | 6 | 20 | 0,37 | 0,32 | 0,54 | 3,38 | 5,38 |
| ASD | | 5 | 1 | 5 | 0,47 | 0,39 | 0,65 | 5,26 | 0,00 |
| OFR | | 5 | 1 | 5 | 0,46 | 0,40 | 0,64 | 5,26 | 0,00 |
| LMS | 0,1 | 45 | 5 | 45 | 0,37 | 0,33 | 0,54 | 4,14 | 2,70 |
| B-LMS | 0,001 | 155 | 26 | 155 | 0,37 | 0,33 | 0,54 | 3,76 | 1,73 |
| PR | 0,001 | 155 | 28 | 155 | 0,37 | 0,33 | 0,53 | 3,76 | 1,86 |
| FR | 0,001 | 150 | 33 | 150 | 0,37 | 0,33 | 0,53 | 3,76 | 1,56 |
| RGS | 0,01 | 65 | 37 | 2 | 0,39 | 0,34 | 0,54 | 3,76 | 7,16 |
| JOR | 0,2 | 15 | 2 | 15 | 0,37 | 0,32 | 0,54 | 4,14 | 5,17 |
| SOR | 1,2 | 20 | 15 | 0 | 0,39 | 0,32 | 0,51 | 3,76 | 5,17 |
| FLARE2 | | | | | | | | | |
| Μέθοδος | Συντελεστής Εκπ/σης-Ωμέγα | Εποχές | Σχετικές Εποχές | Διορθώσεις Συνάψεων | Training Set Error | Validation Set Error | Test Set Error | Test Set Classification | Απόλεια Γενίκευσης |
| RPROP | | 10 | 8 | 10 | 0,44 | 0,45 | 0,31 | 2,26 | 5,34 |
| ASD | | 5 | 1 | 5 | 0,57 | 0,54 | 0,30 | 3,01 | 0,00 |
| OFR | | 5 | 1 | 5 | 0,57 | 0,54 | 0,29 | 3,01 | 0,00 |
| LMS | 0,001 | 160 | 23 | 160 | 0,42 | 0,45 | 0,28 | 2,26 | 4,32 |
| B-LMS | 0,001 | 160 | 19 | 160 | 0,42 | 0,44 | 0,27 | 2,63 | 4,77 |
| PR | 0,001 | 170 | 18 | 170 | 0,42 | 0,44 | 0,27 | 2,63 | 4,95 |
| FR | 0,001 | 160 | 20 | 160 | 0,42 | 0,44 | 0,27 | 2,63 | 4,81 |
| RGS | 0,01 | 70 | 40 | 2 | 0,44 | 0,44 | 0,27 | 2,63 | 8,06 |
| JOR | 0,1 | 65 | 5 | 65 | 0,42 | 0,45 | 0,27 | 2,63 | 3,98 |
| SOR | 0,7 | 50 | 16 | 2 | 0,43 | 0,45 | 0,25 | 2,63 | 5,30 |
| FLARE3 | | | | | | | | | |
| Μέθοδος | Συντελεστής Εκπ/σης-Ωμέγα | Εποχές | Σχετικές Εποχές | Διορθώσεις Συνάψεων | Training Set Error | Validation Set Error | Test Set Error | Test Set Classification | Απόλεια Γενίκευσης |
| RPROP | | 55 | 8 | 55 | 0,39 | 0,45 | 0,36 | 3,38 | 3,09 |
| ASD | | 5 | 1 | 5 | 0,54 | 0,55 | 0,40 | 3,76 | 0,00 |
| OFR | | 5 | 1 | 5 | 0,53 | 0,54 | 0,40 | 3,76 | 0,00 |
| LMS | 0,001 | 155 | 155 | 155 | 0,39 | 0,46 | 0,35 | 3,01 | 0,00 |
| B-LMS | 0,001 | 160 | 160 | 160 | 0,39 | 0,46 | 0,35 | 3,01 | 0,00 |
| PR | 0,001 | 160 | 160 | 160 | 0,39 | 0,46 | 0,35 | 3,01 | 0,00 |
| FR | 0,001 | 145 | 145 | 145 | 0,39 | 0,46 | 0,35 | 3,01 | 0,00 |
| RGS | 0,001 | 180 | 179 | 7 | 0,42 | 0,48 | 0,33 | 2,63 | 0,00 |
| JOR | 0,3 | 30 | 4 | 30 | 0,39 | 0,46 | 0,35 | 3,01 | 1,42 |
| SOR | 1,2 | 45 | 17 | 1 | 0,39 | 0,44 | 0,38 | 3,01 | 5,71 |

Πίνακας 7.27 Καλύτερο Πείραμα για το Πρόβλημα Διάγνωσης Καρδιοπάθειας HEARTA

| HEARTA1 | | | | | | | | | |
|----------------|---------------------------|--------|-----------------|---------------------|--------------------|----------------------|----------------|-------------------------|--------------------|
| Μέθοδος | Συντελεστής Εκπ/σης-Ωμέγα | Εποχές | Σχετικές Εποχές | Διορθώσεις Συνάψεων | Training Set Error | Validation Set Error | Test Set Error | Test Set Classification | Απώλεια Γενίκευσης |
| RPROP | | 40 | 10 | 40 | 3,94 | 4,26 | 4,60 | 12,17 | 8,67 |
| ASD | | 220 | 6 | 220 | 3,84 | 4,34 | 4,69 | 11,30 | 2,96 |
| OFR | | 75 | 6 | 75 | 3,82 | 4,34 | 4,68 | 11,30 | 3,67 |
| LMS | 0,001 | 300 | 86 | 300 | 3,86 | 4,46 | 4,53 | 10,00 | 0,99 |
| B-LMS | 0,001 | 5 | 4 | 5 | 5,37 | 5,29 | 6,50 | 20,00 | 14,77 |
| PR | 0,001 | 5 | 4 | 5 | 5,05 | 4,93 | 5,93 | 18,70 | 11,94 |
| FR | 0,001 | 5 | 4 | 5 | 5,37 | 5,30 | 6,44 | 20,87 | 13,94 |
| RGS | 0,001 | 435 | 424 | 12 | 4,37 | 4,45 | 4,82 | 12,17 | 0,15 |
| JOR | 0,1 | 240 | 65 | 240 | 3,83 | 4,38 | 4,44 | 10,00 | 1,07 |
| SOR | 0,3 | 120 | 110 | 3 | 4,19 | 4,40 | 4,54 | 11,74 | 0,18 |
| HEARTA2 | | | | | | | | | |
| Μέθοδος | Συντελεστής Εκπ/σης-Ωμέγα | Εποχές | Σχετικές Εποχές | Διορθώσεις Συνάψεων | Training Set Error | Validation Set Error | Test Set Error | Test Set Classification | Απώλεια Γενίκευσης |
| RPROP | | 105 | 86 | 105 | 4,16 | 4,20 | 4,19 | 10,87 | 0,34 |
| ASD | | 175 | 175 | 175 | 4,18 | 4,25 | 4,18 | 10,87 | 0,00 |
| OFR | | 70 | 64 | 70 | 4,16 | 4,23 | 4,18 | 10,43 | 0,06 |
| LMS | 0,001 | 225 | 225 | 225 | 4,20 | 4,26 | 4,20 | 9,57 | 0,00 |
| B-LMS | 0,001 | 225 | 225 | 225 | 4,20 | 4,27 | 4,20 | 10,87 | 0,00 |
| PR | 0,001 | 220 | 220 | 220 | 4,20 | 4,27 | 4,21 | 10,87 | 0,00 |
| FR | 0,001 | 225 | 225 | 225 | 4,19 | 4,27 | 4,20 | 10,87 | 0,00 |
| RGS | 0,001 | 510 | 504 | 14 | 4,49 | 4,50 | 4,57 | 8,26 | 0,01 |
| JOR | 0,1 | 170 | 170 | 170 | 4,18 | 4,29 | 4,19 | 11,30 | 0,00 |
| SOR | 0,3 | 160 | 154 | 4 | 4,39 | 4,31 | 4,31 | 9,13 | 0,17 |
| HEARTA3 | | | | | | | | | |
| Μέθοδος | Συντελεστής Εκπ/σης-Ωμέγα | Εποχές | Σχετικές Εποχές | Διορθώσεις Συνάψεων | Training Set Error | Validation Set Error | Test Set Error | Test Set Classification | Απώλεια Γενίκευσης |
| RPROP | | 95 | 57 | 95 | 4,06 | 4,07 | 4,56 | 11,30 | 1,06 |
| ASD | | 140 | 139 | 140 | 4,08 | 4,11 | 4,58 | 12,17 | 0,04 |
| OFR | | 50 | 48 | 50 | 4,06 | 4,06 | 4,52 | 11,30 | 0,02 |
| LMS | 0,01 | 60 | 60 | 60 | 4,09 | 4,11 | 4,53 | 10,87 | 0,00 |
| B-LMS | 0,001 | 185 | 185 | 185 | 4,09 | 4,12 | 4,59 | 11,74 | 0,00 |
| PR | 0,001 | 190 | 190 | 190 | 4,09 | 4,12 | 4,59 | 11,74 | 0,00 |
| FR | 0,001 | 185 | 185 | 185 | 4,09 | 4,11 | 4,59 | 11,74 | 0,00 |
| RGS | 0,001 | 500 | 486 | 14 | 4,42 | 4,25 | 4,76 | 13,91 | 0,18 |
| JOR | 0,1 | 300 | 123 | 300 | 4,16 | 4,11 | 4,56 | 12,17 | 1,33 |
| SOR | 0,2 | 190 | 190 | 5 | 4,28 | 4,09 | 4,56 | 11,74 | 0,00 |

Πίνακας 7.28 Καλύτερο Πείραμα για το Πρόβλημα Διάγνωσης Καρδιοπάθειας HEARTAC

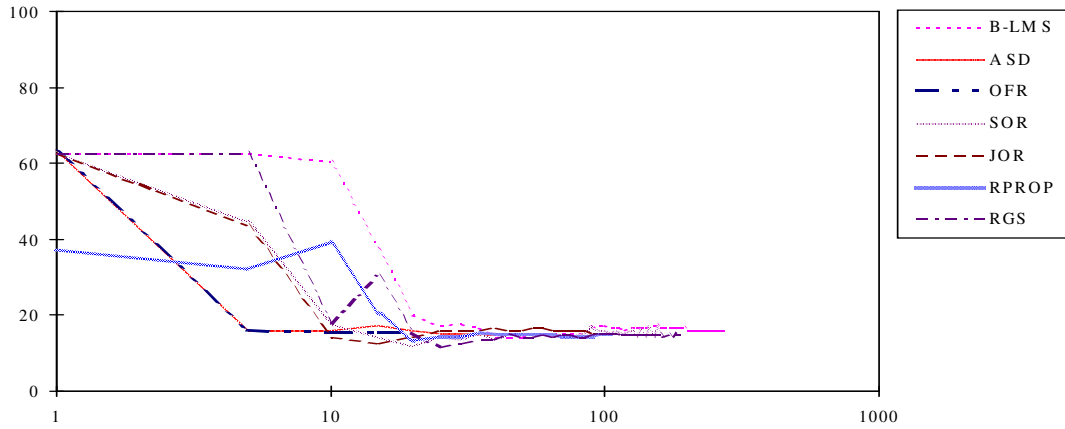
| HEARTAC1 | | | | | | | | | |
|----------|---------------------------|--------|-----------------|---------------------|--------------------|----------------------|----------------|-------------------------|--------------------|
| Μέθοδος | Συντελεστής Εκπ/σης-Ωμέγα | Εποχές | Σχετικές Εποχές | Διορθώσεις Συνάψεων | Training Set Error | Validation Set Error | Test Set Error | Test Set Classification | Απώλεια Γενίκευσης |
| RPROP | | 15 | 9 | 15 | 4,19 | 4,63 | 3,14 | 6,67 | 7,85 |
| ASD | | 125 | 125 | 125 | 4,05 | 4,67 | 2,66 | 5,33 | 0,00 |
| OFR | | 75 | 75 | 75 | 4,04 | 4,62 | 2,72 | 5,33 | 0,00 |
| LMS | 0,001 | 320 | 320 | 320 | 4,09 | 4,77 | 2,58 | 4,00 | 0,00 |
| B-LMS | 0,001 | 335 | 335 | 335 | 4,09 | 4,74 | 2,61 | 4,00 | 0,00 |
| PR | 0,001 | 340 | 340 | 340 | 4,09 | 4,74 | 2,61 | 5,33 | 0,00 |
| FR | 0,001 | 335 | 335 | 335 | 4,09 | 4,74 | 2,61 | 4,00 | 0,00 |
| RGS | 0,01 | 50 | 45 | 1 | 5,00 | 5,51 | 4,07 | 6,67 | 0,19 |
| JOR | 0,1 | 170 | 170 | 170 | 4,07 | 4,81 | 2,62 | 4,00 | 0,00 |
| SOR | 0,5 | 45 | 34 | 1 | 4,59 | 5,59 | 3,77 | 6,67 | 1,06 |
| HEARTAC2 | | | | | | | | | |
| Μέθοδος | Συντελεστής Εκπ/σης-Ωμέγα | Εποχές | Σχετικές Εποχές | Διορθώσεις Συνάψεων | Training Set Error | Validation Set Error | Test Set Error | Test Set Classification | Απώλεια Γενίκευσης |
| RPROP | | 10 | 7 | 10 | 3,74 | 4,49 | 4,28 | 6,67 | 11,97 |
| ASD | | 5 | 2 | 5 | 3,90 | 4,59 | 4,11 | 9,33 | 16,32 |
| OFR | | 5 | 2 | 5 | 3,89 | 4,64 | 4,09 | 9,33 | 15,91 |
| LMS | 0,001 | 110 | 18 | 110 | 3,28 | 5,19 | 3,72 | 8,00 | 5,05 |
| B-LMS | 0,001 | 100 | 18 | 100 | 3,31 | 5,15 | 3,72 | 8,00 | 5,06 |
| PR | 0,001 | 105 | 18 | 105 | 3,29 | 5,18 | 3,70 | 8,00 | 5,08 |
| FR | 0,001 | 105 | 18 | 105 | 3,30 | 5,16 | 3,74 | 8,00 | 5,18 |
| RGS | 0,01 | 220 | 208 | 6 | 5,05 | 5,49 | 4,71 | 8,00 | 0,82 |
| JOR | 0,1 | 25 | 9 | 25 | 3,28 | 4,92 | 3,79 | 8,00 | 6,81 |
| SOR | 0,1 | 630 | 341 | 18 | 3,35 | 4,95 | 3,91 | 9,33 | 4,25 |
| HEARTAC3 | | | | | | | | | |
| Μέθοδος | Συντελεστής Εκπ/σης-Ωμέγα | Εποχές | Σχετικές Εποχές | Διορθώσεις Συνάψεων | Training Set Error | Validation Set Error | Test Set Error | Test Set Classification | Απώλεια Γενίκευσης |
| RPROP | | 20 | 9 | 20 | 2,89 | 5,12 | 5,07 | 12,00 | 7,43 |
| ASD | | 5 | 2 | 5 | 3,23 | 5,00 | 5,43 | 16,00 | 7,04 |
| OFR | | 5 | 2 | 5 | 3,24 | 5,00 | 5,44 | 16,00 | 6,84 |
| LMS | 0,001 | 35 | 12 | 35 | 3,12 | 5,19 | 5,42 | 14,67 | 5,15 |
| B-LMS | 0,001 | 35 | 12 | 35 | 3,12 | 5,11 | 5,34 | 12,00 | 5,45 |
| PR | 0,001 | 35 | 11 | 35 | 3,14 | 5,11 | 5,43 | 16,00 | 5,45 |
| FR | 0,001 | 35 | 12 | 35 | 3,11 | 5,12 | 5,37 | 13,33 | 5,28 |
| RGS | 0,01 | 325 | 312 | 9 | 3,87 | 5,26 | 5,42 | 16,00 | 0,24 |
| JOR | 0,001 | 1450 | 704 | 1450 | 2,91 | 5,34 | 5,22 | 16,00 | 3,48 |
| SOR | 0,2 | 250 | 127 | 7 | 2,93 | 5,49 | 5,24 | 13,33 | 3,98 |

7.3.4.4 Παρατηρήσεις στα Αποτελέσματα των Πινάκων (7.25)-(7.28)

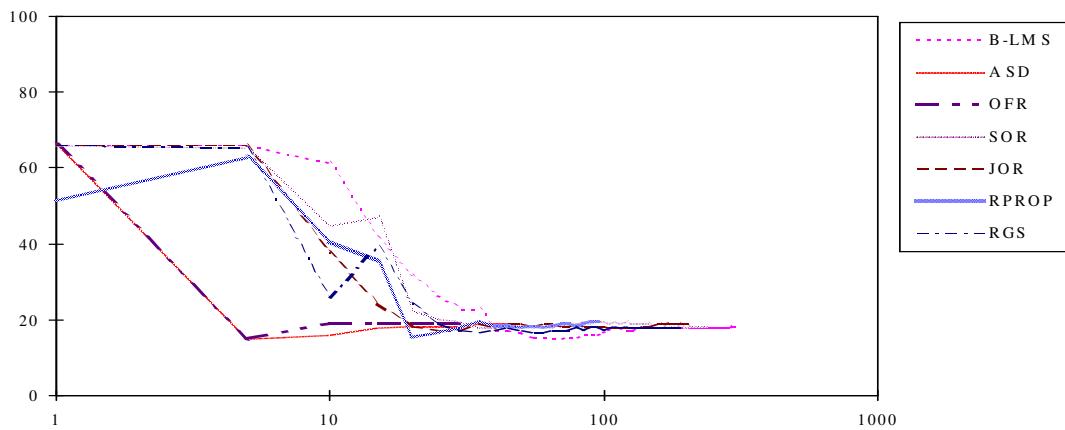
1. Ο συντελεστής εκπαίδευσης του καλύτερου πειράματος, ω_{best} , για τη μέθοδο SOR παίρνει τιμές στο διάστημα $[0.1, 1.5]$, όπου $\omega_{best} = 0.1$ για το πρόβλημα HEARTC2 και $\omega_{best} = 1.5$ για τα προβλήματα BUILDING1 και BUILDING2. Για τη μέθοδο JOR, το ω_{best} παίρνει τιμές στο διάστημα $[0.001, 0.4]$, όπου $\omega_{best} = 0.001$ για το πρόβλημα HEARTC3 και $\omega_{best} = 0.4$ για το πρόβλημα BUILDING. Για τη μέθοδο RGS, το $\omega_{best} = 0.001$ για τα περισσότερα προβλήματα, εκτός των προβλημάτων FLARE και HEARTAC, όπου $\omega_{best} = 0.01$. Για τη μέθοδο LMS το $\omega_{best} = 0.001$ εκτός των προβλημάτων BUILDING3 και HEARTC3, όπου $\omega_{best} = 0.01$ και του προβλήματος FLARE1, όπου $\omega_{best} = 0.1$. Για τις μεθόδους BATCH-LMS, PR και FR το $\omega_{best} = 0.001$ για όλα τα προβλήματα.
2. Δεν υπήρχε μεγάλη απώλεια γενίκευσης (πάνω από 5) σε καμιά μέθοδο για τα προβλήματα HEARTA2, και HEARTA3. Οι καλύτερες μέθοδοι με απώλεια γενίκευσης σε δύο μόνο απ' τα δώδεκα προβλήματα, ήταν οι ASD, OFR, RGS, JOR και LMS και χειρότερες με απώλεια γενίκευσης σε έξι απ' τα δώδεκα προβλήματα, ήταν οι μέθοδοι RPROP, BATCH-LMS, PR και FR. Μεγάλη απώλεια γενίκευσης για τις πιο πολλές μεθόδους είχαν τα προβλήματα HEARTAC2 και HEARTAC3.
3. Μεγάλα σφάλματα παρατηρήθηκαν στο ποσοστιαίο σφάλμα των προτύπων εκπαίδευσης (Training Set Error), το ποσοστιαίο σφάλμα των προτύπων ελέγχου εγκυρότητας, (Validation Set Error) και το ποσοστιαίο σφάλμα των προτύπων ανάκλησης, (Test Set Error) για τις μεθόδους BATCH-LMS, PR και FR για το πρόβλημα BUILDING.
4. Μεγάλα σφάλματα παρατηρήθηκαν στο σφάλμα ταξινόμησης των προτύπων ανάκλησης, (Test Set Classification Error) για τις μεθόδους BATCH-LMS, PR και FR στα προβλήματα BUILDING και HEARTA1. Στα πιο πολλά προβλήματα δε, στο σφάλμα ταξινόμησης των προτύπων ανάκλησης, είναι το ίδιο, για όλες σχεδόν τις μεθόδους.
5. Στα πιο πολλά προβλήματα, οι μέθοδοι OFR και ASD απαιτούν το μικρότερο αριθμό εποχών και σχετικών εποχών που απαιτούνται για τη σύγκλισή τους και το σφάλμα ταξινόμησης των προτύπων ανάκλησης είναι το μικρότερο, ή ισοδύναμο με το σφάλμα ταξινόμησης των προτύπων ανάκλησης της μεθόδου RPROP.
6. Οι χειρότερες μέθοδοι ως προς τις εποχές και τις σχετικές εποχές που απαιτούνται για σύγκλιση ήταν οι μέθοδοι RGS, BATCH-LMS, PR και FR. Η BATCH-LMS εμφανίζει το μικρότερο αριθμό εποχών και σχετικών εποχών στα προβλήματα BUILDING, αλλά το σφάλμα ταξινόμησης των προτύπων ανάκλησης παίρνει τιμές μεταξύ 85% και 100%.
7. Οι μέθοδοι SOR και RGS εμφανίζουν τις μικρότερες τιμές στις επισκέψεις των συνάψεων, επειδή σε κάθε κύκλο εκπαίδευσης διορθώνονται μόνο οι συνάψεις ενός νευρώνα εισόδου, ενώ στις υπόλοιπες μεθόδους διορθώνονται όλες οι συνάψεις.

8. Οι μέθοδοι LMS, PR και FR εμφανίζουν παρόμοια συμπεριφορά με τη μέθοδο B-LMS, όσον αφορά τον αριθμό επαναλήψεων, τα σφάλματα των προτύπων και το σφάλμα ταξινόμησης, στα παραδείγματα FLARE2, FLARE3, HEARTA2, HEARTAC1, HEARTAC2 και HEARTAC3, όπου ο συντελεστής εκπαίδευσης, για τον οποίο συγκλίνουν είναι ο ίδιος. Τα αποτελέσματα της LMS είναι διαφορετικά απ' αυτά των υπολοίπων, για τα παραδείγματα FLARE1 και HEARTA3, όπου ο συντελεστής εκπαίδευσης, για τον οποίο συγκλίνει η LMS έχει μεγαλύτερη τιμή απ' αυτόν της BATCH-LMS, και για τα παραδείγματα BUILDING1, BUILDING2, BUILDING3 και HEARTA1, όπου οι υπόλοιπες μέθοδοι έχουν πολύ κακά αποτελέσματα, λόγω μεγάλης απώλειας γενίκευσης.
9. Σε πολλά προβλήματα, υπάρχουν διαφορές στη συμπεριφορά των αλγορίθμων στους τρεις διαφορετικούς συνδυασμούς δεδομένων για το κάθε πρόβλημα. Έτσι, ο αριθμός επαναλήψεων που απαιτείται για τη σύγκλιση της κάθε μεθόδου στο πρόβλημα HEARTA (HEARTA1, HEARTA2 και HEARTA3) έχει μεγάλη απόκλιση για τις μεθόδους BATCH-LMS, PR και FR και στο πρόβλημα HEARTAC για τις μεθόδους OFR και ASD. Απόκλιση εμφανίζει και το πρόβλημα HEARTAC στο σφάλμα ταξινόμησης των προτύπων ανάκλησης (Test Set Classification Error).
10. Σε κάποια προβλήματα, τα σφάλματα έχουν μια ανώμαλη συμπεριφορά. Π.χ., στα προβλήματα BUILDING3 και FLARE1 το ποσοστιαίο σφάλμα των προτύπων ελέγχου εγκυρότητας είναι μικρότερο απ' το ποσοστιαίο σφάλμα των προτύπων εκπαίδευσης, ενώ στα προβλήματα FLARE2, FLARE3 και HEARTAC1 το ποσοστιαίο σφάλμα των προτύπων ανάκλησης είναι μικρότερο απ' το ποσοστιαίο σφάλμα των προτύπων εκπαίδευσης.

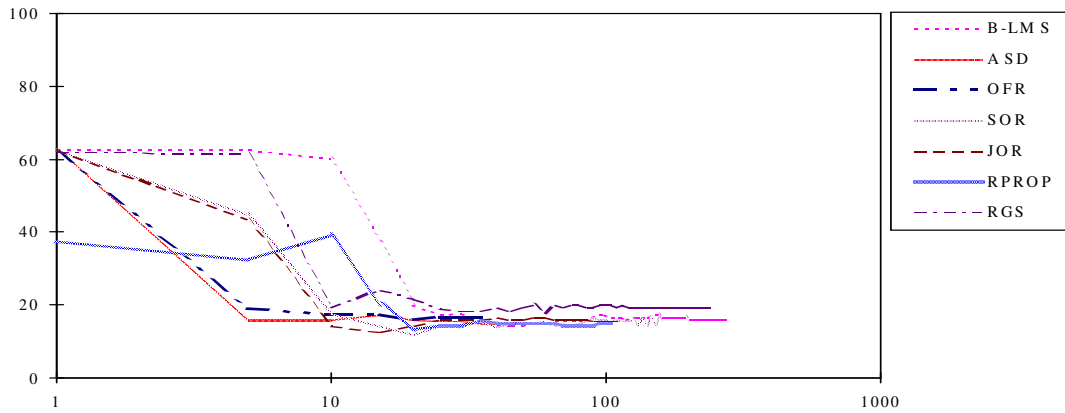
Στα διαγράμματα (7.10-7.19) που ακολουθούν παρουσιάζονται τα αποτελέσματα του καλύτερου πειράματος του κάθε προβλήματος ταξινόμησης, για τις μεθόδους RPROP, BATCH-LMS, ASD, SOR και JOR, όσον αφορά τη μεταβολή του σφάλματος ταξινόμησης των προτύπων ανάκλησης σε σχέση με τις επαναλήψεις. Παραλήφθηκαν οι μέθοδοι PR, FR και LMS, των οποίων η συμπεριφορά είναι παρόμοια με τη BATCH-LMS.



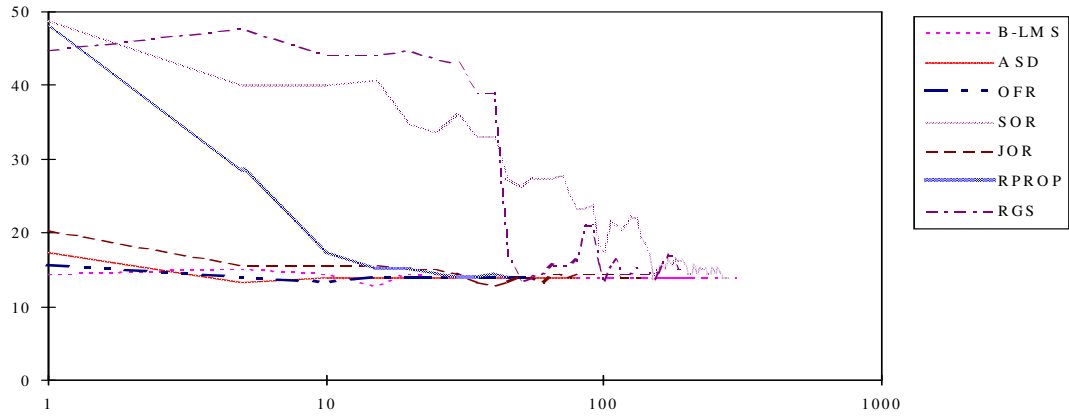
Διάγραμμα 7.10α. Μεταβολή Test Set Classification Error ως προς τις Επαναλήψεις για το Πρόβλημα CANCER1



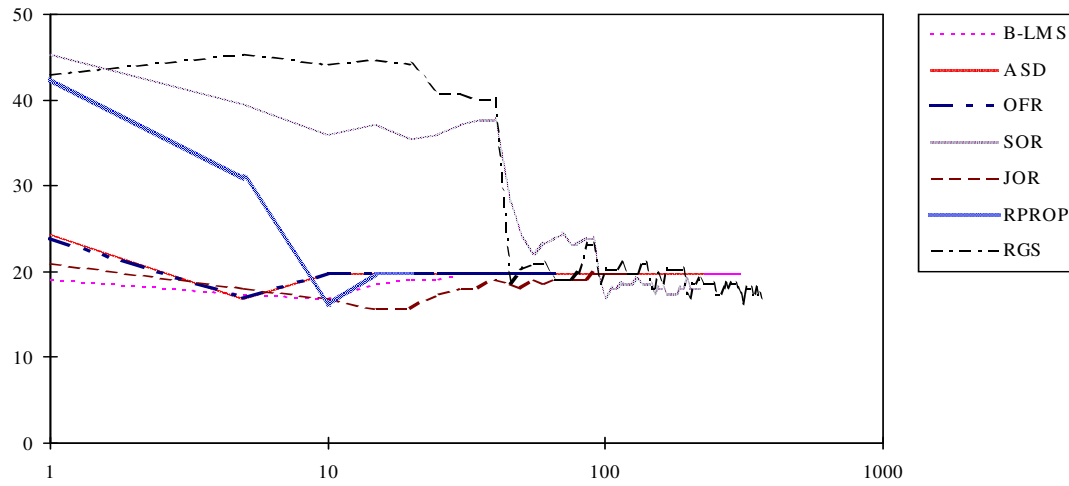
Διάγραμμα 7.10β. Μεταβολή Test Set Classification Error ως προς τις Επαναλήψεις για το Πρόβλημα CANCER2



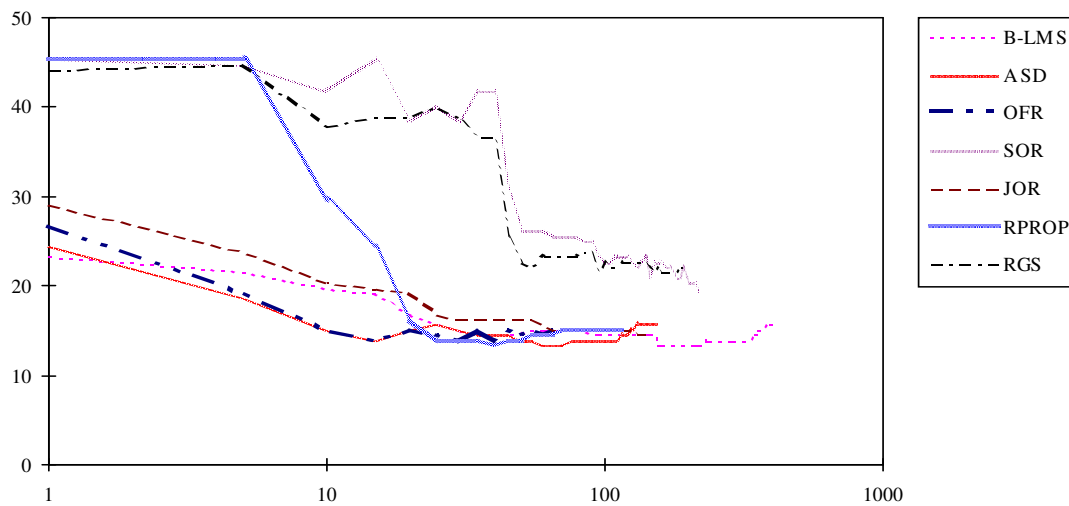
Διάγραμμα 7.10γ. Μεταβολή Test Set Classification Error ως προς τις Επαναλήψεις για το Πρόβλημα CANCER3



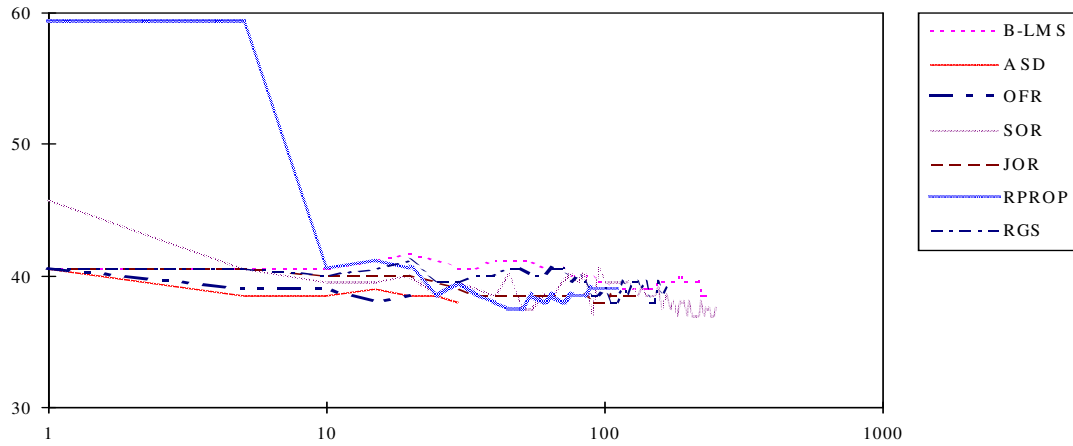
Διάγραμμα 7.11a. Μεταβολή Test Set Classification Error ως προς τις Επαναλήψεις για το Πρόβλημα CARD1



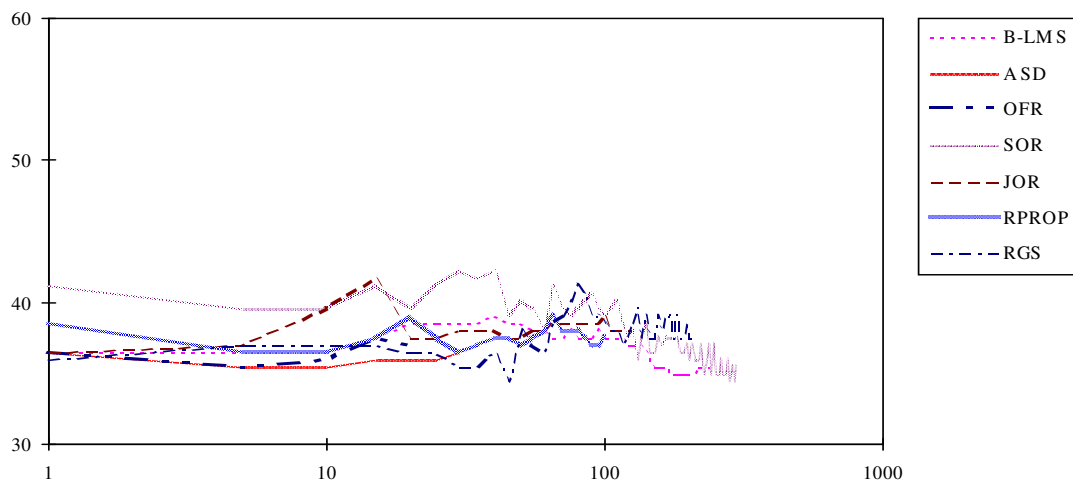
Διάγραμμα 7.11b. Μεταβολή Test Set Classification Error ως προς τις Επαναλήψεις για το Πρόβλημα CARD2



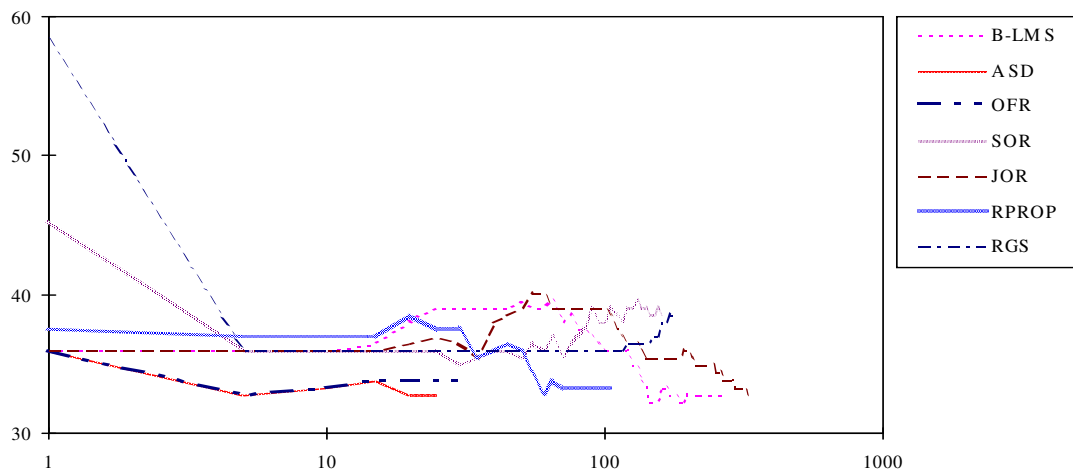
Διάγραμμα 7.11c. Μεταβολή Test Set Classification Error ως προς τις Επαναλήψεις για το Πρόβλημα CARD3



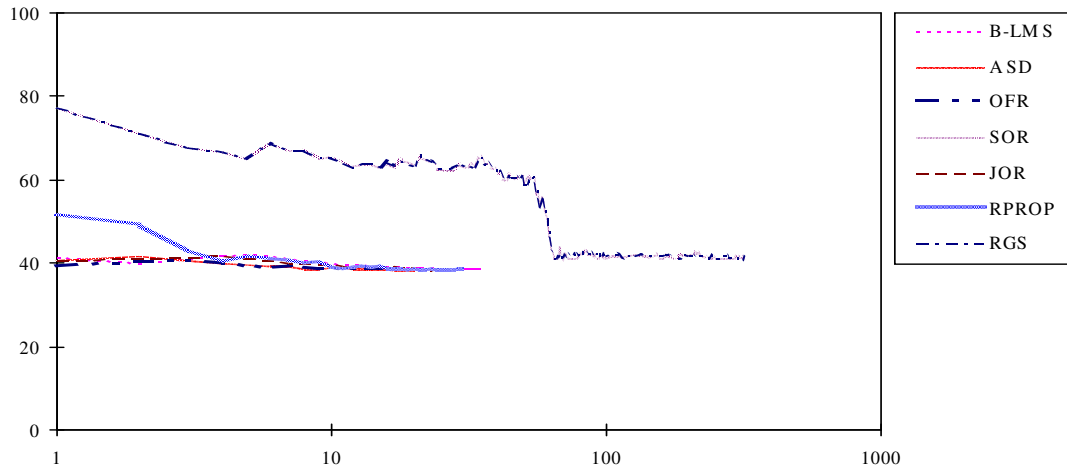
Διάγραμμα 7.12a. Μεταβολή Test Set Classification Error ως προς τις Επαναλήψεις για το Πρόβλημα DIABETES1



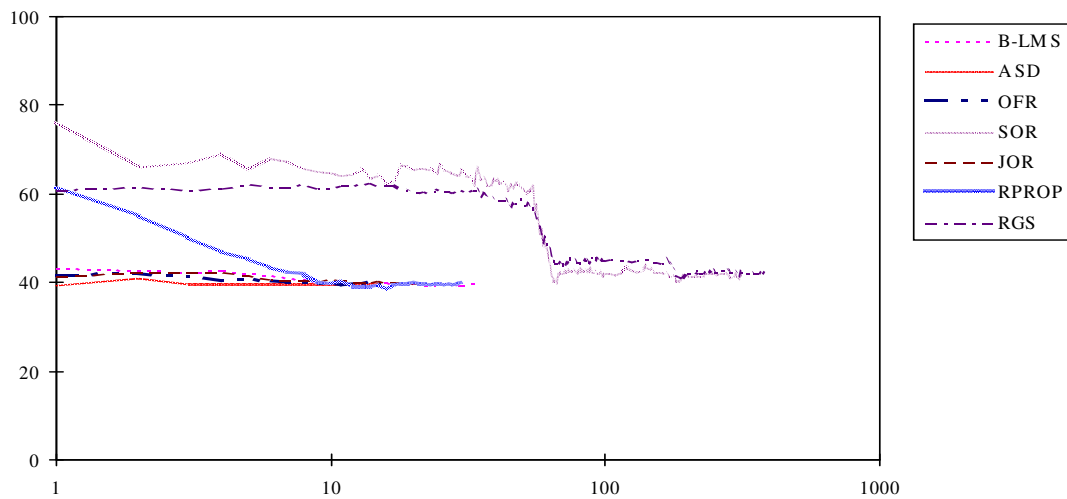
Διάγραμμα 7.12b. Μεταβολή Test Set Classification Error ως προς τις Επαναλήψεις για το Πρόβλημα DIABETES2



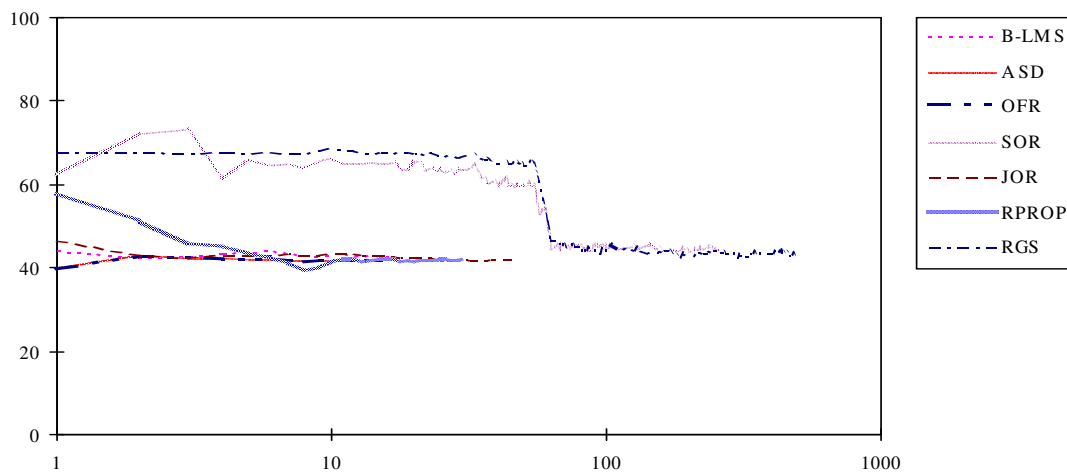
Διάγραμμα 7.12c. Μεταβολή Test Set Classification Error ως προς τις Επαναλήψεις για το Πρόβλημα DIABETES3



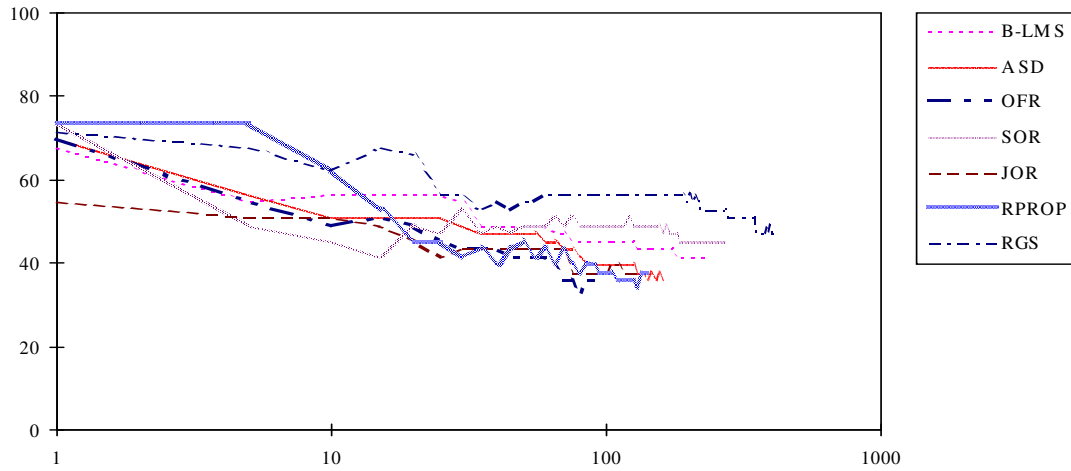
Διάγραμμα 7.13a. Μεταβολή Test Set Classification Error ως προς τις Επαναλήψεις για το Πρόβλημα GENE1



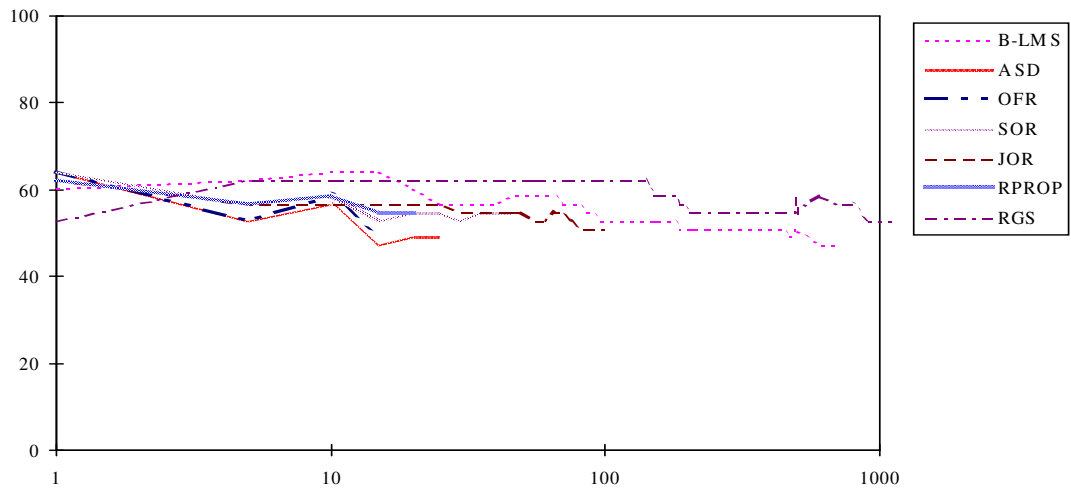
Διάγραμμα 7.13b. Μεταβολή Test Set Classification Error ως προς τις Επαναλήψεις για το Πρόβλημα GENE2



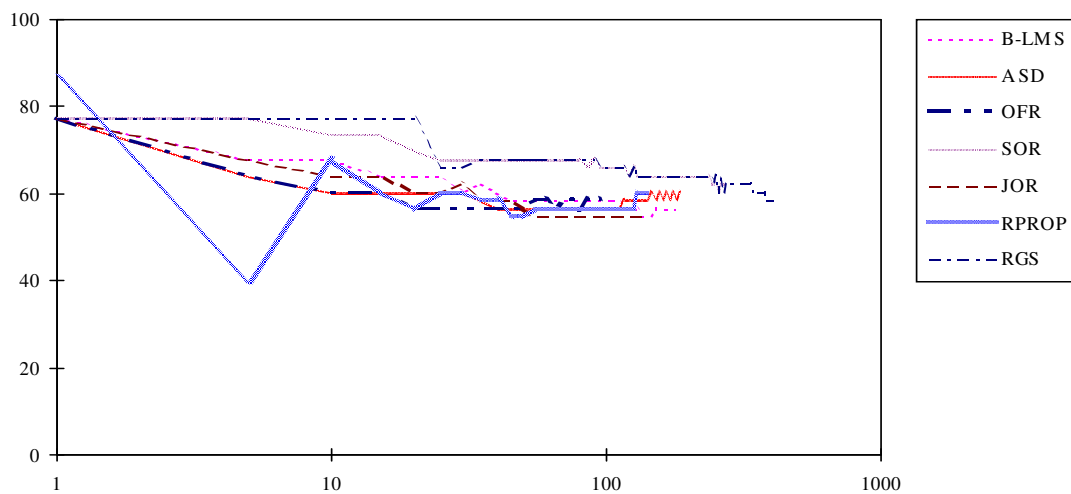
Διάγραμμα 7.13c. Μεταβολή Test Set Classification Error ως προς τις Επαναλήψεις για το Πρόβλημα GENE3



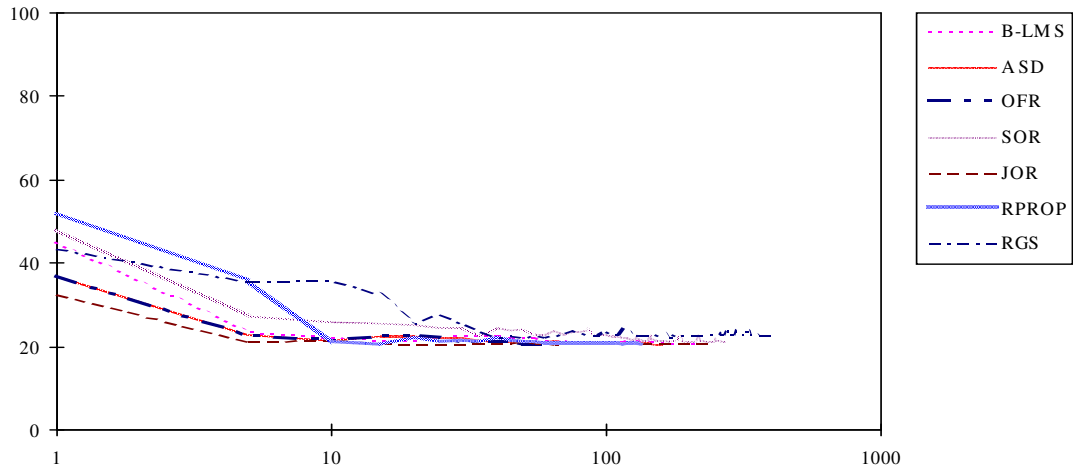
Διάγραμμα 7.14a. Μεταβολή Test Set Classification Error ως προς τις Επαναλήψεις για το Πρόβλημα GLASS1



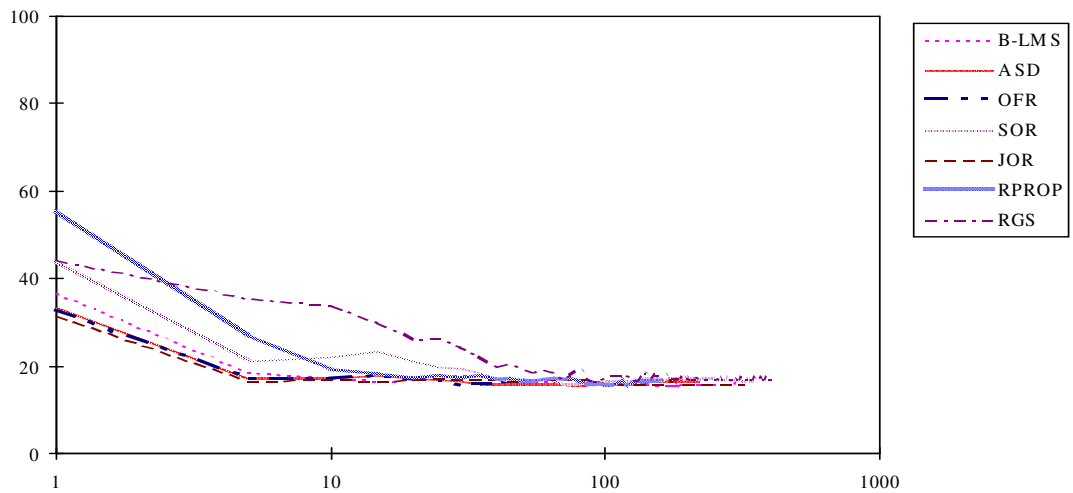
Διάγραμμα 7.14b. Μεταβολή Test Set Classification Error ως προς τις Επαναλήψεις για το Πρόβλημα GLASS2



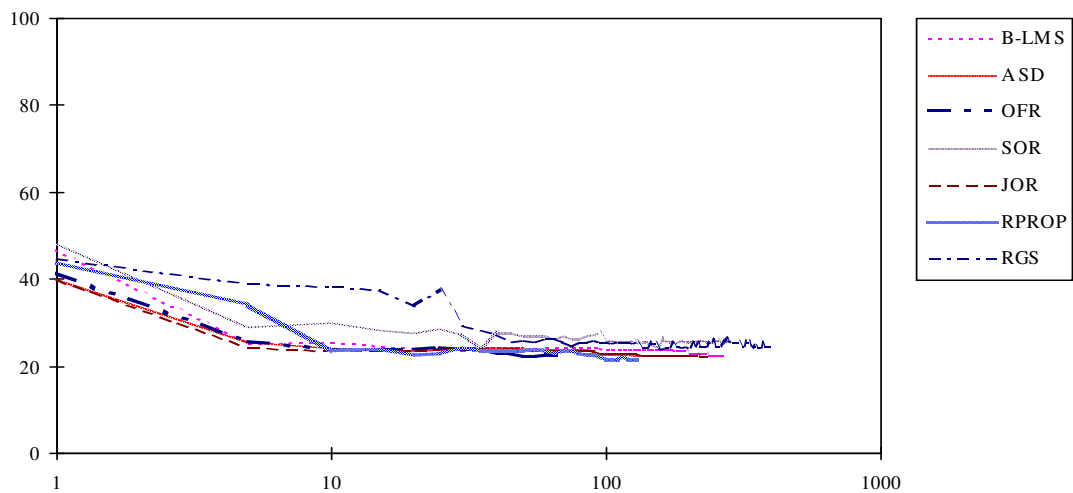
Διάγραμμα 7.14c. Μεταβολή Test Set Classification Error ως προς τις Επαναλήψεις για το Πρόβλημα GLASS3



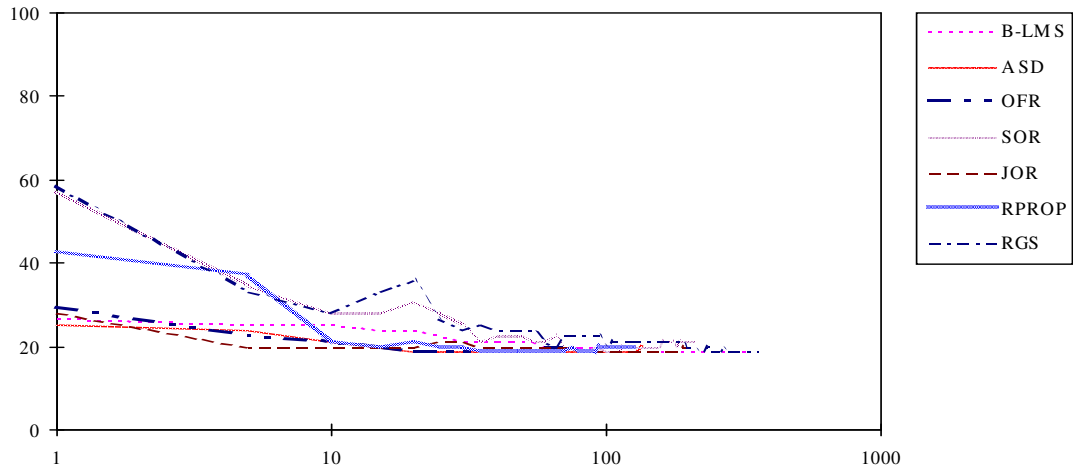
Διάγραμμα 7.15a. Μεταβολή Test Set Classification Error ως προς τις Επαναλήψεις για το Πρόβλημα HEART1



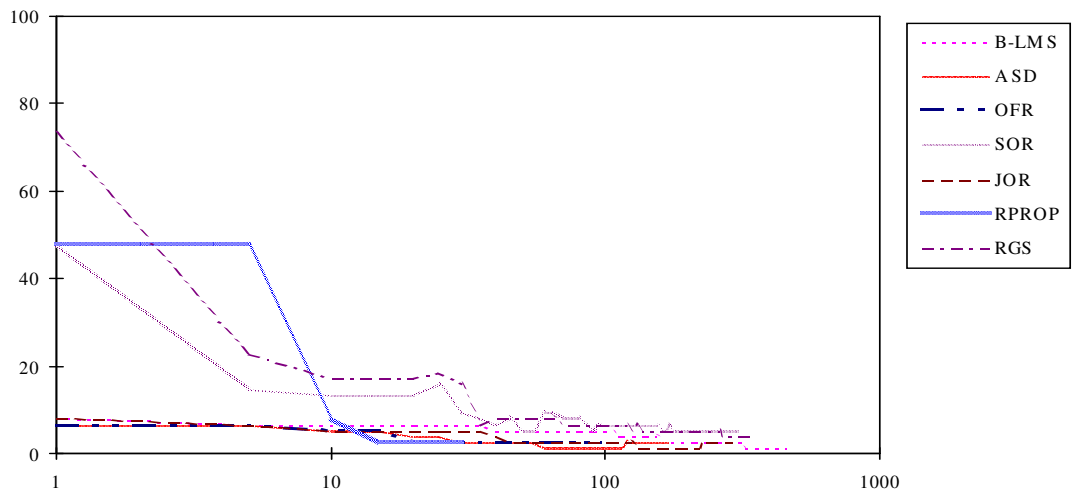
Διάγραμμα 7.15b. Μεταβολή Test Set Classification Error ως προς τις Επαναλήψεις για το Πρόβλημα HEART2



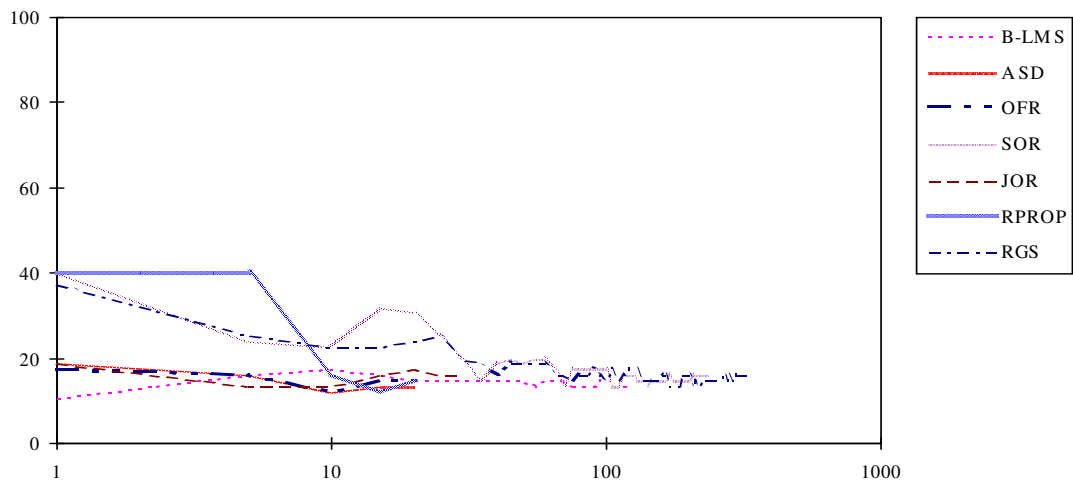
Διάγραμμα 7.15c. Μεταβολή Test Set Classification Error ως προς τις Επαναλήψεις για το Πρόβλημα HEART3



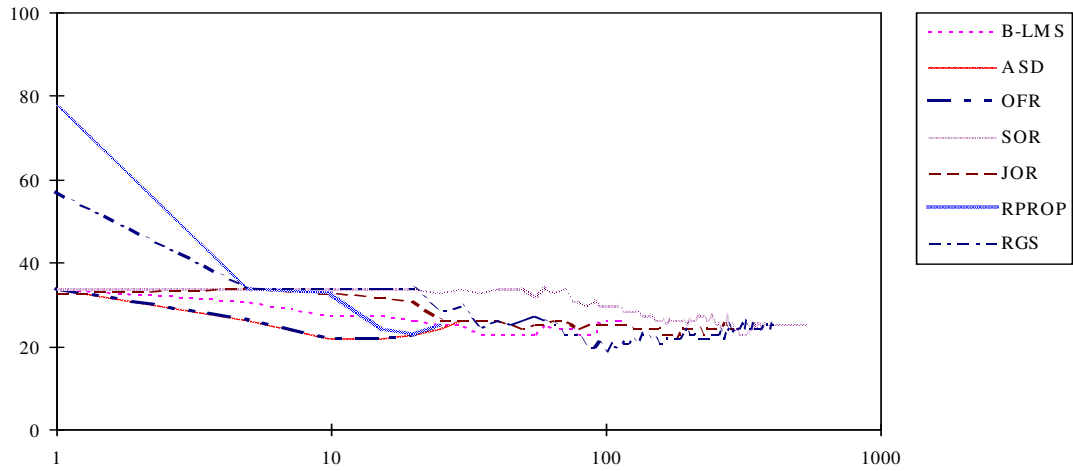
Διάγραμμα 7.16α. Μεταβολή Test Set Classification Error ως προς τις Επαναλήψεις για το Πρόβλημα HEARTC1



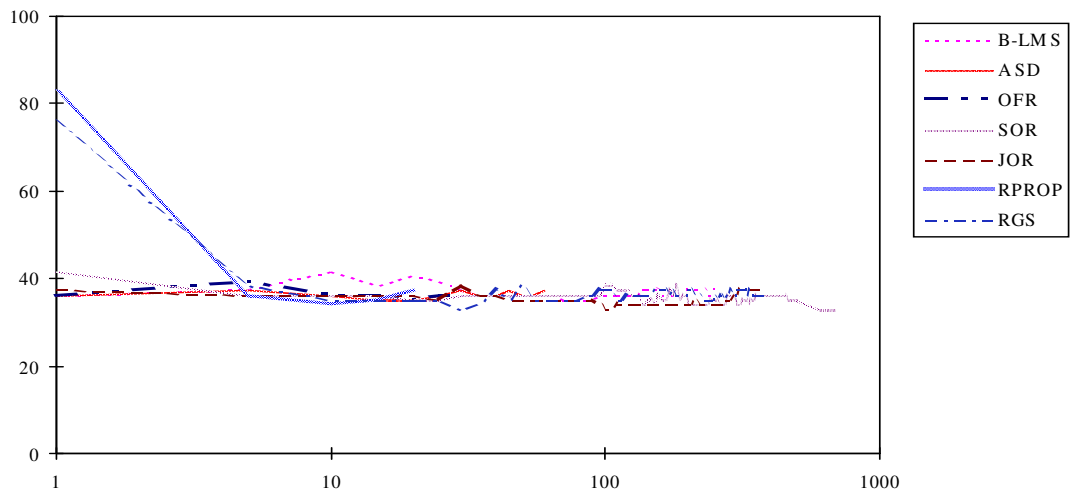
Διάγραμμα 7.16β. Μεταβολή Test Set Classification Error ως προς τις Επαναλήψεις για το Πρόβλημα HEARTC2



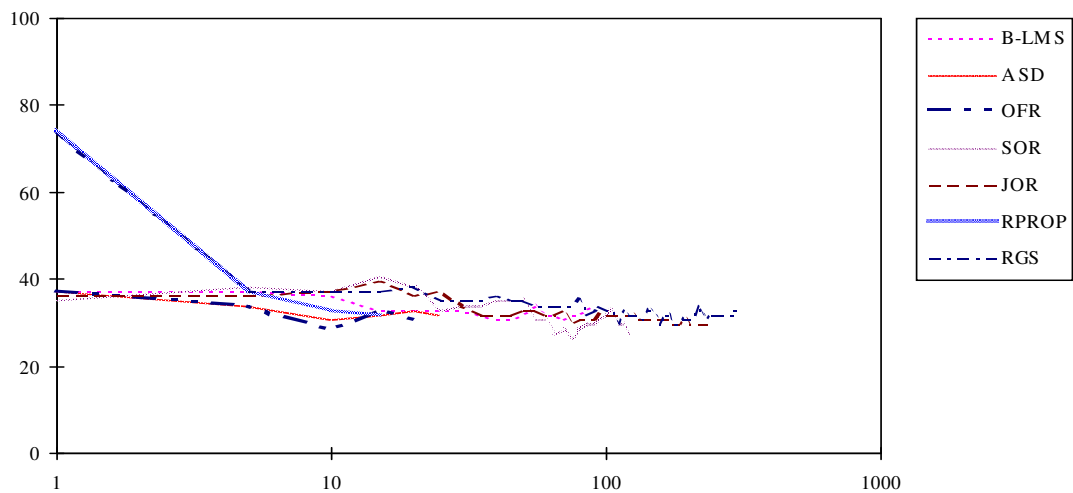
Διάγραμμα 7.16γ. Μεταβολή Test Set Classification Error ως προς τις Επαναλήψεις για το Πρόβλημα HEARTC3



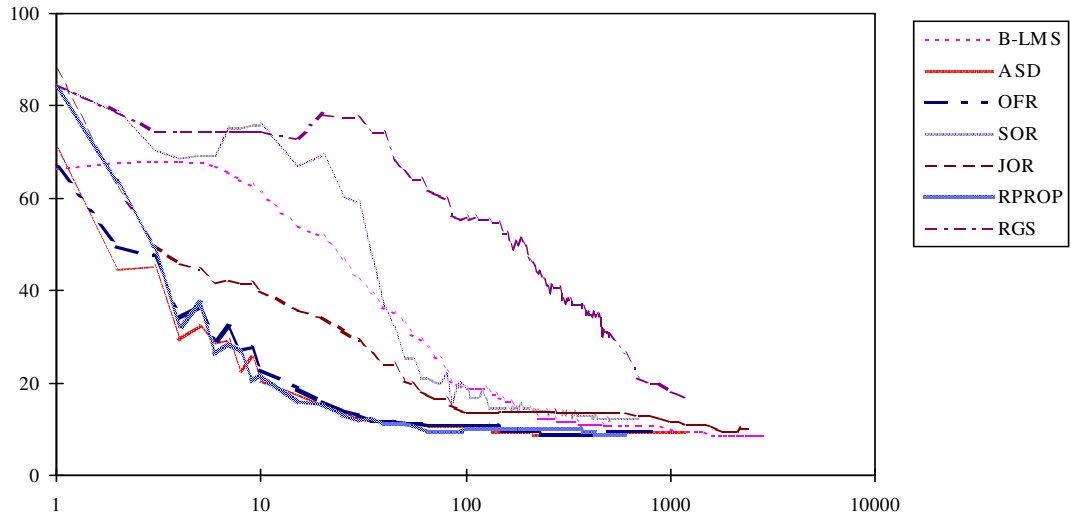
Διάγραμμα 7.17α. Μεταβολή Test Set Classification Error ως προς τις Επαναλήψεις για το Πρόβλημα HORSE1



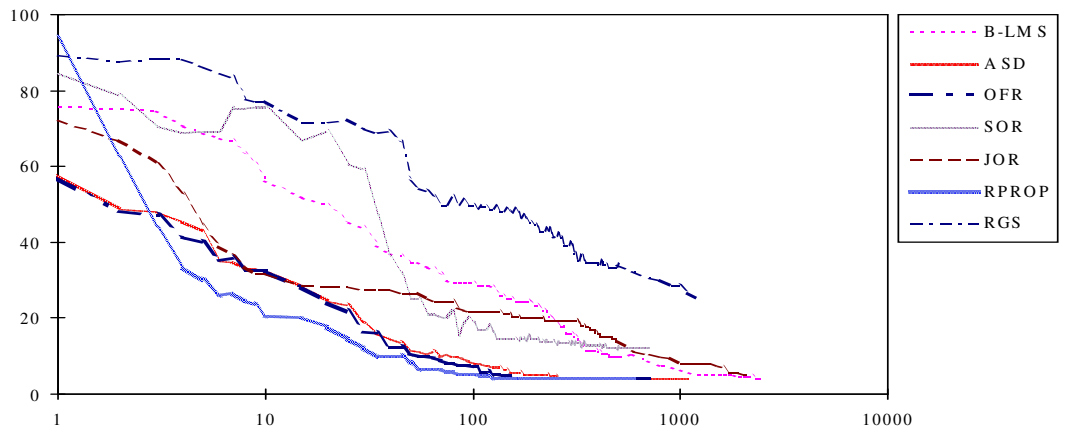
Διάγραμμα 7.17β. Μεταβολή Test Set Classification Error ως προς τις Επαναλήψεις για το Πρόβλημα HORSE2



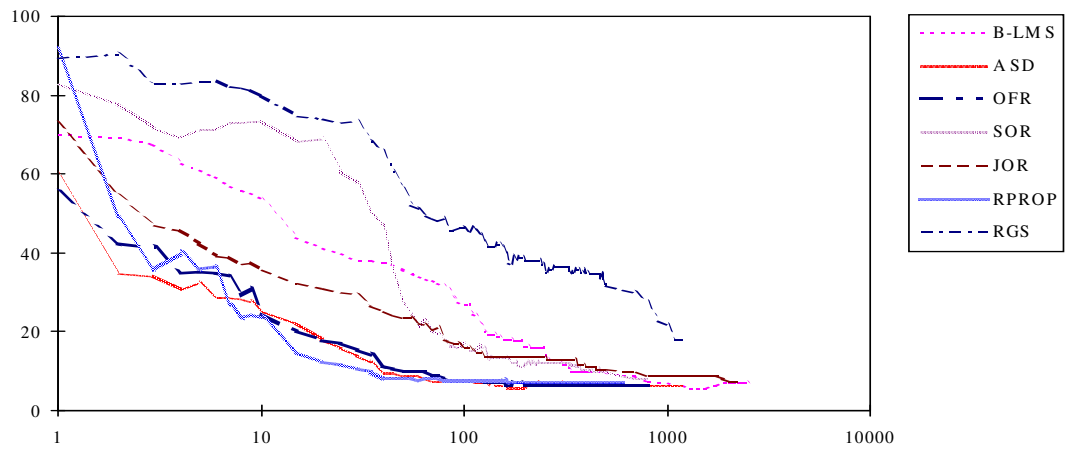
Διάγραμμα 7.17γ. Μεταβολή Test Set Classification Error ως προς τις Επαναλήψεις για το Πρόβλημα HORSE3



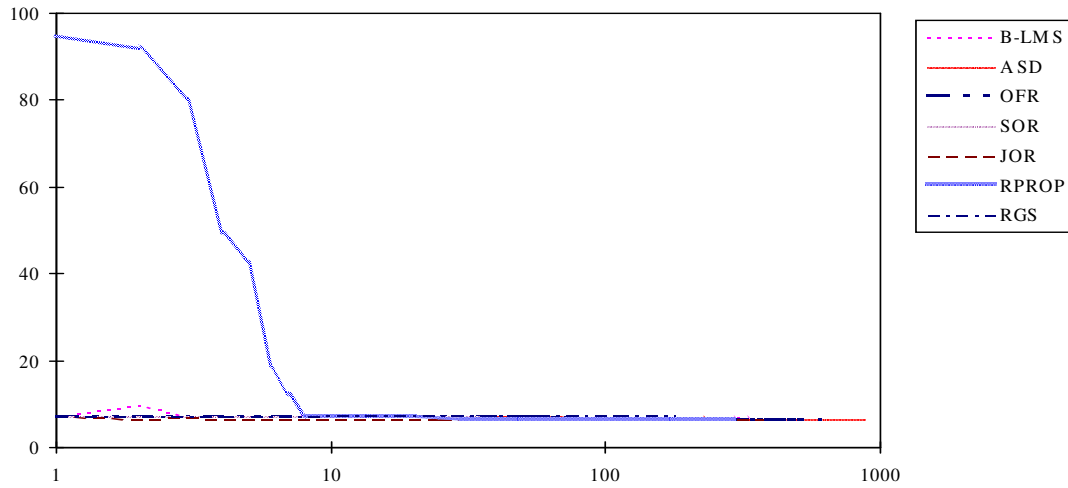
Διάγραμμα 7.18a. Μεταβολή Test Set Classification Error ως προς τις Επαναλήψεις για το Πρόβλημα SOYBEAN1



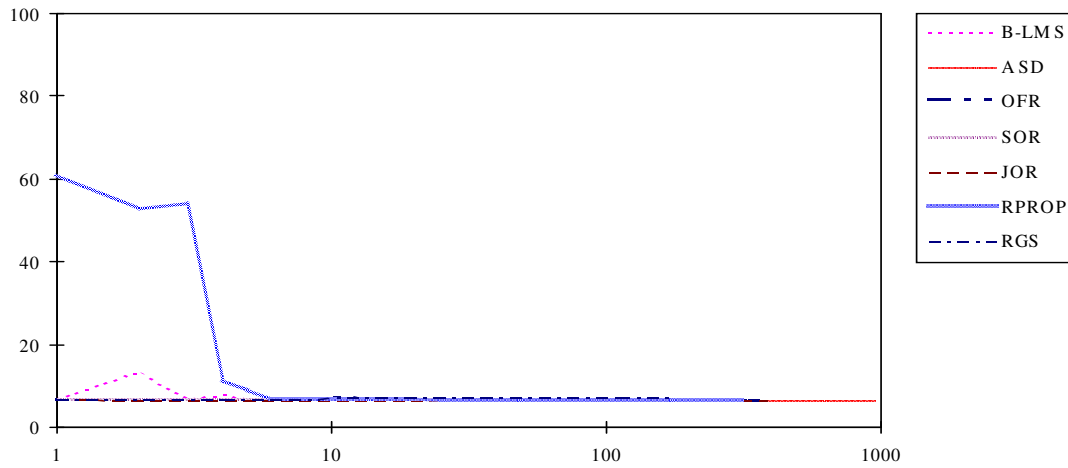
Διάγραμμα 7.18b. Μεταβολή Test Set Classification Error ως προς τις Επαναλήψεις για το Πρόβλημα SOYBEAN2



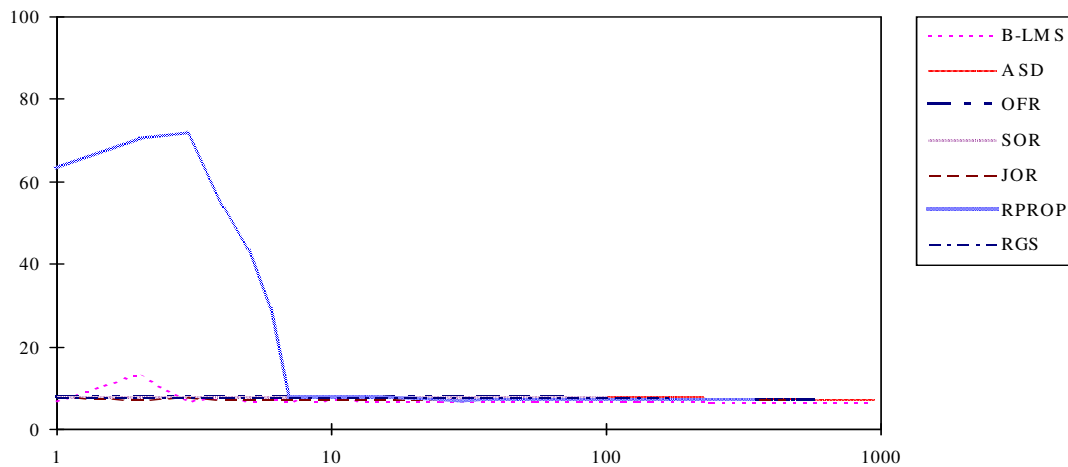
Διάγραμμα 7.18c. Μεταβολή Test Set Classification Error ως προς τις Επαναλήψεις για το Πρόβλημα SOYBEAN3



Διάγραμμα 7.19α. Μεταβολή Test Set Classification Error ως προς τις Επαναλήψεις για το Πρόβλημα THYROID1



Διάγραμμα 7.19β. Μεταβολή Test Set Classification Error ως προς τις Επαναλήψεις για το Πρόβλημα THYROID2

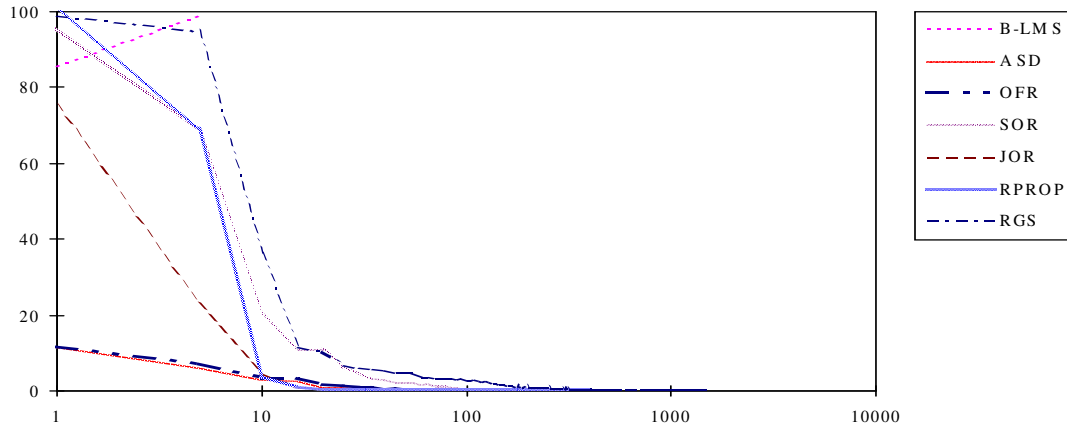


Διάγραμμα 7.19γ. Μεταβολή Test Set Classification Error ως προς τις Επαναλήψεις για το Πρόβλημα THYROID3

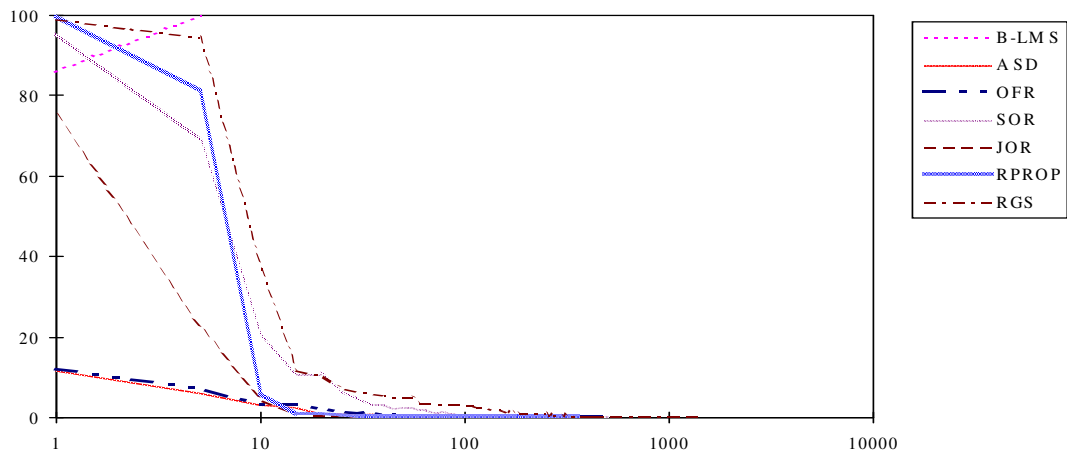
7.3.4.5 Παρατηρήσεις στα Διαγράμματα 7.10-7.19

1. Σε όλα τα προβλήματα (εκτός του SOYBEAN), οι μέθοδοι OFR και ASD, οι οποίες εμφανίζουν μια παρόμοια συμπεριφορά, απαιτούν το μικρότερο αριθμό εποχών και σχετικών εποχών για τη σύγκλισή τους και η μείωση του σφάλματος ταξινόμησης των προτύπων ανάκλησης είναι ραγδαία, η τελική τιμή του οποίου είναι σχεδόν ίδια για όλες τις μεθόδους, στα προβλήματα CANCER, CARD1, CARD2, DIABETES1, DIABETES2, GENE, HEART, HEARTC, HORSE και THYROID.
2. Σε όλα σχεδόν τα προβλήματα, καλύτεροι μέθοδοι αποδεικνύονται οι ASD και OFR και χειρότερη η RGS.
3. Στα προβλήματα THYROID, το σφάλμα ταξινόμησης των προτύπων ανάκλησης για τη μέθοδο RPROP ξεκινάει από μεγάλη τιμή, σε αντίθεση με τις υπόλοιπες μεθόδους, και μικραίνει μετά από κάποιον αριθμό επαναλήψεων, οπότε και παραμένει σταθερό μετά τη δέκατη επανάληψη για όλες τις μεθόδους.
4. Οι χειρότερες μέθοδοι ως προς τις εποχές και τις σχετικές εποχές που απαιτούνται για σύγκλιση ήταν οι μέθοδοι RGS και SOR, στις περιπτώσεις όπου ο συντελεστής εκπαίδευσης του καλύτερου πειράματος, ω_{best} έχει πολύ μικρή τιμή.

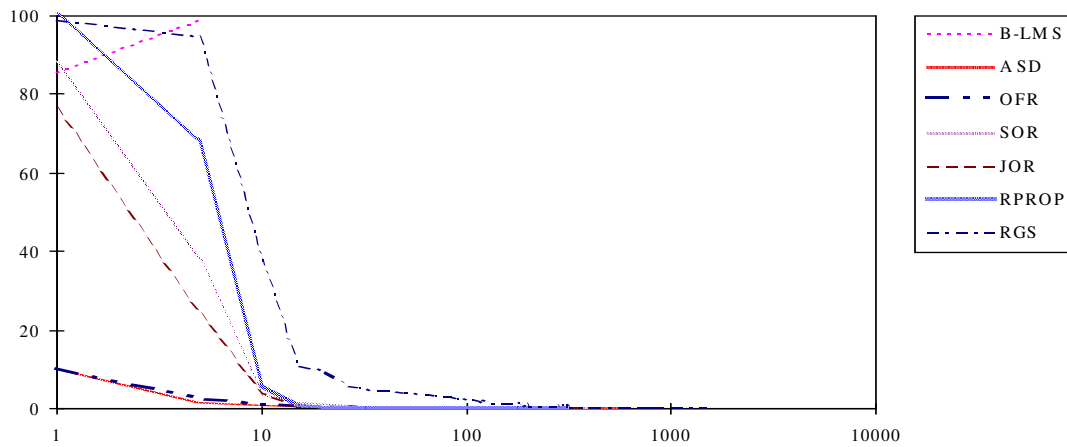
Στα διαγράμματα (7.20-7.23) που ακολουθούν παρουσιάζονται τα αποτελέσματα του καλύτερου πειράματος του κάθε προβλήματος προσαρμογής, για τις μεθόδους RPROP, BATCH-LMS, ASD, SOR και JOR, άσον αφορά τη μεταβολή του σφάλματος ταξινόμησης των προτύπων ανάκλησης σε σχέση με τις επαναλήψεις. Παραλήφθηκαν οι μέθοδοι PR, FR και LMS, των οποίων η συμπεριφορά είναι παρόμοια με τη BATCH-LMS.



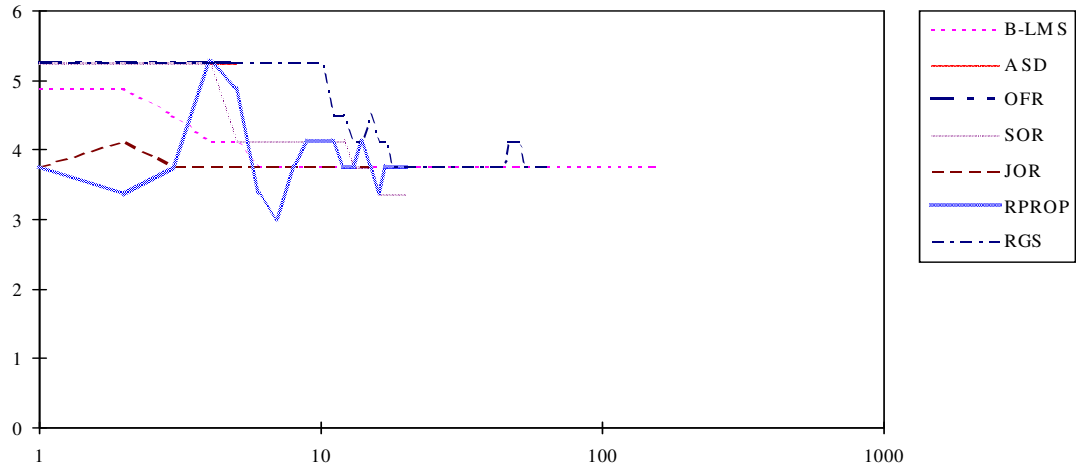
Διάγραμμα 7.20a Μεταβολή Test Set Classification Error ως προς τις Επαναλήψεις για το Πρόβλημα BUILDING1



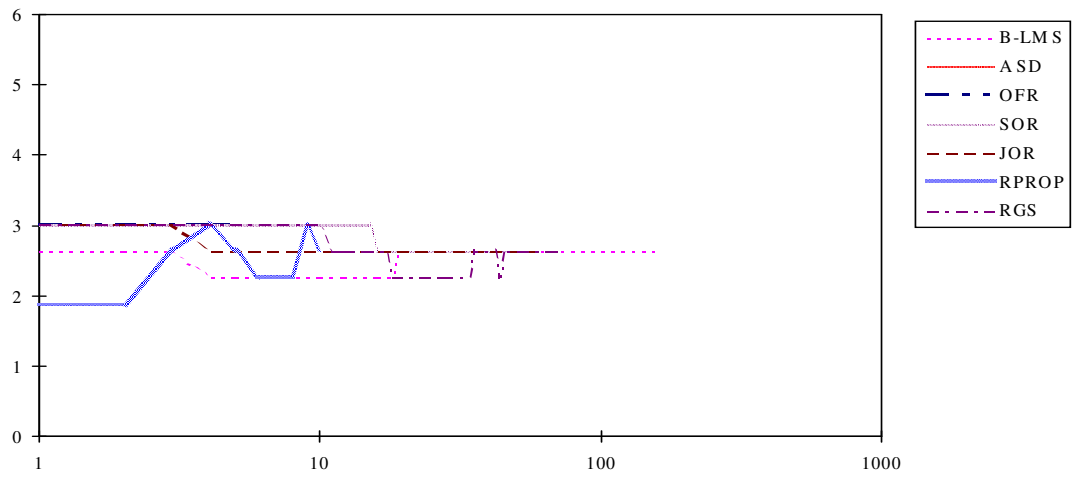
Διάγραμμα 7.20b Μεταβολή Test Set Classification Error ως προς τις Επαναλήψεις για το Πρόβλημα BUILDING2



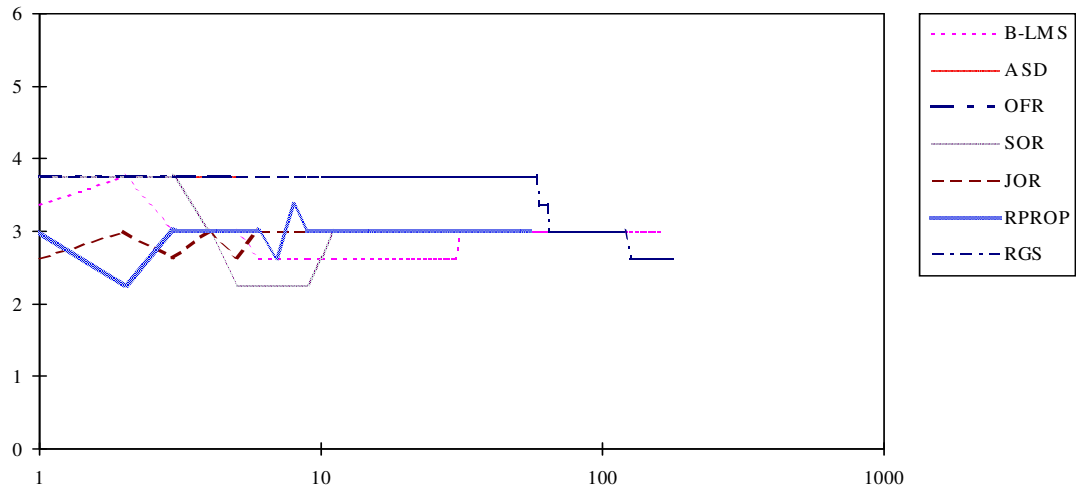
Διάγραμμα 7.20c Μεταβολή Test Set Classification Error ως προς τις Επαναλήψεις για το Πρόβλημα BUILDING3



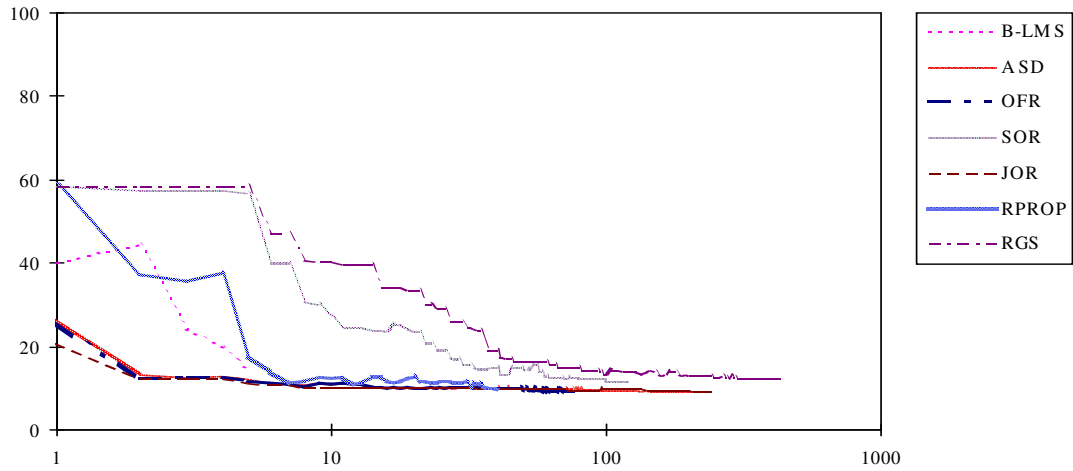
Διάγραμμα 7.21a Μεταβολή Test Set Classification Error ως προς τις Επαναλήψεις για το Πρόβλημα FLARE1



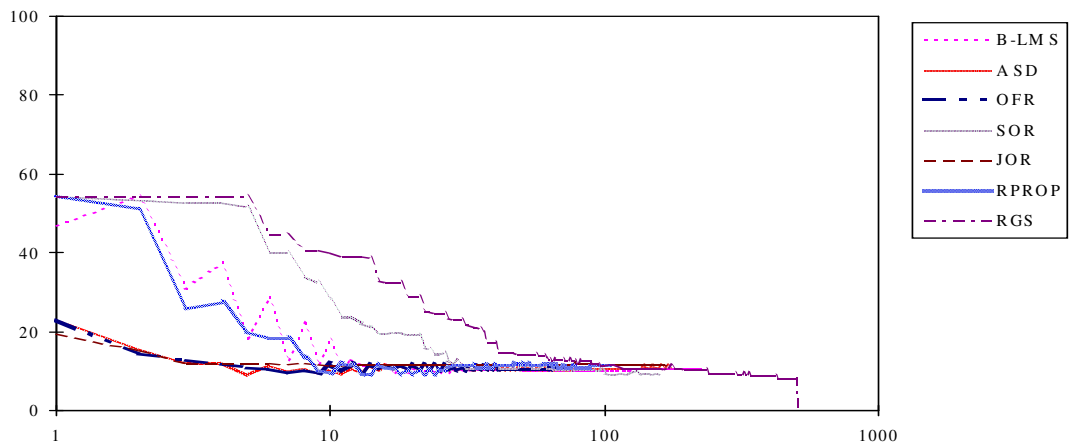
Διάγραμμα 7.21b Μεταβολή Test Set Classification Error ως προς τις Επαναλήψεις για το Πρόβλημα FLARE2



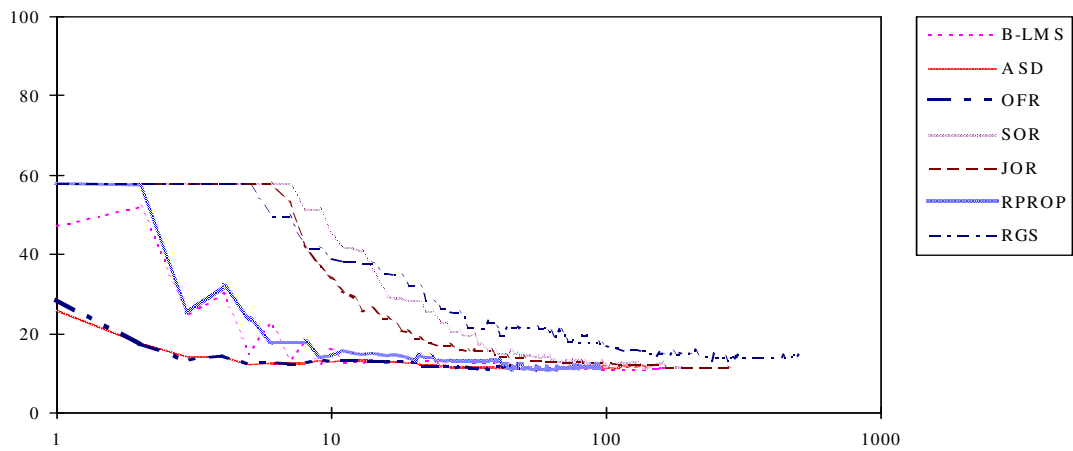
Διάγραμμα 7.21c Μεταβολή Test Set Classification Error ως προς τις Επαναλήψεις για το Πρόβλημα FLARE3



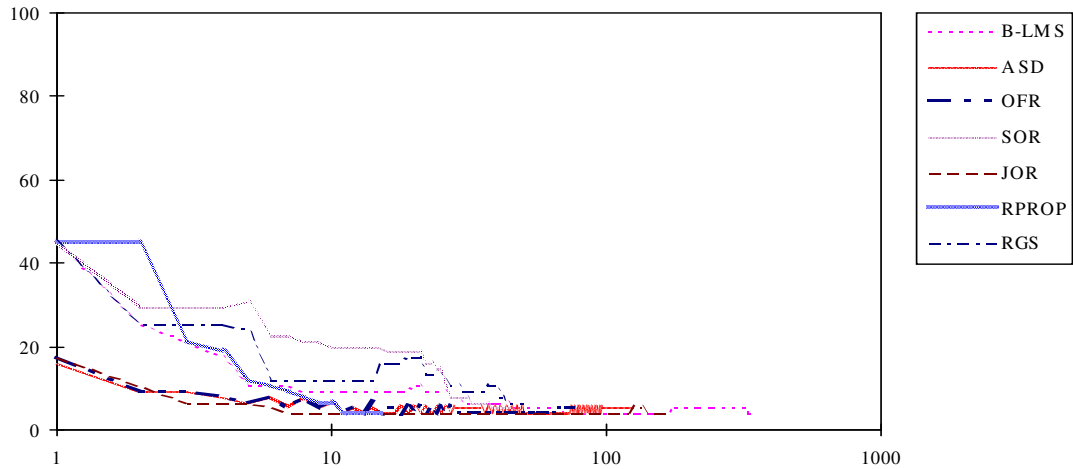
Διάγραμμα 7.22a Μεταβολή Test Set Classification Error ως προς τις Επαναλήψεις για το Πρόβλημα HEARTA1



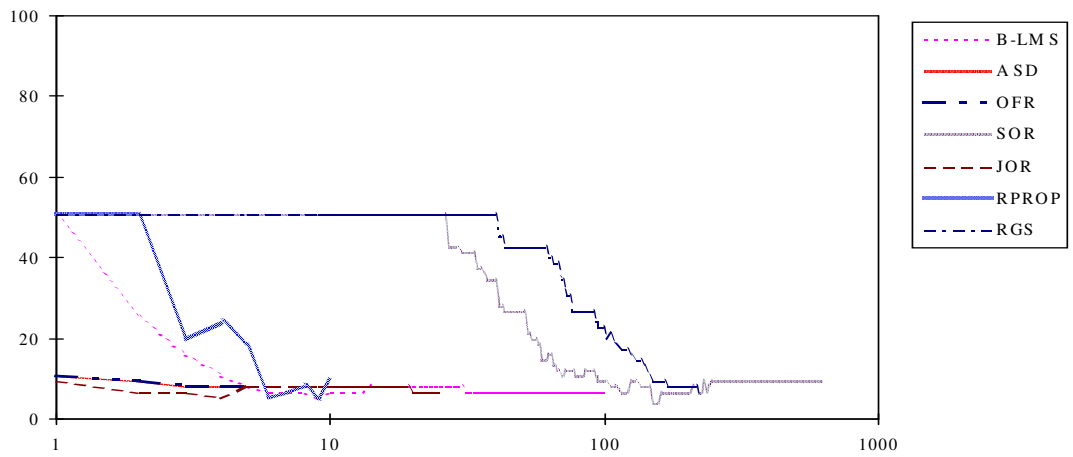
Διάγραμμα 7.22b Μεταβολή Test Set Classification Error ως προς τις Επαναλήψεις για το Πρόβλημα HEARTA2



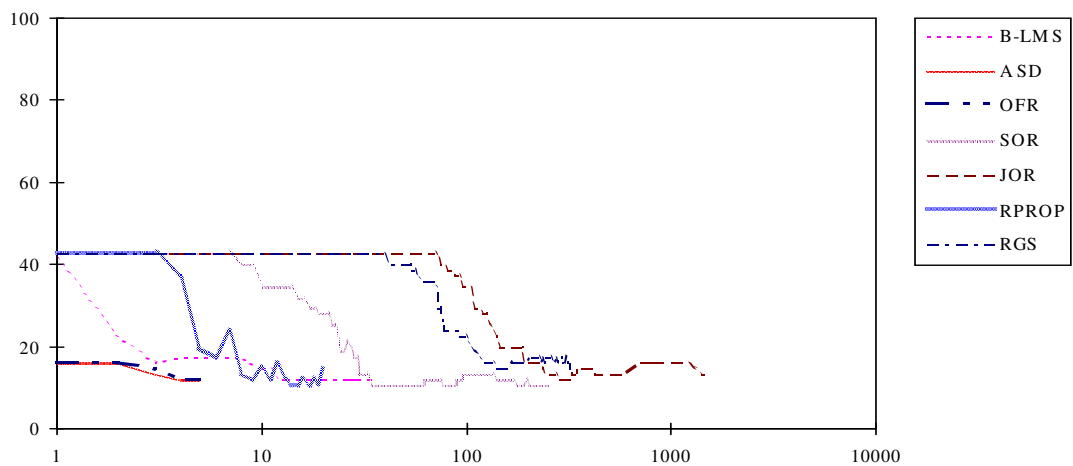
Διάγραμμα 7.22c Μεταβολή Test Set Classification Error ως προς τις Επαναλήψεις για το Πρόβλημα HEARTA3



Διάγραμμα 7.23a Μεταβολή Test Set Classification Error ως προς τις Επαναλήψεις για το Πρόβλημα HEARTAC1



Διάγραμμα 7.23b Μεταβολή Test Set Classification Error ως προς τις Επαναλήψεις για το Πρόβλημα HEARTAC2

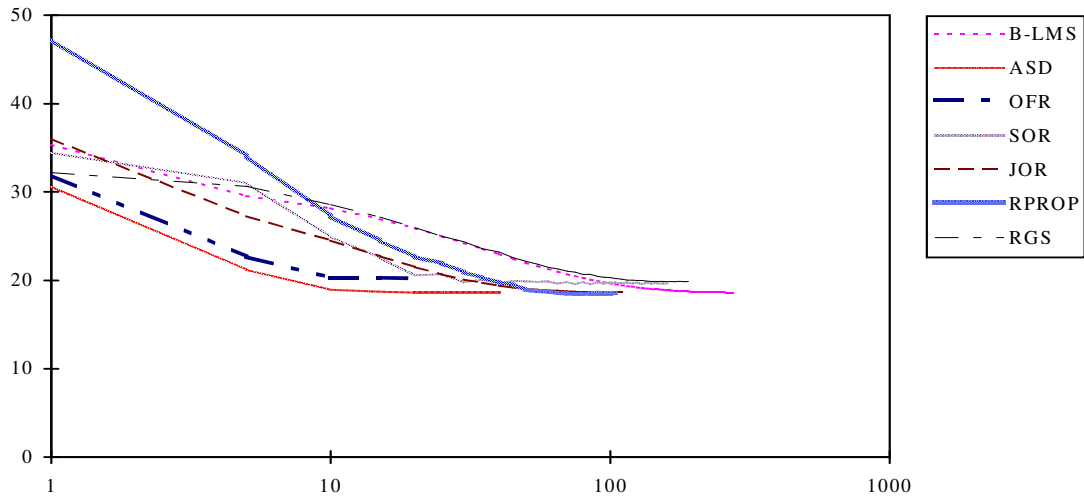


Διάγραμμα 7.23c Μεταβολή Test Set Classification Error ως προς τις Επαναλήψεις για το Πρόβλημα HEARTAC3

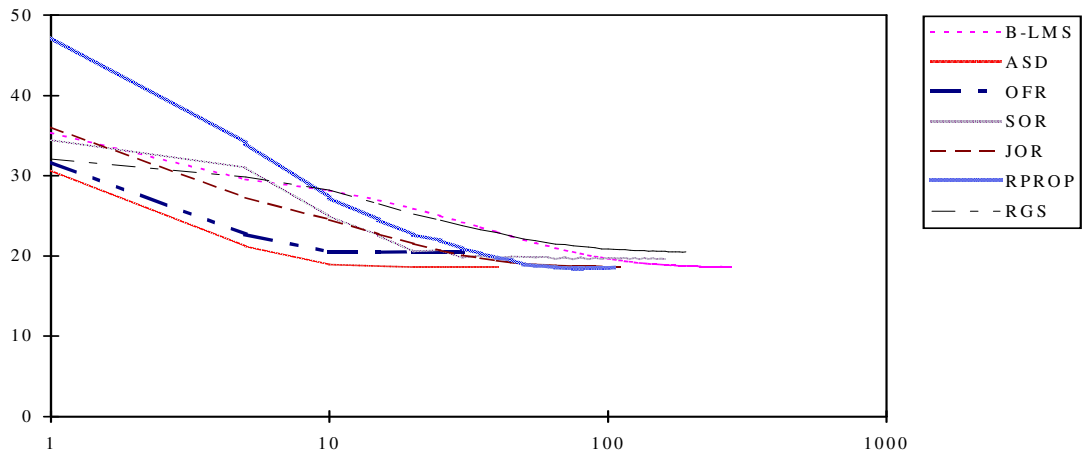
7.3.4.6 Παρατηρήσεις στα Διαγράμματα 7.20-7.23

1. Στα πιο πολλά προβλήματα, οι μέθοδοι OFR και ASD απαιτούν το μικρότερο αριθμό εποχών και σχετικών εποχών για τη σύγκλισή τους και η μείωση του σφάλματος ταξινόμησης των προτύπων ανάκλησης είναι ραγδαία, η τελική τιμή του οποίου είναι σχεδόν ίδια για όλες τις μεθόδους. Μοναδική εξαίρεση αποτελούν τα προβλήματα FLARE, όπου το σφάλμα παίρνει εξ αρχής την ελάχιστη τιμή, η οποία δεν αλλάζει και είναι μεγαλύτερη απ' την ελάχιστη τιμή των υπολοίπων μεθόδων.
2. Στα προβλήματα BUILDING, το σφάλμα ταξινόμησης των προτύπων ανάκλησης για τη μέθοδο BATCH-LMS μεγαλώνει, αντί να μικραίνει και η μέθοδος τερματίζει λόγω απώλειας γενίκευσης.
3. Οι χειρότερες μέθοδοι ως προς τις εποχές και τις σχετικές εποχές που απαιτούνται για σύγκλιση ήταν οι μέθοδοι BATCH-LMS, PR και FR για τα προβλήματα BUILDING, οι μέθοδοι OFR και ASD για τα προβλήματα FLARE και οι μέθοδοι SOR και RGS για τα προβλήματα HEARTA και HEARTC, στις περιπτώσεις όπου ο συντελεστής εκπαίδευσης του καλύτερου πειράματος, ω_{best} έχει πολύ μικρή τιμή.

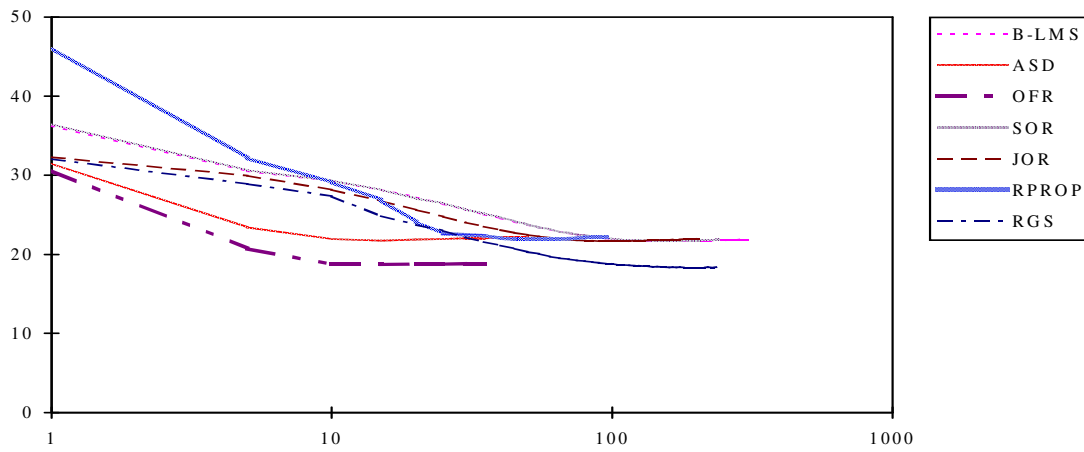
Στα διαγράμματα (7.24-7.33) που ακολουθούν παρουσιάζονται τα αποτελέσματα του καλύτερου πειράματος του κάθε προβλήματος ταξινόμησης, για τις μεθόδους RPROP, BATCH-LMS, ASD, SOR και JOR, άσον αφορά τη μεταβολή του ποσοστιαίου σφάλματος των προτύπων ελέγχου εγκυρότητας των προτύπων ανάκλησης ως προς τον αριθμό των κύκλων εκπαίδευσης. Παραλήφθηκαν οι μέθοδοι PR, FR και LMS, των οποίων η συμπεριφορά είναι παρόμοια με τη BATCH-LMS.



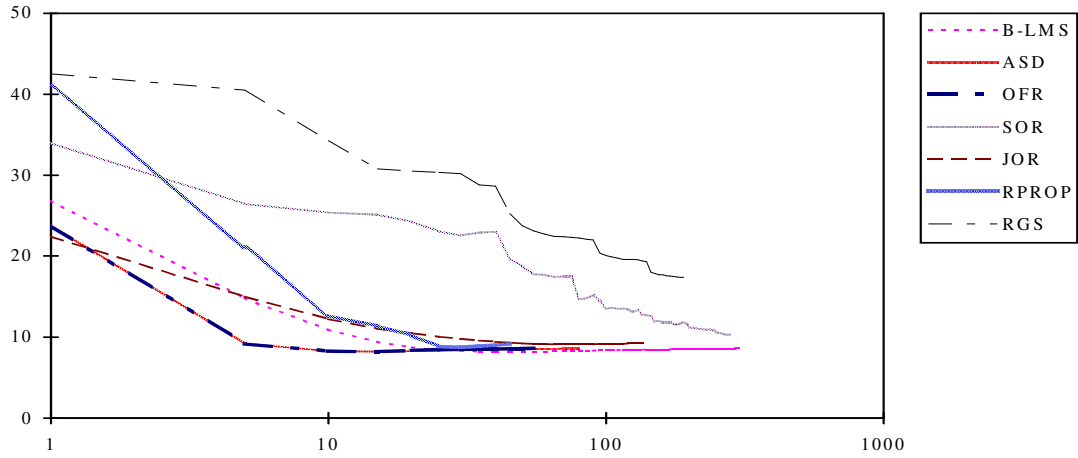
Διάγραμμα 7.24a Μεταβολή Validation Set Error Percentage ως προς τις Επαναλήψεις για το Πρόβλημα CANCER1



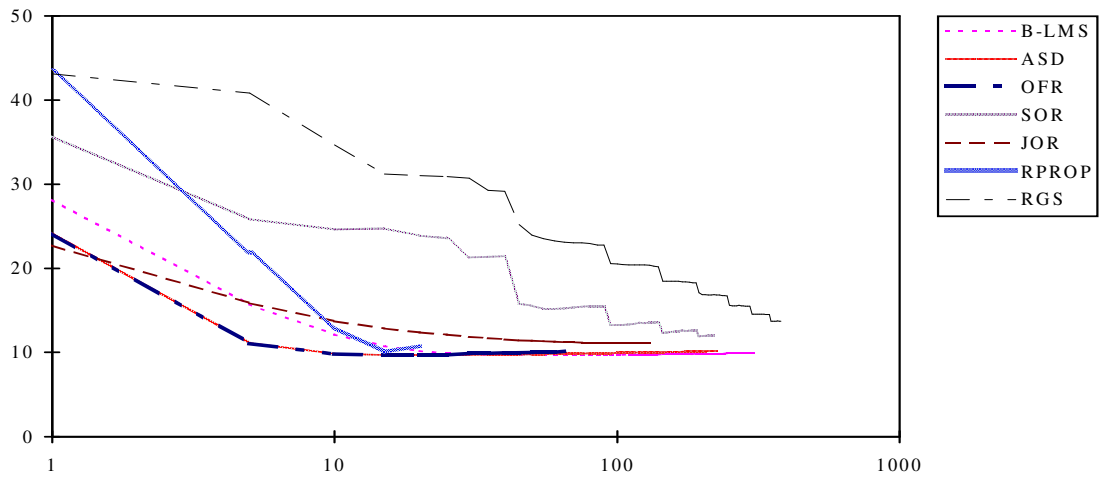
Διάγραμμα 7.24b Μεταβολή Validation Set Error Percentage ως προς τις Επαναλήψεις για το Πρόβλημα CANCER2



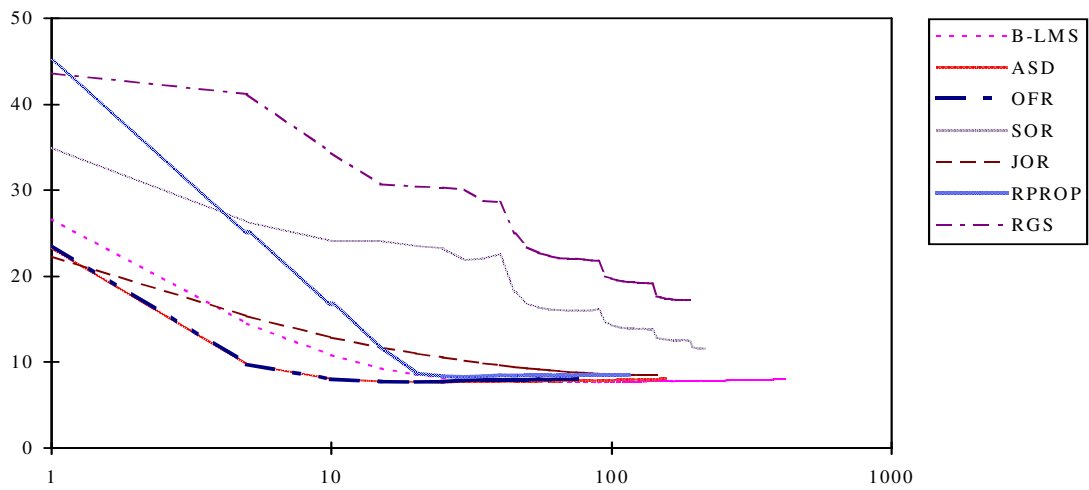
Διάγραμμα 7.24c Μεταβολή Validation Set Error Percentage ως προς τις Επαναλήψεις για το Πρόβλημα CANCER3



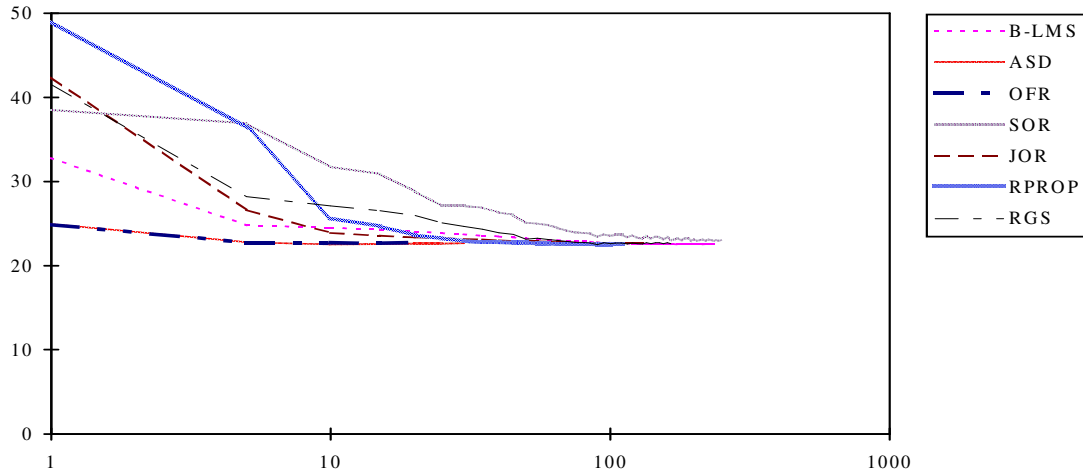
Διάγραμμα 7.25a Μεταβολή Validation Set Error Percentage ως προς τις Επαναλήψεις για το Πρόβλημα CARD1



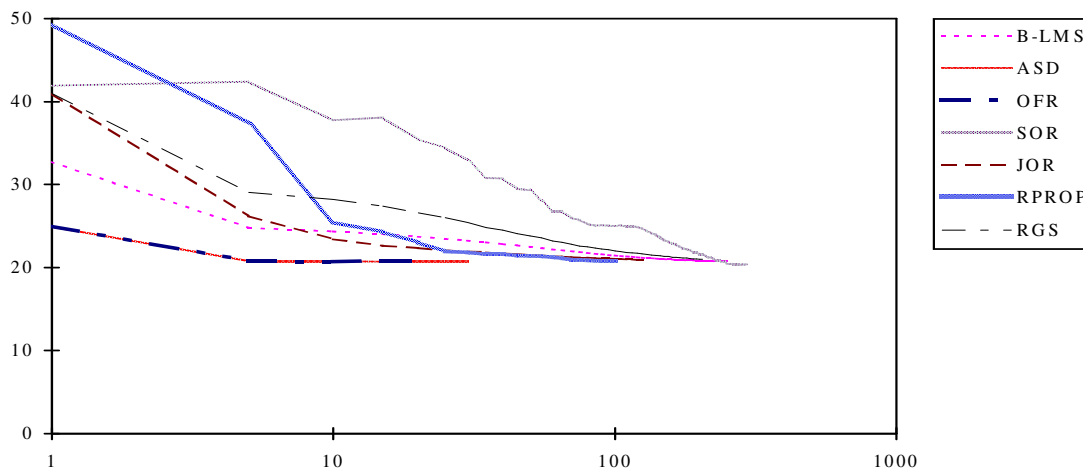
Διάγραμμα 7.25b Μεταβολή Validation Set Error Percentage ως προς τις Επαναλήψεις για το Πρόβλημα CARD2



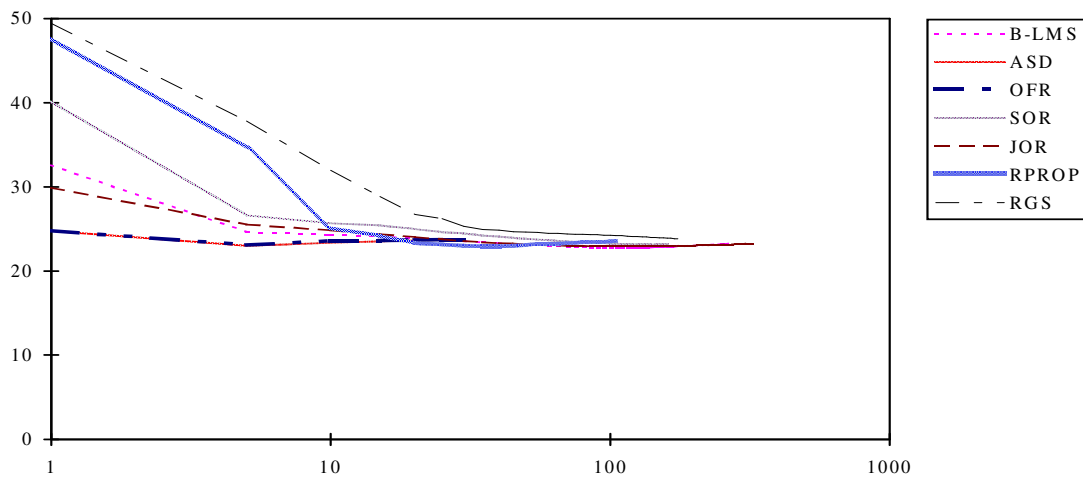
Διάγραμμα 7.25c Μεταβολή Validation Set Error Percentage ως προς τις Επαναλήψεις για το Πρόβλημα CARD3



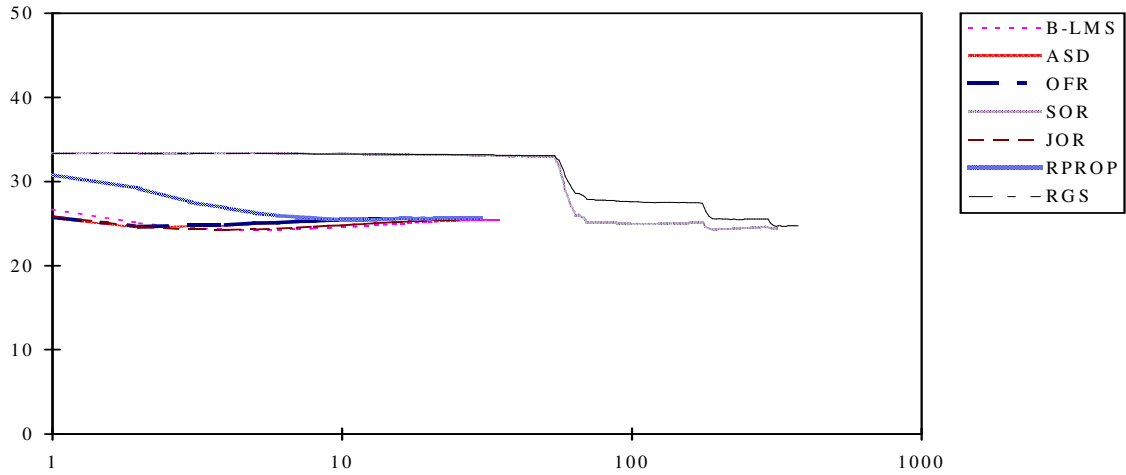
Διάγραμμα 7.26a Μεταβολή Validation Set Error Percentage ως προς τις Επαναλήψεις για το Πρόβλημα DIABETES1



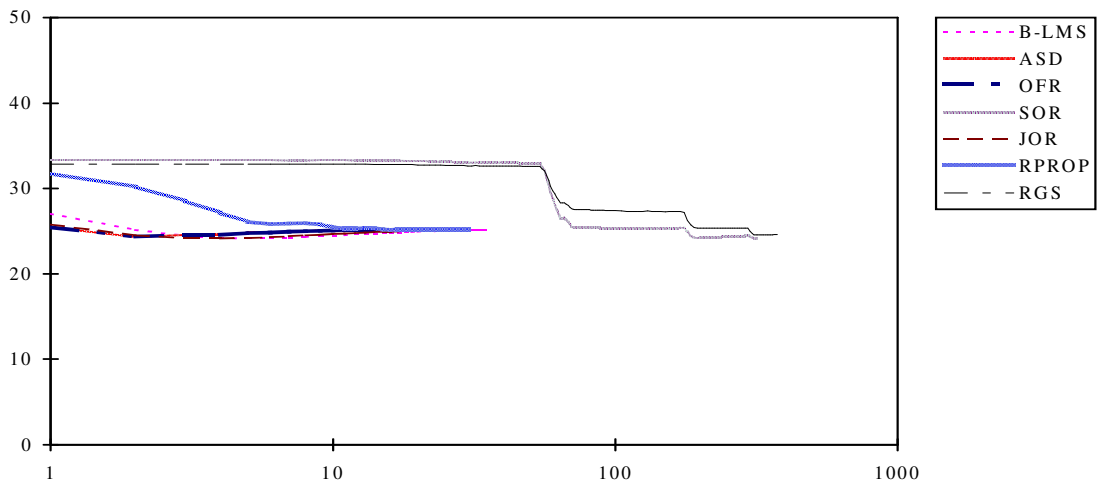
Διάγραμμα 7.26b Μεταβολή Validation Set Error Percentage ως προς τις Επαναλήψεις για το Πρόβλημα DIABETES2



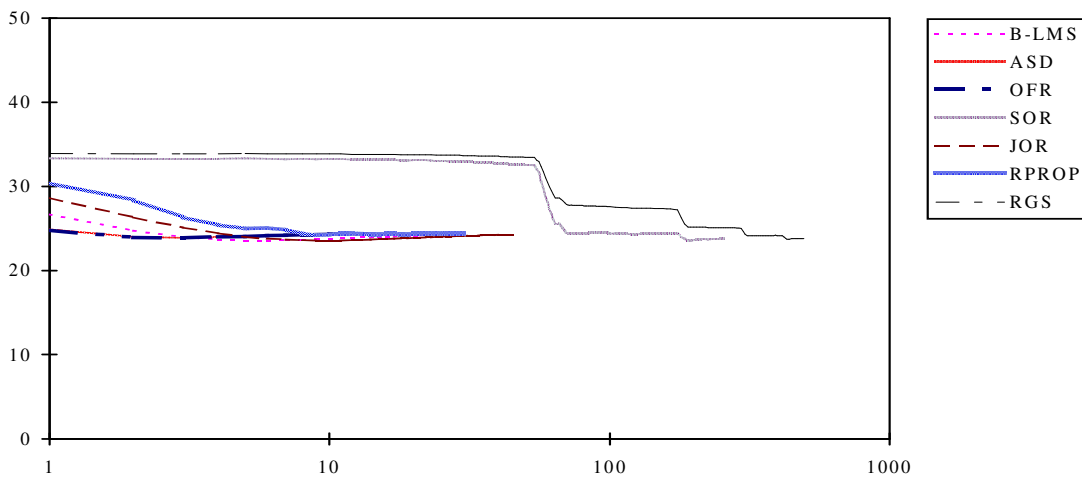
Διάγραμμα 7.26c Μεταβολή Validation Set Error Percentage ως προς τις Επαναλήψεις για το Πρόβλημα DIABETES3



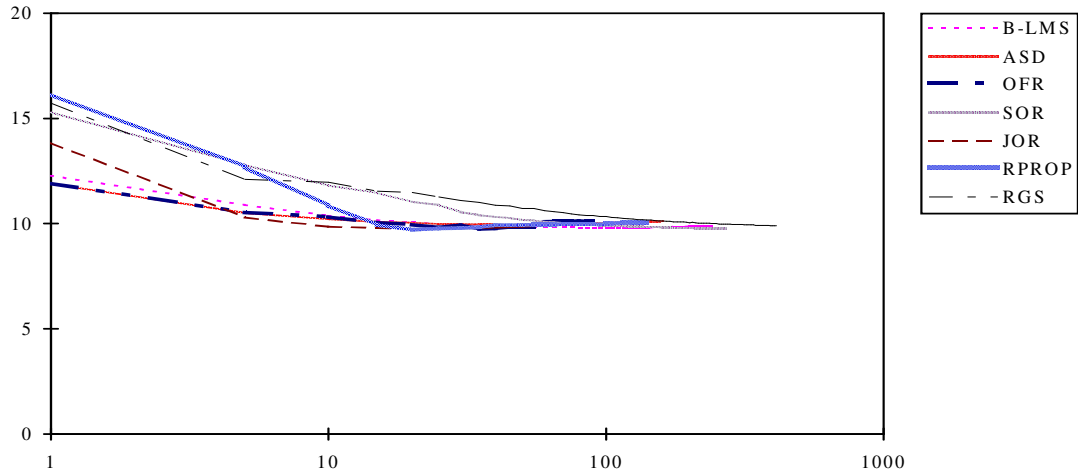
Διάγραμμα 7.27a Μεταβολή Validation Set Error Percentage ως προς τις Επαναλήψεις για το Πρόβλημα GENE1



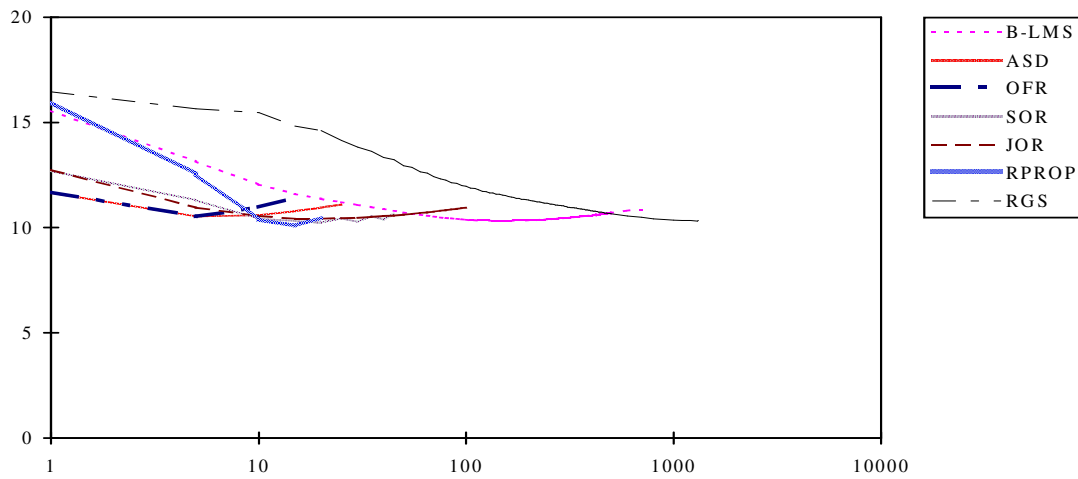
Διάγραμμα 7.27b Μεταβολή Validation Set Error Percentage ως προς τις Επαναλήψεις για το Πρόβλημα GENE2



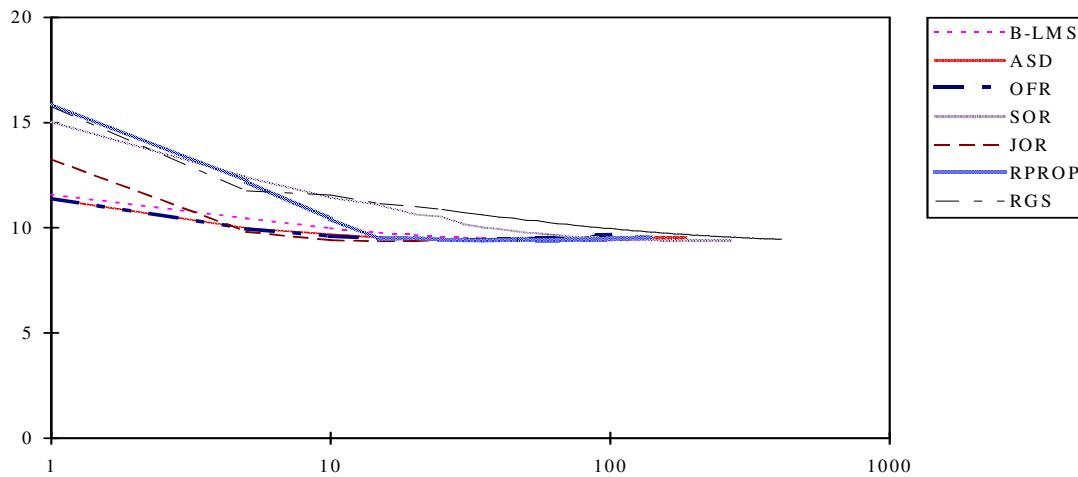
Διάγραμμα 7.27c Μεταβολή Validation Set Error Percentage ως προς τις Επαναλήψεις για το Πρόβλημα GENE3



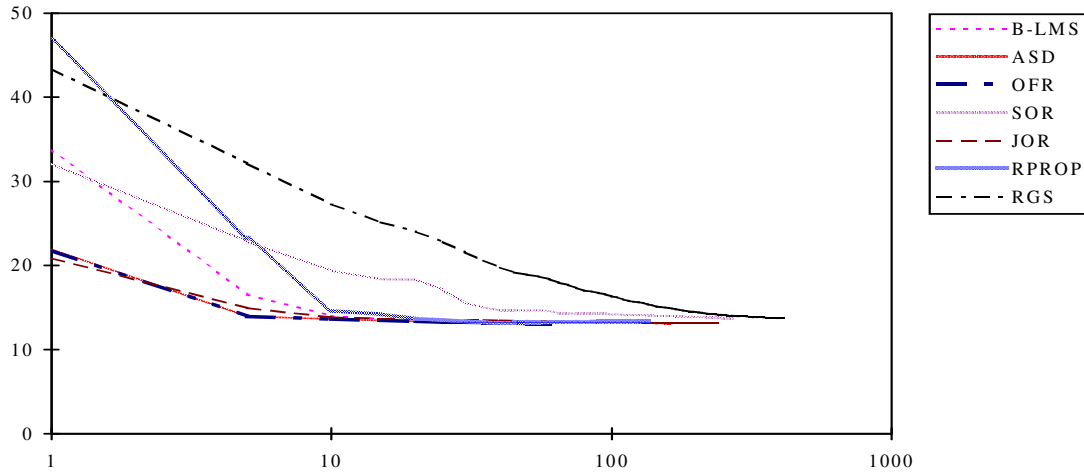
Διάγραμμα 7.28a Μεταβολή Validation Set Error Percentage ως προς τις Επαναλήψεις για το Πρόβλημα GLASS1



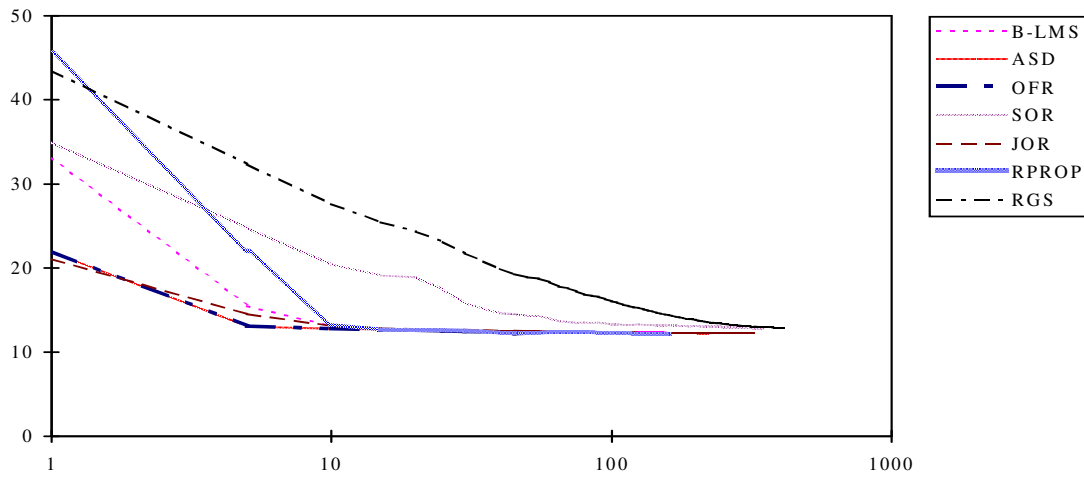
Διάγραμμα 7.28b Μεταβολή Validation Set Error Percentage ως προς τις Επαναλήψεις για το Πρόβλημα GLASS2



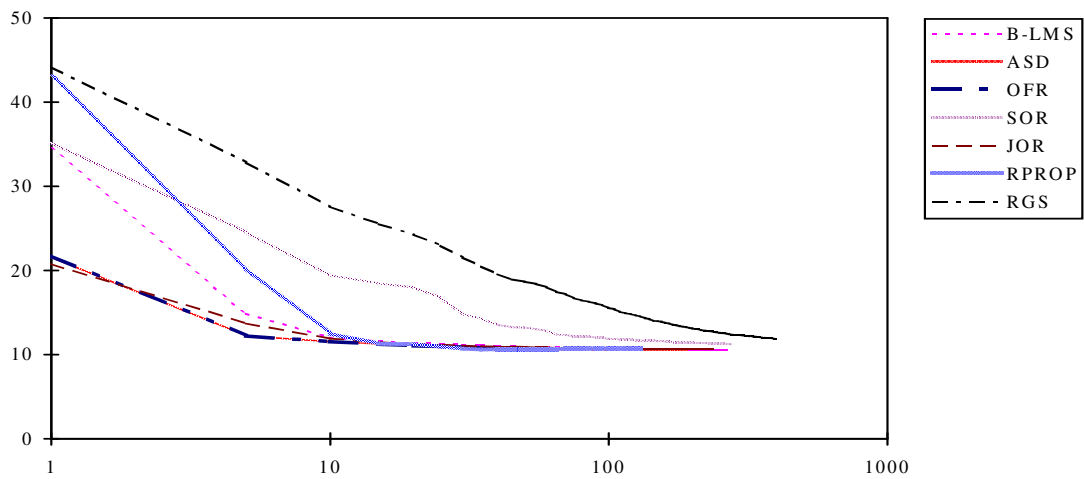
Διάγραμμα 7.28c Μεταβολή Validation Set Error Percentage ως προς τις Επαναλήψεις για το Πρόβλημα GLASS3



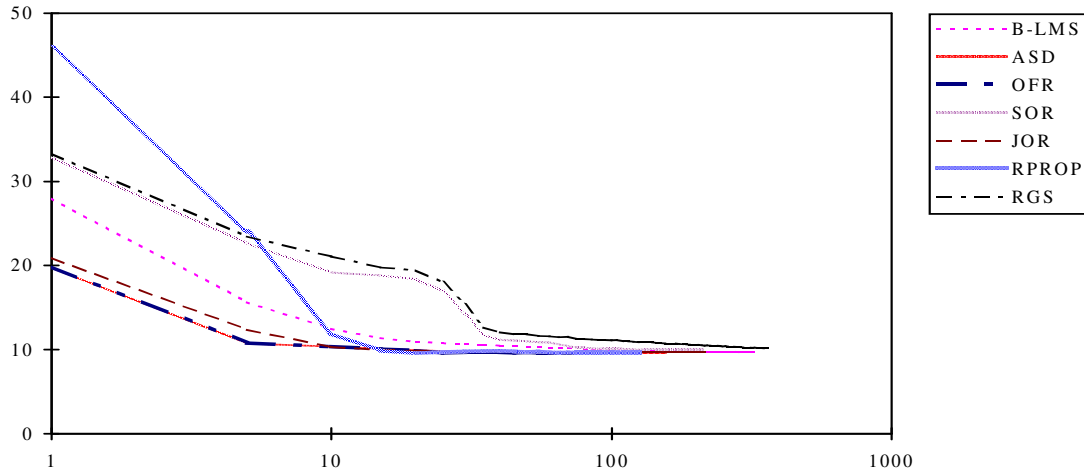
Διάγραμμα 7.29a Μεταβολή Validation Set Error Percentage ως προς τις Επαναλήψεις για το Πρόβλημα HEART1



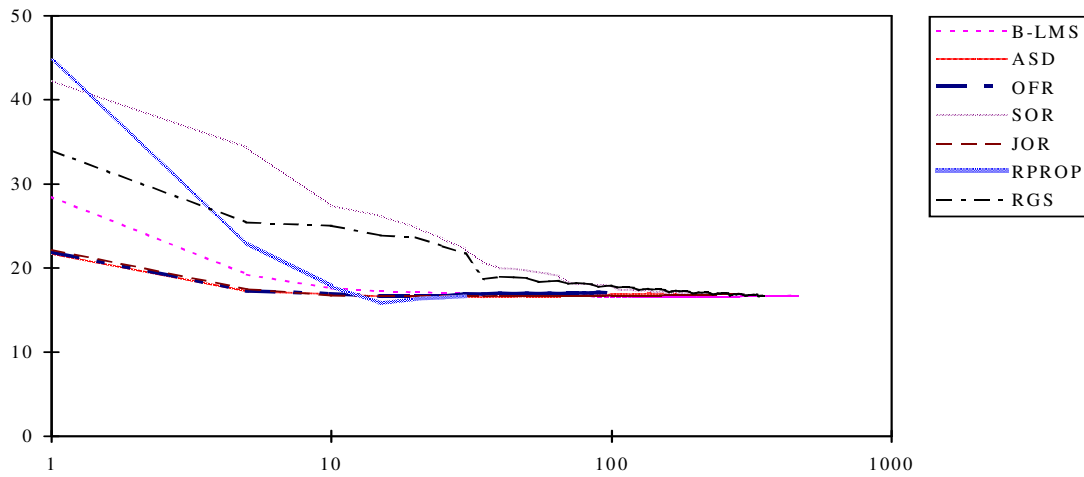
Διάγραμμα 7.29b Μεταβολή Validation Set Error Percentage ως προς τις Επαναλήψεις για το Πρόβλημα HEART2



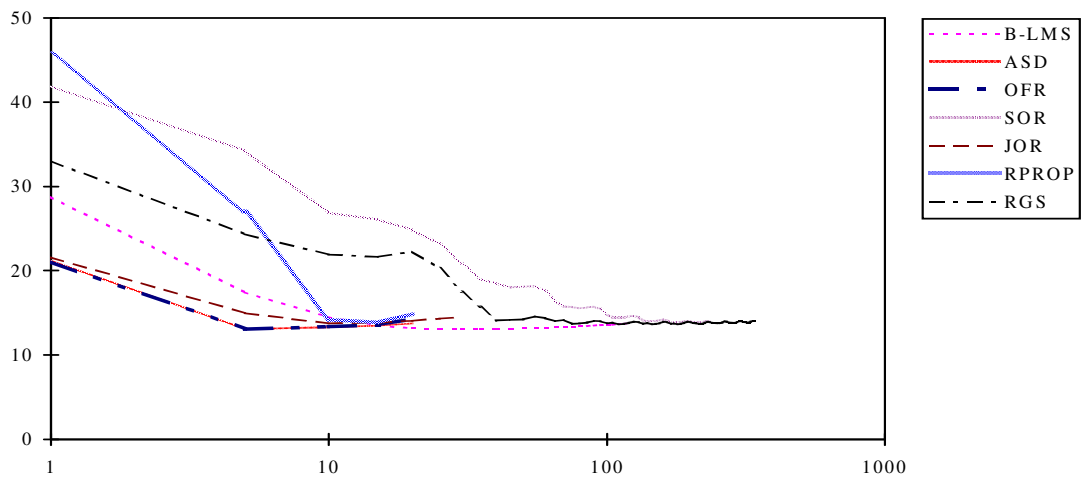
Διάγραμμα 7.29c Μεταβολή Validation Set Error Percentage ως προς τις Επαναλήψεις για το Πρόβλημα HEART3



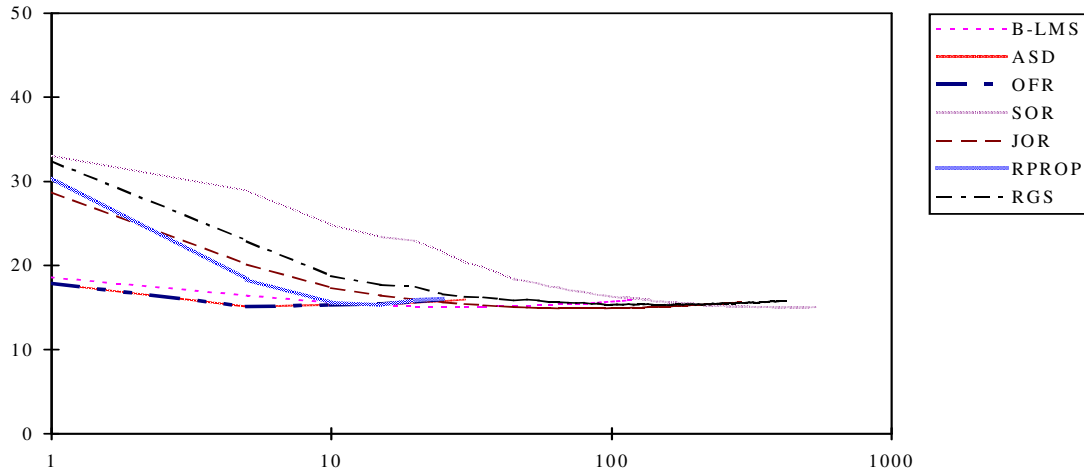
Διάγραμμα 7.30a Μεταβολή Validation Set Error Percentage ως προς τις Επαναλήψεις για το Πρόβλημα HEARTC1



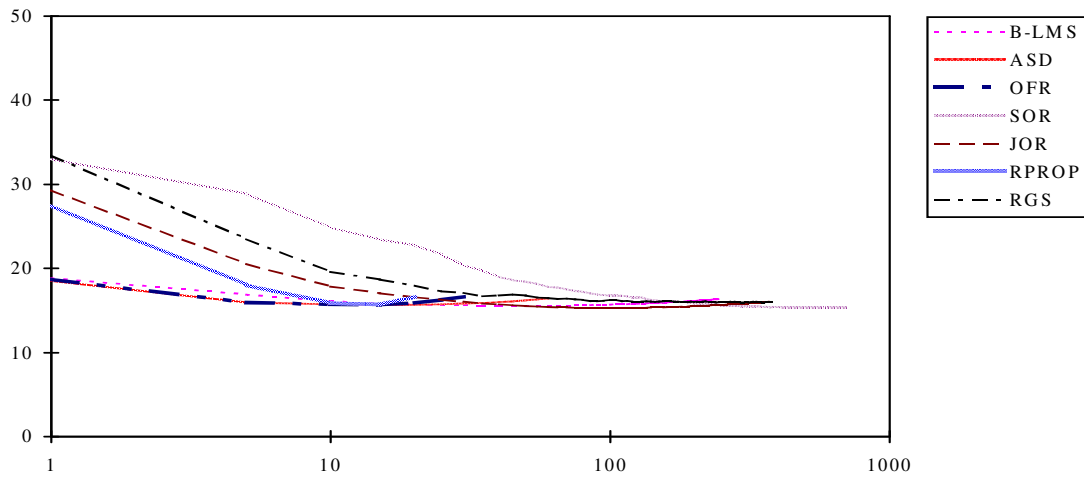
Διάγραμμα 7.30b Μεταβολή Validation Set Error Percentage ως προς τις Επαναλήψεις για το Πρόβλημα HEARTC2



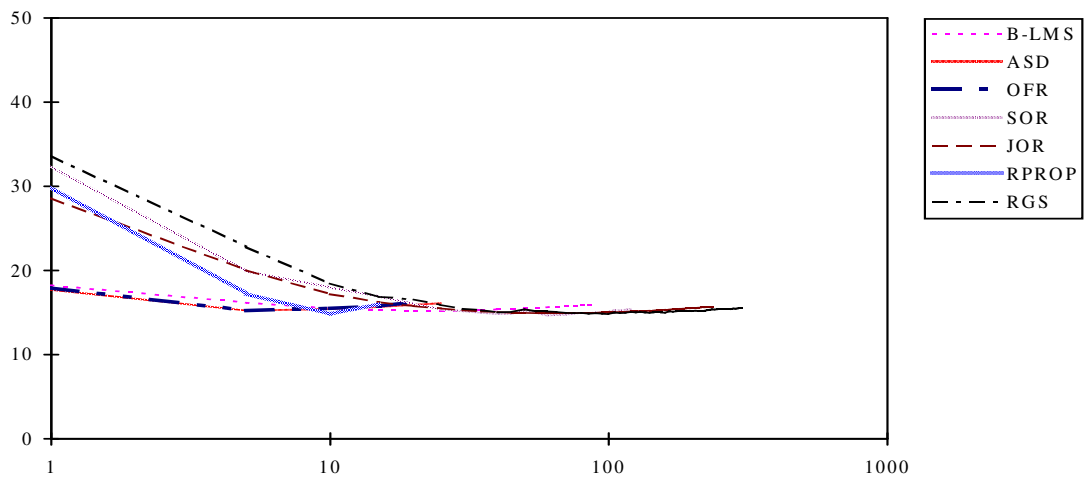
Διάγραμμα 7.30c Μεταβολή Validation Set Error Percentage ως προς τις Επαναλήψεις για το Πρόβλημα HEARTC3



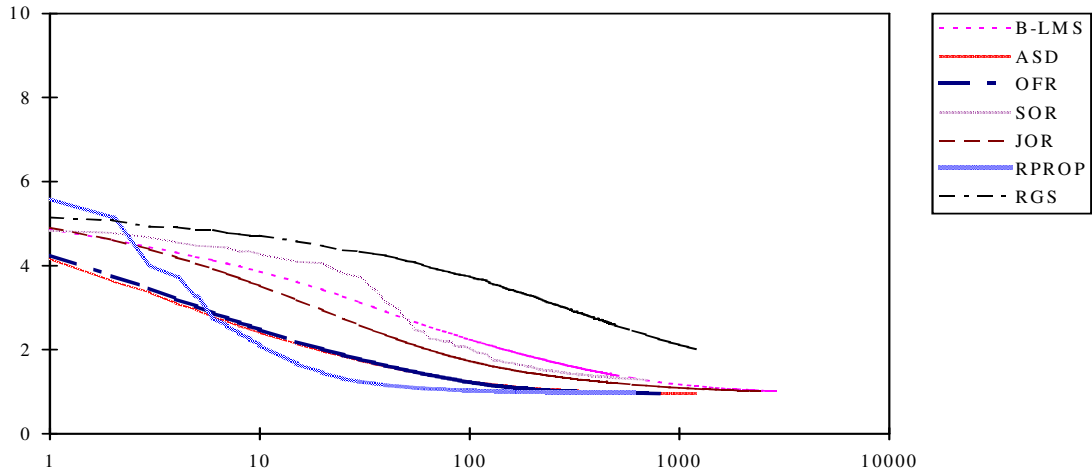
Διάγραμμα 7.31a Μεταβολή Validation Set Error Percentage ως προς τις Επαναλήψεις για το Πρόβλημα HORSE1



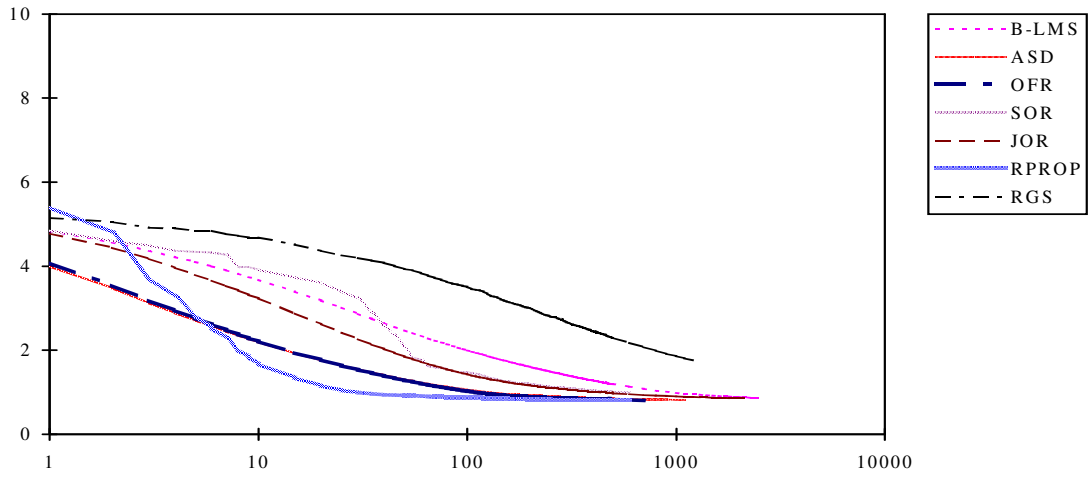
Διάγραμμα 7.31b Μεταβολή Validation Set Error Percentage ως προς τις Επαναλήψεις για το Πρόβλημα HORSE2



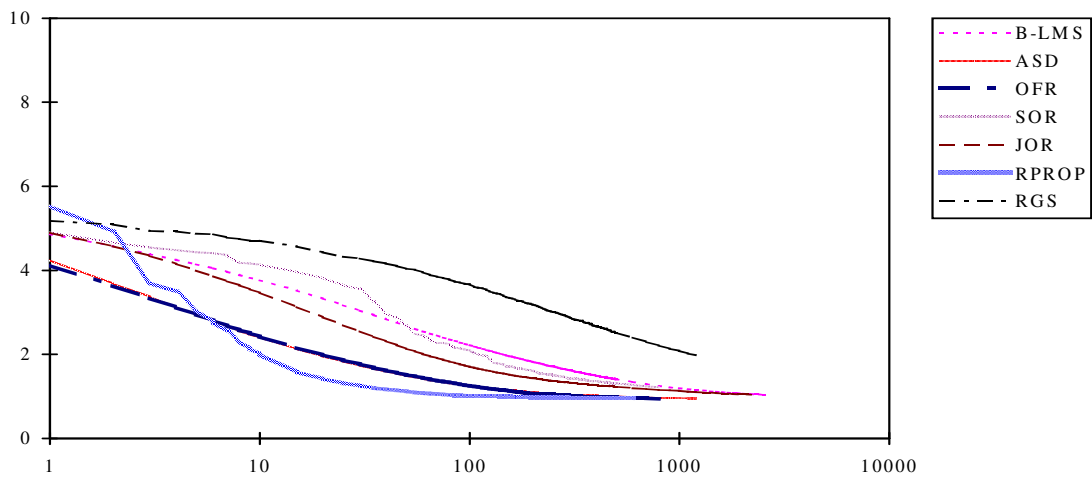
Διάγραμμα 7.31c Μεταβολή Validation Set Error Percentage ως προς τις Επαναλήψεις για το Πρόβλημα HORSE3



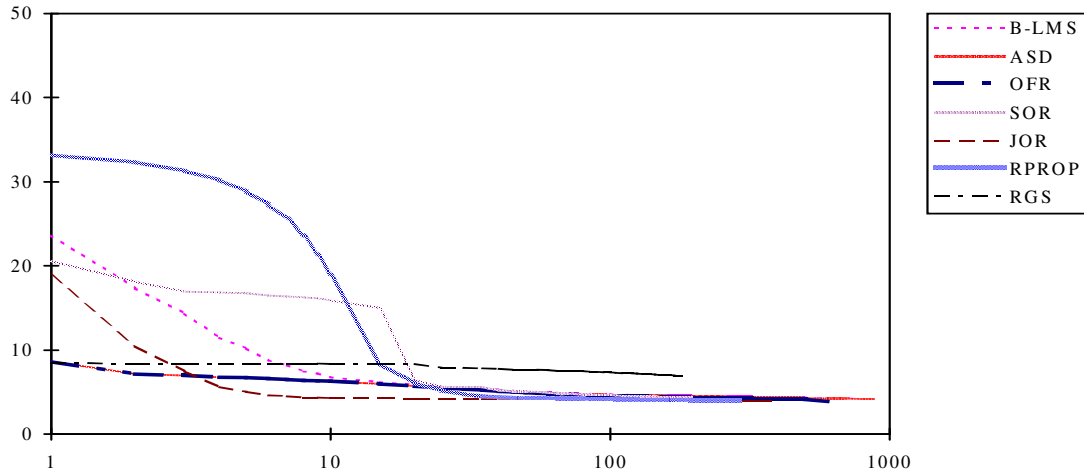
Διάγραμμα 7.32a Μεταβολή Validation Set Error Percentage ως προς τις Επαναλήψεις για το Πρόβλημα SOYBEAN1



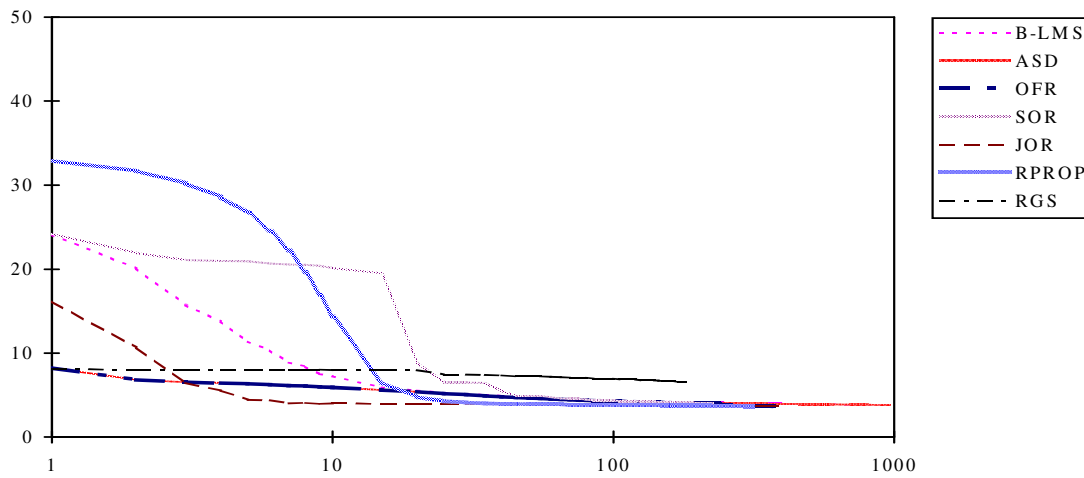
Διάγραμμα 7.32b Μεταβολή Validation Set Error Percentage ως προς τις Επαναλήψεις για το Πρόβλημα SOYBEAN2



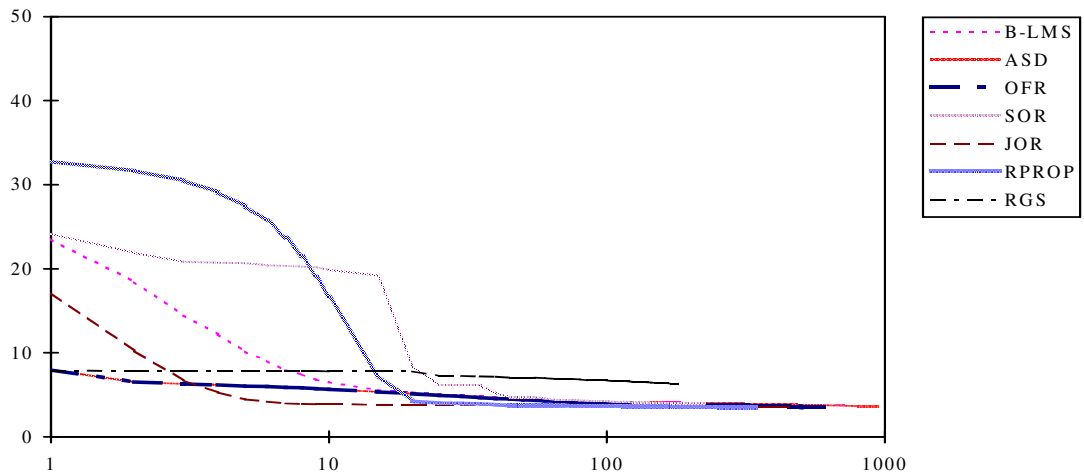
Διάγραμμα 7.32c Μεταβολή Validation Set Error Percentage ως προς τις Επαναλήψεις για το Πρόβλημα SOYBEAN3



Διάγραμμα 7.33a Μεταβολή Validation Set Error Percentage ως προς τις Επαναλήψεις για το Πρόβλημα THYROID1



Διάγραμμα 7.33b Μεταβολή Validation Set Error Percentage ως προς τις Επαναλήψεις για το Πρόβλημα THYROID2

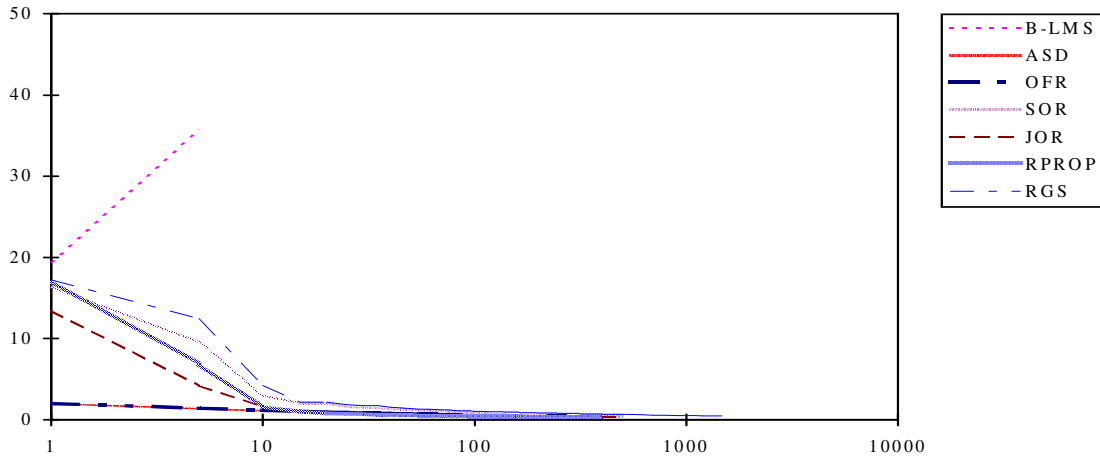


Διάγραμμα 7.33c Μεταβολή Validation Set Error Percentage ως προς τις Επαναλήψεις για το Πρόβλημα THYROID3

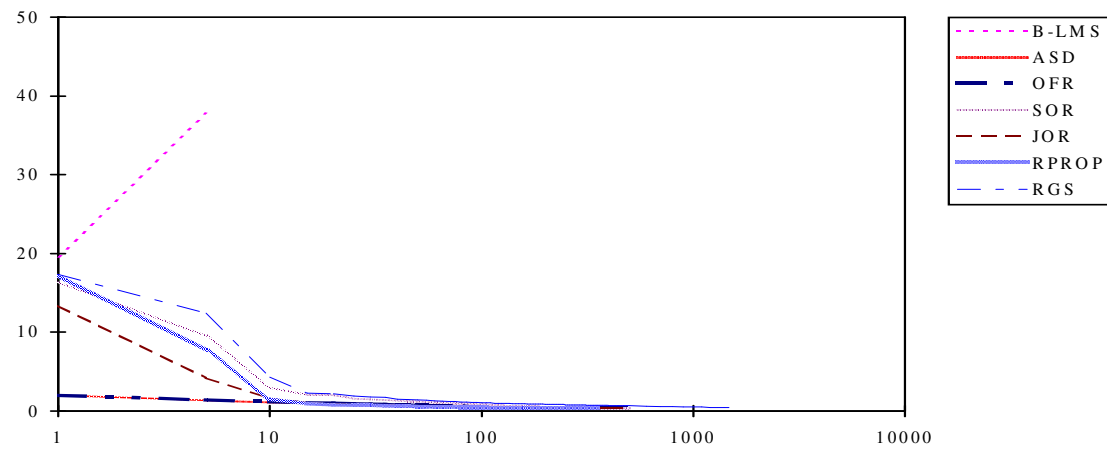
7.3.4.7 Παρατηρήσεις στα Διαγράμματα 7.24-7.33

1. Στα πιο πολλά προβλήματα, οι μέθοδοι OFR και ASD απαιτούν το μικρότερο αριθμό εποχών και σχετικών εποχών για τη σύγκλισή τους και η μείωση του ποσοστιαίου σφάλματος εγκυρότητας (Validation Set Error Percentage) είναι ραγδαία, η τελική τιμή του οποίου είναι σχεδόν ίδια για όλες τις μεθόδους, στα προβλήματα CANCER, DIABETES, GENE, HEART, HEARTC, HORSE και THYROID.
2. Οι χειρότερες μέθοδοι ως προς τις εποχές και τις σχετικές εποχές που απαιτούνται για σύγκλιση ήταν οι μέθοδοι SOR και RGS, στις περιπτώσεις όπου ο συντελεστής εκπαίδευσης του καλύτερου πειράματος, ω_{best} έχει πολύ μικρή τιμή.

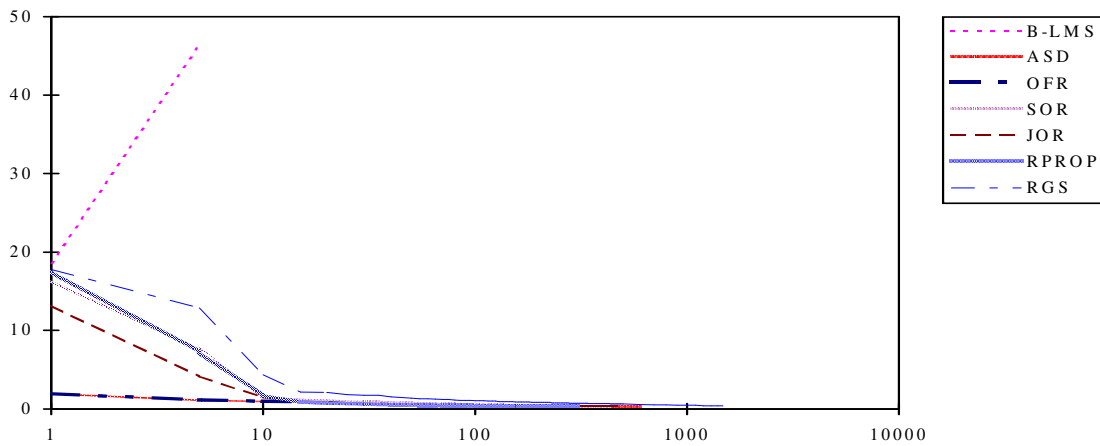
Στα διαγράμματα (7.34-7.37) που ακολουθούν παρουσιάζονται τα αποτελέσματα του καλύτερου πειράματος του κάθε προβλήματος προσαρμογής, για τις μεθόδους RPROP, BATCH-LMS, ASD, SOR και JOR, άσον αφορά τη μεταβολή του σφάλματος ταξινόμησης των προτύπων ανάκλησης σε σχέση με τις επαναλήψεις. Παραλήφθηκαν οι μέθοδοι PR, FR και LMS, των οποίων η συμπεριφορά είναι παρόμοια με τη BATCH-LMS.



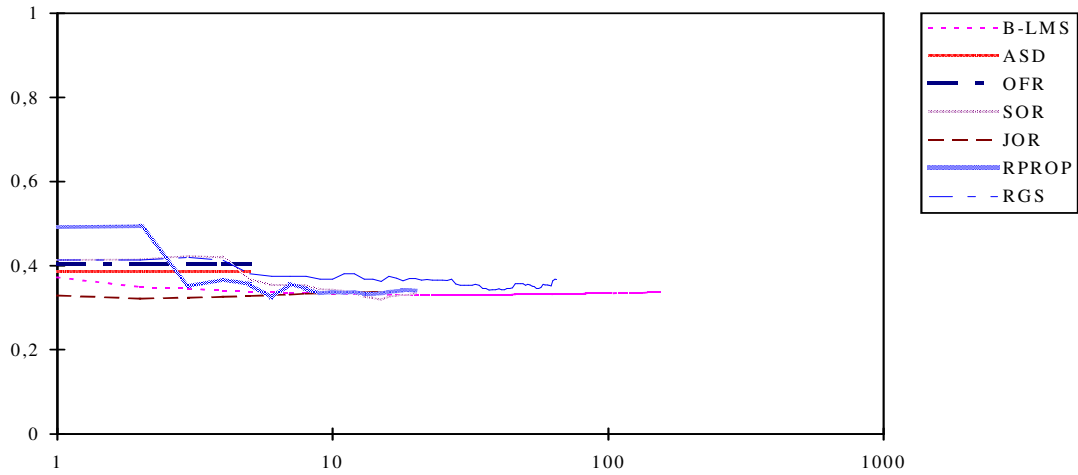
Διάγραμμα 7.34a Μεταβολή Validation Set Error Percentage ως προς τις Επαναλήψεις για το Πρόβλημα BUILDING1



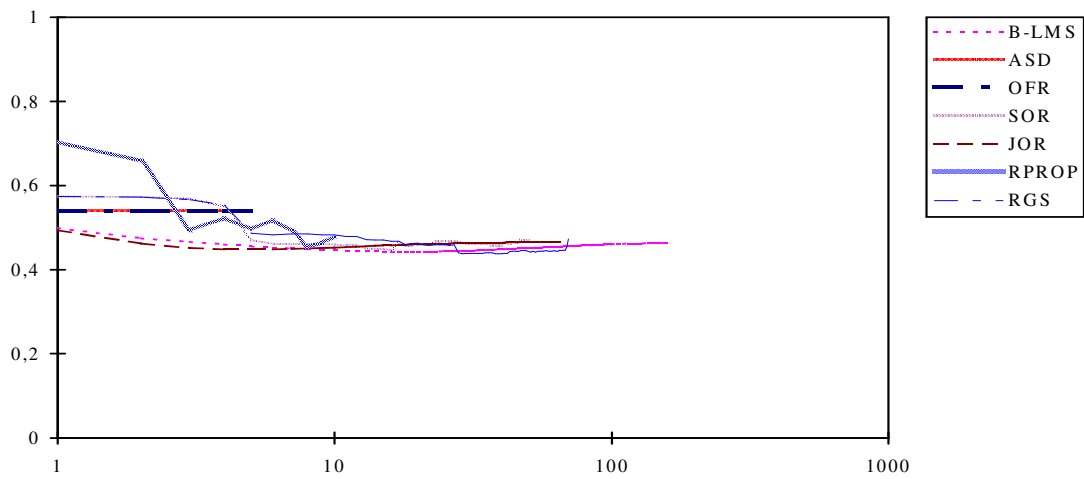
Διάγραμμα 7.34b Μεταβολή Validation Set Error Percentage ως προς τις Επαναλήψεις για το Πρόβλημα BUILDING2



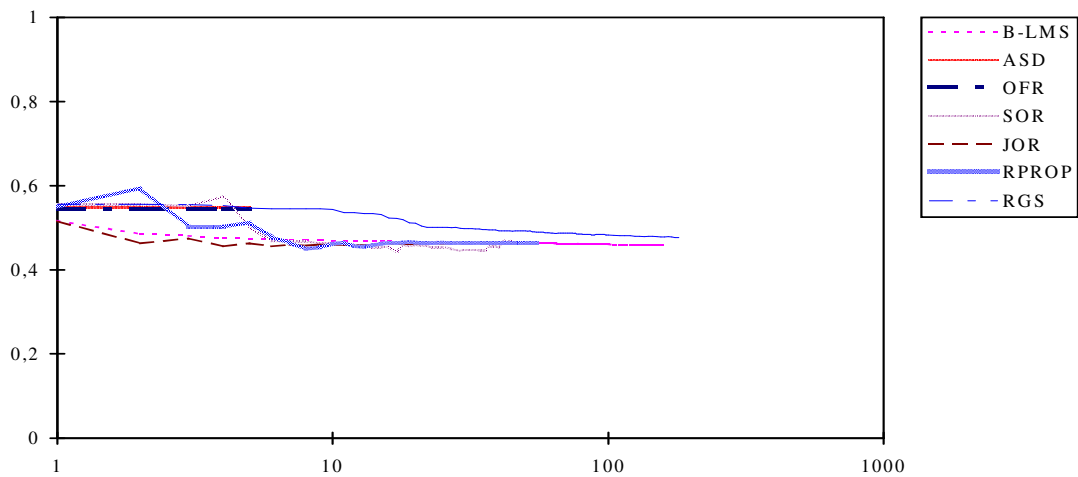
Διάγραμμα 7.34c Μεταβολή Validation Set Error Percentage ως προς τις Επαναλήψεις για το Πρόβλημα BUILDING3



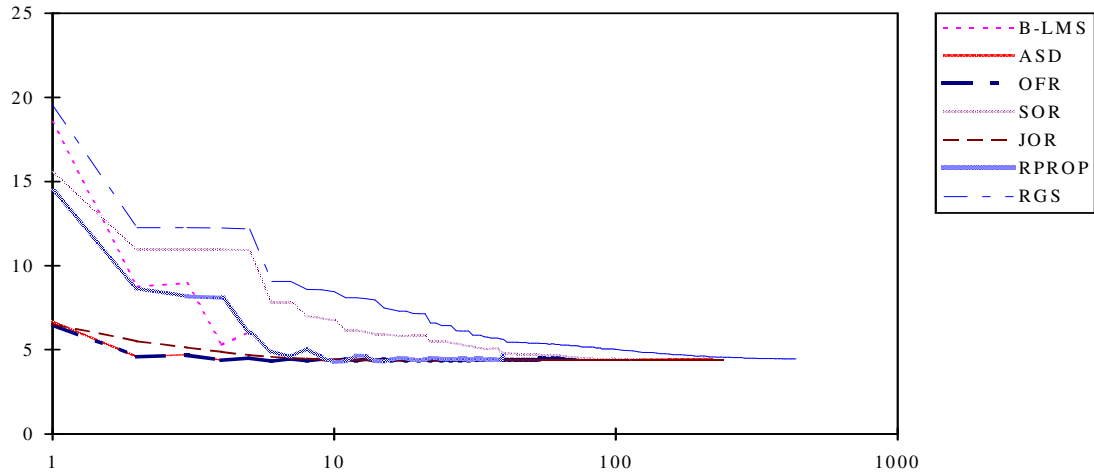
Διάγραμμα 7.35a Μεταβολή Validation Set Error Percentage ως προς τις Επαναλήψεις για το Πρόβλημα FLARE1



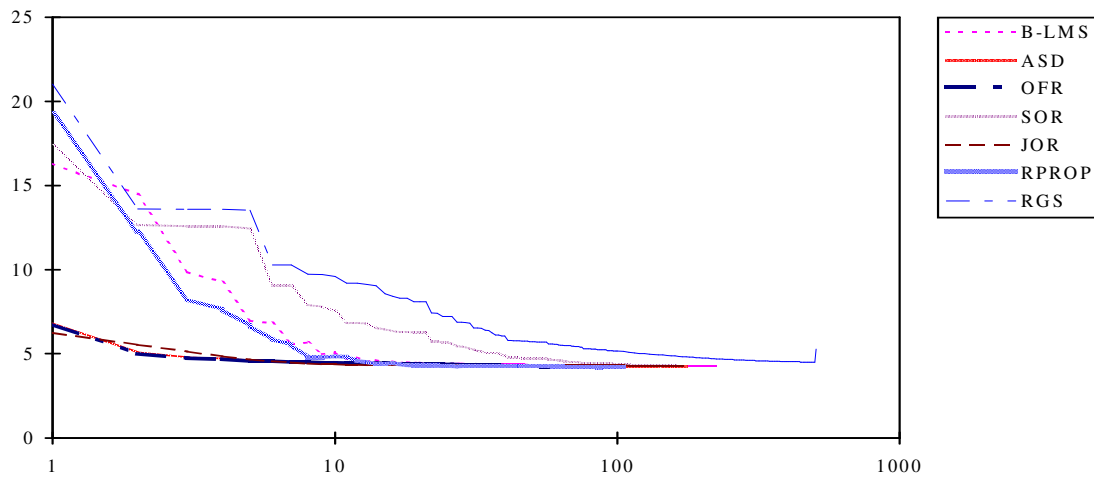
Διάγραμμα 7.35b Μεταβολή Validation Set Error Percentage ως προς τις Επαναλήψεις για το Πρόβλημα FLARE2



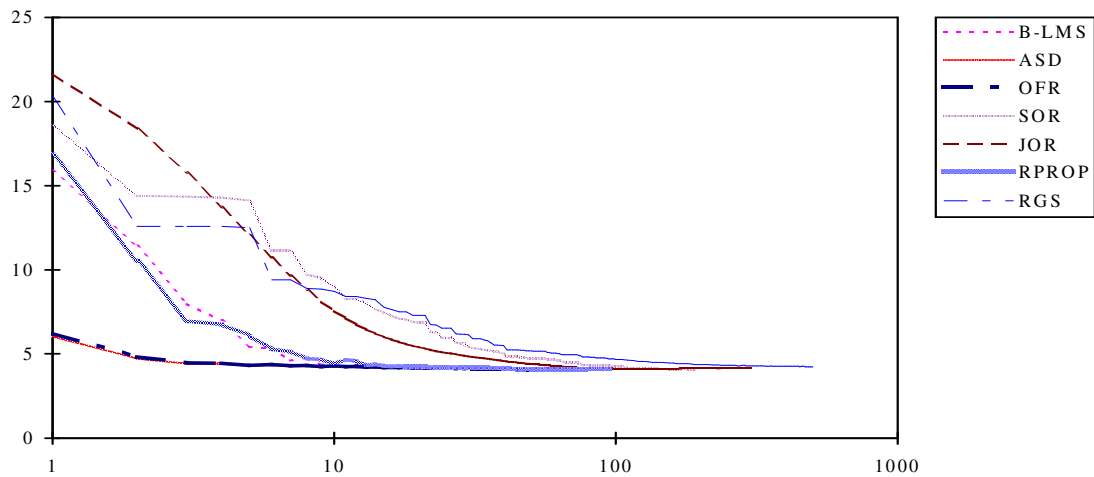
Διάγραμμα 7.35c Μεταβολή Validation Set Error Percentage ως προς τις Επαναλήψεις για το Πρόβλημα FLARE3



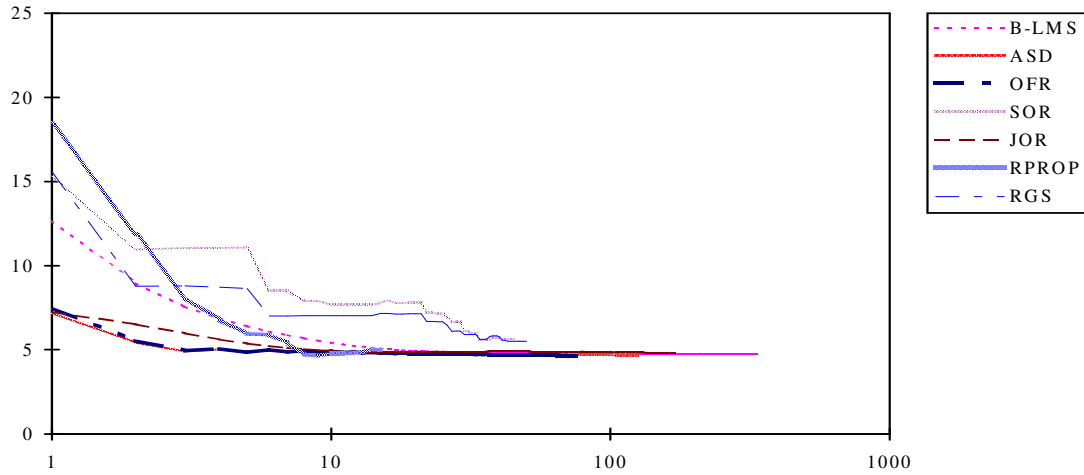
Διάγραμμα 7.36a Μεταβολή Validation Set Error Percentage ως προς τις Επαναλήψεις για το Πρόβλημα HEARTA1



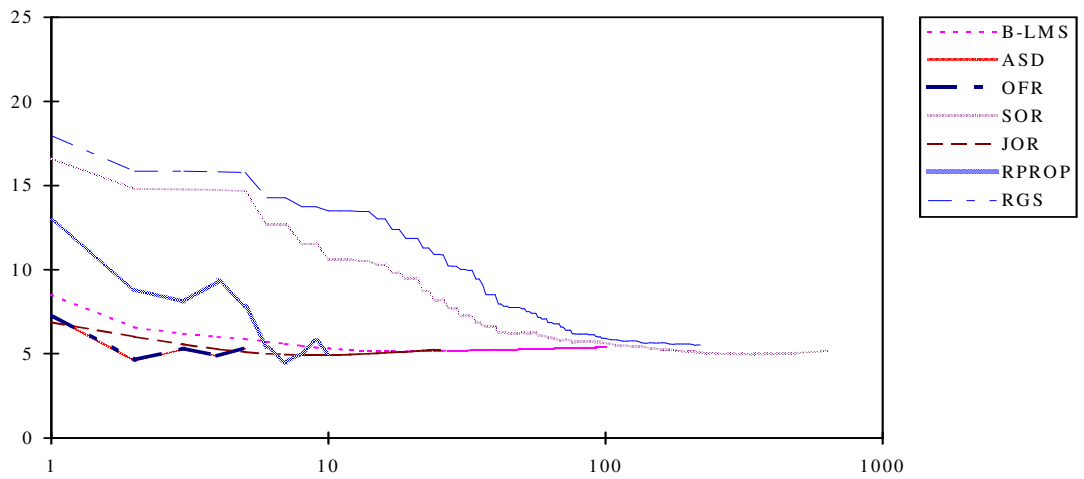
Διάγραμμα 7.36b Μεταβολή Validation Set Error Percentage ως προς τις Επαναλήψεις για το Πρόβλημα HEARTA2



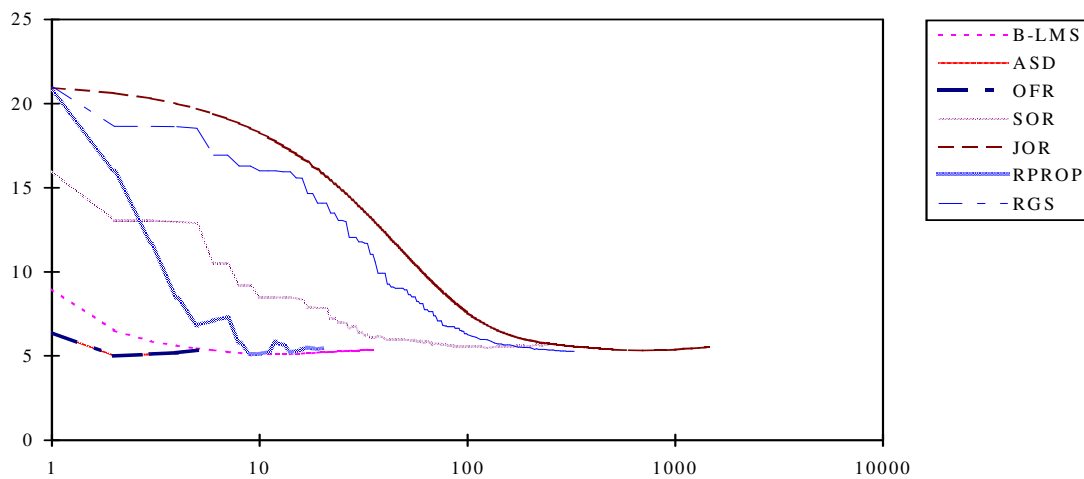
Διάγραμμα 7.36c Μεταβολή Validation Set Error Percentage ως προς τις Επαναλήψεις για το Πρόβλημα HEARTA3



Διάγραμμα 7.37a Μεταβολή Validation Set Error Percentage ως προς τις Επαναλήψεις για το Πρόβλημα HEARTAC1



Διάγραμμα 7.37b Μεταβολή Validation Set Error Percentage ως προς τις Επαναλήψεις για το Πρόβλημα HEARTAC2



Διάγραμμα 7.37c Μεταβολή Validation Set Error Percentage ως προς τις Επαναλήψεις για το Πρόβλημα HEARTAC3

7.3.4.8 Παρατηρήσεις στα Διαγράμματα 7.34-7.37

1. Στα πιο πολλά προβλήματα, οι μέθοδοι OFR και ASD απαιτούν το μικρότερο αριθμό εποχών και σχετικών εποχών για τη σύγκλισή τους και η μείωση του σφάλματος εγκυρότητας είναι ραγδαία, η τελική τιμή του οποίου είναι σχεδόν ίδια για όλες τις μεθόδους, στα προβλήματα BUILDING, HEARTA και HEARTAC.
2. Στα προβλήματα BUILDING, το σφάλμα ταξινόμησης των προτύπων ανάκλησης για τη μέθοδο BATCH-LMS μεγαλώνει, αντί να μικραίνει και η μέθοδος τερματίζει λόγω απώλειας γενίκευσης.
3. Οι χειρότερες μέθοδοι ως προς τις εποχές και τις σχετικές εποχές που απαιτούνται για σύγκλιση ήταν οι μέθοδοι BATCH-LMS, PR και FR για τα προβλήματα BUILDING, οι μέθοδοι SOR και RGS για τα προβλήματα HEARTA, η μέθοδος RGS για τα προβλήματα HEARTAC και οι μέθοδοι OFR και ASD για τα προβλήματα FLARE.

8

Συμπεράσματα

Οι αρχιτεκτονικές ΤΝΔ και οι αλγόριθμοι εκπαίδευσης που προτείνουμε στην παρούσα διατριβή προσπαθούν να εκμεταλλευτούν τα πλεονεκτήματα των επαναληπτικών μεθόδων της Αριθμητικής Ανάλυσης για την επίλυση συστημάτων γραμμικών εξισώσεων και γενικά προβλημάτων γραμμικής άλγεβρας, όπως είναι η δυνατότητα παραλληλοποίησής τους, η ταχύτητα σύγκλισης και η δυνατότητα υλοποίησής τους με τη χρήση Τεχνητών Νευρωνικών Δικτύων (ΤΝΔ), με την ανάπτυξη που υπάρχει στον τομέα των VLSI κυκλωμάτων. Με ένα απλό δίκτυο με προώθηση πληροφοριών προς τα εμπρός (Feedforward Neural Network) επιτυγχάνεται η ενοποίηση των διαφόρων αρχιτεκτονικών ΤΝΔ για την επίλυση συστημάτων γραμμικών εξισώσεων. Με την προσαρμογή του αλγορίθμου εκπαίδευσης LMS και Back-propagation, καθίσταται δυνατή η προσομοίωση των κλασικών μεθόδων της Αριθμητικής Ανάλυσης όπως η Jacobi Extrapolated (JOR), η οποία προσφέρει τη δυνατότητα παραλληλοποίησης, η μέθοδος Successive OverRelaxation (SOR), μια μέθοδος που παρέχει τη δυνατότητα να χρησιμοποιήσουμε μια μεγάλη γκάμμα τιμών (μεταξύ 0 και 2) για το συντελεστή εκπαίδευσης, η μέθοδος Richardson (ισοδύναμη της Batch-LMS), η μέθοδος Richardson Gauss Seidel (RGS), όπως και μεθόδων κλιμακωτής μείωσης, όπως η Adaptive Steepest Descent (ASD) και μεθόδων συζυγούς κλίσης, όπως η Polak-Ribiere (PR), Fletcher-Reeves (FR) και Optimal-Fletcher-Reeves (OFR). Οι τέσσερις τελευταίες μέθοδοι έχουν σαν μειονέκτημα το μεγάλο υπολογιστικό κόστος, για τον υπολογισμό, κάθε φορά, του συντελεστή εκπαίδευσης. Απαιτείται, ως εκ τούτου, η χρήση κάποιων επι πλέον νευρώνων για τους υπολογισμούς αυτούς, μιας και οι αντίστοιχες πράξεις απαιτούν πολλαπλασιασμό πίνακα επί διάνυσμα ή διάνυσμα επί διάνυσμα, οι οποίες μπορούν εύκολα να υλοποιηθούν.

Παρά το γεγονός ότι οι παραπάνω αλγόριθμοι είναι γραμμικοί, όταν χρησιμοποιήθηκαν για την επίλυση συστημάτων γραμμικών εξισώσεων έδειξαν μια υπεροχή σε σχέση με τους κλασικούς αλγόριθμους εκπαίδευσης Γραμμικών Νευρωνικών Δικτύων, όπως η μέθοδοι LMS και Batch-LMS.

Από τα αποτελέσματα της σύγκρισης των μεθόδων SOR και JOR με τη μέθοδο LMS [MAGE94], για την επίλυση όλων των ειδών συστημάτων γραμμικών εξισώσεων που μελετήθηκαν, φαίνεται η υπεροχή των μεθόδων SOR και JOR, όσον αφορά τον αριθμό των επαναλήψεων που απαιτούνται για να συγκλίνουν οι αλγόριθμοι.

Από τα αποτελέσματα της σύγκρισης της μεθόδου ASD σε σχέση με τη μέθοδο Batch-LMS [GoAM97], για την επίλυση όλων των ειδών συστημάτων γραμμικών

εξισώσεων που μελετήθηκαν, φαίνεται η υπεροχή της μεθόδου ASD, όσον αφορά τον αριθμό των επαναλήψεων που απαιτούνται για να συγκλίνουν οι αλγόριθμοι, καθώς και η ακρίβεια της λύσης που παρέχουν.

Η εφαρμογή των αλγορίθμων στην επίλυση προβλημάτων ταξινόμησης (classification) προσφέρει τη δυνατότητα να λειτουργούν τα παραπάνω δίκτυα σαν συσχετιστικές μνήμες (associative memories). Η εκπαίδευση του δικτύου ανάγεται σε ένα πρόβλημα βελτιστοποίησης, όπου ζητείται να βρεθεί ο κατάλληλος συνδυασμός συνάψεων και παραμέτρων, έτσι ώστε η αρχιτεκτονική να παράγει την καλύτερη απεικόνιση μεταξύ των δεδομένων.

Η συμπεριφορά και απόδοση των μεθόδων SOR, JOR, ASD και OFR στην επίλυση των προβλημάτων ταξινόμησης και προσαρμογής της συλλογής Proben1, μιας συλλογής προβλημάτων με πραγματικά δεδομένα, ήταν εφάμιλλη και πολλές φορές καλύτερη μη γραμμικών αλγορίθμων εκπαίδευσης, όπως η μέθοδος RPROP, η οποία χρησιμοποιήθηκε με τη μη γραμμική της μορφή, με σκοπό να συγκριθούν τα αποτελέσματα αυτών των μεθόδων, με αποτελέσματα που έχουν δημοσιευθεί χρησιμοποιώντας τη μέθοδο RPROP [Prec94].

Σημαντικό ρόλο στη συμπεριφορά των μεθόδων SOR, JOR και RGS, παίζει ο συντελεστής υπερχαλάρωσης ω , ο οποίος χρησιμοποιείται για τον υπολογισμό του συντελεστή εκπαίδευσης. Έτσι, η απόδοση των μεθόδων για πολύ μεγάλες ή πολύ μικρές τιμές του συντελεστή υπερχαλάρωσης ω δεν είναι καλή, ενώ η τιμή του ω , ω_{opt} με τα καλύτερα αποτελέσματα μπορεί να βρεθεί με ευρετικούς αλγόριθμους.

Ένας επίσης σημαντικός παράγοντας για όλες τις μεθόδους, ήταν και το σύνολο δεδομένων που χρησιμοποιήθηκε για την εκπαίδευση και τον έλεγχο της εκπαίδευσης. Τα αποτελέσματα που έδωσαν τα τρία διαφορετικά σύνολα δεδομένων για το κάθε πρόβλημα της συλλογής Proben1, είχαν σε πολλές περιπτώσεις μεγάλη απόκλιση μεταξύ των τριών συνόλων, πράγμα που οφείλεται στον εκάστοτε πίνακα των προτύπων.

Με τη μελέτη των ιδιοτιμών του κάθε πίνακα, θα μπορούσαμε να χρησιμοποιήσουμε ένα καλύτερο σύνολο δεδομένων για την εκπαίδευση του δικτύου, το οποίο μπορεί να δώσει μικρότερο σφάλμα ταξινόμησης των προτύπων.

Για περαιτέρω έρευνα θα μπορούσε επίσης να εξετασθεί η συμπεριφορά των αλγορίθμων που προτείναμε σε δίκτυα με μη γραμμικούς νευρώνες ενεργοποίησης.

ΑΝΑΦΟΡΕΣ

- [Amar77] Amari S. I., “Neural Theory of Association and Concept-Formation”, *Biological Cybernetics*, Vol. 26, pp. 175-185, 1977.
- [Amar89] Amari S. I., “Characteristics of Sparsely Encoded Associative Memory”, *Neural Networks*, Vol. 2, No. 6, pp. 451-457, 1989.
- [AMMFY95] S. Amari, N.Murata, K.-R. Muller, M. Finke, H. Yang “Asymptotic Statistical Theory of Overtraining and Cross-Validation. METR 95-06, 1995, Department of Mathematical Engineering and Information Physics, University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113, Japan.
- [Ande72] Anderson J. A., A Simple Neural Network Generating Interactive Memory, *Mathematical Biosciences*, No. 14, pp. 197-220, 1972.
- [Ande88] Anderson D. Z., *Neural Information Processing Systems*, John Wiley & Sons, 1988.
- [Atki89] Atkinson K., “An Introduction to Numerical Analysis”, New York, *American Institute of Physics*, 1989.
- [Batt92] Battiti, R., "First- and Second-Order Methods for Learning: Between Steepest Descent and Newton's Method", *Neural Computation* (1992), Vol. 4, pp. 141-166.
- [BeLC88] Becker S. and Le Cun, Y., "Improving the Convergence of Back-Propagation Learning with Second Order Methods", *Proceedings of the 1988 Connectionist Models Summer School, (Pittsburgh, 1988)*, D. Touretzky, G. Hinton, and T. Sejnowski, ed., pp. 29-37, Morgan Kaufmann, San Mateo, Calif..
- [BeTs89] Bertsekas, D. and Tsitsiklis J., "Parallel and Distributed Computation: Numerical Methods", (1989), *Prentice-Hall*, Englewood Cliffs, N. J..
- [BeTs96] Bertsekas, D. and Tsitsiklis J., "Neuro-Dynamic Programming", *Athena Scientific*, (1996), Belmont, Mass..
- [CaBa92] Carvalho, L. A., Barbosa, V., “Fast Linear System Solution by Neural Networks”, *Operations Research Letters*, Vol. 11, pp. 141-145, 1992.
- [ChHs92] Chu, Y. P., and Hsieh, C. M., “An Artificial Neural Network Model with Modified Perceptron Algorithm”, *Parallel Computing*, Vol. 18, pp. 983-996, 1992.
- [ChFa87] Chan, L. W., and Fallside, F., “An Adaptive Training Algorithm for Backpropagation Networks”, *Computers Speech and Language*, Vol. 2, pp. 205-218, 1987.
- [CiUn92a] Cichocki, A., Unbehauen R., “Neural Networks for Solving Systems of Linear Equations and Related Problems”, *IEEE Transactions on Circuits and Systems-1: Fundamental Theory and Applications*, Vol. 39, No. 2, pp. 124-138, 1992.
- [CiUn92b] Cichocki, A., Unbehauen R., “Neural Networks for Solving Systems of Linear Equations-Part II: Minimax and Least Absolute Values”, *IEEE*

- Transactions on Circuits and Systems-II: Analog and Digital Signal Processing*, Vol. 39, No. 9, pp. 619-633, 1992.
- [CiUn94a] Cichocki, A. and Unbehauen, R., "Simplified Neural Networks for Solving Linear Least Squares and Total Least Squares Problems in Real Time", *IEEE Transactions on Neural Networks*, (1994), Vol. 5, no. 6, pp. 910-923.
- [CiUn94b] Cichocki, A., Unbehauen R., "Simplified Neural Networks for Solving Linear Least Squares and Total Least Squares Problems in Real Time", *IEEE Transactions on Neural Networks*, Vol. 5, No. 6, pp. 910-923, 1994.
- [CULW95] Cichocki, A., Unbehauen R., Lendl, M., Weinzierl, K., "Neural Networks for Linear Inverse Problems with Incomplete Data Especially in Applications to Signal and Image Reconstruction", *Neurocomputing*, Vol. 8, pp. 7-41, 1995.
- [DaBi74] Dahlquist G. and Biorg A., "Numerical Methods", *Prentice Hall*, 1974.
- [DaMo91] Darken, C. and Moody, J., "Note on learning Rate Schedules for Stochastic Optimization", *Neural Information Processing Systems*, Lippmann R. P. and Moody J. E. and Touretzky D. S. (Editors), pp. 832-838, 1991.
- [DeMa76] Demidovich B and Maron I., "Computational Mathematics", *MIR Publishers*, 1976.
- [Fahl88] Fahlman Scott, "An Empirical Study of Learning Speed in Back-Propagation Networks ", *Technical Report*, CMU-CS-88-162, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, 1988.
- [FiHZ93] Finnof, W., Hergert, F., and Zimmermann, H.G. (1993), "Improving Model Selection by Nonconvergent Methods", *Neural Networks*, Vol. 6, pp. 771-783.
- [Fran87] Franzini, M. A., "Speech Recognition with Back Propagation", *Proceedings in 9th Annual Conference of IEEE Engineering in Medicine and Biology Society*, 1987.
- [GeBD92] Geman, S., Bienestock, E., and Doursat, R., "Neural Networks and the Bias/Variance Dilemma", *Neural Computation*, Vol. 4, pp. 1-58, 1992.
- [GiMW91] Gill P., Murray W. and Wright M., "Numerical Linear Algebra and Optimization", Volume 1, *Addison Wesley*, 1991.
- [GoAM97] Goulianas K., Adamopoulos M., and Margaritis K.G., "Structured Artificial Neural Networks for Fast Batch LMS Algorithms", *Neural, Parallel & Scientific Computation*, Vol. 5, No 4, pp. 549-562, 1997..
- [GoSe88] Gorman , R. P., and Sejnowski, T. J., "Analysis of Hidden Units in a Layered Network Trained to Classify Sonar Targets", *Neural Networks*, Vol. 1, pp. 75-89, 1988.
- [GoVL83] Golub, G. H., and Van Loan C.F., "Matrix Computations", *The John Hopkins University Press*, Baltimore, MD, 1983.
- [HaYo81] Hageman, L., and Young, D., "Applied Iterative Methods", *Academic Press*, 1981.
- [Hass95] Hassoun, M. H., "Fundamentals of Artificial Neural Networks", *The MIT Press*, Cambridge, Massachusetts 02142, 1995.
- [Hebb49] D. O. Hebb. *The Organization of Behavior*, Wiley, New York, 1949.

- [Hopf82] John Hopfield. Neural Networks and Physical Systems with Emergent Collective Computation Abilities. Proceedings in *National Academic Sciences-Biophysics*, 79, pp. 2554-2558, 1982.
- [Koho72] Kohonen T, "Correlation Matrix Memories", *IEEE Transactions on Computers*, C-21 (4), pp. 353-359, 1972.
- [Koho74] Kohonen T, "An Adaptive Associative Memory Principle", *IEEE Transactions on Computers*, pp. 444-445, 1974.
- [Koho84] Kohonen T, "Self-Organisation and Associative Memory", *Springer-Verlag*, 1984.
- [Koho89] Teuvo Kohonen, Self-Organisation and Associative Memory, *Springler-Verlag*, 3rd edition, 1989.
- [KrKr94] Krishnamurthy E. V., and Krishnamurthy, V., "An ANN Model Perceptron Algorithm Using Generalized Matrix Inversion", *Parallel Computing*, Vol. 20, pp. 799-806, 1994.
- [KrSV89] Kramer, A. H., and Sangiovanni-Vincentelli, A., "Efficient Parallel Learning Algorithms for Neural Networks", *Advances in Neural Information Processing Systems 1*, Morgan Kaufmann, San Mateo, Calif, pp. 40-48, 1989.
- [LCSP93] Le Cun, Y., Simard, P. Y., and Pearlmutter, B., "Automatic Learning Rate Maximization by On-line Estimation of the Hessian's Eigenvectors", *Advances in Neural Information Processing Systems 5*, Morgan Kaufmann, San Mateo, Calif, pp. 156-163, 1993.
- [LuBa92] Luo, F. L., Bao, Z., "Neural Network Approach to Computing Matrix Inversion", *Applied Mathematics and Computation*, Vol. 47, pp. 109-120, 1992.
- [Luen73] Luenberger, "Introduction to Linear and Non-linear Programming", *Addison-Wesley*, Reading, MA, 1973.
- [LuLi93] Luo, F.L., Li, Y.D., "A 2D Neural Network with Application to Solving a Class of Matrix Equations", *International Journal of Circuit Theory and Applications*, Vol. 21, pp. 539-549, 1993.
- [LuZQ91] Luo Zhi-Quan, "On the Convergence of the LMS Algorithm with Adaptive Learning Rate for Linear Feedforward Networks", *Neural Computation* (1991), Vol. 3, pp. 226-245.
- [MAGE94] Margaritis K.G., Adamopoulos M., Goulianas K., and Evans D.J., "Artificial Neural Networks and Iterative Linear Algebra Methods", *Parallel Algorithms and Applications* (1994), Vol. 3, pp. 31-44.
- [MaHP90] Maren A.J., Harston C.T., Pap R.M., "Handbook of neural computing applications", *Academic Press*, 1990.
- [MALE92] Margaritis, K. G., Adamopoulos, M., Loukakis, M., Evans, D.J., "On The Solution of General Linear Systems and General Matrix Inversion Using Neural Network Architectures", *Transputing in Numerical and Neural Network Applications*, G.L. Reijns and J. Luo, Eds, IOS Press, pp. 231-243, 1992.

- [McPi43] Warren S. McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5, pp. 115-133, 1943.
- [MiPa69] Marvin Minsky and Seymour Papert. *Perceptrons: An Introduction to Computational Geometry*, MIT Press, 1969.
- [Moll93] Martin Moller, "A Scaled Conjugate Gradient Algorithm for Fast Supervised Learning", *Neural Networks*, Vol. 6, No. 4, pp. 525-533, 1993.
- [MoBo90] Morgan, N., and Bourland, H., "Generalisation and Parameter Estimation in Feedforward Nets. Some Experiments", *Touretzky*, pp. 630-637, 1990.
- [Naka72] Nakano, K., "Associatron: A Model of Associative Memory", *IEEE Transactions on Systems, Man and Cybernetics*, SMC-2, pp. 380-388, 1972.
- [NeIl91] Nelson, M.C. and Illingworth, W.T. (1991), "A Practical Guide to Neural Nets", Reading, MA: Addison-Wesley.
- [PaoY89] Pao, Y-H., *Adaptive Pattern Recognition and Neural Networks*, Addison-Wesley, 1989.
- [Pflu90] Pflug, G. Ch., "Non-asymptotic Confidence Bounds for Stochastic Approximation Algorithms with Constant Stepsize", *Mathematik*, Vol. 110, pp. 297-314, 1990.
- [PoIo92] Polycarpou M. and Ioannou P., "Learning and Convergence Analysis of Neural-Type Structured Networks", *IEEE Transactions on Neural Networks* (1992), Vol. 3, no. 1, pp. 39-50.
- [Prec94] Prechelt Lutz, "Proben1 - A Set of Neural Network Benchmark Problems and Benchmarking Rules", *Technical Report*, Universitat Karlsruhe 21/94 (1994).
- [RiBr93] Riedmiller M. and Braun H., "A Direct Adaptive Method for Faster Backpropagation Learning: The RPROP Algorithm", *Proceedings of the IEEE International Conference on Neural Networks*, San Francisco, CA, April 1993, IEEE.
- [Rose59] Frank Rosenblatt. "Two theorems of statistical separability in the perceptron". *Symposium on the Mechanization of Thought Process*, pp. 421-456, 1959.
- [Rose62] Rosenblatt F, "Principles of neurodynamics", *Spartan Books*, 1962.
- [RuHW86] Rumelhart D.E., Hinton G.E., and Williams R.J., "Learning Internal Representation by Error Propagation", *Parallel Distributed Processing I*, Rumelhart D.E., and McClelland J, Eds., Cambridge, MA, MIT Press, 1986.
- [RuML86] Rumelhart D.E., McClelland J, "Parallel distributed processing: Explorations in the microstructure of cognition", *MIT Press*, 1986.
- [Sar195] Sarle, W.S. (1995), "Stopped Training and Other Remedies for Overfitting," *Proceedings of the 27th Symposium on the Interface of Computing Science and Statistics*, pp. 352-360.
- [SiAl90] Silva, F. M., and Almeida, L. B., "Acceleration Techniques for the Back-propagation Algorithm in Neural Networks", *Europe Lecture Notes in Computer Science*, Springer-Verlag, Berlin, pp. 110-119, 1990.

- [SjLj92] Sjoberg, J. and Ljung, L. (1992), "Overtraining, Regularization, and Searching for Minimum in Neural Networks" *Technical Report LiTH-isy-I-1297*, Department of Electrical Engineering, Linkoping University, S-581 83 Linkoping, Sweden.
- [Stur88] Sturges R. H. Jr, "Analog Matrix Inversion", *IEEE J. Robotics Automat.*, Vol. 4, pp. 157-162, 1988.
- [Sutt86] Sutton, R., "Two Problems with Back-propagation and Other Steepest-Descent Learning Procedures for Networks", *Proceedings of the 8th Annual Conference on the Cognitive Science Society*, Erlbaum, Hillsdale, N.J., pp. 823-831, 1986.
- [Toll90] Tollenaere, T., "SuperSAB: Fast Adaptive Back-propagation with Good Scaling Properties", *Neural Networks*, Vol. 3, pp. 561-573, 1990.
- [Tsy71] Tsytkin, Ya. Z., "Adaptation and Learning in Automatic Systems", translated by Z. J. Nikolic, Academic Press, New York, 1971.
- [Varg62] Varga, R. S., "Matrix Iterative Analysis", *Prentice Hall*, 1962.
- [WaLi94] Wang, J., Li, H., "Solving Simultaneous Linear Equations Using Recurrent Neural Networks", *Information Sciences*, Vol. 76, pp. 255-277, 1994.
- [WaMe91] Wang L. X. and Mendel, J. M., "Three-Dimensional Structured Networks for Matrix Equation Solving", *IEEE Transactions on Computers* (1991), Vol. 40, no. 12, pp. 1337-1346.
- [WaMe92] Wang L. X. and Mendel, J.M., "Parallel Structured Networks for Solving a Wide Variety of Matrix Algebra Problems", *Journal of Parallel and Distributed Computing* (1992), Vol. 14, pp. 236-247.
- [Wan92a] Wang, J., "Recurrent Neural Network for Solving Systems of Complex-Valued Linear Equations", *Electronics Letters*, Vol. 28, No. 18, pp. 1751-1753, 1992.
- [Wan92b] Wang, J., "Electronic Realization of Recurrent Neural Network for Solving Simultaneous Linear Equations", *Electronics Letters*, Vol. 28, No. 5, pp. 493-495, 1992.
- [Wang93] Wang, J., "A Recurrent Neural Network for Real-Time Matrix Inversion", *Applied Mathematics and Computation*, Vol. 55, pp. 89-100, 1993.
- [Wang94] Wang, C. (1994), "A Theory of Generalisation in Learning Machines with Neural Network Application", *Ph.D. thesis*, University of Pennsylvania.
- [WaVJ94] Wang, C., Venkatesh, S.S., and Judd, J.S. (1994), "Optimal Stopping and Effective Machine Complexity in Learning", *NIPS6*, pp. 303-310.
- [Weig94] Weigend, A. (1994), "On Overfitting and the Effective Number of Hidden Units", *Proceedings of the 1993 Connectionist Models Summer School*, pp. 335-342.
- [Werb74] Werbos, P. (1974), "Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences", *Ph D Dissertation, Committee on Applied Mathematics*, Harvard University, Cambridge, Mass.
- [WiLe90] Widrow B, and Lehr M, "30 Years of Adaptive Neural Networks: Perceptron, Madaline and Backpropagation", *Proceedings of the IEEE*, Vol. 78, No. 9, pp. 1415-1442, 1990.

- [WiHo60] Widrow B, and Hoff M. E. Jr., "Adaptive Switching Circuits", *IRE Western Electric Show and Convention Record*, Part 4, pp. 96-104, 1960.
- [WiSt85] Widrow B, and Stearns, "Adaptive Signal Processing", *Prentice-Hall*, 1985.
- [WiWB87] Widrow B, Winter R., and Bexter R., "Learning Phenomena in Neural Nets for Adaptive Filtering and Adaptive Pattern Recognition", *Proceedings First International Conference on Neural Networks*, Vol. II, pp. 411-429, IEEE, 1987.
- [WiWi88] Widrow B, and Winter R., "Neural Nets for Adaptive Filtering and Adaptive Pattern Recognition", *Computer*, pp. 25-39, 1988.
- [WuWH94] Wu, G., Wang, J., Hootman, J., "A Recurrent Neural Network for Computing Pseudoinverse Matrices", *Mathl. Comput. Modeling*, Vol. 20, No. 1, pp. 13-21, 1994.
- [Youn71] Young, D. M., "Iterative Solution of Large Linear Systems", *Academic Press*, 1971.