ΠΡΟΣΕΓΓΙΣΗ ΚΑΙ ΑΝΑΠΑΡΑΣΤΑΣΗ ΔΕΔΟΜΕΝΩΝ ΜΙΑΣ ΔΙΑΣΤΑΣΗΣ ΜΕ ΣΥΝΔΥΑΣΜΟ MLP ΚΑΙ RBF ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ

Η
ΜΕΤΑΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ ΕΞΕΙΔΙΚΕΥΣΗΣ

Υποβάλλεται στην

ορισθείσα από την Γενική Συνέλευση Ειδικής Σύνθεσης
του Τμήματος Μηχανικών Ηλεκτρονικών Υπολογιστών και Πληροφορικής
Εξεταστική Επιτροπή

από την

Ερίσα Σκέντα

ως μέρος των Υποχρεώσεων

για τη λήψη

του

ΜΕΤΑΠΤΥΧΙΑΚΟΥ ΔΙΠΛΩΜΑΤΟΣ ΣΤΗΝ ΠΛΗΡΟΦΟΡΙΚΗ

ΜΕ ΕΞΕΙΔΙΚΕΥΣΗ ΣΤΟΥΣ ΕΠΙΣΤΗΜΟΝΙΚΟΥΣ ΥΠΟΛΟΓΙΣΜΟΥΣ

Πανεπιστήμιο Ιωαννίνων

Ιούλιος 2018

Εξεταστική Επιτροπή:

- Ισαάκ Λαγαρής, Καθηγητής, Τμήμα Μηχανικών Η/Υ και Πληροφορικής, Πανεπιστήμιο Ιωαννίνων
- Κωνσταντίνος Μπλέκας, Αναπλ. Καθηγητής, Τμήμα Μηχανικών Η/Υ και Πληροφορικής, Πανεπιστήμιο Ιωαννίνων
- Κωνσταντίνος Παρσόπουλος, Αναπλ. Καθηγητής, Τμήμα Μηχανικών Η/Υ και Πληροφορικής, Πανεπιστήμιο Ιωαννίνων

ΑΦΙΕΡΩΣΗ

Στην οικογένειά μου και σε όλους τους φίλους μου για όλη τη στήριξή τους σε όλο αυτό το διάστημα εκπόνησης της διατριβής και του μεταπτυχιακού αυτού.

ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα να ευχαριστήσω τον επιβλέποντα μου για την ευκαιρία, τις συμβουλές και την καθοδήγηση του καθόλη τη διάρκεια εκπόνησης της διατριβής.

ΠΕΡΙΕΧΟΜΕΝΑ

	Σελ
ΑΦΙΕΡΩΣΗ	iii
ΕΥΧΑΡΙΣΤΙΕΣ	iv
ΠΕΡΙΕΧΟΜΕΝΑ	v
ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ	viii
ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ	ix
ΠΕΡΙΛΗΨΗ	1
EXTENDED ABSTRACT	2
ΚΕΦΑΛΑΙΟ 1. ΕΙΣΑΓΩΓΗ	3
1.1. Στόχοι	3
1.2. Δομή της Διατριβής	3
ΚΕΦΑΛΑΙΟ 2. ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΕ ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ	5
2.1. Βελτιστοποίηση	5
2.1.1. Τοπική και Ολική Βελτιστοποίηση	7
2.1.2. Μαθηματική Αναπαράσταση	8
2.2. Νευρωνικό Δίκτυο	9
2.3. Τεχνητοί Νευρώνες	11
2.3.1. Νευρώνας	12
2.3.2. Συναρτήσεις Ενεργοποίησης	14
2.3.3. Συνάρτηση Ακτίνας Βάσης (Radial Basis Function)	16
2.4. Μοντέλα Νευρωνικών Δικτύων	17
2.4.1. Perceptron ενός Επιπέδου	19
2.4.2. Πολυεπίπεδο Δίκτυο Perceptron (MLP)	21
2.4.3. Δίκτυο Ακτινικής Συνάρτησης Βάσης (Radial Basis Function(RBF))	23
2.5. Τρόποι Εκπαίδευσης Νευρωνικών Δικτύων	24
2.5.1. Εκπαίδευση με Επίβλεψη	25
2.5.2. Εκπαίδευση Χωρίς Επίβλεψη	26
2.5.3. Εκπαίδευση με Ενίσχυση (Reinforcement Learning)	27
2.6. Εκπαίδευση Δικτύου MLP	27
2.6.1. Εκπαίδευση ανά Πρότυπο (Stochastic Training)	28
2.6.2. Μαζική Εκπαίδευση (Batch Mode)	29
2.6.3. Συνεχομένη Εκπαίδευση (On-Line Training)	29
2.6.4. Αλγόριθμος Backpropagation	29
2.6.5. Κριτήρια Τερματισμού Εκπαίδευσής	35
2.6.6. Μάθηση με Επίβλεψη ως Πρόβλημα Βελτιστοποίησης	35
2.7. Εκπαίδευση Δικτύου RBF	37
2.7.1. Εκπαίδευση Δύο Σταδίων	38
2.7.2. Ενιαία Εκπαίδευση	39
2.8. Ικανότητα Γενίκευσης (Generalization)	40

2.9. Μέθοδος Cross Validation 41
2.9.1. Διασταυρωμένη Επικύρωση k-τμημάτων(k-fold cross-validation) 41
2.9.2. Leave One Out 42
ΚΕΦΑΛΑΙΟ 3. ΣΧΕΔΙΑΣΤΙΚΗ ΔΙΑΔΙΚΑΣΙΑ ΔΙΚΤΥΩΝ 43
3.1. Προσέγγιση Συναρτήσεων με Νευρωνικά Δίκτυα 43
3.2. Ευρετικές Μέθοδοι για την Βελτίωση Απόδοσης του Αλγορίθμου Backpropagation 45
3.2.1. Αρχικοποίηση Βαρών 45
3.2.2. Ρυθμός Μάθησης 46
3.2.3. Συνάρτηση Ενεργοποίησης 47
3.3. Δομή των Πολυεπίπεδων Νευρωνικών Δικτύων(MLP) 48
3.4. Δομή Υβριδικού Δικτύου RBF 51
3.4.1. Απλό Δίκτυο MLP 51
3.4.2. Δίκτυο RBF του Υβριδικού Δικτύου 52
3.4.3. Αλγόριθμος K-Means 53
3.5. Εκπαίδευση Δυο Σταδίων του Υβριδικού Δικτύου RBF 54
ΚΕΦΑΛΑΙΟ 4. ΠΡΟΣΟΜΟΙΩΣΗ ΔΙΚΤΥΩΝ ΚΑΙ ΑΞΙΟΛΟΓΗΣΗ ΔΕΔΟΜΕΝΩΝ 55
4.1. Ανάλυση του Προγραμματιστικού Μοντέλου που Εκπονήθηκε 55
4.2. Ο Ρόλος των Παραμέτρων στο Δίκτυο MLP 57
4.2.1. Ρόλος του Αριθμού Νευρώνων 58
4.2.2. Ο Ρόλος του Αριθμού των Εποχών 64
4.2.3. Ο Ρόλος του Ρυθμού Εκμάθησης 69
4.2.4. Ο Ρόλος της Αρχικοποίησης των Βαρών 71
4.2.5. Διακύμανση του Χρόνου Προσομοίωσης και του Σφάλματος Ανάλογα με την Παράμετρο που Μεταβάλλονταν σε Κάθε Περίπτωση 72
4.3. Απόδοση του MLP στις 4 Συναρτήσεις που Επιλέξαμε 73
4.3.1. Συνάρτηση Α 74
4.3.2. Συνάρτηση Β 75
4.3.3. Συνάρτηση Γ 76
4.3.4. Συνάρτηση Δ 77
4.4. Απόδοση του RBF στις 4 Συναρτήσεις που Επιλέξαμε 78
4.4.1. Συνάρτηση Α 78
4.4.2. Συνάρτηση Β 80
4.4.3. Συνάρτηση Γ 82
4.4.4. Συνάρτηση Δ 83
4.4.5. Παρουσίαση Μέσου Τετραγωνικού Σφάλματος για τις Παραπάνω Περιπτώσεις 85
4.5. Σύγκριση του Υβριδικού Δικτύου με το Εκτεταμένο MLP 85
4.5.1. Συνάρτηση Α 87
4.5.2. Συνάρτηση Β 89
4.5.3. Συνάρτηση Γ 91
4.5.4. Συνάρτηση Δ 93
4.5.5. Σχολιασμός Αποτελεσμάτων 95
4.6. Προσομοίωση του Υβριδικού Δικτύου για αριθμό νευρώνων RBF = 30 96
4.6.1. Συνάρτηση Α 96
4.6.2. Συνάρτηση Β 99
4.6.3. Συνάρτηση Γ 101
4.6.4. Συνάρτηση Δ 103

4.6.5. Σχολιασμός Αποτελεσμάτων 105
4.7. Διαφοροποίηση των Δεδομένων Εκπαίδευσης με τα Δεδομένα Ελέγχου 105
4.7.1. Συνάρτηση Α 106
4.7.2. Συνάρτηση Β 108
4.7.3. Συνάρτηση Γ 110
4.7.4. Συνάρτηση Δ 112
4.7.5. Σχολιασμός Αποτελεσμάτων 114
ΚΕΦΑΛΑΙΟ 5. ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΠΡΟΤΑΣΕΙΣ 115
5.1. Συμπεράσματα 115
5.2. Προτάσεις 116
ΒΙΒΛΙΟΓΡΑΦΙΑ 117
ΣΥΝΤΟΜΟ ΒΙΟΓΡΑΦΙΚΟ 121

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

Πίνακας	Σελ
Πίνακας 3.1: Ευρετικές Τιμές για Σταθερό Ρυθμό Μάθησης	46
Πίνακας 3.2: Ψευδοκώδικας Εκτεταμένου Δικτύου MLP	50
Πίνακας 4.1: Τιμές για Ελάχιστο Σφάλμα Σύμφωνα με τις Εποχές	72
Πίνακας 4.2: Τιμές για Ελάχιστο Σφάλμα Σύμφωνα με τους Νευρώνες	72
Πίνακας 4.3: Τιμές για Ελάχιστο Σφάλμα Σύμφωνα με το Ρυθμό μάθησης	73
Πίνακας 4.4: Τιμές για Σφάλμα για Διάφορες Τιμές της Μεταβλητής σ	85
Πίνακας 4.5: Συγκεντρωτικός Πίνακας Σφαλμάτων για νευρώνες RBF = 60	96
Πίνακας 4.6: Συγκεντρωτικός Πίνακας Σφαλμάτων για νευρώνες RBF = 30	105
Πίνακας 4.7: Συγκεντρωτικός Πίνακας Σφαλμάτων για train ≠ test	114

ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ

Σχήμα Σελ
Σχήμα 2.1 Μοντέλο Βιολογικού Νευρώνα 10
Σχήμα 2.2 Μοντέλο των McCulloch και Pitts για τον Νευρώνα 11
Σχήμα 2.3 Μοντέλο Τεχνητού Νευρώνα 13
Σχήμα 2.4 Ενναλακτική Μορφή ενός Μοντέλου Τεχνητού Νευρώνα 13
Σχήμα 2.5 Βηματική Συνάρτηση 14
Σχήμα 2.6 Γραμμική Συνάρτηση 15
Σχήμα 2.7 Σιγμοειδής Συνάρτηση 15
Σχήμα 2.8 Συναρτήσεις Ακτινικού Τύπου 17
Σχήμα 2.9 Αρχιτεκτονικές Δικτύων 18
Σχήμα 2.10 Νευρώνας Perceptron 19
Σχήμα 2.11 Ταξινόμηση Προτύπων σε Δυο Κλάσεις με Βάση την Οριακή Γραμμή 20
Σχήμα 2.12 Δίκτυο Πρόσθιας Τροφοδότησης με ένα Μεμονωμένο Επίπεδο Νευρώνων 21
Σχήμα 2.13 Πολυεπίπεδο Δίκτυο Perceptron 23
Σχήμα 2.14 Δίκτυο Ακτινικής Συνάρτησης Βάσης 24
Σχήμα 2.15 Ροή της Διαδικασίας Μάθησης με Επίβλεψη 26
Σχήμα 2.16 Αναπροσαρμογή Βαρών στην Επιφάνεια Σφάλματος 36
Σχήμα 2.17 Παραδείγματα Ικανότητας Γενίκευσης (Α) και (β) 40
Σχήμα 2.18 Διάγραμμα Διαδικασίας k-fold-Cross-Validation 42
Σχήμα 3.1 Σιγμοειδής Συνάρτηση 48
Σχήμα 3.2 Δομή Υβριδικού Δικτύου 51
Σχήμα 4.1 Γραφική Συνάρτησης $y=x\sin(x^2)$ 58
Σχήμα 4.2 Σφάλμα για MLP Δίκτυο με 1 Νευρώνα 59
Σχήμα 4.3 Προσέγγιση Συνάρτησης $y=x\sin(x^2)$ με MLP με1 Νευρώνα 59
Σχήμα 4.4 Σφάλμα για MLP Δίκτυο με 2 Νευρώνες 60
Σχήμα 4.5 Προσέγγιση Συνάρτησης $y=x\sin(x^2)$ με MLP με 2 Νευρώνες 60
Σχήμα 4.6 Σφάλμα για MLP Δίκτυο με 4 Νευρώνες 60
Σχήμα 4.7 Προσέγγιση Συνάρτησης $y=x\sin(x^2)$ με MLP με 4 Νευρώνες 61
Σχήμα 4.8 Σφάλμα για MLP Δίκτυο με 8 Νευρώνες 61
Σχήμα 4.9 Προσέγγιση Συνάρτησης $y=x\sin(x^2)$ με MLP με 8 Νευρώνες 61
Σχήμα 4.10 Σφάλμα για MLP Δίκτυο με 16 Νευρώνες 62
Σχήμα 4.11 Προσέγγιση Συνάρτησης $y=x\sin(x^2)$ με MLP με 16 Νευρώνες 62
Σχήμα 4.12 Σφάλμα για MLP Δίκτυο με 32 Νευρώνες 63
Σχήμα 4.13 Προσέγγιση Συνάρτησης $y=x\sin(x^2)$ με MLP με 32 Νευρώνες 63
Σχήμα 4.14 Σφάλμα για MLP Δίκτυο με 64 Νευρώνες 63
Σχήμα 4.15 Προσέγγιση Συνάρτησης $y=x\sin(x^2)$ με MLP με 64 Νευρώνες 64
Σχήμα 4.16 Σφάλμα για MLP Δίκτυο για 500 Εποχές 64

Σχήμα 4.17 Προσέγγιση Συνάρτησης $y=x\sin(x^2)$ με MLP με 500 Εποχές 65
Σχήμα 4.18 Σφάλμα για MLP Δίκτυο για 1000 Εποχές 65
Σχήμα 4.19 Προσέγγιση Συνάρτησης $y=x\sin(x^2)$ με MLP με 1000 Εποχές 65
Σχήμα 4.20 Σφάλμα για MLP Δίκτυο για 3000 Εποχές 66
Σχήμα 4.21 Προσέγγιση Συνάρτησης $y=x\sin(x^2)$ με MLP με 3000 Εποχές 66
Σχήμα 4.22 Σφάλμα για MLP Δίκτυο για 5000 Εποχές 67
Σχήμα 4.23 Προσέγγιση Συνάρτησης $y=x\sin(x^2)$ με MLP με 5000 Εποχές 67
Σχήμα 4.24 Σφάλμα για MLP Δίκτυο για 10000 Εποχές 68
Σχήμα 4.25 Προσέγγιση Συνάρτησης $y=x\sin(x^2)$ με MLP με 10000 Εποχές 68
Σχήμα 4.26 Σφάλμα για MLP Δίκτυο με Learning Rate 0.01 και 0.001 69
Σχήμα 4.27 Προσέγγιση Συνάρτησης $y=x\sin(x^2)$ με MLP με Learning Rate 0.01 και 0.001 69
Σχήμα 4.28 Σφάλμα για MLP Δίκτυο με Learning Rate 0.1 και 0.01 70
Σχήμα 4.29 Προσέγγιση Συνάρτησης $y=x\sin(x^2)$ με MLP με Learning Rate 0.1 και 0.01 70
Σχήμα 4.30 Σφάλμα για MLP Δίκτυο με Learning Rate 0.3 και 0.03 71
Σχήμα 4.31 Προσέγγιση Συνάρτησης $y=x\sin(x^2)$ με MLP με Learning Rate 0.3 και 0.03 71
Σχήμα 4.32 Σφάλμα Συνάρτησης $y=\sin(x/10) + (x/50)^2$ 74
Σχήμα 4.33 Προσέγγιση Συνάρτησης $y=\sin(x/10) + (x/50)^2$ 74
Σχήμα 4.34 Σφάλμα Συνάρτησης $y=\sin(2\pi x)$ 75
Σχήμα 4.35 Προσέγγιση Συνάρτησης $y=\sin(2\pi x)$ 75
Σχήμα 4.36 Σφάλμα Συνάρτησης $y=2x^2+e^{(\pi/x)}*\sin(2\pi x)$ 76
Σχήμα 4.37 Προσέγγιση Συνάρτησης $y=2x^2+e^{(\pi/x)}*\sin(2\pi x)$ 76
Σχήμα 4.38 Σφάλμα Συνάρτησης $y=x*\sin(x^2)$ 77
Σχήμα 4.39 Προσέγγιση Συνάρτησης $y=x*\sin(x^2)$ 77
Σχήμα 4.40 Προσέγγιση Συνάρτησης $y=\sin(x/10) + (x/50)^2$ για $\sigma=10$ 78
Σχήμα 4.41 Προσέγγιση Συνάρτησης Συνάρτησης $y=\sin(x/10) + (x/50)^2$ για $\sigma=20$ 79
Σχήμα 4.42 Προσέγγιση Συνάρτησης Συνάρτησης $y=\sin(x/10) + (x/50)^2$ για $\sigma=50$ 79
Σχήμα 4.43 Προσέγγιση Συνάρτησης Συνάρτησης $y=\sin(x/10) + (x/50)^2$ $\sigma=100$ 80
Σχήμα 4.44 Προσέγγιση Συνάρτησης $y=\sin(2\pi x)$ για $\sigma=0.5$ 80
Σχήμα 4.45 Προσέγγιση Συνάρτησης $y=\sin(2\pi x)$ για $\sigma=1$ 81
Σχήμα 4.46 Προσέγγιση Συνάρτησης $y=\sin(2\pi x)$ για $\sigma=2$ 81
Σχήμα 4.47 Προσέγγιση Συνάρτησης $y=\sin(2\pi x)$ για $\sigma=5$ 81
Σχήμα 4.48 Προσέγγιση Συνάρτησης $y=2x^2+e^{(\pi/x)}*\sin(2\pi x)$ για $\sigma=0.5$ 82
Σχήμα 4.49 Προσέγγιση Συνάρτησης $y=2x^2+e^{(\pi/x)}*\sin(2\pi x)$ για $\sigma=1$ 82
Σχήμα 4.50 Προσέγγιση Συνάρτησης $y=2x^2+e^{(\pi/x)}*\sin(2\pi x)$ για $\sigma=2$ 83
Σχήμα 4.51 Προσέγγιση Συνάρτησης $y=2x^2+e^{(\pi/x)}*\sin(2\pi x)$ για $\sigma=5$ 83
Σχήμα 4.52 Προσέγγιση Συνάρτησης $y=x*\sin(x^2)$ για $\sigma=0.5$ 84
Σχήμα 4.53 Προσέγγιση Συνάρτησης $y=x*\sin(x^2)$ για $\sigma=1$ 84
Σχήμα 4.54 Προσέγγιση Συνάρτησης $y=x*\sin(x^2)$ για $\sigma=2$ 84
Σχήμα 4.55 Προσέγγιση Συνάρτησης $y=x*\sin(x^2)$ για $\sigma=5$ 85
Σχήμα 4.56 Προσέγγιση Συνάρτησης $y=\sin(x/10) + (x/50)^2$ για Εκτεταμένο MLP 87
Σχήμα 4.57 Προσέγγιση Συνάρτησης $y=\sin(x/10) + (x/50)^2$ για Απλό MLP 88
Σχήμα 4.58 Προσέγγιση Συνάρτησης $y=\sin(x/10) + (x/50)^2$ με RBF 88
Σχήμα 4.59 Σύγκριση Μεταξύ Πραγματικού, Εκτεταμένου MLP και Υβριδικού Μοντέλου 89
Σχήμα 4.60 Προσέγγιση Συνάρτησης $y=\sin(2\pi x)$ για Εκτεταμένο MLP 90
Σχήμα 4.61 Προσέγγιση Συνάρτησης $y=\sin(2\pi x)$ για Απλό MLP 90

Σχήμα 4.62 Προσέγγιση Συνάρτησης y=sin(2πx) για RBF 91
Σχήμα 4.63 Σύγκριση Μεταξύ Πραγματικού, Εκτεταμένου MLP και Υβριδικού Μοντέλου 91
Σχήμα 4.64 Προσέγγιση Συνάρτησης $y=2x^2+e^{(\pi/x)}*\sin(2\pi x)$για Εκτεταμένο MLP 92
Σχήμα 4.65 Προσέγγιση Συνάρτησης $y=2x^2+e^{(\pi/x)}*\sin(2\pi x)$για Απλό MLP 92
Σχήμα 4.66 Προσέγγιση Συνάρτησης $y=2x^2+e^{(\pi/x)}*\sin(2\pi x)$για RBF 93
Σχήμα 4.67 Προσέγγιση Σύγκριση Μεταξύ Πραγματικού, Εκτεταμένου MLP και Υβριδικού μοντέλου 93
Σχήμα 4.68 Προσέγγιση Συνάρτησης $y=x*\sin(x^2)$ για Εκτεταμένο MLP 94
Σχήμα 4.69 Προσέγγιση Συνάρτησης y=x*sin(x2) για Απλό MLP 94
Σχήμα 4.70 Προσέγγιση Συνάρτησης $y=x*\sin(x^2)$ για RBF 95
Σχήμα 4.71 Σύγκριση Μεταξύ Πραγματικού, Εκτεταμένου MLP και Υβριδικού Μοντέλου 95
Σχήμα 4.72 Προσέγγιση Συνάρτησης $y=\sin(x/10) + (x/50)^2$ με Χρήση Συνόλου Ελέγχου για Εκτεταμένο MLP 97
Σχήμα 4.73 Προσέγγιση Συνάρτησης $y=\sin(x/10) + (x/50)^2$ με Χρήση Συνόλου Ελέγχου για Απλό MLP 97
Σχήμα 4.74 Προσέγγιση Συνάρτησης Σφάλματος της $y=\sin(x/10) + (x/50)^2$ με Χρήση Συνόλου Ελέγχου για RBF 98
Σχήμα 4.75 Σύγκριση Μεταξύ Πραγματικού, Εκτεταμένου MLP και Υβριδικού Μοντέλου με Χρήση Συνόλου Ελέγχου της Συνάρτησης Α 98
Σχήμα 4.76 Προσέγγιση Συνάρτησης y=sin(2πx) με Χρήση Συνόλου Ελέγχου για Εκτεταμένο MLP 99
Σχήμα 4.77 Προσέγγιση Συνάρτησης y=sin(2πx) με Χρήση Συνόλου Ελέγχου για Απλό MLP 99
Σχήμα 4.78 Προσέγγιση Συνάρτησης Σφάλματος της y=sin(2πx) με Χρήση Συνόλου Ελέγχου RBF 100
Σχήμα 4.79 Σύγκριση Μεταξύ Πραγματικού, Εκτεταμένου MLP και Υβριδικού Μοντέλου με Χρήση Συνόλου Ελέγχου της Συνάρτησης της Β 100
Σχήμα 4.80 Προσέγγιση Συνάρτησης $y=2x^2+e^{(\pi/x)}*\sin(2\pi x)$ με Χρήση Συνόλου Ελέγχου για Εκτεταμένο MLP 101
Σχήμα 4.81 Προσέγγιση Συνάρτησης $y=2x^2+e^{(\pi/x)}*\sin(2\pi x)$ με Χρήση Συνόλου Ελέγχου για Απλό MLP 102
Σχήμα 4.82 Προσέγγιση Συνάρτησης Σφάλματος της $y=2x^2+e^{(\pi/x)}*\sin(2\pi x)$ με Χρήση Συνόλου Ελέγχου με RBF 102
Σχήμα 4.83 Σύγκριση Μεταξύ Πραγματικού, Εκτεταμένου MLP και Υβριδικού Μοντέλου με Χρήση Συνόλου Ελέγχου της Συνάρτησης Γ 102
Σχήμα 4.84 Προσέγγιση Συνάρτησης $y=x*\sin(x^2)$ με Χρήση Συνόλου Ελέγχου για Εκτεταμένο MLP 103
Σχήμα 4.85 Προσέγγιση Συνάρτησης $y=x*\sin(x^2)$ με Χρήση Συνόλου Ελέγχου για Απλό MLP 103
Σχήμα 4.86 Προσέγγιση Συνάρτησης Σφάλματος της $y=x*\sin(x^2)$ με Χρήση Συνόλου Ελέγχου RBF 104
Σχήμα 4.87 Σύγκριση Μεταξύ Πραγματικού, Εκτεταμένου MLP και Υβριδικού Μοντέλου με Χρήση Συνόλου Ελέγχου της Συνάρτησης Δ 104
Σχήμα 4.88 Προσέγγιση Συνάρτησης $y=\sin(x/10) + (x/50)^2$ με Χρήση Συνόλου Ελέγχου για Εκτεταμένο MLP 106
Σχήμα 4.89 Προσέγγιση Συνάρτησης $y=\sin(x/10) + (x/50)^2$ με Χρήση Συνόλου Ελέγχου για Απλό MLP 107

Σχήμα 4.90 Προσέγγιση Συνάρτησης Σφάλματος της $y=\sin(x/10) + (x/50)^2$ με Χρήση Συνόλου Ελέγχου για RBF 107
Σχήμα 4.91 Σύγκριση Μεταξύ Πραγματικού, Εκτεταμένου MLP και Υβριδικού Μοντέλου με Χρήση Συνόλου Ελέγχου της Συνάρτησης Α 107
Σχήμα 4.92 Προσέγγιση Συνάρτησης $y=\sin(2\pi x)$ με Χρήση Συνόλου Ελέγχου για Εκτεταμένο MLP 108
Σχήμα 4.93 Προσέγγιση Συνάρτησης $y=\sin(2\pi x)$ με Χρήση Συνόλου Ελέγχου για Απλό MLP 109
Σχήμα 4.94 Προσέγγιση Συνάρτησης Σφάλματος της $y=\sin(2\pi x)$ με Χρήση Συνόλου Ελέγχου RBF 109
Σχήμα 4.95 Σύγκριση Μεταξύ Πραγματικού, Εκτεταμένου MLP και Υβριδικού Μοντέλου με Χρήση Συνόλου Ελέγχου της Συνάρτησης της Β 110
Σχήμα 4.96 Προσέγγιση Συνάρτησης $y=2x^2+e^{(\pi/x)}*\sin(2\pi x)$ με Χρήση Συνόλου Ελέγχου για Εκτεταμένο MLP 110
Σχήμα 4.97 Προσέγγιση Συνάρτησης $y=2x^2+e^{(\pi/x)}*\sin(2\pi x)$ με Χρήση Συνόλου Ελέγχου για Απλό MLP 111
Σχήμα 4.98 Προσέγγιση Συνάρτησης Σφάλματος της $y=2x^2+e^{(\pi/x)}*\sin(2\pi x)$ με Χρήση Συνόλου Ελέγχου με RBF 111
Σχήμα 4.99 Σύγκριση Μεταξύ Πραγματικού, Εκτεταμένου MLP και Υβριδικού Μοντέλου με Χρήση Συνόλου Ελέγχου της Συνάρτησης Γ 112
Σχήμα 4.100 Προσέγγιση Συνάρτησης $y=x*\sin(x^2)$ με Χρήση Συνόλου Ελέγχου για Εκτεταμένο MLP 112
Σχήμα 4.101 Προσέγγιση Συνάρτησης $y=x*\sin(x^2)$ με Χρήση Συνόλου Ελέγχου για Απλό MLP 113
Σχήμα 4.102 Προσέγγιση Συνάρτησης Σφάλματος της $y=x*\sin(x^2)$ με Χρήση Συνόλου Ελέγχου RBF 113
Σχήμα 4.103 Σύγκριση Μεταξύ Πραγματικού, Εκτεταμένου MLP και Υβριδικού Μοντέλου με Χρήση Συνόλου Ελέγχου της Συνάρτησης Δ 114

ΠΕΡΙΛΗΨΗ

Ερίσα Σκέντα, Μ.Δ.Ε., Τμήμα Μηχανικών Η/Υ και Πληροφρικής, Πανεπιστήμιο Ιωαννίνων, Ιούλιος 2018.
Προσέγγιση και αναπαράστασή δεδομένων μιας διάστασης με συνδυασμό MLP και RBF Νευρωνικών δικτύων.
Επιβλέπων: Ισαάκ Λαγαρής, Καθηγητής.

Η παρούσα διατριβή πραγματεύεται την σύγκριση ενός εκτεταμένου Νευρωνικου Δικτύου(τύπου Multilayer Perceptron-MLP) με ένα υβριδικό σύστημα Νευρωνικών Δικτυών που αποτελείται από την εν σειρά σύνδεση μιας απλούστερης μορφής του πρώτου(MLP) και ενός RBF δικτύου. Τα Τεχνητά Νευρωνικά Δίκτυα μπορούν να λύσουν ορισμένης φύσεως προβλήματα σε διάφορους επιστημονικούς κλάδους. Το ενδιαφέρον για αυτά παραμένει μετά από τόσα χρόνια μελέτης τους, αυξημένο χάριν στην υπολογιστική τους ισχύ και στην ικανότητα τους να μαθαίνουν και να γενικεύουν.
Το υβριδικό μας δίκτυο εκπαιδεύεται συνδυάζοντας τη μέθοδο μάθησης με επίβλεψη και την μάθηση χωρίς επίβλεψη. Με την πρώτη μέθοδο χωρίς επίβλεψη και πιο συγκεκριμένα με τον αλγόριθμο ομαδοποίησης k-means ορίζονται τα κέντρα του RBF ενώ για την αναπροσαρμογή των βαρών στα δυο δίκτυα τύπου MLP χρησιμοποιείται ο αλγόριθμος backpropagation.
Ο αλγόριθμος backpropagation πραγματοποίει την αναπροσαρμογή των βαρών σύμφωνα με την τεχνική βελτιστοποιήσης απότομης καθόδου (gradient decent). Δηλαδή, υπολογίζει την κλίση του σφάλματος. Σε κάθε βήμα του αλγορίθμου, ενημερώνονται τα βάρη με βάση την κλίση του σφάλματος και το ρυθμό μάθησης. Η ενημέρωση των βαρών επαναλαμβάνεται μέχρις ότου να βρεθούν εκείνα τα βάρη που ελαχιστοποιούν το σφάλμα.
Η επιδίωξη είναι να γίνει σύγκριση του υβριδικού δικτύου και ενός εκτεταμένου MLP όσον αφορά την ποιότητα της γενίκευσης. Για την επίτευξη της σύγκρισης αναπτύξαμε κατάλληλο κώδικα για την υλοποίηση των δικτυών και την διαδικασία της εκπαίδευσης τους. Παρατηρώντας τα αποτελέσματα, παρατηρήσαμε ότι ανεξαρτήτως συνάρτησης που εξετάσαμε, η γενικευτική ικανότητα του υβριδικού δικτύου ήταν ισάξια ή και καλύτερη από το εκτεταμένο MLP μιας και προσέγγιζε πολύ καλά τη συνάρτηση και παρουσίαζε μικρότερο σφάλμα σε σχέση πάντα με το σύνθετο MLP

EXTENDED ABSTRACT

Erisa Skenda, M.Sc., Department of Computer Science & Engineering, University of Ioannina, Greece, September 2018.
Approximation and representation of data of one dimension with combination of MLP and RBF Neural network.
Advisor: Isaak Lagaris, Professor.

One of the most rapidly growing scientific fields is the one of the artificial networking. During the past years more and more neural-network based applications are being developed and applied in many and different environments. This exponential growth, arises from the fact that these artificial networks require a small computational power and at the same time they present a great capability in learning and generalize.

In this thesis, we compare an extended Neural Network (Multilayer Perceptron-MLP) with a hybrid network system, which consists of a small MLP(less number of nodes than extended MLP) and a RBF neural network. To train this hybrid network, we combine a supervised with an unsupervised learning method.

The unsupervised learning method, and more specifically the k-means algorithm is used to find the centers of the RBF nodes. On the other hand, MLP networks utilize a supervised learning technique called bagkpropagation for adjusting the weights. Backpropagation is used by the gradient descent optimization algorithm to adjust the weight of neurons by calculating the gradient of the loss function. In every step of the algorithm, the weights are updated on the basis of the learning rate and the slope of error. The weights are updated until the algorithm find the ones which minimize the error.

The main target is to compare the hybrid network with an extended MLP, in terms of generalizing ability. In order to achieve this goal, we've developed a specific code, for the implementation of networks and their learning procedure. After examination of the results, we noticed that the capability of the Hybrid network, to approximate each function examined, was almost similar or better to the one of the extended MLP. The approximation of the functions was similarly good whereas the error of the Hybrid was less than the extended MLP

ΚΕΦΑΛΑΙΟ 1.
ΕΙΣΑΓΩΓΗ

1.1 Στόχοι
1.2 Δομή της Διατριβής

Στο κεφάλαιο αυτό παρουσιάζονται οι στόχοι και η δομή της διατριβής:

1.1. Στόχοι

Ο γενικός στόχος της διατριβής είναι η δημιουργία μιας ολοκληρωμένης μελέτης σύγκρισης ενός εκτεταμένου Νευρωνικού δικτύου(τύπου Multilayer Perceptron-MLP) με ένα υβριδικό σύστημα νευρωνικών δικτυών που αποτελείται από τον συνδυασμό ενός μικρού(ως προς αριθμό κομβων) MLP και ενός RBF δικτύου. Πιο αναλυτικά οι επί μέρους στόχοι καταμερίζονται ως εξής:

- Να αναλυθεί ο τρόπος επιλογής των παραμέτρων σχεδίασης των νευρωνικων δικτυών MLP και RBF.
- Να αναπτυχθεί κώδικας που υλοποιεί τις αρχιτεκτονικές των δικτυών καθώς και την διαδικασία εκπαίδευσης των δικτύων
- Να συγκριθεί η απόδοση του εκτεταμένου νευρωνικού δικτύου με το υβριδικό σύστημα νευρωνικών δικτυών, πάνω σε διάφορες συναρτήσεις, ως προς κάποιες μετρικές αξιολόγησης απόδοσης.
- Να εξαχθούν συμπεράσματα από τα αποτελέσματα για περαιτέρω έρευνα

1.2. Δομή της Διατριβής

Η διατριβή περιέχει 5 Κεφάλαια: Το Κεφάλαιο 1 είναι το τρέχον και περιέχει τους στόχους και τη δομή της διατριβής. Το Κεφάλαιο 2 αναφέρεται στην διερεύνηση και ανάλυση του θεωρητικού υποβάθρου των νευρωνικών δικτυών ξεκινώντας από την θεωρία της βελτιστοποίησης και ολοκληρώνεται με την θεωρία για τα νευρωνικά δίκτυα. Το Κεφάλαιο 3 παρουσιάζει την μεθοδολογία που προτείνουμε

για την σχεδίαση και υλοποίηση του δικτύου. Το Κεφάλαιο 4 παρουσιάζει το πρακτικό κομμάτι υλοποίησης του δικτύου .Τέλος, το Κεφάλαιο 5 παρουσιάζει τα συμπεράσματα και τις προτάσεις που εξήλθαν κατά την εκπόνηση και συγγραφή της παρούσας διατριβής.

ΚΕΦΑΛΑΙΟ 2.
ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΕ ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

2.1. Βελτιστοποίηση
2.2. Νευρωνικό Δίκτυο
2.3. Τεχνητοί Νευρώνες
2.4. Μοντέλα Νευρωνικών Δικτύων
2.5. Τρόποι Εκπαίδευσης Νευρωνικών Δικτύων
2.6. Εκπαίδευση Δικτύου MLP
2.7. Εκπαίδευση Δικτύου RBF
2.8. Ικανότητα Γενίκευσης
2.9. Η Μέθοδος Cross-Validation

2.1. Βελτιστοποίηση

Ως βελτιστοποίηση αναφέρεται, όχι απλά η προσπάθεια επίλυσης προβλημάτων, αλλά η επίλυσή τους με τον καλύτερο δυνατό τρόπο. Λόγω της ευρύτητάς της, χρησιμοποιείται σε πολλά επιστημονικά, τεχνολογικά και πρακτικά προβλήματα όπως για παράδειγμα στην διαχείρισή αμοιβαίων κεφαλαίων (βελτιστοποίηση χαρτοφυλακίου), τη σχεδίαση της λειτουργίας βιομηχανικών εγκαταστάσεων με σκοπό τη μεγιστοποίηση της παραγωγής ή ελαχιστοποίηση απωλειών ούτως ώστε εν τέλει να μεγιστοποιείται το κέρδος από αυτή. Επίσης, η βελτιστοποίηση χρησιμοποιείται στην μελέτη συνθέτων μορίων με τη μέθοδο της ‘’’Μοριακής Μηχανικής’’ που ελαχιστοποιεί την δυναμική ενέργεια του μορίου, τη σχεδίαση δικτυών υπολογιστών που να εγγυώνται την ελαχιστοποίηση του χρόνου καθυστέρησης ή άλλων δεικτών ποιότητας της επικοινωνίας ή ακόμα και για την επίλυση αλγεβρικών συστημάτων και διαφορικών εξισώσεων ή την εκπαίδευση ‘’Νευρωνικών Δικτυών’’ καθώς και σε πολλούς άλλους τομείς.

Το πρώτο βήμα για την επίτευξη της βελτιστοποίησης αποτελεί η μαθηματική περιγραφή της λεκτικής διατύπωσης ενός προβλήματός. Στην πλειονότητα των περιπτώσεων, η μαθηματική περιγραφή μας οδηγεί στην βελτιστοποίηση μιας συνάρτησής όπου οι παράμετροι μεταξύ τους δεν είναι ανεξάρτητες, άλλα συνδέονται μέσω κάποιων συναρτησιακών σχέσεων, ισοτήτων ή και ανισοτήτων. Η συνάρτηση που επιθυμούμε να βελτιστοποιήσουμε καλείται αντικειμενική συνάρτηση και οι συναρτησιακές σχέσεις ονομάζονται περιορισμοί. [1]

Η διαδικασία καθορισμού της αντικειμενικής συνάρτησής, των παραμέτρων και περιορισμών ενός δοθέντος προβλήματος ονομάζεται μοντελοποίηση (modeling). Η κατασκευή ενός κατάλληλου μοντέλου είναι το πρώτο βήμα και το πιο σημαντικό στην διαδικασία βελτιστοποιήσεις.

Κάθε μαθηματική περιγραφή ενός προβλήματος βελτιστοποίησης αποτελεί ουσιαστικά απλούστευση του πραγματικού προβλήματος επιλογής της καλύτερης λύσης απ' ένα σύνολο πιθανών λύσεων. Το όρισμα (είσοδος) της αντικειμενικής συνάρτησης αναπαριστά την επιλογή των δεδομένων εισόδου της. Οι περιορισμοί στους οποίους υπόκειται το πρόβλημα αποτελούν τις προδιάγραφες του προβλήματος ή /και συνθήκες που πρέπει να ικανοποιούν οι λύσεις. Τέλος, η αντικειμενική συνάρτηση δεν είναι τίποτα άλλο από το κόστος υιοθέτησης κάποιας λύσης. Η επίλυση ενός προβλήματος βελτιστοποίησης ισοδυναμεί με την εύρεση εκείνης της λύσης που οδηγεί στο ελάχιστο (ή μέγιστο) κόστος, σε σχέση με όλες τις πιθανές λύσεις που ικανοποιούν τις προδιάγραφες λειτουργίας.

Υπάρχουν διάφορες κατηγορίες προβλημάτων και αντίστοιχα διάφορες κατηγορίες μεθόδων βελτιστοποίησης. Μόλις το μοντέλο δημιουργηθεί κάποιος αλγόριθμος βελτιστοποίησης πρέπει να χρησιμοποιηθεί ώστε να βρει τη λύση. Δεν υπάρχει ενιαία μέθοδος για όλα τα προβλήματα βελτιστοποίησης. Ως εκ τούτου, έχουν αναπτυχθεί πολλές μέθοδοι για την επίλυση διάφορων τύπων από προβλήματα βελτιστοποίησης. Εξαιτίας του μεγάλου φάσματος προβλημάτων αλλά και των μεθόδων βελτιστοποίησης καθώς και σε συνάρτηση με την ολοένα αυξανομένη πολυπλοκότητα των προβλημάτων είναι δύσκολο να υπάρχει εξάρτηση μόνο από μια μέθοδο.

Οι μέθοδοι βελτιστοποιήσεις διαχωρίζονται σε κατηγορίες με βάση τον τύπο της αντικειμενικής τους συνάρτησης, τις τιμές των παραμέτρων και τους περιορισμούς. Παρακάτω παρουσιάζεται η κατηγοριοποίηση αυτή πιο αναλυτικά:

- Γραμμική βελτιστοποίηση(Linear optimization or linear programming): Μελετάει περιπτώσεις όπου η αντικειμενική συνάρτηση και οι περιορισμοί είναι γραμμική.
- Μη Γραμμική βελτιστοποίηση (Nonlinear optimization or nonlinear programming): Εφαρμόζεται σε προβλήματα βελτιστοποίησης όπου τουλάχιστον μια συνάρτηση δεν είναι γραμμική
- Κυρτή Βελτιστοποίηση (Convex optimization): Μελετάει προβλήματα όπου η αντικειμενική συνάρτηση είναι κυρτή και το σύνολο των εφικτών λύσεων(λύσεις που ικανοποιούν τους περιορισμούς) κυρτό.
- Τετραγωνική Βελτιστοποίηση (Quadratic optimization or quadratic optimization): Περιλαμβάνει την ελαχιστοποίηση της τετραγωνικής συνάρτησης και των γραμμικών περιορισμών
- Στοχαστική Βελτιστοποίηση (Stochastic optimization): Αναφέρεται στη βελτιστοποίηση υπό την παρουσία της τυχαιότητας η οποία εισάγεται στη συνάρτηση εκτίμησης είτε ως θόρυβος είτε ως πιθανοτική επιλογή των μεταβλητών και παραμέτρων βασιζομένη σε στατιστικές κατανομές. [2]

Ένας διαφορετικός τρόπος κατηγοριοποίησης βασίζεται στη φύση του χώρου αναζήτησης και τις παραμέτρους του προβλήματος. Υπάρχουν περιπτώσεις όπου οι επιτρεπτές τιμές των παραμέτρων είναι διακριτές ή ακέραιες, άλλες όπου οι παράμετροι είναι συνεχείς μεταβλητές και μικτά προβλήματα όπου μερικές παράμετροι παίρνουν διακριτές και οι υπόλοιπες συνεχείς τιμές. Για προβλήματα βελτιστοποιήσης συνεχών συναρτήσεων επιπλέον συνθήκες όπως η ύπαρξη πρώτων η δευτέρων παράγωγων επιτρέπουν την ανάπτυξη αποτελεσματικών μέθοδών υψηλών επιδόσεων.

2.1.1. Τοπική και Ολική Βελτιστοποίηση

Δυο είναι οι κυρίες γενικές στρατηγικές αντιμετώπισης προβλημάτων μη γραμμικής βελτιστοποίησης. Η τοπική και η ολική βελτιστοποίηση. Στις μεθόδους της τοπικής βελτιστοποίησης δεν αναζητούμε τη λύση που ελαχιστοποιεί (ή μεγιστοποιεί) την αντικειμενική συνάρτηση ως προς το σύνολο των εφικτών λύσεων. Αντίθετα εστιάζουμε την προσοχή μας στην εύρεση τοπικά βέλτιστης λύσης που ελαχιστοποιεί (ή μεγιστοποιεί) την αντικειμενική συνάρτηση ως προς

τις εφικτές λύσεις που ανήκουν σε μια γειτονιά. Βέβαια η τοπικά βέλτιστη λύση δεν ταυτίζεται κατ' ανάγκη με το ολικό βέλτιστο. Εν αντίθεσή στην ολική βελτιστοποίηση βρίσκουμε την ολικά βέλτιστη λύση του προβλήματος από το σύνολο των εφικτών λύσεων.

2.1.2. Μαθηματική Αναπαράσταση

Ένα πρόβλημα βελτιστοποίησης μπορεί μαθηματικά να αναπαρασταθεί με πάρα πολλούς τρόπους. Γνωρίζουμε, όμως, ότι ένα πρόβλημα μεγιστοποίησης είναι ισοδύναμο με ένα πρόβλημα ελαχιστοποίησης εάν στην αντικειμενική συνάρτηση βάλουμε αρνητικό πρόσημο ($max_x f(x) = min_x - f(x)$). Συνήθως με τον ορό βελτιστοποίηση, καθώς επίσης και στα πλαίσια της παρούσης, εννοείται η ελαχιστοποίηση της αντικειμενικής συνάρτησής. Εφεξής, για την μαθηματική αναπαράσταση ενός προβλήματος βελτιστοποίησης χρησιμοποιούνται οι συμβολισμοί που ακολουθούν:

χ είναι διάνυσμα μεταβλητών οι οποίες λέγονται επίσης άγνωστες ή παράμετροι

$f(x)$ είναι η αντικειμενική συνάρτηση

$c(x)$ είναι η διανυσματική συνάρτηση των περιορισμών που πρέπει να ικανοποιούν οι μεταβλητές

Και το πρόβλημα βελτιστοποίησης περιγράφεται από το τύπο

$$\min_{x \in D} f(x) \qquad \text{Εξ. 2.1}$$

Όπου $x \in R^n$ είναι το διάνυσμα των μεταβλητών με R^n να είναι n-διαστατός ευκλείδειος χώρος και $D \subseteq R^n$ η εφικτή περιοχή που προκύπτει από το σύνολο περιορισμών του προβλήματος. Σε περίπτωση όμως όπου $D = R^n$ το πρόβλημα βελτιστοποιήσεις ονομάζεται πρόβλημα βελτιστοποίησης χωρίς περιορισμούς και περιγράφεται από τη σχέση 2.2

$$\min_{x \in R^n} f(x) \qquad \text{Εξ. 2.2}$$

Συνοψίζοντας, όταν $D \subset R^n$ το πρόβλημα ονομάζεται πρόβλημα βελτιστοποίησης με περιορισμούς και περιγράφεται από την εξής σχέση

$$\min_{x \in R^n} f(x) \ \text{υπο περιορισμούς} \begin{cases} \boldsymbol{c_i}(\boldsymbol{x}) = 0, & i \in \mathcal{E} \\ \boldsymbol{c_i}(\boldsymbol{x}) \geq 0, & i \in \boldsymbol{I} \end{cases} \ \boldsymbol{i = 1, 2, .., k} \qquad \text{Εξ. 2.3}$$

Όπου η μεταβλητή I και το $\mathcal{E}$ είναι σύνολα με δείκτες των περιορισμών ισότητας(equalities) και ανισότητας(inequalities) αντίστοιχα και k ο αριθμός των περιορισμών

Η εκπαίδευση ενός νευρωνικού δικτύου πρόσθιας τροφοδότησης θεωρείται ότι είναι ένα πρόβλημα βελτιστοποίησης χωρίς περιορισμούς όπου στόχος είναι να ελαχιστοποιηθεί η συνάρτηση κόστους γνωστή ως άθροισμα τετραγωνικού σφάλματος(sum square error(SSE)) ως προς τις παραμέτρους του δικτύου, τα συναπτικά βάρη. [3] Τους όρους νευρωνικό δίκτυο πρόσθιας τροφοδότησης, συναπτικά βάρη και άθροισμα τετραγωνικού σφάλματος θα τους αναλύσουμε και μελετήσουμε στα επόμενα υποκεφάλαια

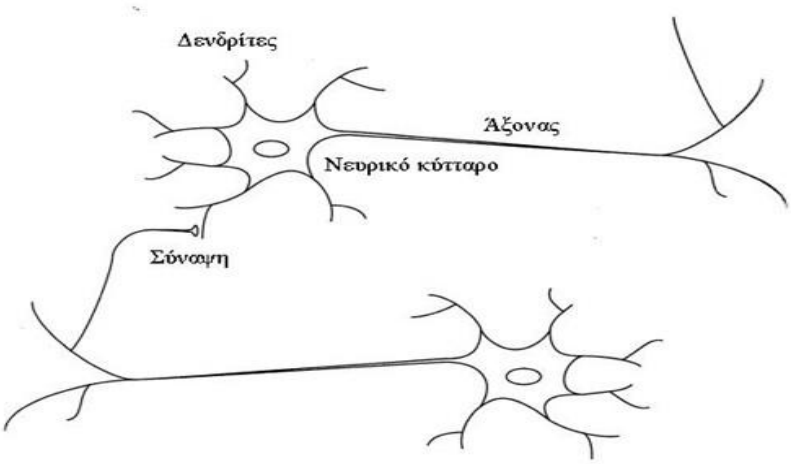
2.2. Νευρωνικό Δίκτυο

Ο ανθρώπινος εγκέφαλος είναι ένα εξαιρετικά εκτεταμένο, μη γραμμικό, παράλληλο σύστημα επεξεργασίας πληροφορίας (υπολογιστική μονάδα). Η δομή του εγκέφαλου είναι τέτοια ώστε επιτρέπει την παράλληλη επεξεργασία δεδομένων και συνεχούς μάθησης μέσω αλληλεπίδρασης με το περιβάλλον. Ο ανθρώπινος εγκέφαλος, αποτελεί ένα Βιολογικό Νευρωνικό Δίκτυο, όπου οι κόμβοι είναι στην πραγματικότητα τα νευρικά κύτταρα (“νευρώνες”).

Το 1943 ο ψυχίατρος McCulloch και ο μαθηματικός Pitts στην μελέτη τους συνδυάσανε την κεκτημένη γνώση από την επιστήμη της νευροψυχολογίας και των μαθηματικών αντίστοιχα ώστε να κατασκευάσουν ένα τεχνητό νευρώνα ο οποίος θα είχε ομοιότητες με τους νευρώνες του ανθρωπίνου εγκέφαλου. Το μοντέλο αυτό ονομάστηκε McCulloch and Pitts [4] .

Ένα <<Τεχνητό Νευρωνικό δίκτυο >>(Artificial neural network), χάριν συντομίας αποκαλείται καθ’ όλη τη διάρκεια της παρούσας διατριβής ΤΝΔ ή <<νευρωνικό δίκτυο>>, είναι ένα απλοποιημένο μοντέλο του κεντρικού νευρικού συστήματος το οποίο προσπαθεί να προσομοίωση το κεντρικό νευρωνικό σύστημα. Τα Τεχνητά Νευρωνικά Δίκτυα μιμούνται τα Βιολογικά Νευρωνικά Δίκτυα (για λόγους συντομίας θα τα αναφέρουμε ως ΒΔΝ) ως προς τη δομή τους και την λειτουργεία τους. Ο ανθρώπινος εγκέφαλος αποτελείται κατά προσέγγιση από 10 δισεκατομμύρια νευρώνες και σε κάθε νευρώνα αντιστοιχούν κατά μέσο ορό περίπου 1000 συνάψεις. [5]

Ο εγκέφαλος οργανώνει τα δομικά του στοιχεία, γνωστά ως νευρώνες με τρόπο ώστε να εκτελούν συγκεκριμένους υπολογισμούς (π,χ αναγνώριση προτύπων και αντίληψη). Κάθε νευρώνας είναι ένα εξειδικευμένο κύτταρο το οποίο έχει τη δυνατότητα μετάδοσης ενός ηλεκτροχημικού σήματος. Ο νευρώνας έχει μια διακλαδωτική διάρθρωση εισροών, τους δενδρίτες (dendrites), ένα κυτταρικό σώμα και μια διακλαδωτική δομή εκροών (τον άξονα). Οι άξονες ενός νευρικού κυττάρου συνδέονται με τους δενδρίτες ενός άλλου, μέσω μιας σύναψης. Όταν, λοιπόν, ένας άξονας ενεργοποιηθεί, πυροδοτεί ένα ηλεκτροχημικό σήμα κατά μήκος του άξονα. Ένας νευρώνας εκτελεί αυτή τη διαδικασία μόνο όταν το συνολικό σήμα το οποίο λήφθηκε από τους δενδρίτες, υπερβεί ένα συγκεκριμένο επίπεδο, δηλαδή, το κατώφλι ενεργοποίησης (firing threshold).



Σχήμα 2.1 Μοντέλο Βιολογικού Νευρώνα

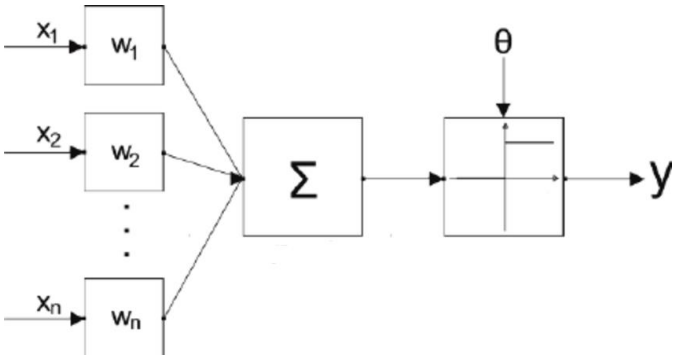
Τα ΤΝΔ όπως αναφέραμε και παραπάνω δημιουργήθηκαν έχοντας ως πρότυπο τα ΒΝΔ. Οπότε ένα ΤΝΔ αποτελείται από ένα σύνολο τεχνητών νευρώνων(απλοί υπολογιστικοί κόμβοι) οι οποίοι είναι διασυνδεδεμένοι μεταξύ τους και αλληλοεπιδρούν δια μέσου των συνάψεων. Ένας απ' τους ορισμούς που δίνεται στη βιβλιογραφία είναι

Ένα νευρωνικό δίκτυο είναι ένας τεράστιος παράλληλος επεξεργαστής με κατανεμημένη αρχιτεκτονική, που αποτελείται από απλές μονάδες επεξεργασίας και έχει από τη φύση του τη δυνατότητα αποθήκευσης των εμπειρικών γνώσεων και την ανακάλεση των γνώσεων για χρήση. Το ΤΝΔ Μοιάζει με τον εγκέφαλο σε δυο σημεία:

1. Το δίκτυο αποκτά τη γνώση από το περιβάλλον μέσω της διαδικασίας μάθησης-εκπαιδεύσεις.
2.Οι δυνάμεις των συνδέσεων μεταξύ των νευρώνων, γνωστές ως συναπτικά βάρη, χρησιμοποιούνται για την αποθήκευση της επιθυμητής γνώσης. [6]
Εδώ αξίζει να σημειωθεί ότι, παρόλο που κυριολεκτικά ο όρος νευρώνας αναφέρεται στον βιολογικό νευρώνα(νευρικό κύτταρο), ως νευρώνας στα επόμενα κεφάλαια θα αναφέρεται ο τεχνητός νευρώνας.

2.3. Τεχνητοί Νευρώνες

Σε αυτό το υποκεφάλαιο θα παρουσιάσουμε κάποια μοντέλα τεχνητών νευρώνων καθώς και σημαντικά χαρακτηριστικά τους όπως για παράδειγμα τη συνάρτηση ενεργοποίησής. Πριν προχωρήσουμε στην επεξήγηση λειτουργίας και δομής των νευρώνων, θα παρουσιάσουμε το πρώτο μοντέλο για τεχνητό νευρώνα των McCulloch and Pitts, το οποίο ήταν το θεμέλιο για την εξέλιξη στο σχεδιασμό καινούργιων μοντέλων για νευρώνες.



Σχήμα 2.2 Μοντέλο των McCulloch και Pitts για τον Νευρώνα

Το μοντέλο όπως παρατηρούμε στο σχήμα 2.2 ξεκινώντας από τα αριστερά προς τα δεξιά αποτελείται από τις εξωτερικές εισόδους οι οποίες είναι οι $\chi_1,\chi_2,..\chi_n$. Το κόμβο αθροιστή του μοντέλου ο οποίος υπολογίζει ένα γραμμικό συνδυασμό των εξωτερικών εισόδων με τα συναπτικά βάρη($w_1,..w_n$)(είναι σημαντικό να σημειωθεί ότι οι συνάψεις του βιολογικού νευρώνα περιγράφονται από τα συναπτικά βάρη) των αντίστοιχων συνδέσεων και επίσης προσθέτει την εξωτερικά εφαρμοζόμενη πόλωση. Το άθροισμα που προκύπτει περιγράφεται από τον τύπο

$$u = \sum_{i=1}^{n} w_i x_i - \theta \qquad \text{Εξ. 2.4}$$

Τέλος το άθροισμα u περνά από τη βηματική συνάρτηση κατωφλιού f(…). Ο νευρώνας παράγει μια έξοδο +1 αν η είσοδος της βηματική είναι θετική και -1 αν η είσοδος είναι αρνητική.

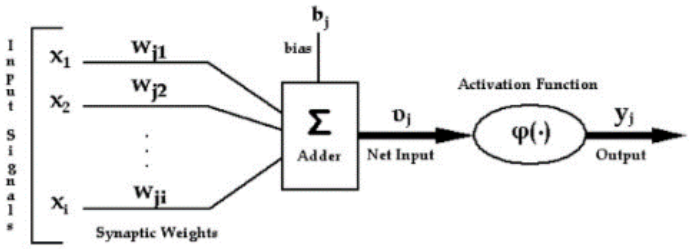
2.3.1. Νευρώνας

Ο τεχνητός νευρώνας είναι μια μονάδα επεξεργασίας πληροφορίας και αποτελεί το βασικό δομικό στοιχείο για την λειτουργεία ενός τεχνητού νευρωνικού δικτύου. Η μαθηματική αναπαράσταση του νευρώνα j που απεικονίζεται στο σχήμα 2.3 περιγράφεται από το ζεύγος των εξισώσεων και 2.5 και 2.6

$$u_j = \sum_{j=1}^{n} w_j x_j \qquad \text{Εξ. 2.5}$$

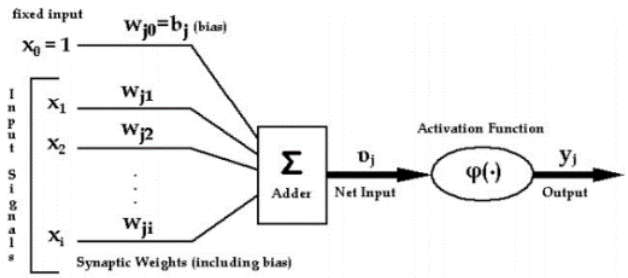
$$\mathrm{y} = \varphi\left(\sum_{j=1}^{n} w_j x_j + b\right) \qquad \text{Εξ. 2.6}$$

όπου το y είναι η έξοδος του, w=[w_1 , …,w_n] είναι το διάνυσμα με τα συναπτικά βάρη , το x=[x_1 , …,x_n] είναι το διάνυσμα εισόδου, το n είναι το πλήθος των σημάτων εισόδου, u_j είναι η έξοδος του αθροιστή (τοπικό πεδίο) που προκύπτει από τα σήματα εισόδου, b είναι η πόλωση (bias) και φ(…) η συνάρτηση ενεργοποιήσης. Η πόλωση είναι ένας εξωτερικός παράγοντας (bias) που επηρεάζει την δικτυακή διέγερση της συνάρτησης ενεργοποιήσης, είτε την αυξάνει είτε την μειώνει ανάλογα με τον εάν είναι θετική η αρνητική, αντίστοιχα. Επιπλέον η πόλωση καθορίζει την έξοδο του νευρώνα όταν η είσοδος είναι μηδέν.



Σχήμα 2.3 Μοντέλο Τεχνητού Νευρώνα

$$y = \varphi\left(\sum_{j=0}^{n} w_j x_j\right) \qquad \text{Εξ. 2.7}$$



Σχήμα 2.4 Ενναλακτική Μορφή ενός Μοντέλου Τεχνητού Νευρώνα

Το σχήμα 2.3 παρουσιάζει το μοντέλο του τεχνητού νευρώνα όπου μπορεί να διαπιστώσει κανείς ότι απαρτίζεται από τρία μέρη

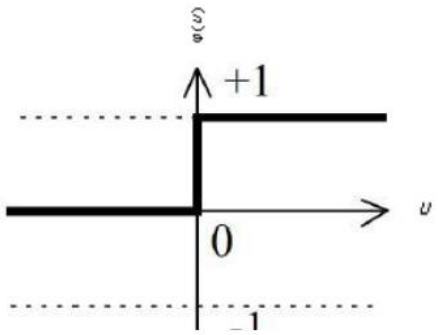
- Τα συναπτικά βάρη(w) Οι είσοδοι συνδέονται στον αθροιστή με ένα σύνολο από διασυνδέσεις (συνάψεις) και κάθε μια από τις συνάψεις έχει το δικό της συναπτικό βάρος. Το βάρος μπορεί να έχει θετική η αρνητική τιμή για χάριν συντομίας όμως και απλότητας στη παρούσα διατριβή θα τα ονομάζουμε βάρη.
- Τον αθροιστή (Σ) στον οποίο τα επηρεασμένα απ' τα βάρη σήματα εισόδου και η πόλωση προστίθενται και παράγουν το τοπικό πεδίο(το τοπικό πεδίο μπορεί να οριστεί και ως γραμμικός συνδυαστής, οι δυο όροι είναι ταυτόσημοι) το οποίο είναι ένας γραμμικός συνδυασμός των εισόδων με συντελεστές τα προσαρμοζόμενα συναπτικά βάρη.
- Τη Συνάρτηση ενεργοποίησης (φ) ή γνωστή και ως συνάρτηση μεταφοράς. Η συνάρτηση ενεργοποίησης δέχεται ως είσοδο το τοπικό πεδίο που υπολογίζεται από

τον αθροιστή και παράγει το σήμα έξοδού του νευρώνα. Η έξοδος που παράγεται υφίσταται από την συνάρτηση ενεργοποιήσης περιορισμό του πλάτους σε κάποια πεπερασμένη τιμή.

Στο μοντέλο που απεικονίζεται στο σχήμα 2.2 η πόλωση εφαρμόζεται εξωτερικά. Το μοντέλο του νευρώνα μπορεί να τροποποιηθεί πραγματοποιώντας δυο αλλαγές, πρώτα προσθέτοντας ένα νέο σήμα εισόδου σταθερό στο +1 και έπειτά προσθέτοντας ένα νέο συναπτικό βάρος ίσο με την πόλωση b. Η μαθηματική αναπαράσταση του νευρώνα αλλάζει και παίρνει την μορφή όπως δείχνει η Εξ. 2.7 ενώ το μοντέλο του νευρώνα αλλάζει και γίνεται όπως απεικονίζεται στο σχήμα 2.4. Τα μοντέλα που απεικονίζονται στο σχήμα 2.3 και στο σχήμα 2.4 παρόλο που διαφέρουν είναι μαθηματικά ισοδύναμα

2.3.2. Συναρτήσεις Ενεργοποίησης

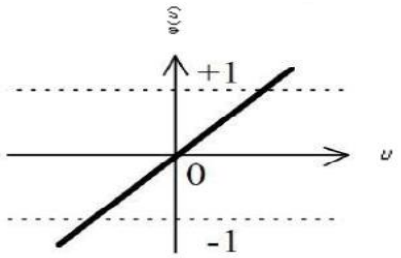
Οι συναρτήσεις ενεργοποίησης όπως προαναφέραμε δέχονται μια είσοδο και σε συνάρτηση με αυτή παράγουν μια έξοδο, ο συμβολισμός της συνάρτησης ενεργοποίησης είναι φ(u). Υπάρχει πληθώρα συναρτήσεων ενεργοποίησης από τις οποίες μπορεί να επιλεχθεί κάποια για το νευρώνα ενός ΤΝΔ. Σε ένα ΤΝΔ δεν είναι υποχρεωτικό όλοι οι νευρώνες να έχουν την ίδια συνάρτηση ενεργοποίησης. Αυτό όμως υπάγεται στο θέμα σχεδιασμού ενός ΤΝΔ και θα μας απασχολήσει σε επόμενο κεφάλαιο. Ένα κριτήριο επιλογής της συναρτήσεις ενεργοποίησης είναι το διάστημα τιμών από το οποίο θέλουμε να πάρουμε τιμές για την έξοδο του νευρώνα. Μερικές βασικές συναρτήσεις ενεργοποίησης παρουσιάζονται παρακάτω.



Σχήμα 2.5 Βηματική Συνάρτηση

$$\varphi(\upsilon) = \begin{cases} 0, \upsilon < 0 \\ 1, \upsilon \geq 0 \end{cases}$$ Εξ. 2.8

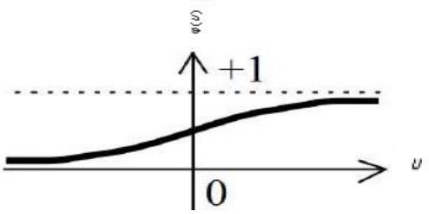
Οι νευρώνες που χρησιμοποιούν βηματική συνάρτηση ως συνάρτηση ενεργοποιήσης αναφέρονται ως μοντέλο McCulloch-Pitts. Το μοντέλο αυτό χρησιμοποιείται όταν η έξοδος πρέπει να είναι δυαδική δηλαδή λαμβάνει τιμή 0 ή 1 .



Σχήμα 2.6 Γραμμική Συνάρτηση

$$\varphi(\upsilon) = \upsilon b$$ Εξ. 2.9

Η γραμμική συνάρτηση ενεργοποίησης χρησιμοποιείται σε νευρώνες ενός νευρωνικού δικτύου οι οποίοι πρέπει να είναι γραμμικοί. Το b δηλώνει την κλίση της ευθείας.



Σχήμα 2.7 Σιγμοειδής Συνάρτηση

$$\varphi(\upsilon) = \frac{1}{1 + e^{-\alpha\upsilon}}$$ Εξ. 2.10

Η σιγμοειδής συνάρτηση μεταφοράς είναι η συνάρτηση που χρησιμοποιείται κατά κόρον. Το σχήμα της γραφικής της παράστασης μοιάζει σαν ένα S και μπορεί να πάρει τιμές στο συνεχές διάστημα τιμών μεταξύ 0 και 1. Η συνάρτηση είναι μη συμμετρική, μη γραμμική, είναι διαφορίσιμη και αυξάνεται μονοτονικά. Η παράμετρος κλίσης=α, όταν αλλάζει η παράμετρος κλίσης η κλίση της συνάρτησης μεταβάλλεται ενώ όταν η παράμετρος κλίσης τείνει στο άπειρό η σιγμοειδής συνάρτηση τείνει προς τη βηματική συνάρτηση άρα προκύπτει πάλι το μοντέλο McCulloch-Pitts [7].

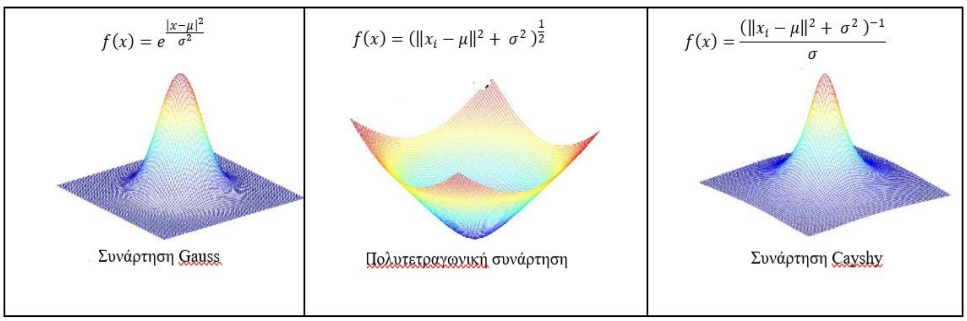
2.3.3. Συνάρτηση Ακτίνας Βάσης (Radial Basis Function)

Η συνάρτηση Ακτίνας Βάσης μπορεί να χρησιμοποιηθεί ως συνάρτηση ενεργοποίησης σε ένα νευρώνα. Η διαφορά από το μοντέλο McCulloch-Pitts έγκειται στα διανύσματα εισόδου, (ο όρος διάνυσμα εισόδου και σήμα εισόδου είναι ταυτόσημοι) τα οποία δεν είναι επηρεασμένα απ' τα βάρη. Για κάθε διάνυσμα είσοδού μετράμε την απόσταση του από ένα σταθερό σημείο το οποίο ονομάζεται κέντρο. Ως κέντρο μπορεί να οριστεί ένα οποιοδήποτε από τα διανύσματα εισόδου. Για τον ορισμό του κέντρου υπάρχουν πολλές μεθοδολογίες τις οποίες όμως δεν θα τις αναλύσουμε στο παρόν υποκεφάλαιο. Για τα διανύσματα εισόδου χ που έχουν μικρή απόσταση από το κέντρο, οι συναρτήσεις ακτίνας βάσης δίνουν υψηλές τιμές ενώ όταν το διάνυσμα εισόδου ταυτίζεται με το κέντρο (χ=μ οι παράμετροι χ και μ εξηγούνται παρακάτω τι δηλώνουν), δίνουν τη μέγιστη τιμή. Καθώς απομακρυνόμαστε ακτινικά από το κέντρο η τιμή της συνάρτησης μειώνεται εκθετικά, σχεδόν εκμηδενίζεται για τα διανύσματα x που είναι μακριά από το κέντρο. Για την εύρεση της απόστασης συνήθως χρησιμοποιείται η ευκλείδεια απόσταση η οποία δίνεται από τον παρακάτω τύπο

$$\gamma(x) = \varphi\left(\sum_{i=1}^{n} \|x_i - \mu\|\right) \qquad \text{Εξ. 2.11}$$

όπου το φ(..) είναι η συνάρτηση ακτίνας βάσης, $\chi=[\chi_1,\ldots,\chi_n]$ είναι το διάνυσμα εισόδου και μ είναι το κέντρο. Η ακτινική συνάρτηση χαρακτηρίζεται επίσης από τη παράμετρο σ_j η οποία καθορίζει την ακτίνα της. Από την τιμή της ακτίνας σ_j καθορίζεται ο ρυθμός μείωσης της συνάρτησης . Για μικρές τιμές της ακτίνας ο

ρυθμός μείωσης είναι μεγάλος, ενώ το αντίθετο συμβαίνει για μεγάλες τιμές της ακτίνας. Μερικές συναρτήσεις ακτινικού τύπου είναι



Σχήμα 2.8 Συναρτήσεις Ακτινικού Τύπου

Από τις συναρτήσεις ακτίνας βάσης αυτή που χρησιμοποιείται ευρέως είναι η Gauss. Από εδώ και πέρα στη παρούσα διατριβή όταν θα αναφερόμαστε σε συνάρτηση ακτίνας βάσης θα εννοούμε την Gauss

2.4. Μοντέλα Νευρωνικών Δικτύων

Όπως προαναφέραμε ένα τεχνητό νευρωνικό δίκτυο(ΤΝΔ) είναι μία αρχιτεκτονική δομή (δίκτυο) αποτελούμενη από ένα πλήθος διασυνδεδεμένων μονάδων επεξεργασίας (τεχνητούς νευρώνες). Τα ΤΝΔ κατηγοριοποιούνται σύμφωνα με την αρχιτεκτονική δομή τους. Η αρχιτεκτονική δομή του δικτύου καθορίζεται από τον τρόπο διασυνδέσεις των νευρώνων, τον τύπο των νευρώνων και το πλήθος των επίπεδων. Σύμφωνα με τον τρόπο διασυνδέσεις των νευρώνων τα ΤΝΔ μπορούν να χαρακτηριστούν ως πλήρως συνδεδεμένα ΤΝΔ ή μερικώς συνδεδεμένα ΤΝΔ. Τα τεχνητά νευρωνικά δίκτυα διαχωρίζονται στις παρακάτω κατηγορίες με βάση την αρχιτεκτονική τους:

- ΤΝΔ Πρόσθιας Τροφοδότησης με ένα επίπεδο (feedforward)

Το πιο απλό δίκτυο που μπορεί να κατασκευαστεί είναι ένα δίκτυο με ένα επίπεδο εισόδου (αποτελούμενο από πηγαίους κόμβους) το οποίο συνδέεται απευθείας με ένα επίπεδο νευρώνων εξόδου (υπολογιστικοί κόμβοι) .Το διάνυσμα εισόδου μεταδίδεται από το επίπεδο εισόδου στο επίπεδο εξόδου αλλά όχι αντίστροφα. Ο χαρακτηρισμός ενός τέτοιου δικτύου ως δίκτυο ενός επιπέδου οφείλεται στο

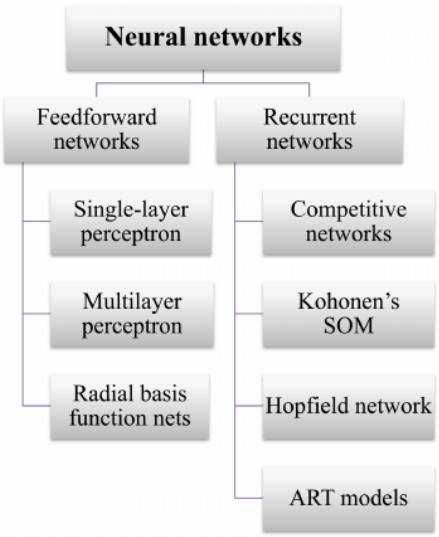
επίπεδο εξόδου το οποίο αποτελείται από υπολογιστικούς κόμβους . Το επίπεδο εισόδου δεν προσμετρείται επειδή δεν εκτελείται κανένας υπολογισμός σ' αυτό.

- Πολυεπιπεδα ΤΝΔ Πρόσθιας Τροφοδότησης (feedforward)

Τα συγκεκριμένα δίκτυα αποτελούνται από ένα η περισσότερα κρυφά επίπεδα των οποίων οι νευρώνες (υπολογιστικοί κόμβοι) ονομάζονται κρυφοί νευρώνες. Οι πηγαίοι κόμβοι στο επίπεδο εισόδου παρέχουν το διάνυσμα εισόδου (σήμα εισόδου) στους νευρώνες (υπολογιστικούς κόμβους) του δευτέρου επιπέδου (πρώτου κρυφού επιπέδου). Οι έξοδοι (σήματα εξόδου) από τους νευρώνες του δευτέρου επιπέδου χρησιμοποιούνται ως είσοδοι στο τρίτο επίπεδο, και αυτό συνεχίζεται μέχρι το τελευταίο επίπεδο του δικτύου δηλαδή το επίπεδο εξόδου.

- ΤΝΔ με ανατροφοδότηση (feedback ή recurrent)

Τα δίκτυα με ανατροφοδότηση έχουν τουλάχιστον ένα βρόχο ανάδρασης .Σε ένα αναδρομικό δίκτυο μπορεί κάθε νευρώνας να τροφοδοτεί το σήμα εξόδου πίσω στις εισόδους όλων των άλλων νευρώνων.



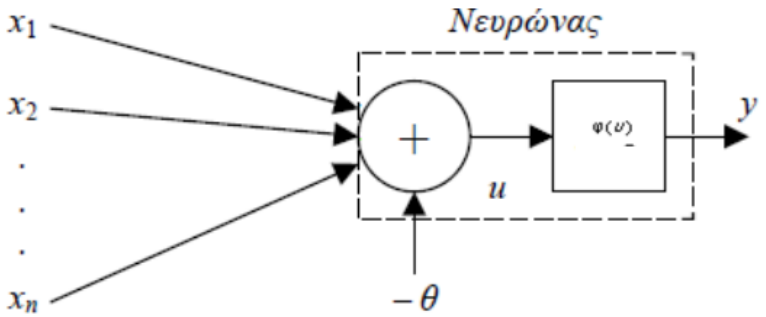
Σχήμα 2.9 Αρχιτεκτονικές Δικτύων

Στο σχήμα 2.10 παρουσιάζονται μερικά ΤΝΔ που ανήκουν στην κατηγορία των ΤΝΔ πρόσθιας τροφοδότησης και κάποια άλλα που ανήκουν στην κατηγορία ΤΝΔ με ανατροφοδότηση. Στην πρώτη κατηγορία ανήκουν τα εξής δίκτυα. Το Δίκτυο Perceptron με ένα επίπεδο (Single Layer Perceptron (SLP)), το Πολυεπιπεδο Perceptron (MultiLayerPerceptron (MLP)) και τα Δίκτυα Ακτινικών Συναρτήσεων

βάσης. Καθόλη τη διάρκεια της παρούσας διατριβής θα μας απασχολήσουν μόνο τα δίκτυα πρόσθιας τροφοδότησης επομένως αυτά τα δίκτυα θα αναλύσουμε και μελετήσουμε πως λειτουργούν στα επόμενα υποκεφάλαια. Για τα δίκτυα με ανατροφοδότηση θα αναφέρουμε επιγραμματικά την ονομασία μερικών απ' αυτά, Ανταγωνιστικά δίκτυα, το δίκτυο SOM , το δίκτυο Hopfield και τα μοντέλα Art.

2.4.1. Perceptron ενός Επιπέδου

Ο Rosenblatt το 1958 εισήγαγε τον νευρώνα (αισθητήρα) perceptron το οποίο ήταν το πρώτο νευρωνικό δίκτυο το οποίο μπορούσε να εκπαιδευτεί. Ο αλγόριθμος εκπαίδευσης που χρησιμοποιείται για τον καθορισμό των τιμών των ελευθέρων παραμέτρων που είναι τα βάρη και η πόλωση του δικτύου, αναπτύχθηκε από τον ίδιο τον Rosenblatt(1958,1962). Ο αισθητήρας (Perceptron) βασίζεται στο μοντέλο McCulloch-Pitts και αποτελεί το πιο απλό νευρωνικό δίκτυο που μπορεί να σχεδιαστεί. Στο σχήμα 2.19 παρουσιάζεται το μοντέλο του νευρώνα Perceptron με n εισόδους το οποίο δέχεται το διάνυσμα εισόδου (σήμα εισόδου) $\chi=(\chi_1,\chi_2,..\chi_n)$ και πρέπει να υπολογίσει την έξοδο y του δικτύου. Οι παράμετροι $(w_1,w_2,..w_n)$ είναι τα συναπτικά βάρη του νευρώνα perceptron ενώ η παράμετρος θ δηλώνει την εξωτερικά εφαρμοζόμενη πόλωση. Ο υπολογισμός του τοπικού πεδίου(γραμμικός αθροιστης) u του νευρώνα δίνεται από τη σχέση 2.4. Το τοπικό πεδίο είναι η είσοδος της συνάρτησης ενεργοποίησης φ(u) η οποία με την σειρά της δίδει την έξοδο y του νευρώνα



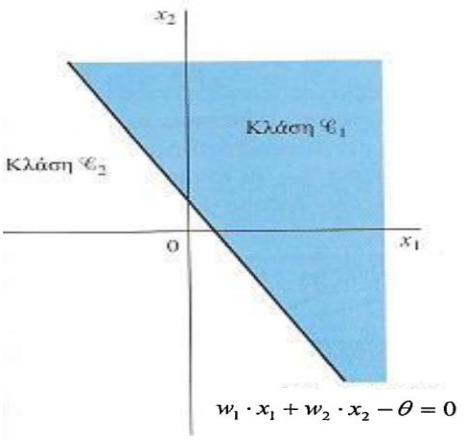
Σχήμα 2.10 Νευρώνας Perceptron

$$u = \sum_{i=1}^{n} w_i x_i - \theta$$ Εξ. 2.12

Στόχος στο δίκτυο Pereptron είναι να ταξινομήσει την είσοδο $\chi=(\chi_1,\chi_2,..\chi_n)$ σωστά σε μια από τις δυο κατηγορίες C_1 ή C_2. Ο κανόνας απόφασης ορίζει ότι η είσοδος χ θα εκχωρηθεί στην κατηγορία C_1 αν η έξοδος y είναι +1 και στην κατηγορία C_2 αν είναι -1. Για να είναι κατανοητός ο τρόπος λειτουργείας του ταξινομητή προτύπων ,απεικονίζονται σε χάρτη οι τελικές περιοχές αποφάσεις που προκύπτουν. Στην περίπτωση του απλούστερου δικτύου perceptron (ο όρος δίκτυο σε αυτή την περίπτωση χρησιμοποιείται καταχρηστικά εφόσον το δίκτυο απαρτίζεται από ένα και μόνο νευρώνα) υπάρχουν δυο περιοχές αποφάσεις διαχωριζόμενες από ένα υπερεπίπεδο το οποίο ορίζεται από τη σχέση

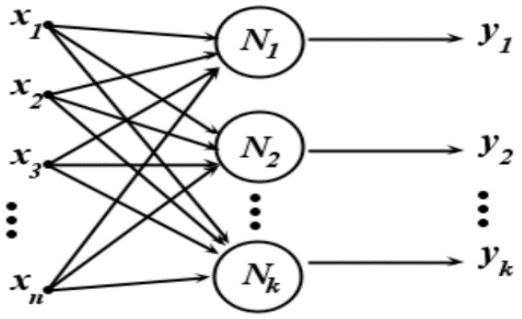
$$\sum_{i=1}^{n} w_i x_i - \theta = 0$$ Εξ. 2.13

Για την περίπτωση δυο μεταβλητών εισόδου χ_1 και χ_2, το διαχωριστικό όριο παίρνει τη μορφή ευθείας γραμμής. Ένα σημείο (χ_1,χ_2) το οποίο βρίσκεται πάνω από την οριακή γραμμή ανήκει στην κατηγορία C1 αντιθέτως το σημείο (χ_1,χ_2) το οποίο βρίσκεται κάτω από τη οριακή γραμμή ανήκει στην κατηγορία C2. Η απόσταση του όριου απόφασης από τη αρχή των αξόνων επηρεάζεται από την τιμή της πόλωσης .



Σχήμα 2.11 Ταξινόμηση Προτύπων σε Δυο Κλάσεις με Βάση την Οριακή Γραμμή

Για να είναι εφικτή η ταξινόμηση προτύπων που ανήκουν σε περισσότερες από δυο κλάσεις το επίπεδο εξόδου του perceptron επεκτείνεται προσθέτοντας νευρώνες έτσι ώστε να περιλαμβάνει περισσοτέρους του ενός νευρώνες. Ωστόσο οι κλάσεις πρέπει να είναι γραμμικά διαχωρίσιμες για να δουλέψει σωστά ο perceptron. Ένα perceptron δίκτυο ενός επιπέδου (SLP) το οποίο έχει περισσότερο απ' ένα νευρώνα στο επίπεδο εξόδου απεικονίζεται στο σχήμα 2.11.



Σχήμα 2.12 Δίκτυο Πρόσθιας Τροφοδότησης με ένα Μεμονωμένο Επίπεδο Νευρώνων

Το απεικονιζόμενο SLP αποτελείται από το επίπεδο εισόδου(αποτελούμενο από n πηγαίους κόμβους) στο οποίο δέχεται τα σήματα εισόδου και το επίπεδο νευρώνων εξόδου. Στο επίπεδο εισόδου δεν πραγματοποιείται κανένας υπολογισμός, το σήμα μεταφέρεται σε μια κατεύθυνση από την είσοδο προς την έξοδο. Οι νευρώνες εξόδου είναι νευρώνες perceptron οπότε η επέκταση της θεωρίας στην περίπτωση περισσοτέρων του ενός νευρώνων είναι απλή.

2.4.2. Πολυεπίπεδο Δίκτυο Perceptron (MLP)

Μια άλλη δομή νευρωνικων δικτυών είναι τα πολυεπίπεδα δίκτυα perceptron (Multilayer Perceptron ή MLP) τα οποία απαρτίζονται από πολλά επίπεδα με νευρώνες (αισθητήρες) Perceptron. Από εδώ και μέχρι το τέλος της παρούσας διατριβής θα αναφέρονται ως "MLP" για χάρη συντομίας. Τα MLP είναι δίκτυα πρόσθιας τροφοδότησης (feedforward) αυτό σημαίνει ότι τροφοδοτούνται από μπροστινές εισόδους μόνο.

Οι νευρώνες είναι οργανωμένοι σε επίπεδα ή στρώματα (layers) και διασυνδέονται μεταξύ τους με τα βάρη, αλλά οι νευρώνες του ίδιου επιπέδου δεν διασυνδέονται μεταξύ τους. Τα επίπεδα συνήθως είναι τα εξής:

- Επίπεδο εισόδου το οποίο απλά στέλνει τα σήματα εισόδου σε όλους τους νευρώνες του κρυφού επιπέδου
- Ένα ή περισσότερα κρυμμένα επίπεδα μη γραμμικών νευρώνων εσωτερικού γινομένου
- Επίπεδο εξόδου το οποίο μπορεί να αποτελείται είτε από γραμμικούς ή από μη γραμμικούς νευρώνες (ή από συνδυασμό και των δυο).

Εφόσον τα δίκτυα MLP είναι δίκτυα πρόσθιας τροφοδότησης οι υπολογισμοί είναι όλοι στην ίδια κατεύθυνση συγκεκριμένα ισχύει ότι η έξοδος από τους νευρώνες ενός επιπέδου τροφοδοτείται ως είσοδος σε νευρώνες του επομένου επιπέδου ενώ η έξοδος από τους νευρώνες του επιπέδου εξόδου κατευθύνεται στο περιβάλλον.

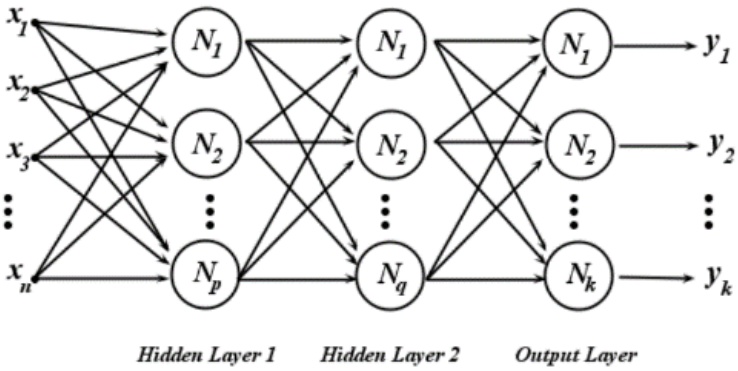
Οι νευρώνες που εμπεριέχονται στο κρυφό επίπεδο ή τα κρυφά επίπεδα καθώς και οι νευρώνες που βρίσκονται στο επίπεδο εξόδου «παράγουν» το σήμα εξόδου χρησιμοποιώντας κάποια από τις συναρτήσεις ενεργοποίησης. Η συνάρτηση ενεργοποίησης είναι συνεχής και παραγωγίσιμη σε κάθε σημείο. [8]

Οι νευρώνες δύο διαδοχικών επιπέδων είναι πλήρη διασυνδεμένα μεταξύ τους αυτό συνεπάγεται ότι κάθε νευρώνας συνδέεται με όλους τους νευρώνες του επόμενου επιπέδου. Συνήθως δεν επιτρέπονται συνδέσεις μεταξύ νευρώνων που ανήκουν σε επίπεδα που δεν είναι διαδοχικά.

Η υπολογιστική ισχύς του πολυεπίπεδου perceptron ωφελείται στην συνάρτηση ενεργοποίησής η οποία είναι διαφορισιμη και στην μεγάλη διασυνδεσιμότητά του δικτύου. Το μειονέκτημα είναι ότι η παρουσία κατανεμημένης μορφής μη γραμμικότητας και η μεγάλη διασυνδεσιμότητά του δικτύου perceptron καθιστούν υπερβολικά δύσκολη τη θεωρητική ανάλυση τους. Δεύτερον η χρήση κρυμμένων νευρώνων δυσκολεύει την κατανόηση και την εύκολη αναπαράσταση της διαδικασίας μάθησης.

Η διαδικασία μάθησης στα MLP υλοποιείται με τη χρήση του δημοφιλούς αλγορίθμου εκπαίδευσης backpropagation [9] [10] [11]. Ο τελευταίος περιλαμβάνει τον αλγόριθμο ελάχιστων τετράγωνων LMS σαν ειδική περίπτωση.

Στο παρακάτω σχήμα απεικονίζεται ένα MLP με δυο κρυμμένα επίπεδα και ένα επίπεδο εξόδου.



Σχήμα 2.13 Πολυεπίπεδο Δίκτυο Perceptron

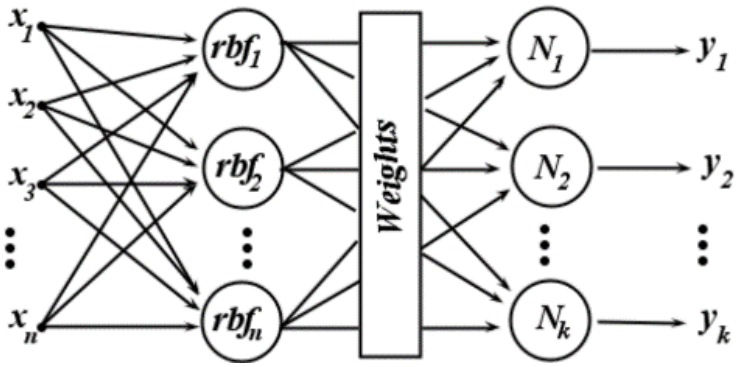
2.4.3. Δίκτυο Ακτινικής Συνάρτησης Βάσης (Radial Basis Function(RBF))

Το δίκτυο ακτινικής συνάρτησης Βάσης (Radial Basis Function ή RBF εφεξής και μέχρι το πέρας της παρούσας διατριβής) προτάθηκε από τον Broomhead Lowe το 1988. Τα δίκτυα ακτινικής Συνάρτησης Βάσης είναι τεχνητά νευρωνικά δίκτυα τα οποία ανήκουν στην κατηγορία δικτυών πρόσθιας τροφοδότησης [12], [13]. Τα δίκτυα αυτά έχουν ισοδύναμες δυνατότητες με τα δίκτυα MLP αλλά έχουν ένα πλεονέκτημα σε σύγκριση με τα τελευταία, έχουν μεγαλύτερη ταχύτητα εκπαίδευσης και αυτό τα καθιστά μια καλή εναλλακτική επιλογής νευρωνικου δικτύου. Το δίκτυο RBF απαρτίζεται συνολικά από δυο επίπεδα, ένα κρυφό επίπεδο και ένα επίπεδο εξόδου, να θυμίσουμε ότι το επίπεδο εισόδου δεν προσμετράτε. Οι νευρώνες στο κρυφό επίπεδο υπολογίζουν την συνάρτηση ακτίνας βάσης, με όρισμα το διάνυσμα εισόδου x. Στο επίπεδο εξόδου οι νευρώνες είναι τυπικοί νευρώνες εσωτερικού γινομένου (όπως και στο MLP) με γραμμική συνάρτηση ενεργοποίησης. Κάθε RBF νευρώνας έχει ισάριθμες εισόδους με τις διαστάσεις του διανύσματος εισόδου, ενώ κάθε νευρώνας του επιπέδου εξόδου έχει ισάριθμες εισόδους με τους RBF νευρώνες. Το RBF είναι πλήρως διασυνδεμένο νευρωνικό δίκτυο. Η έξοδος του δικτύου y_i για ένα διάνυσμα εισόδου $x=(x_1, \ldots, x_d)^T$ υπολογίζεται ως εξής:

$$y_i(x) = w_{oi} + \sum_{j=1}^{n} w_{ij} h_j(x) \qquad \text{Εξ. 2.14}$$

Όλες οι έξοδοι του δικτύου y_i (i=1,..,k) υπολογίζονται με την παραπάνω σχέση, όπου η παράμετρος w_{ij} δηλώνει το βάρος της σύνδεσης από τον κρυμμένο νευρώνα j προς το νευρώνα εξόδου i, n είναι οι κρυμμένοι νευρώνες οι οποίοι υπολογίζουν την συνάρτηση ακτίνας βάσης h_j και w_{oi} είναι η πόλωση του νευρώνα εξόδου i. Οι πολώσεις των νευρώνων που ανήκουν στο επίπεδο εξόδου μπορούν να υλοποιηθούν προσθέτοντας έναν επιπλέον νευρώνα στο κρυφό επίπεδο ο οποίος έχει πάγια τιμή στη συνάρτηση ενεργοποίησης (h_o =1). Κατά την εκπαίδευση ενός δικτύου RBF πρέπει να καθοριστούν οι παράμετροι (γνωστά και με την ονομασία βάρη) του δικτύου, οι οποίες είναι τα κέντρα μ και οι ακτίνες σ από του νευρώνες RBF του δικτύου. Επίσης, πρέπει να καθοριστούν οι παράμετροι w_{ij} του γραμμικού αθροίσματος.

Το δίκτυο RBF (όπως και το MLP) χαρακτηρίζεται από την ιδιότητα της καθολικής προσέγγισης (universal approximation). Η ιδιότητα αυτή μας εξασφαλίζει ότι ένα δίκτυο RBF μπορεί να προσεγγίσει οποιαδήποτε συνάρτηση με οποιαδήποτε ακρίβεια, αυξάνοντας επαρκώς τον αριθμό των νευρώνων RBF. Ωστόσο αυτή η ιδιότητα δεν είναι πρακτικά εκμεταλλεύσιμη, διότι δεν μας προτείνει πόσους RBF νευρώνες να χρησιμοποιήσουμε δοθέντος ενός συνόλου παραδειγμάτων εκπαίδευσης. Το σχήμα 2.13 απεικονίζει ένα δίκτυο RBF με ένα n-διαστατο διάνυσμα εισόδου χ, με n RBF νευρώνες καθώς k νευρώνες εξόδου.



Σχήμα 2.14 Δίκτυο Ακτινικής Συνάρτησης Βάσης

2.5. Τρόποι Εκπαίδευσης Νευρωνικών Δικτύων

Το κύριο χαρακτηριστικό των νευρωνικών δικτύων είναι η εγγενής ικανότητα μάθησης. Ως μάθηση μπορεί να οριστεί η σταδιακή βελτίωση της ικανότητας του

δικτύου να επιλύει κάποιο πρόβλημα (π.χ. η σταδιακή προσέγγιση μίας συνάρτησης). Η μάθηση επιτυγχάνεται μέσω της εκπαίδευσης, μίας επαναληπτικής διαδικασίας σταδιακής προσαρμογής των παραμέτρων του δικτύου (συνήθως των βαρών και της πόλωσής του) σε τιμές κατάλληλες ώστε να επιλύεται με επαρκή επιτυχία το προς εξέταση πρόβλημα. Η εκπαίδευση πραγματοποιείται με τη χρήση ενός αλγορίθμου εκπαίδευσης [14].

Αφού ένα δίκτυο εκπαιδευτεί, οι παράμετροί του συνήθως «παγώνουν» στις κατάλληλες τιμές και από εκεί κι έπειτα είναι σε λειτουργική κατάσταση. Το ζητούμενο είναι το λειτουργικό δίκτυο να χαρακτηρίζεται από ικανότητα γενίκευσης: αυτό σημαίνει πως το δίκτυο δίνει ορθές εξόδους για καινοφανείς εισόδους δηλαδή εισόδους που διαφέρουν από αυτές με τις οποίες εκπαιδεύτηκε.

Ο τρόποι εκπαίδευσης μπορούν να διαχωριστούν στις εξής κατηγορίες [15]

- Εκπαίδευση με επίβλεψη (supervised learning)
- Εκπαίδευση χωρίς επίβλεψη (unsupervised learning)
- Μάθηση με Ενίσχυση (reinforcement learning)

Στη διατριβή αυτή θα εστιάσουμε στη προσέγγιση συναρτήσεων με νευρωνικά δίκτυα. Η εκπαίδευσή των νευρωνικων δικτυών για αυτή τη κατηγορία προβλημάτων πραγματοποιείται με τη μέθοδο, εκπαίδευση με επίβλεψη. Συνεπώς στις παρακάτω παραγράφους θα επικεντρωθούμε και αναλύσουμε κυρίως αυτή τη κατηγορία.

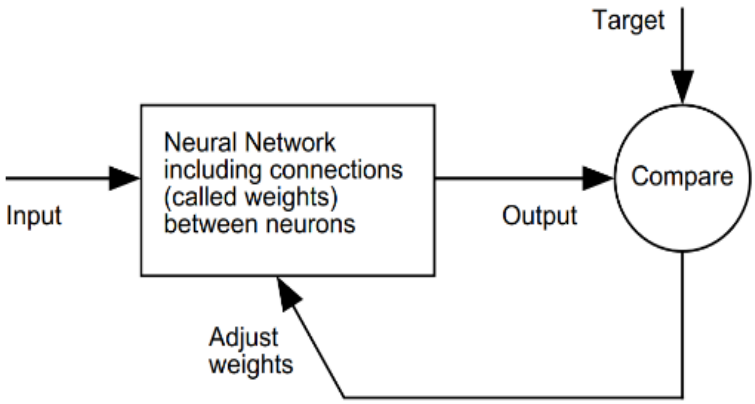
2.5.1. Εκπαίδευση με Επίβλεψη

Στην εκπαίδευση με επίβλεψη για κάθε διάνυσμα εισόδου χ του δικτύου πρέπει να είναι γνωστή η αντίστοιχη επιθυμητή έξοδος t. Η επιθυμητή έξοδος t για κάθε διάνυσμα είσοδού χ ορίζεται από τον εκπαιδευτή του δικτύου. Τα στοιχεία του συνόλου των παραδειγμάτων που χρησιμοποιούνται στη εκπαίδευση είναι ζεύγη της μορφής: (είσοδος, επιθυμητή έξοδος) ($X=\{x_i, t_i\}$, $i=1,\ldots,N$). Αυτά τα ζεύγη ονομάζονται πρότυπα εκπαίδευσης. Αυτή η μέθοδος εκπαίδευσης είναι κατάλληλη για προσέγγιση συναρτήσεων επίσης, είναι κατάλληλη και για προβλήματα ταξινόμησης ή κατηγοριοποίησης διότι και στις δυο περιπτώσεις η επιθυμητή έξοδος t είναι γνωστή.

Σε περίπτωση που εφαρμόσουμε την εκπαίδευση με επίβλεψη για την προσέγγιση συναρτήσεων η επιθυμητή έξοδος t δηλώνει την αναμενομένη τιμή της συνάρτησης ενώ στην περίπτωση που έχουμε πρόβλημα ταξινόμησης η παράμετρος t δηλώνει την ετικέτα κατηγορίας (class label). Κατά τη διαδικασία της εκπαίδευσης ο εκάστοτε αλγόριθμος εκπαίδευσης που χρησιμοποιείται, αναζητά τα βάρη με τα οποία το δίκτυο θα παράγει εκείνες τις εξόδους που ΄΄πλησιάζουν΄΄ πιο πολύ τις επιθυμητές εξόδους. Η διαφορά ανάμεσά στις επιθυμητές εξόδους και τις εξόδους του δικτύου ονομάζεται σφάλμα(error).

Ο αλγόριθμος εκπαίδευσης έχοντας ως κριτήριο το σφάλμα σε κάθε επανάληψη του προσαρμόζει τα βάρη, έως ότου η απόκριση του δικτύου στη βέλτιστη περίπτωση να ταυτιστεί με την επιθυμητή έξοδο ή το σφάλμα να πάρει τη μικρότερη δυνατή τιμή. Μέσω της εκπαίδευσης το δίκτυο αποκτά γνώση η οποία αποθηκεύεται με τη μορφή στα συναπτικών βαρών.

Τέλος η αποκτώμενη γνώση χρησιμοποιείται κατά την ανάκληση του δικτύου. Ως ανάκληση ορίζεται η διαδικασία υπολογισμού της εξόδου του δικτύου για ένα συγκεκριμένο διάνυσμα εισόδου κάνοντας χρήση των βέλτιστων βαρών που βρέθηκαν στην διαδικασία εκπαίδευσης.



Σχήμα 2.15 Ροή της Διαδικασίας Μάθησης με Επίβλεψη

2.5.2. Εκπαίδευση Χωρίς Επίβλεψη

Στην εκπαίδευση χωρίς επίβλεψη είναι γνωστές μόνο οι είσοδοι. Τα στοιχεία του συνόλου των παραδειγμάτων που χρησιμοποιούνται στη εκπαίδευση είναι της

μορφής: (είσοδος) ($X=\{x_i\}$), $i=1,\ldots,N$). Σε αυτή τη κατηγορία εκπαίδευσης το δίκτυο πρέπει μόνο του να ανακαλύψει συσχετίσεις ή ομάδες σε ένα σύνολο δεδομένων, δημιουργώντας πρότυπα, χωρίς να είναι γνωστό αν υπάρχουν, πόσα και ποια είναι. Η εκπαίδευση αυτή είναι κατάλληλη για επίλυση των παρακάτω προβλημάτων, ομαδοποίηση (clustering), μείωση της διάστασης των δεδομένων (dimensionality reduction), οπτικοποίηση (visualisation) και τοπογραφικό Χάρτη Δεδομένων (topographic data map).

2.5.3. Εκπαίδευση με Ενίσχυση (Reinforcement Learning)

Η έννοια της ενισχυτικής μάθησης είναι εμπνευσμένη από τα αντίστοιχη της μάθησης με επιβράβευση και τιμωρία που συναντώνται ως μοντέλα μάθησης των έμβιων όντων. Στην εκπαίδευση με ενίσχυση το δίκτυο γνωρίζει μόνο τις εισόδους οι αποκρίσεις του δικτύου(επιθυμητές έξοδοι) για αυτές τις εισόδους δεν παρέχονται. Τα παραδείγματα που χρησιμοποιούνται στη εκπαίδευση είναι ζεύγη της μορφής: (είσοδος, σήμα ενίσχυσης) ($X=\{x_i, r_i\}$), $i=1,\ldots,N$). Με βάση το σήμα ενίσχυσης για κάθε είσοδο η απόκριση του δικτυού χαρακτηρίζεται ως σωστή ή λάθος και τα βάρη αναπροσαρμόζονται με βάση το σήμα αυτό. Στην περίπτωση που το σύστημα μάθησης παρέχει λάθος έξοδο, ενημερώνεται ότι απάντησε λανθασμένα, αλλά δεν πληροφορείται σχετικά με το ποια είναι η σωστή έξοδος. Αυτή η μέθοδος μάθησης χρησιμοποιείται συνήθως σε εφαρμογές στη ρομποτική και τα παιχνίδια(games).

2.6. Εκπαίδευση Δικτύου MLP

Υπάρχει πληθώρα αλγορίθμων εκπαίδευσης που χρησιμοποιείται για την εκπαίδευση ενός δικτύου MLP, εμείς όμως θα επεξηγήσουμε και εφαρμόσουμε τον αλγόριθμο οπισθοδρομικής διάδοσης (backpropagation), ο οποίος εισηγήθηκε από τον Rumelhart,Hinton και Williams 1986. Κατά την εκπαίδευση ενός MLP σύμφωνα με τον αλγόριθμο οπισθοδρομικής διάδοσης στο δίκτυο παρουσιάζεται και εφαρμόζεται ένα σύνολο διανυσμάτων εκπαίδευσης η γνωστά και ως πρότυπα [16]. Υπάρχουν δυο τρόποι που παρουσιάζονται τα διανύσματα εκπαίδευσης στο νευρωνικό δίκτυο. Η παρουσίαση όλων των προτύπων στο δίκτυο από μια φορά το καθένα ονομάζεται κύκλος εκπαίδευσης ή εποχή.

Στον ένα τρόπο τα διανύσματα εκπαίδευσης παρουσιάζονται σειριακά μέχρι να παρουσιαστούν όλα. Η παρουσίαση του ενός προτύπου μετά το άλλο εξασφαλίζει συνήθως μικρότερο σφάλμα εκπαίδευσης όμως έχει το μειονέκτημα ότι το δίκτυο <<αποστηθίζει >> τη σειρά παρουσιάσεις των προτύπων κατά συνέπεια τα σφάλμα αξιολογήσεις είναι μεγαλύτερο σε σύγκριση με το σφάλμα αξιολογήσεις που προκύπτει από την τυχαία παρουσίαση προτύπων.

Στον δεύτερο τρόπο τα διανύσματα εκπαίδευσης παρουσιάζονται τυχαία και μόνο μια φορά το καθένα ανά εποχή. Με αυτό τον τρόπο επιτυγχάνεται στοχαστική εκπαίδευση του δικτύου όμως έχει το μειονέκτημα των συνεχών ταλαντώσεων του σφάλματος εκπαίδευσης και αξιολογήσεις αυτό συνεπάγεται με πολύ αργή σύγκλιση του αλγόριθμου εκπαίδευσης .

Στην διαδικασία εκπαίδευσης του δικτύου για την ενημέρωση των συναπτικών βαρών υπάρχουν τρεις τρόποι, εκπαίδευση ανά πρότυπο, μαζική εκπαίδευση και συνεχομένη μάθηση.

2.6.1. Εκπαίδευση ανά Πρότυπο (Stochastic Training)

Όταν εφαρμόζεται αυτός ο τρόπος εκπαίδευσης η ενημέρωση των βαρών γίνεται αμέσως μετά την παρουσίαση ενός προτύπου εκπαίδευσης, από το σύνολο των προτύπων εκπαίδευσης, στο δίκτυο. Όταν παρουσιαστεί ένα πρότυπο από το σύνολο εκπαίδευσης στο δίκτυο, πρώτα γίνεται ένα ευθύ πέρασμα προς τα μπρος στο δίκτυο, δηλαδή, από τα αριστερά προς τα δεξιά έπειτα υπολογίζεται το σφάλμα το οποίο διαδίδεται προς τα πίσω (ανάστροφο πέρασμα). Τέλος αναπροσαρμόζονται τα βάρη. Η διαδικασία αυτή γίνεται για κάθε πρότυπο εκπαίδευσης του συνόλου εκπαίδευσης, όταν όλα τα πρότυπα παρουσιαστούν στο δίκτυο τότε ολοκληρώνεται ένας κύκλος εκπαίδευσης. Κατά συνέπεια στην εκπαίδευση ανα πρότυπο η αντικειμενική συνάρτηση που πρέπει να ελαχιστοποιηθεί είναι η στιγμιαία ενέργεια του σφάλματος E_n. Ο Rumelhart,Hinton και Williams 1986 καθώς και Haykin το 1999 έχουν αποδείξει ότι η εκπαίδευση ανά πρότυπο είναι υπολογιστικά πιο γρήγορη και απαιτεί λιγότερη μνήμη από ότι η μαζική μάθηση. Επιπλέον στην εκπαίδευση ανά πρότυπο ο αλγόριθμος αποκτά στοχαστικό χαρακτήρα και αποφεύγεται η περίπτωση παγίδευσης του αλγόριθμου σε τοπικά ελάχιστα

2.6.2. Μαζική Εκπαίδευση (Batch Mode)

Στη μαζική εκπαίδευση επιβλεπομένης μάθησης, τα βάρη ενημερώνονται εφόσον στο δίκτυο παρουσιαστούν όλα τα πρότυπα του συνόλου εκπαίδευσης. Τα πρότυπα παρουσιάζονται με τυχαία σειρά ανά εποχή. Συνεπώς όλα τα βάρη αναπροσαρμόζονται μόνο μια φορά στο τέλος της εποχής με βάση είτε τον κανόνα δέλτα είτε με κάποιον απ' τους άλλους τρόπους. Συνεπάγεται ότι η αντικειμενική συνάρτηση που πρέπει να βελτιστοποιηθεί κατά τη μαζική εκπαίδευση είναι το μέσο τετραγωνικό σφάλμα (E_{av}) του δικτύου. Τα πλεονεκτήματα της μεθόδου μαζικής μάθησης είναι ότι ο υπολογισμός του διανύσματος κλίσης(δηλαδή της παραγωγού της αντικειμενικής συνάρτησης E_{av})υλοποιείται με μεγαλύτερη ακρίβεια. Επίσης, είναι δυνατή η παράλληλη υλοποίηση της εκπαίδευσης.

2.6.3. Συνεχομένη Εκπαίδευση (On-Line Training)

Στη συνεχομένη εκπαίδευση το κάθε διάνυσμα εκπαίδευσης παρουσιάζεται μια φορά στο δίκτυο. Αυτή η μεθοδολογία εκπαίδευσης προτιμάται όταν το πλήθος των διανυσμάτων εκπαίδευσης είναι αρκετά μεγάλο. Επίσης, η συνεχομένη εκπαίδευση χρησιμοποιείται σπάνια οι δυο συνήθης τρόποι εκπαίδευσης είναι δυο προηγούμενοι.

2.6.4. Αλγόριθμος Backpropagation

Ο αλγόριθμος Backpropagation [17] τμηματοποιείται σε δυο <<φάσεις>> σύμφωνα με την εκτέλεση τον υπολογισμών που επιτελεί. [18] Η πρώτη φάση είναι γνωστή ως ευθύ πέρασμα (forward pass) και η δεύτερη ως αντίστροφο πέρασμα(reverse pass). Στην αρχή τα βάρη είναι αρχικοποιημένα σε τυχαίες τιμές στο διάστημα [-1,1], στο ευθύ πέρασμα τα βάρη δεν τροποποιούνται καθόλου. Το σύνολο εκπαιδεύσεις αποτελείται από ζεύγη τιμών D=(x(n),t(n)), n=1,…N όπου x(n) είναι το διάνυσμα είσοδού όπως είδαμε και παραπάνω και το t(n) δηλώνει το επιθυμητό-αναμενόμενο διάνυσμα εξόδου.

Κατά το ευθύ πέρασμα για κάθε διάνυσμα είσοδού χ(n) υπολογίζεται η έξοδός y για κάθε νευρώνα του δικτύου. Συγκεκριμένα το διάνυσμα είσοδού παρουσιάζεται στους νευρώνες-κόμβους που απαρτίζουν το πρώτο επίπεδό, σε αυτούς τους

νευρώνες το σήμα δεν δέχεται καμία επεξεργασία και το σήμα έξοδού που παράγουν είναι [19]

$$y_i(n) = x_i(n) \qquad \text{Εξ.2.15}$$

Έπειτα οι νευρώνες που ανήκουν στο κρυφό επίπεδό και στο επίπεδό έξοδού, λαμβάνουν το σήμα από τους νευρώνες που ανήκουν στο προηγούμενο επίπεδό. Με άλλα λόγια, η έξοδός του ενός επιπέδου γίνεται είσοδος στο επόμενο επίπεδό. Το τοπικό πεδίο για οποιοδήποτε νευρώνα ο οποίος είτε ανήκει στο κρυφό επίπεδό είτε στο επίπεδό έξοδού υπολογίζεται από τη σχέση 2.15

$$Z_j(n) = \sum_{i=0}^{m} w_{ji}^{l}(n) y_i^{l-1}(n) \qquad \text{Εξ.2.16}$$

όπου το l=[I,II,…,0] δηλώνει το σύνολο των επιπέδων, το m δηλώνει το σύνολο των νευρώνων του προηγουμένου επιπέδου και το y^{l-1} την έξοδο i-οστου νευρώνα του προηγουμένου επιπέδου l-1 και η παράμετρος w_{ji} δηλώνει το συναπτικό βάρος της διασύνδεσης που ενώνει τον νευρώνα i και j. Η τιμή z_j ονομάζεται τοπικό πεδίο του νευρώνα j. Φερ ειπείν σε περίπτωση όπου ο νευρώνας j είναι νευρώνας έξοδού τότε αυτός τροφοδοτείται από τα σήματα έξοδού όλων των νευρώνων που ανήκουν στο προηγούμενο επίπεδό. Η έξοδος των νευρώνων που ανήκουν στο κρυφού επιπέδου η στο επιπέδου έξοδού y_j υπολογίζεται από τον τύπο [6]

$$y_j(n) = \varphi(z_i(n)) \qquad \text{Εξ.2.17}$$

όπου η παράμετρος φ(…) δηλώνει τη συνάρτηση ενεργοποίησης. Όταν το σήμα φτάσει στους νευρώνες που ανήκουν στο τελευταίο επίπεδό αρά στο επίπεδό έξοδού κάθε νευρώνας υπολογίζει το διάνυσμα έξοδού του δικτύου. Οι έξοδοι των νευρώνων του τελευταίου επιπέδου συγκροτούν τις εξόδους του νευρωνικού δικτύου.

Το σφάλμα στην έξοδο του j-στου νευρώνα , ο οποίος είναι νευρώνας έξοδού, όταν παρουσιάζεται το n-ιοστο πρότυπό χ(n){n=1,2,..N), υπολογίζεται από τον τύπο

$$e_j(n) = t_j(n) - y_j(n) \qquad \text{Εξ.2.18}$$

όπου $t_j(n)$ είναι η επιθυμητή/αναμενόμενη έξοδός για το νευρώνα j και $y_j(n)$ είναι η έξοδός που υπολογίστηκε από τον νευρώνα j. Με τον ίδιο τύπο υπολογίζεται το σφάλμα στην έξοδο οποιοδήποτε νευρώνα που ανήκει στο επίπεδο έξοδού. Το άθροισμα των τετραγωνικών σφαλμάτων όλων των νευρώνων έξοδού ορίζεται από τον τύπο 2.19

$$E(n)=\frac{1}{2}\sum_{j=1}^{k}\left(t_j\,(n) - y_j(n)\right)^2 \qquad \text{Εξ.2.19}$$

ή

$$E(n)=\frac{1}{2}\sum_{j=1}^{k} e_j^2(n) \qquad \text{Εξ.2.20}$$

όπου η παράμετρος k δηλώνει το σύνολο των κόμβων του επιπέδου έξοδού ενώ η ποσότητα $\frac{1}{2}\ \left(d_j\,(n) - y_j(n)\right)^2$ δηλώνει τη στιγμιαία τιμή του τετραγωνικού σφάλματος για τον νευρώνα j. Η μέση τιμή των σφαλμάτων για όλα τα N πρότυπα δίνεται από τον τύπο [20]

$$E_{av}(n)=\frac{1}{N}\sum_{n=1}^{N} E(n) = \frac{1}{2N}\sum_{n=1}^{N}\sum_{j=1}^{k} e_j^2(n) \qquad \text{Εξ.2.21}$$

Το άθροισμα των τετραγωνικών σφαλμάτων E(n) καθώς και η μέση τιμή των τετραγωνικών σφαλμάτων $E_{av}(n)$ εξαρτώνται από τις πολώσεις και τα συναπτικά βάρη. Κατά την εκπαίδευση στόχος είναι να ελαχιστοποιηθεί η μέση τιμή των σφαλμάτων, αυτό επιτυγχάνεται με την προσαρμογή των βαρών σε κάθε βήμα επανάληψης του αλγόριθμου. Τα βάρη σε κάθε βήμα επανάληψής του αλγόριθμου διαφέρουν από την διόρθωση που επιδέχονται. Πιο συγκεκριμένα για το πρότυπό x(n) που παρουσιάζεται στο δίκτυο κατά την νιοστή επανάληψή, στις εξόδους του δικτύου υπολογίζεται το σφάλμα E(w(n)) με βάση τα βάρη w(n). Στην αμέσως

επομένη επανάληψη (n+1) του αλγόριθμου σύμφωνα με τα βάρη που διορθωθήκαν στο νιοστό βήμα επανάληψης, για το πρότυπό χ(n+1) υπολογίζεται το νέο σήμα σφάλματος στις εξόδους του δικτύου. Η διόρθωση βαρών γίνεται με σκοπό, την ελαχιστοποίηση του τετραγωνικού σφάλματος E(w) και σε κάθε βήμα επανάληψης του αλγόριθμου εκπαίδευσής να ισχύει ότι

$$E(w(n+1)) <= E(w(n)) \qquad \text{Εξ.2.22}$$

όπου το w (n) δηλώνει το διάνυσμα των συναπτικών βαρών στη νιοστή (n) επανάληψή του αλγόριθμου ενώ στο w (n+1) το διάνυσμα των βαρών με τις νέες διορθωμένες τιμές βαρών .Εφόσον ολοκληρωθεί το πέρασμα προς τα μπρος ξεκινάει η δεύτερή φάση, το αντίστροφο πέρασμα.

Η διόρθωση στο βάρος w_{ji} (n) κατά τη νιοστη(n) επανάληψή του αλγόριθμου που συνδέει το νευρώνα j ο οποίος ανήκει στο επίπεδο έξοδού με τον νευρώνα i ο οποίος ανήκει στο τελευταίο κρυφό επίπεδο δίνεται από τον τύπο [21]

$$\Delta w_{ji}(n) = -\eta \frac{\partial E(n)}{\partial w_{ji}(n)} \qquad \text{Εξ.2.23}$$

Η παραπάνω σχέση ονομάζεται κανόνας δέλτα. Για να υπολογίσουμε το Δw_{ji} πρέπει να γνωρίζουμε την κλίση $\frac{\partial E(n)}{\partial w_{ji}(n)}$ οποία καθορίζει την κατεύθυνση ελαχιστοποίησης στο χώρο των βαρών και τη σταθερά η(ητα) η οποία δηλώνει το ρυθμό μάθησης. Σύμφωνα με τον κανόνα της αλυσίδας του διαφορικού λογισμού η μερική παράγωγος $\frac{\partial E(n)}{\partial w_{ji}(n)}$ γράφεται ως εξής

$$\frac{\partial E(n)}{\partial w_{ji}(n)} = \frac{\partial E(n)}{\partial e_j(n)} \cdot \frac{\partial e_j(n)}{\partial y_j(n)} \cdot \frac{\partial y_j(n)}{\partial z_j(n)} \cdot \frac{\partial z_j(n)}{\partial w_{ji}(n)} \qquad \text{Εξ.2.24}$$

όπου το e_j (n) δηλώνει το σφάλμα στην έξοδο του νευρώνα j, το y_j (n) την έξοδο του νευρώνα j στη νιοστη επανάληψή του αλγόριθμου. Από τη παραγώγιση των σχέσεων 2.25, 2.26, 2.27, 2.28 γνωρίζουμε ότι

$$\frac{\partial E(n)}{\partial e_j(n)} = e_j(n) \qquad \text{Εξ.2.25}$$

$$\frac{\partial e_j(n)}{\partial y_j(n)} = -1 \qquad \text{Εξ.2.26}$$

$$\frac{\partial y_j(n)}{\partial z_j(n)} = \varphi'(z_j(n)) \qquad \text{Εξ.2.27}$$

$$\frac{\partial z_j(n)}{\partial w_{ji}\,(n)} = y_i(n) \qquad \text{Εξ.2.28}$$

Η τοπική κλίση $\delta_j(n)$ για τον νευρώνα j δίνεται από την σχέση 2.29

$$\delta_j(n) = \frac{\partial E(n)}{\partial e_j(n)} \cdot \frac{\partial e_j(n)}{\partial y_j(n)} \cdot \frac{\partial y_j(n)}{\partial z_j(n)} = e_j(n) \cdot \varphi'_j\left(z_j(n)\right) \qquad \text{Εξ.2.29}$$

Εν τέλει προκύπτει

$$\Delta w_{ji}(n) = \eta \cdot \delta_j\,(n) \cdot\, y_i(n) \qquad \text{Εξ.2.30}$$

Με βάση τους παραπάνω τύπους διαπιστώνουμε ότι για να υπολογίσουμε τη διόρθωση Δw_{ji} (n) στο βάρος $w_{ji}(n)$ ο πιο σημαντικός παράγοντας είναι το σφάλμα $e_j(n)$ στον νευρώνα j. Όταν ο νευρώνας j ανήκει στο επίπεδό έξοδού επειδή η αναμενομένη έξοδός του νευρώνα είναι γνωστή ευκολά υπολογίζεται το σφάλμα. Επιπλέον για να είναι εφικτή η διόρθωση Δw_{ji} (n) πρέπει πρώτα να υπολογιστεί η τοπική κλίση στο νευρώνα. [20]

Στην περίπτωση που ο νευρώνας αποτελεί εσωτερικό κόμβο του νευρωνικού δικτύου ο υπολογισμός του σφάλματος πραγματοποιείται διαφορετικά διότι για τον

κόμβο αυτό δεν υπάρχει κάποια αναμενομένη έξοδός και το σφάλμα καθορίζεται αναδρομικά σε σχέση με τα σφάλματα όλων των νευρώνων με τους οποίους αυτός συνδέεται απευθείας. Συνεπώς η τοπική κλίση για ένα νευρώνα που ανήκει στο κρυφό επίπεδό δεν μπορεί να υπολογιστεί με την ίδια σχέση που υπολογίζεται για ένα νευρώνα που ανήκει στο εξωτερικό επίπεδό. Η τοπική κλίση για το νευρώνα i ο οποίος ανήκει στο κρυφό επίπεδό δίνεται από την σχέση

$$\delta_i\,(n) = \varphi'_i\left(z_i(n)\right) \cdot \sum_j \delta_j\,(n)\ \cdot\ w_{ji}\,(n) \qquad \text{Εξ.2.31}$$

Παρατηρώντας τη σχέση 2.30 διαπιστώνουμε ότι, η κλίση δ_i για τον νευρώνα i ισούται με το γινόμενο της σχετιζόμενης παραγωγού $\varphi'_i\left(z_i(n)\right)$ και του σταθμισμένου αθροίσματος των τοπικών κλίσεων δ. Οι τελευταίες έχουν υπολογιστεί για τους νευρώνες του επομένου κρυφού επιπέδου ή επιπέδου έξοδού οι οποίοι συνδέονται με τον νευρώνα i. Εφόσον η τοπική κλίση για τον νευρώνα i είναι γνωστή μπορεί να υπολογιστεί η διόρθωση $\Delta w_{il}(n)$ στο βάρος

$$\Delta w_{il}(n) = \eta \cdot \delta_i\,(n) \cdot\ y_l(n) \qquad \text{Εξ.2.32}$$

Η απόδειξη για το πως προέκυψε ο τύπος 2.31 από την αρχική σχέση $\Delta w_{il}(n) = -\eta \frac{\partial E(n)}{\partial w_{il}\,(n)}$ δεν αναφέρεται [20]

Για να συνοψίσουμε στη δεύτερη φάση του αλγόριθμου είδαμε ότι σε κάθε νευρώνα έξοδού του δικτύου υπολογίζεται ένα σφάλμα. Τα σφάλμα είναι απαραίτητο για να υπολογιστεί η τοπική κλίση στον νευρώνα. Σύμφωνα με τον κανόνα δέλτα η τοπική κλίση είναι αναγκαίος παράγοντας για να γίνει αναπροσαρμογή στα βάρη. Στη συνέχεια οι τοπικές κλίσεις απ' τους νευρώνες έξοδού χρησιμοποιούνται ώστε να υπολογιστούν οι τοπικές κλίσεις στους νευρώνες του τελευταίου κρυμμένου επιπέδου, αυτές του τελευταίου κρυμμένου επιπέδου χρησιμοποιούνται για αυτές του προτελευταίου κρυμμένου επιπέδου (εάν υπάρχουν) και ούτω καθεξής, έως το πρώτο κρυφό επίπεδό του δικτύου. Με λίγα λόγια το αντίστροφο πέρασμα ξεκινά από το επίπεδο έξοδού μεταδίδει τα σήματα σφάλματος προς τα πίσω και υπολογίζει αναδρομικά την τιμή της τοπικής κλίσης του κάθε νευρώνα. Οι δυο φάσεις του

αλγόριθμου επαναλαμβάνονται μέχρις ότου τα κριτήρια τερματισμού που έχει ορίσει ο εκπαιδευτής του δικτύου επιτευχθούν.

2.6.5. Κριτήρια Τερματισμού Εκπαίδευσής

Ο πρωτότυπος αλγόριθμος Backpropagation η μέθοδος που χρησιμοποιεί για την ελαχιστοποίηση του σφάλματος είναι η μέθοδος απότομης καθόδου (gradient decent). Στην εκπαίδευση του δικτύου MLP με gradient decent ο εκπαιδευτής, ανάλογα με τη φύση του προβλήματος, αποφασίζει τι κριτήρια τερματισμού του αλγόριθμου θα εφαρμόσει. [17] Πριν προχωρήσουμε στην απαρίθμηση των κριτήριων πρέπει να προσφύγουμε στις μοναδικές ιδιότητες που έχει ένα τοπικό ή καθολικό ελάχιστο της επιφάνειας σφάλματος. Έστω ότι το διάνυσμα βαρών w^* συμβολίζει ένα τοπικό ελάχιστο ή ένα καθολικό ελάχιστο. Μια αναγκαία συνθήκη για να είναι το w^* ένα ελάχιστο είναι ότι το διάνυσμα κλίσης g(w)(δηλαδή η μερική παράγωγος πρώτης τάξης)της επιφάνειας σφάλματος σε σχέση με το διάνυσμα βαρών(w) πρέπει να είναι μηδέν στο $w = w^*$. Συνεπώς τα κριτήρια τερματισμού που προκύπτουν είναι [21]

- Ο αλγόριθμος Backpropagation μπορεί να τερματιστεί όταν η ευκλείδεια νόρμα του διανύσματος κλίσης φτάσει σε ένα επαρκώς μικρό κατώφλι κλίσης.
- Ο αλγόριθμος θεωρείτε ότι μπορεί να τερματιστεί όταν ο απολυτός ρυθμός μεταβολής του μέσου τετραγωνικού σφάλματος ανά εποχή είναι επαρκώς μικρός
- Για να μπορέσει τερματιστεί, έλεγχος της γενικευτικής ικανότητας του δικτύου μετά από κάθε επανάληψη της διαδικασίας μάθησης.
- Καθορισμός μέγιστου αριθμού εποχών ώστε ο αλγόριθμος να σταματήσει μόλις αυτές ολοκληρωθούν
- Ένα τελευταίο κριτήριο είναι η τιμή του ολικού σφάλματος $E_{av}(w)$ να πέσει κάτω από μια επιθυμητή τιμή

2.6.6. Μάθηση με Επίβλεψη ως Πρόβλημα Βελτιστοποίησης

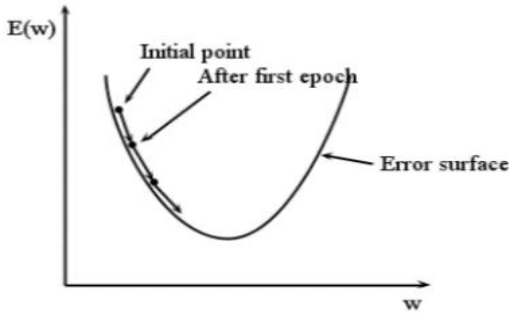
Η μάθηση με επίβλεψη ενός δικτύου MLP μπορεί να αντιμετωπιστεί ως πρόβλημα βελτιστοποίησης. Στα MLP το σφάλμα αναπαριστάνεται ως μια επιφάνεια και είναι μη γραμμική συνάρτηση ενός διανύσματος βαρών w. Η επιφάνεια του σφάλματος E_{av}

μπορεί να αναπτυχθεί χρησιμοποιώντας τις σειρές Taylor γύρο από το τρέχων σημείο λειτουργείας στην επιφάνεια σφάλματος [6].

$$E_{av}(w(n) + \Delta w(n)) = E_{av}(w(n) + g^T(n)\,\Delta w(n) + \frac{1}{2}\,\Delta w^T(n) H(n)\,\Delta w(n)\,, \qquad \text{Εξ. 2.33}$$

Όπου $g(n) = \frac{\partial E(w)}{\partial w}$ είναι η τοπική κλίση και $H(n) = \frac{\partial^2 E(w)}{\partial^2 w}$ είναι ο εσσιανός πίνακας που αντιπροσωπεύει την <<καμπυλότητα>> της επιφάνειας σφάλματος. Η διαδικασία βελτιστοποιήσεις που εκτελείται αναζητά το κατώτατο σημείο της επιφάνειας σφάλματος. Η μέθοδος βελτιστοποιήσεις που εφαρμόζεται είναι η μέθοδος απότομης καθόδου όπου η ενημέρωση/αναπροσαρμογή βαρών γίνεται από τον τύπο

$$\Delta w(n) = -\eta g(n) \qquad \text{Εξ.2.34}$$



Σχήμα 2.16 Αναπροσαρμογή Βαρών στην Επιφάνεια Σφάλματος

Η παράμετρος η(ητα) δηλώνει το ρυθμό μάθησης. Η μέθοδος απότομης καθόδου λειτουργεί στη βάση μιας γραμμική προσέγγισης της συνάρτησης σφάλματος σε μια « τοπική γειτονιά » του σημείου λειτουργείας w(n) . Άρα η τοπική κλίση g(n) είναι η μονή πληροφορία για την επιφάνεια σφάλματος, αυτό σημαίνει ότι σύγκλιση προς το βέλτιστο διάνυσμα βάρους είναι πολύ αργή. Για να βελτιωθεί η ταχύτητα σύγκλισης χρειαζόμαστε για την επιφάνεια σφάλματος πληροφορία μεγαλύτερης τάξης. Αυτό μπορεί να επιτευχθεί χρησιμοποιώντας τετραγωνική προσέγγιση στην επιφάνεια σφάλματος γύρω από το τρέχον σημείο w(n). Συνεπώς προκύπτει ότι η βέλτιστη τιμή

της προσαρμογής των Δw(n) που εφαρμόζεται στο διάνυσμα συναπτικών βαρών υπολογίζεται από τη σχέση 2.35

$$w^*(n) = H^{-1}(n)g(n) \qquad \text{Εξ. 2.35}$$

όπου $H^{-1}(n)$ δηλώνει τον αντίστροφο εσσιανό πίνακα, υποθέτοντας ότι υπάρχει. Αυτή η εξίσωση μπορεί να επιλυθεί με τη μέθοδο Newton χρησιμοποιώντας μόνο μια επανάληψη εάν η συνάρτηση σφάλματος E_{av} είναι δευτεροβάθμια. Όμως αυτό είναι αδύνατο να υλοποιηθεί στην πράξη για τρεις λόγους [6] .

- H^{-1} είναι υπολογιστικά ακριβό να επιλυθεί
- Δεν είναι βέβαιο ότι ο αντίστροφος εσσιανός υπάρχει.
- Όταν η συνάρτηση σφάλματος Eav δεν είναι δευτεροβάθμια, δεν είναι εγγυημένο ότι θα υπάρχει σύγκλιση με τη μέθοδο Newton

Η μέθοδος Quasi-Newton η οποία χρειάζεται μόνο προσέγγισή του διανύσματος κλίσεων g, μπορεί να ξεπεράσει τους τρεις παραπάνω περιορισμούς. Αλλά η πολυπλοκότητα της μεθόδου Quasi-Newton είναι $O(W^2)$, όπου το W είναι το μέγεθος του διανύσματος βαρών w, κατά συνέπεια είναι μη πρακτικό για μεγάλα προβλήματα. Μια άλλη μέθοδος η οποία ονομάζεται Conjugate-gradient προσπαθεί να συνδυάσει δυο στοιχεία, το πρώτο είναι η γρηγορότερη σύγκλιση της μεθόδου steepest decent , δεύτερον η επιτάχυνση της σύγκλισης να επιτευχθεί χωρίς την παρουσία υπολογισμών που απαιτούνται για τον εσσιανο πίνακα. Η μέθοδος αυτή έχει πολυπλοκότητα O(W). Επομένως από όλες τις μεθόδους είναι η πιο κατάλληλη μέθοδος για μεγάλα προβλήματα άρα και η καταλληλότερη για την εκπαίδευση του MLP δικτύου [6].

2.7. Εκπαίδευση Δικτύου RBF

Η εκπαίδευση του δικτύου RBF απαιτεί να καθοριστούν τα κέντρα των συναρτήσεων ακτίνας βάσης και τα βάρη. Η επιλογή των κέντρων των νευρώνων/κόμβων RBF είναι το πιο κρίσιμο σημείο στην υλοποίηση του RBF δικτύου. Η επιλογή των κέντρων μπορεί να γίνει είτε επιλέγοντας τυχαία ως κέντρο κάποιο πρότυπο από το σύνολο εκπαίδευσης είτε με τη μέθοδο της ομαδοποίησης. Για τα δίκτυα RBF που στο κρυφό επίπεδό ως συνάρτηση ενεργοποιήσης έχουν την Gaussian RBF, πρέπει επίσης να

καθοριστεί η παράμετρος σ. Οι περισσότεροι αλγόριθμοι εκπαίδευσης που υπάρχουν για να εκπαιδεύσουν RBF δίκτυα έχουν σχεδιαστεί για Gaussian RBF δίκτυα. Στόχος στην εκπαίδευση ενός RBF δικτύου όπως και στα δίκτυα MLP είναι να ελαχιστοποιηθεί το μέσο του τετραγωνικού σφάλματος μεταξύ των πραγματικών και των αναμενομένων εξόδων του δικτύου. Οι κυριότερες τεχνικές εκπαίδευσης που χρησιμοποιούνται είναι [13].

2.7.1. Εκπαίδευση Δύο Σταδίων

Η εκπαίδευση δυο σταδίων είναι αποτέλεσμα της παρατήρησης ότι επιθυμούμε τα κέντρα των νευρώνων RBF να τοποθετούνται σε κέντρα ομάδων που υπάρχουν στα δεδομένα εκπαίδευσης. Στο πρώτο στάδιο, από το σύνολο των προτύπων εκπαίδευσης $X=\{x^n\}$ καθορίζονται τα κέντρα και η τυπική απόκλιση των συναρτήσεων βάσης, η τυπική απόκλιση είναι γνωστή ως εύρος ή ακτίνα σ_i. Να θυμίσουμε ότι σε κάθε νευρώνα RBF αντιστοιχεί ένα μοναδικό κέντρο. Ο καθορισμός των κέντρων και των ακτινών γίνεται χρησιμοποιώντας τεχνικές μάθησης χωρίς επίβλεψη. Στις τεχνικές αυτές αγνοείται η πληροφορία σχετικά με την κατηγορία κάθε προτύπου εκπαίδευσης. Η τεχνική μάθησης χωρίς επίβλεψη που εφαρμόζεται, είναι η ομαδοποίηση (clustering). Η μέθοδος της ομαδοποιήσεις μπορεί να υλοποιηθεί με τον αλγόριθμο των k-μέσων ή τον αλγόριθμο LQV ή C-means.

Όταν εφαρμόζεται ομαδοποίηση το σύνολο των προτύπων εκπαίδευσης διαχωρίζεται στις κατάλληλες ομάδες- συστάδες (clusters). Το κεντροειδές ή μεσαίο σημείο της κάθε ομάδας χρησιμοποιείται ως κέντρο της συνάρτησης ακτίνας βάσης (π.χ Gaussς συνάρτησης). Ο αριθμός των ομάδων-συστάδων μπορεί να προκαθοριστεί ή να υπολογιστεί αυτόματα απ' τον εκάστοτε αλγόριθμο ομαδοποιήσεις που θα χρησιμοποιηθεί. [13]

Για την εύρεση των τιμών των ακτινών σ_i μπορούν να εφαρμοστούν διάφορες μέθοδοι για παράδειγμα ένας τρόπος είναι, για κάθε RBF νευρώνα του κρυφού επιπέδου το σ_i να δίνεται από την σχέση

$$\sigma = \frac{1}{N}\sum_{i=1}^{N}\|x_i - \mu\| \qquad \text{Εξ.2.36}$$

Όπου η παράμετρος N δηλώνει το πλήθος των προτύπων εισόδου στην συστάδα. Κατά συνέπεια βλέπουμε ότι το σ ισούται με τη μέση απόσταση όλων των σημείων της συστάδας από το κέντρο της. Στην βιβλιογραφία δίνονται πολλές σχέσεις για την εύρεση της τιμής της ακτίνας. Ένας άλλος τρόπος για να υπολογιστεί η παράμετρός σ δίνεται από την σχέση

$$\sigma = \frac{D_{max}}{\sqrt{2M}} \qquad \text{Εξ.2.37}$$

όπου η παράμετρος M δηλώνει τον αριθμό των κεντροειδών και η D_{max} δηλώνει τη μέγιστη απόσταση μεταξύ τους. Με τον τύπο αυτό διασφαλίζεται ότι οι Γκαουσιανες συναρτήσεις των μεμονωμένων κόμβων δεν θα είναι ούτε πολύ <<αιχμηρές>> ούτε πολύ <<επίπεδες>>. Όταν η παράμετρός σ είναι ίση για όλους τους νευρώνες RBF $\sigma_1 = \ldots = \sigma_M = \sigma$ (εξ. 2.37) τότε το δίκτυο RBF(με Gaussian RBF νευρώνες) χαρακτηρίζεται από την ιδιότητα της παγκόσμιας προσέγγισης.

Στο δεύτερο στάδιο εφόσον το πρώτο στάδιο ολοκληρωθεί και τα κέντρα πλέον είναι γνωστά καθώς και οι ακτίνες σ_i πρέπει υπολογιστούν τα βάρη των διασυνδέσεων μεταξύ των RBF νευρώνων και των νευρώνων του επιπέδου εξόδου. Ο υπολογισμός τους μπορεί να γίνει με τη μέθοδο ελαχιστοποίησης του μέσου τετραγωνικού σφάλματος δηλαδή τον κανόνα δέλτα.

2.7.2. Ενιαία Εκπαίδευση

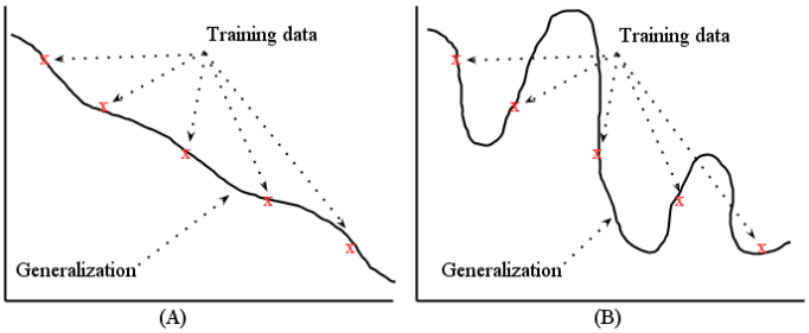
Σε αυτή τη μέθοδο μπορεί να εφαρμοστεί gradient decent με ομαδική ή σειριακή ενημέρωση για τις παραμέτρους του δικτύου σ_j,μ_j, και ν_j (Η παράμετρος σ δηλώνει την ακτίνα ,η μ δηλώνει το κέντρο και ν τα βάρη) όπως και στο MLP. Τα βάρη αρχικοποούνται τυχαία στο διάστημα [-1,1] αλλά ο τρόπος που υπολογίζονται οι μερικές παραγωγοί των εξόδων ως προς τις παραμέτρους σ,μ,ν είναι διαφορετικός από το MLP. [13]

Η ενιαία εκπαίδευση έχει το μειονέκτημα ότι όταν τα κέντρα αρχικοποούνται εντελώς τυχαία η μέθοδος gradient decent δεν ελαχιστοποιεί σε ικανοποιητικό βαθμό το σφάλμα διότι τα τοπικά ελάχιστα είναι αρκετά και δύσκολα να αποφευχθούν. Στην περίπτωση όμως που τα κέντρα και οι ακτίνες για να αρχικοποιηθούν χρησιμοποιηθεί κάποια μέθοδος μάθησης χωρίς

επίβλεψη(clustering) και πάρουμε καλές τιμές τότε η μέθοδος gradient decent για την ενημέρωση όλων των παραμέτρων βρίσκει πολύ καλές λύσεις.

2.8. Ικανότητα Γενίκευσης (Generalization)

Γενίκευση ονομάζεται η ικανότητα του ΤΝΔ που παρέχει σωστές αποφάσεις για παραδείγματα που δεν έχουν χρησιμοποιηθεί κατά την δημιουργία ή εκπαίδευση του. Απώτερος στόχος κατά την σχεδίαση και επιλογή της αρχιτεκτονικής ενός ΤΝΔ είναι να έχει καλή ικανότητα γενίκευσης. Ένα ΤΝΔ με μεγάλο αριθμό κρυμμένων νευρώνων μπορεί να εκπαιδευτεί ώστε να απεικονίζει με μεγάλη ακρίβεια όλα τα παραδείγματα του συνόλου εκπαίδευσης [22]. Όμως, ένα τέτοιο δίκτυο συνήθως έχει μικρή ικανότητα γενίκευσης διότι <<απομνημονεύει>> και δεν παρουσιάζει καλές επιδόσεις σε νέα δεδομένα διότι λόγω, της μεγάλης ευελιξίας του, δημιουργεί απεικονίσεις οι οποίες είναι συνήθως περισσότερο <<πολύπλοκες>> απ' ότι χρειάζεται [21].



Σχήμα 2.17 Παραδείγματα Ικανότητας Γενίκευσης (Α) και (β)

Στο σχήμα 2.23 (Α) παρατηρούμε μια ομαλότερη συνάρτηση(fitted data) η οποία είναι καλύτερη λύση γενίκευσης από την συνάρτηση που απεικονίζεται στο παράδειγμα (Β). Η συνάρτηση που απεικονίζεται στο παράδειγμα (Β) παρότι έχει μηδενικό σφάλμα εκπαίδευσης, είναι περισσότερο πολύπλοκη από ότι χρειάζεται στην περίπτωση αυτή έχουμε υπερεκπαίδευση [23](overfitted data).

Εντούτοις ένα ΤΝΔ με πολύ λίγους κρυμμένους νευρώνες πιθανόν να μην έχει την απαιτούμενη 'ευελιξία' ώστε να μπορεί να ορίσει πολύπλοκες περιοχές απόφασης ή να προσεγγίσει συναρτήσεις με πολύπλοκη γραφική παράσταση Αυτό το φαινόμενο αναφέρεται ως υποεκπαίδευση του δικτύου. Το κριτήριο επιλογής ενός

μοντέλου ΤΝΔ σύμφωνα με την βασική εμπειρική αρχή μηχανικής μάθησης (Occam's razor) είναι η πολυπλοκότητα του δικτύου. Σύμφωνα με αυτή προτιμούμε το απλούστερο μοντέλο που μπορεί να μάθει επαρκώς τα παραδείγματα εκπαίδευσης επειδή μια τέτοια επιλογή απαιτεί γενικά τους ελάχιστους δυνατούς υπολογιστικούς πόρους.

2.9. Μέθοδος Cross Validation

Η μέθοδος Cross Validation(διασταυρωμένη επικύρωση) είναι μια στατιστική μέθοδος αξιολόγησης και επικύρωσης των αλγορίθμων εκπαίδευσης. [24] Στη μέθοδο αυτή το σύνολο δεδομένων τμηματοποιείται σε σύνολο εκπαίδευσης το οποίο χρησιμοποιείται για τον καθορισμό των βαρών του ΤΝΔ και σύνολο ελέγχου το οποίο χρησιμοποιείται για τον υπολογισμό του σφάλματος γενίκευσης του δικτύου που προκύπτει από την εκπαίδευση. Το σύνολο(δείγμα) εκπαίδευσης εν συνέχεια διαχωρίζεται σε

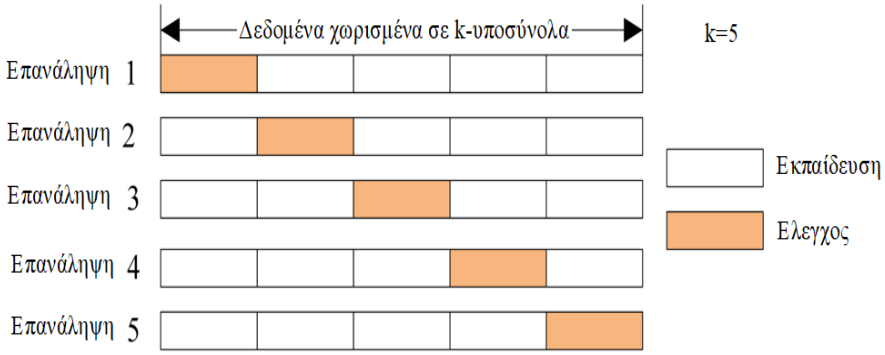
- Υποσύνολο εκπαίδευσης, το οποίο χρησιμοποιείται για την επιλογή του μοντέλου του ΤΝΔ
- Υποσύνολο επικύρωσης, το οποίο χρησιμοποιείται για τον έλεγχο(την επικύρωση) του μοντέλου του ΤΝΔ

Ο διαχωρισμός αυτός γίνεται με σκοπό η επικύρωση του μοντέλου να γίνεται από ένα σύνολο δεδομένων το οποίο είναι διαφορετικό από το σύνολο δεδομένων που θα χρησιμοποιηθεί για την εκπαίδευση του δικτύου. Για τη μέθοδο Cross-Validation υπάρχουν δυο εκδοχές, η πρώτη είναι διασταυρωμένη επικύρωση K-τμημάτων(K-fold cross-validation) και η δεύτερη, η οποία είναι ειδική περίπτωση της διασταυρωμένης επικύρωσης K τμημάτων, είναι ένα παράδειγμα ελέγχου κάθε φορά(Leave one out(LOT)).

2.9.1. Διασταυρωμένη Επικύρωση k-τμημάτων(k-fold cross-validation)

Στη μέθοδο k-fold cross validation τα δεδομένα χωρίζονται σε k-υποσύνολα $D_1,..,D_k$ τα οποία είναι σχεδόν ίσα αν όχι ίσα μεταξύ τους. Ο αριθμός k των υποσυνόλων καθορίζει και τον αριθμό των επαναλήψεων αξιολογήσεις και εκπαίδευσης. Σε κάθε επανάληψη i=1,2,..k ένα από τα υποσύνολα θα χρησιμοποιείται για την αξιολογήσει και τα εναπομένοντα k-1 υποσύνολα θα

χρησιμοποιηθούν για την εκπαίδευση. Με βάση το υποσύνολο δεδομένων D_i χρησιμοποιώντας το ως σύνολο ελέγχου υπολογίζεται το σφάλμα γενίκευσης σε κάθε επανάληψη. Στο τέλος των k επαναλήψεων κάθε υποσύνολο δεδομένων θα έχει χρησιμοποιηθεί μια φορά ως σύνολο ελέγχου. Ενώ το σφάλμα γενίκευσης όλου του ΤΝΔ προκύπτει από το μέσο όρο των επιμέρους σφάλματων που υπολογίζονται για κάθε υποσύνολο D_i.



Σχήμα 2.18 Διάγραμμα Διαδικασίας k-fold-Cross-Validation

Στο παραπάνω σχήμα βλέπουμε το διάγραμμα της μεθόδου k-fold cross validation. Παρατηρούμε στις k- επαναλήψεις όπου το k=5, πως χωρίζονται τα δεδομένα σε υποσύνολα εκπαιδεύσεις και υποσύνολο ελέγχου. Τα ανοικτόχρωμα(άσπρα) κουτάκια αναπαριστούν τα k-1 υποσύνολα εκπαίδευσης και το σκουρόχρωμο(πορτοκαλί) το σύνολο ελέγχου.

2.9.2. Leave One Out

Η μέθοδος ένα παράδειγμα ελέγχου κάθε φορά (leave-one-out) είναι μια ειδική περίπτωση της διασταυρωμένης επικύρωσης K τμημάτων. Στη μέθοδο αυτή ο αριθμός των τμημάτων K είναι ίσος με τον αριθμό N όλων των παραδειγμάτων του συνόλου D. Το σύνολο D πλην ενός παραδείγματος (x^i,t^i) του, θεωρείται ολόκληρο ως σύνολο εκπαίδευσης, Το συγκεκριμένο παράδειγμα που εξαιρείται κάθε φορά από την εκπαίδευση, χρησιμοποιείται για να εκτιμηθεί το σφάλμα γενίκευσης στο ΤΝΔ, όταν σε αυτό εφαρμόζεται το συγκεκριμένο παράδειγμα. Η διαδικασία αυτή επαναλαμβάνεται για όλα τα παραδείγματα (x^i,t^i) (i=1,..,N) ώστε να εκτιμηθεί το σφάλμα γενίκευσης ως μέσος όρος όλων των σφάλματων που υπολογίστηκαν πρωτύτερα.

ΚΕΦΑΛΑΙΟ 3.
ΣΧΕΔΙΑΣΤΙΚΗ ΔΙΑΔΙΚΑΣΙΑ ΔΙΚΤΥΩΝ

3.1. Προσέγγιση Συναρτήσεων με Νευρωνικά Δίκτυα
3.2. Ευρετικές Μέθοδοι για την Βελτίωση Απόδοσης του αλγορίθμου Backpropagation
3.3. Δομή των Πολυεπίπεδων Νευρωνικών Δικτύων
3.4. Δομή Υβριδικού Δικτύου RBF
3.5. Εκπαίδευση δυο Σταδίων του Υβριδικού Δικτύου RBF

3.1. Προσέγγιση Συναρτήσεων με Νευρωνικά Δίκτυα

Στόχος της παρούσας διατριβής , όπως ειπώθηκε στην εισαγωγή , είναι η σύγκριση ενός πολυεπίπεδου νευρωνικού δικτύου (MLP) με ένα υβριδικό δίκτυο το οποίο απαρτίζεται από ένα πολυεπίπεδο νευρωνικό δίκτυο (MLP) πιο απλό από το πρώτο, σε σειρά με ένα δίκτυο ακτινικής συνάρτησης βάσης (RBF). Για χάριν συντομίας, από το σημείο αυτό και έως το πέρας της παρούσας διατριβής, θα ονομάσουμε ως «εκτεταμένο» το αυτόνομο MLP και ως «απλό» εκείνο το MLP που συμμετέχει στο υβριδικό δίκτυο. Αναλυτικά τα δίκτυα που υλοποιήθηκαν θα αναφερθούν σε επόμενες υποπαραγράφους του παρόντος κεφαλαίου. Σκοπός μας είναι να συγκρίνουμε στα δύο δίκτυα την μετρική της γενίκευσης. Η απόδοση των δύο αλγορίθμων θα παρουσιαστεί αναλυτικότερα στο επόμενο κεφάλαιο. Για να πραγματοποιηθεί αυτή η σύγκριση πρέπει να σχεδιαστούν τα δίκτυα, πριν προχωρήσουμε στη σχεδίαση πρέπει να οριστεί το μέγεθος του κάθε δικτύου, οι συναρτήσεις ενεργοποίησης που θα έχουν στους κόμβους του κρυμμένου επιπέδου και του επιπέδου εξόδου. Η λειτουργία των δικτυών που θα φτιάξουμε είναι να προσεγγίζουν ποικίλες συναρτήσεις.

Ένα νευρωνικό δίκτυο πολλών επίπεδων μπορούμε να το παρομοιάσουμε με ένα «εργαλείο» το οποίο εκτελεί μη γραμμική απεικόνιση της εισόδου, στην έξοδο. Έστω η συνάρτηση f(.) η οποία περιγράφεται από τη σχέση 3.3 [25]

$$f(\mathrm{x}) = d \qquad d\epsilon R^1, x\epsilon R^m \qquad \text{Εξ. 3.1}$$

Όπου χ το διάνυσμα , τα οποίο είναι m-διαστατο και η παράμετρος d δηλώνει την έξοδο της συνάρτησης. Η συνάρτηση f(x) είναι η συνάρτηση που θέλουμε να προσεγγίσουμε. Γνωρίζοντας το σύνολο προτύπων εκπαίδευσης N, όπου το κάθε πρότυπο έχει τη μορφή $\{x_i, d_i\}$ μπορούμε να σχεδιάσουμε ένα νευρωνικό δίκτυο το οποίο θα προσεγγίζει την άγνωστη συνάρτηση f(.). Δοθείσας οποιασδήποτε συνάρτησης f υπάρχουν ένας ακέραιος m_1 και σύνολα πραγματικών σταθερών α_i , b_i και w_{ij} όπου i=1,...,k και j=1,…,m τέτοια ώστε να μπορούμε να ορίσουμε την σχέση 3.4 ως μια προσεγγιστική υλοποίηση της συνάρτησής f(.) [19]

$$F(x_1, \dots, x_m) = \sum_{i=1}^{k} a_i \varphi \left(\sum_{j=1}^{m} w_{ij} x_j + b_i \right) \qquad \text{Εξ. 3.2}$$

Στόχος είναι, η προσεγγιστική υλοποίηση της συνάρτησης f(.), δηλαδή, η συνάρτηση F(.) που περιγράφει την αντιστοίχιση εισόδου -εξόδου που υλοποιείται από το νευρωνικό δίκτυο να διαφοροποιείται ελάχιστα από την f(.) ως προς την ευκλείδεια απόσταση. Η διαφοροποίηση αυτή ορίζεται από τη παρακάτω σχέση και πρέπει να ισχύει για όλα τα $x_1, x_2, .., x_m$ που βρίσκονται στο χώρο εισόδου

$$\|F(x_1, \dots x_m) - f(x_1, \dots, x_m)\| < \varepsilon \qquad \text{Εξ. 3.3}$$

Όπου η παράμετρος ε δηλώνει το σφάλμα προσέγγισης και είναι ένας θετικός αριθμός. Σκοπός είναι να ελαχιστοποιηθεί όσο γίνεται περισσότερο το σφάλμα προσέγγισης, το τελευταίο είναι πολύ πιθανόν να συμβεί εάν ο αριθμός των ελευθέρων παραμέτρων του δικτύου είναι ικανοποιητικά μεγάλο καθώς και το μέγεθος N του συνόλου εκπαίδευσης .

Η εξίσωση 3.4 αντιπροσωπεύει την έξοδο ενός νευρωνικού δικτύου που χαρακτηρίζεται από την ιδιότητα καθολικής προσέγγισης.

- Το δίκτυο έχει m κόμβους εισόδου και ένα μεμονωμένο κρυφό επίπεδο, αποτελούμενο από k νευρώνες, οι είσοδοι συμβολίζονται ως x_1, x_2, x_m
- Ο κρυφός νευρώνας i έχει συναπτικά βάρη $w_i, .., w_m$ και πόλωση b_i

- Η έξοδος του δικτύου είναι ένας γραμμικός συνδυασμός των εξόδων των κρυφών νευρώνων με τα $a_1,..,a_k$ να ορίζουν τα συναπτικά βάρη του επιπέδου εξόδου

3.2. Ευρετικές Μέθοδοι για την Βελτίωση Απόδοσης του Αλγορίθμου Backpropagation

Πριν προχωρήσουμε στην υλοποίηση των δικτυών MLP με χρήση του αλγορίθμου backpropagation πρέπει να επιλεχθούν κάποιοι παράγοντες οι οποίοι συμβάλλουν ιδιαίτερα στην απόδοση και τη σύγκλιση του αλγορίθμου . Οι παράγοντες αυτοί είναι, η συνάρτηση ενεργοποίησης των κρυφών κόμβων, η αρχικοποίηση βαρών και ο ρυθμός μάθησης. Στα επόμενα υποκεφάλαια ακολουθεί μια σύντομη περιγραφή των μεθόδων που χρησιμοποιούνται για την βελτίωση της απόδοσης του αλγορίθμου backpropagation.

3.2.1. Αρχικοποίηση Βαρών

Η αρχικοποίηση βαρών μπορεί να θεωρείται ότι είναι ένα τυπικό θέμα στην διαδικασία σχεδίασης ενός νευρωνικού δικτύου. Η κατάλληλη αρχικοποίηση των βαρών επιταχύνει τον αλγόριθμο backpropagation [26], γενικά ισχύει ότι, μια καλή επιλογή για τις αρχικές τιμές των συναπτικών βαρών του δικτύου, βοηθάει πολύ στην επιτυχημένη σχεδίαση ενός δικτύου. Επιπλέον, η αρχικοποίηση των βαρών σε υψηλές τιμές προκαλεί πρόωρο κορεσμό στους νευρώνες, το οποίο έχει ως συνέπεια οι τοπικές κλίσεις στον αλγόριθμο backpropagation να πάρουν μικρές τιμές, και αυτές με τη σειρά τους επηρεάζουν την διαδικασία μάθησης, κάνοντας την πιο αργή [27], για το λόγο αυτό προτείνονται μικρές αρχικές τιμές των βαρών. Απ' ότι άλλη έχει αποδειχθεί ότι οι μικρές τιμές στα βάρη δημιουργούν μια επίπεδη επιφάνεια σφάλματος και η εκπαίδευση γίνεται αργή [28]. Το ερώτημα είναι ποιες είναι οι κατάλληλες αρχικές τιμές για τα βάρη. Ένας τύπος για τον υπολογισμό των αρχικών βαρών που απαντάται στη βιβλιογραφία [29] είναι αυτός που παρουσιάζεται στην εξίσωση 3.6 [11].

$$\sigma_z^2 = w\sigma_w^2 \qquad \text{Εξ. 3.4}$$

Όπου το σ_z είναι η τυπική απόκλιση των εισόδων των νευρώνων, η παράμετρος w είναι ο αριθμός των βαρών και σ_w είναι η απόκλιση των αρχικοποιημένων βαρών.

Στην περίπτωση που χρησιμοποιηθεί κάποια σιγμοειδής συνάρτηση ενεργοποίησης τότε, $\sigma_z = 1$ και η εξίσωση 3.6_αλλάζει και παίρνει τη μορφή.

$$\sigma_w = \frac{1}{\sqrt{w}} \qquad \text{Εξ. 3.5}$$

Ο αριθμός των μεθόδων που έχουν προταθεί στην βιβλιογραφία για τον υπολογισμό του διαστήματος των αρχικών βαρών είναι αξιοσημείωτος. Έχει αποδειχθεί ότι [30] ότι ένα καλό διάστημα τιμών για να αρχικοποιήθουν τα βάρη είναι [-0.3,0.3]. Η τυχαία αρχικοποίηση βαρόν για τα δίκτυα MLP είναι η απλούστερη και πιο διαδεδομένη μέθοδος [3].

3.2.2. Ρυθμός Μάθησης

Ο ρυθμός μάθησης επηρεάζει την ταχύτητα σύγκλισης. Μεγάλος ρυθμός μάθησης μπορεί να οδηγήσει σε γρηγορότερη σύγκλιση και σε ταλάντωση γύρω από τις βέλτιστες τιμές βαρών. Μικρός ρυθμός μάθησης έχει ως αποτέλεσμα πιο αργή σύγκλιση, ενώ μπορεί να οδηγήσει με μεγαλύτερη πιθανότητα σε παγίδευση σε τοπικά ακρότατα. Στην βιβλιογραφία υπάρχει πληθώρα μεθόδων για την επιλογή μιας καλής τιμής για το ρυθμό μάθησης. Η πιο απλή λύση είναι να έχεις μια σταθερή τιμή για το ρυθμό μάθησης. Μερικές λύσεις για σταθερό ρυθμό μάθησης παρουσιάζονται στον πίνακα 3.1

Πίνακας 3.1: Ευρετικές Τιμές για Σταθερό Ρυθμό Μάθησης

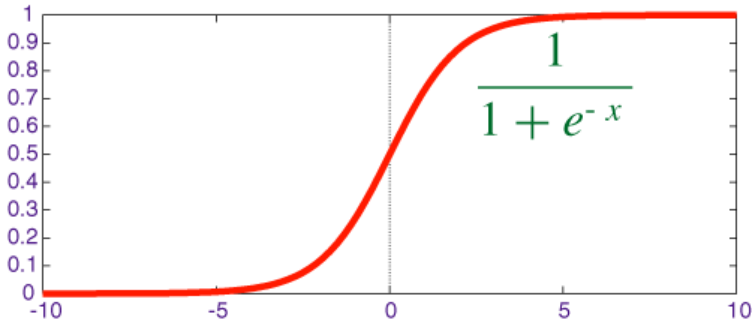
Ευερετικές τιμές για το ρυθμό μάθησης	Αναφορές
0-1	Walczak and Cerpa 1999 [30]
0.1-0.9	Gasteiger and Zupan 1993 [14]
0.01	Bengio 2012

Η άλλη λύση που προτείνεται είναι ο ρυθμός μάθησης να μεταβάλλεται. Οι τεχνικές που χρησιμοποιούνται για μεταβαλλόμενο ρυθμό μάθησης είναι: [31]

- Κάθε συναπτικό βάρος και κάθε πόλωση έχει δικό του, της αντίστοιχα ρυθμό μάθησης
- Ο ρυθμός μάθησης ανά εποχή πρέπει να μπορεί να μεταβάλλεται
- Αν η παράγωγος της συνάρτησης κόστους ως προς ένα συναπτικό βάρος έχει το ίδιο αλγεβρικό πρόσημο σε πολλές διαδοχικές εποχές του αλγόριθμου τότε ο ρυθμός μάθησης για το συγκεκριμένο συναπτικό βάρος πρέπει να αυξηθεί.
- Αν η παράγωγος της συνάρτησης κόστους ως προς ένα συγκεκριμένο συναπτικό βάρος έχει διαφορετικό αλγεβρικό πρόσημο σε διαδοχικές εποχές τότε ο ρυθμός μάθησης για αυτό το βάρος πρέπει να μειωθεί.

3.2.3. Συνάρτηση Ενεργοποίησης

Η συναρτήσεις ενεργοποίησης επίσης γνωστές ως συναρτήσεις μεταφοράς είναι συνήθως μη γραμμικές συναρτήσεις οι οποίες μετατρέπουν το προσαρμοσμένο από τα βάρη άθροισμα σε μια τιμή εξόδου. [7] Ανάλογα με είδος του νευρωνικό δικτύου απαιτείται και η αντίστοιχη συνάρτηση ενεργοποίησης ώστε να παράγει καλυτέρα αποτελέσματα [32] [33]. Στους κόμβους στα νευρωνικά δίκτυα MLP είναι απαραίτητη μια συνάρτηση ενεργοποίησης για να εισάγει τη μη γραμμικότητα στο δίκτυο. Υπάρχουν όπως γνωρίζουμε δυο είδη συναρτήσεων ενεργοποίησης για τα νευρωνικά δίκτυα οι γραμμικές η μη γραμμικές. Η επιλογή της συνάρτησης ενεργοποίησης είναι υψίστης σημασίας για την απόδοση του αλγόριθμου εκπαίδευσης. Μια συνάρτηση ενεργοποίησης για ένα δίκτυο που χρησιμοποίει το αλγόριθμο backpropagation για την εκπαίδευση πρέπει να διαθέτει κάποια σημαντικά χαρακτηριστικά. Πρέπει να είναι συνεχείς, διαφορίσιμη και μονότονα αύξουσα. [34] Όταν χρησιμοποιείται ο αλγόριθμος backpropagation η συνάρτηση ενεργοποίησης που εφαρμόζεται πρέπει να είναι διαφορίσιμη ώστε η συνάρτηση να φράσσεται σε συγκεκριμένο διάστημα τιμών. Για τα περισσότερα δίκτυα η συνάρτηση ενεργοποίησης που προτείνεται είναι η σιγμοειδής ή υπερβολική εφαπτομένη. Στα δίκτυα MLP που θα σχεδιάσουμε θα χρησιμοποιήσουμε τη σιγμοειδής



Σχήμα 3.1 Σιγμοειδής Συνάρτηση

3.3. Δομή των Πολυεπίπεδων Νευρωνικών Δικτύων(MLP)

Το βασικό μοντέλο για ένα νευρωνικό δίκτυο είναι το Single Layer Perceptron, αλλά αυτό μπορεί να χρησιμοποιηθεί μόνο για να ταξινομήσει γραμμικά διαχωρίσιμα προβλήματα, οπότε σε προβλήματα του πραγματικού κόσμου δύσκολα μπορεί να εφαρμοσθεί. Ένα MLP με ένα κρυμμένο επίπεδο έχει την ιδιότητα της καθολικής προσέγγισης, αυτό συνεπάγεται ότι το νευρωνικό δίκτυο MLP μπορεί να προσεγγίσει οποιαδήποτε συνεχή συνάρτηση. Η ιδιότητα της καθολικής προσέγγισης δεν μπορεί όμως να απάντηση στο ερώτημα, αν ένα νευρωνικό δίκτυο με ένα κρυμμένο επίπεδο θα έχει καλή γενικευτική ικανότητα, ούτε μπορεί να ορίσει πως θα είναι η δομή ενός δικτύου, ούτε να εξασφαλίσει ότι ο χρόνος εκπαίδευσής του δικτύου είναι βέλτιστος.

Ένα MLP με δυο κρυμμένα επίπεδα μπορεί να προσεγγίσει μια συνάρτηση με μικρότερο αριθμό νευρώνων από ένα MLP με ένα κρυμμένο επίπεδο. Ο αριθμός των κρυμμένων επιπέδων είναι ένα trade-off μεταξύ της ομαλότητας της προσεγγισμένης συνάρτησης και της ακρίβειας της. Εάν επιλεχθεί ο αριθμός των κρυμμένων επιπέδων να είναι μικρός τότε αυξάνεται η ομαλότητα της προσεγγισμένης συνάρτησης σε περίπτωση που επιλεχθεί μεγαλύτερος αριθμός κρυμμένων επιπέδων αυξάνεται η ακρίβεια [22]. Ο αριθμός των κρυμμένων επιπέδων επιδρά στην πολυπλοκότητα ενός προβλήματος κατά συνέπεια όταν η πολυπλοκότητα του προβλήματος αυξάνεται και ο αριθμός των κρυμμένων επιπέδων πρέπει να αυξηθεί [35]. Ο αριθμός των κρυμμένων επιπέδων που έχει ένα ΤΝΔ καθορίζει τον τύπο των υπερπίπεδων που μπορεί να προσεγγίσει το δίκτυο .

- Ένα MLP με ένα κρυμμένο επίπεδο μπορεί να δημιουργήσει ένα υπερεπίπεδο
- Ένα MLP με 2 κρυμμένα επίπεδα μπορεί να συνδυάσει υπερεπίπεδα από κυρτές περιοχές αποφάσεις.
- Ένα MLP με τρία κρυμμένα επίπεδα μπορεί να συνδυάσει κυρτές περιοχές αποφάσεις ώστε να δημιουργήσει κυρτές περιοχές αποφάσεις οι οποίες περιλαμβάνουν κοιλότητες. [36]

Οι περισσότερες συναρτήσεις αν όχι όλες μπορούν να προσεγγιστούν με ένα ΤΝΔ με ένα κρυμμένο επίπεδο [37] για το λόγο αυτό και τα τρία δίκτυα, δηλαδή το εκτεταμένο MLP, το απλό MLP καθώς και το RBF, που θα υλοποιηθούν θα έχουν ένα μόνο κρυμμένο επίπεδο.

Εφόσον ο αριθμός των κρυμμένων επίπεδων καθοριστεί σε ένα ΤΝΔ πρέπει να καθοριστεί και ο αριθμός των νευρώνων που εμπεριέχονται σε αυτό. Ο αριθμός των κρυμμένων νευρώνων είναι ένα trade-off μεταξύ του χρόνου εκπαίδευσης και της ακρίβειας. [10] Όσο αυξάνει ο αριθμός των κρυμμένων νευρώνων αυξάνει και ο χρόνος εκπαίδευσης του δικτύου, το οποίο έχει μεγαλύτερη ακρίβεια από ότι θα είχε με μικρότερο αριθμό κρυμμένων νευρώνων. Ο μεγάλος αριθμός κρυμμένων νευρώνων μπορεί να οδηγήσει σε υπερεκπαίδευση(overtraining) του δικτύου με αποτέλεσμα η γενικευτική ικανότητα να μην είναι καλή [38]. Στον Πίνακα 3.1 παρουσιάζεται σε μορφή ψευδοκώδικα ο αλγόριθμος τον οποίο χρησιμοποιήσαμε για την υλοποίηση του εκτεταμένου δικτύου της παρούσας διατριβής.

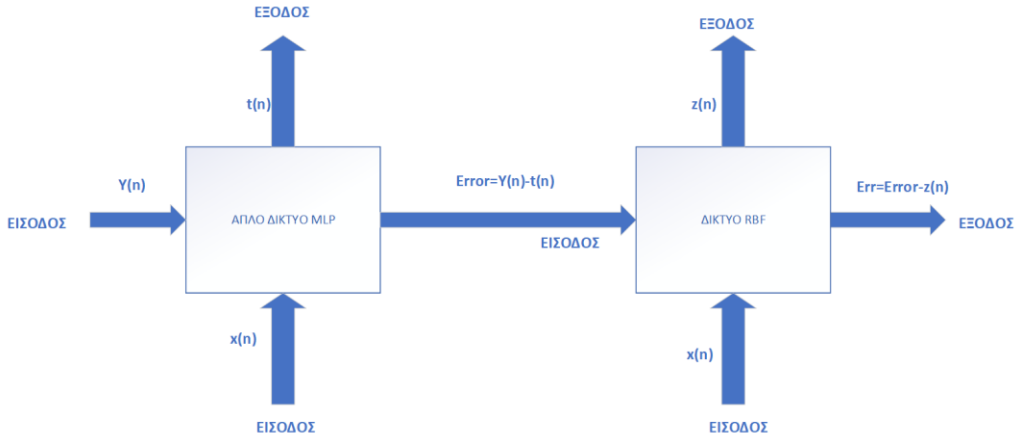
Πίνακας 3.2: Ψευδοκώδικας Εκτεταμένου Δικτύου MLP

```
Initialize X
Y_REAL = create_real_values(X)

CONSTANT DECLARATION FOR BIG MLP
nnB=neurons_nunber
enB = epoch_number
lr = learning_rate
parametersB =(nnB,enB,lr,)
CALL MLP
[y_pred_BIG, error_BIG]= predict_mlp(x,y_real, parametersB)
CONSTANT DECLARATION FOR SMALL MLP
nnS=neurons_nunber
enS = epoch_number
lr = input learning_rate
parametersS =(nnS,enS,lr)
CALL MLP
[y_pred_SMALL, error_SMALL]= predict_mlp(x,y_real,
parametersS)

CONSTANTS DECLARATION FOR RBF
nnRBF = neurons number for RBF
sigma
parametersRBF=(nnRBF,sigma)
y_diff= y_pred_SMALL – y_real
[y_pred_RBF, error_RBF] = rbfnetwork(x, y_diff, parametersRBF)
```

3.4. Δομή Υβριδικού Δικτύου RBF



Σχήμα 3.2 Δομή Υβριδικού Δικτύου

Στο υποκέφαλαιο 3.3 είδαμε την μεθοδολογία για το σχεδιασμό του εκτεταμένου δίκτυού MLP. Στο υποκεφάλαιο αυτό θα αναλύσουμε τις σχεδιαστικές επιλογές για το υβριδικό δίκτυο που θα υλοποιήσουμε. Το υβριδικό δίκτυο θα αποτελείται από ένα πιο απλό δίκτυο MLP, σε σύγκριση με το εκτεταμένου δίκτυο MLP, σε σειρά με ένα δίκτυο RBF. Το απλό δίκτυο MLP θα δέχεται το διάνυσμα εισόδων x(n) επίσης στο απλό δίκτυο θα εφαρμόζεται και το διάνυσμα επιθυμητών αποκρίσεων y(n). Ως έξοδο θα βγάζει την απόκριση του δικτύου t(n) και το σήμα σφάλματος(συνάρτηση σφάλματος) το οποίο είναι η διαφορά της επιθυμητής απόκρισης με την απόκριση του δικτύου. Το RBF δίκτυο με τη σειρά του θα δέχεται ως είσοδό το σήμα σφάλματος και το διάνυσμα εισόδων χ(n). Στόχος του RBF δικτύου είναι προσεγγίσει στο βέλτιστο το σήμα σφάλματος. Το προσεγγισμένο σήμα σφάλματος, είναι η έξοδος του δικτύου RBF το οποίο ελέγχουμε πόσο διαφέρει από το σήμα σφάλματος, που δέχτηκε το δίκτυο RBF ως είσοδό.

3.4.1. Απλό Δίκτυο MLP

Για να δημιουργήσουμε το απλό δίκτυο MLP μειώσαμε τον αριθμό των κρυφών κόμβων ώστε να απλοποιηθεί το εκτεταμένου MLP κατά τα άλλα η λειτουργεία του είναι αναλλοίωτη όπως στο 3.3

3.4.2. Δίκτυο RBF του Υβριδικού Δικτύου

Στόχος του RBF δικτύου είναι να προσεγγίσουμε την συνάρτηση σφάλματος του απλού MLP δικτύου, όπως αναφέραμε και προηγουμένως. Επίσης γνωρίζουμε ότι το δίκτυο RBF είναι καθολικός προσεγγιστής [39]άρα αυτό που πρέπει να προσέξουμε είναι να σχεδιάσουμε ένα δίκτυο RBF που θα προσεγγίσει στο βέλτιστο τη συνάρτηση σφάλματος. Το δίκτυο που θα σχεδιάσουμε θα αποτελείται δυο επίπεδα, να θυμίσουμε ότι το επίπεδο εισόδου δεν προσμετράτε, το επίπεδο εισόδου ,το κρυφό επίπεδο και το επίπεδο εξόδου. Το επίπεδο εισόδου το οποίο θα απαρτίζεται από m πηγαίους κόμβους, το m είναι ίσο με τη διαστατικότητα του διανύσματος εισόδου. Το κρυφό επίπεδο θα αποτελείται από αριθμό νευρώνων ο οποίος αριθμός θα καθοριστεί από το K-means [40] αλγόριθμο τον οποίο θα επεξηγήσουμε στην συνέχεια. Κάθε κρυφός νευρώνας περιγράφεται μαθηματικά από την Gauss ακτινική συνάρτηση βάσης που δίνεται από την σχέση

$$\varphi_j(x) = \varphi(\|x - x_j\|) = exp\left(-\frac{1}{2\sigma_j^2}\|x - x_j\|^2\right),\ \ j = 1,2,..N \qquad \text{Εξ. 3.6}$$

Οπού το j-οστο σημείο δεδομένων εισόδου x_j, είναι το κέντρο της ακτινικής συνάρτησης βάσης και το διάνυσμα χ είναι το πρότυπο(σήμα) που εφαρμόζεται στο επίπεδο εισόδου. Η παράμετρος σ_j είναι το μέτρο του πλάτους(τυπική απόκλιση) της j-οστης Gauss συνάρτησης με κέντρο x_jΤο επίπεδο εξόδου θα αποτελείται από έναν μόνο νευρώνα και χαρακτηρίζεται από το διάνυσμα βαρών w. Η διαστασικότητα του επιπέδου εξόδου είναι ίση με τον αριθμό των νευρώνων στο κρυφό επίπεδο. Τέλος η μαθηματική συνάρτηση που θα προσεγγιστεί F(x) από το δίκτυο θα έχει την εξής μαθηματική μορφή

$$F(x) = \sum_{j=1}^{k} w_j \varphi(x, x_j) \qquad \text{Εξ. 3.7}$$

Η παράμετρος k δηλώνει τον αριθμό των νευρώνων που χαρακτηρίζονται από την ακτινική συνάρτηση βάσης Gauss στο κρυφό επίπεδο.

3.4.3. Αλγόριθμος K-Means

Και στη ενιαία εκπαίδευση και στην εκπαίδευση δυο επίπεδων ενός RBF δικτύου η μέθοδος ομαδοποίησης είναι απαραίτητη ώστε να βρεθούν καλές λύσεις . Η ομαδοποίηση είναι μια από τις τεχνικές μη επιβλεπομένης μάθησης . Εμείς έως τώρα στη διατριβή μας έχουμε αναφερθεί κυρίως σε τεχνικές επιβλεπόμενης μάθησης. Στην μη επιβλεπομένη μάθηση μας δίνεται ένα σύνολο δεδομένων, χωρίς τις αντίστοιχες κλάσεις-ετικέτες κάθε στοιχείου και στόχος είναι η χρήση κάποιου αλγορίθμου, ώστε αυτόματα να ανακαλύψουμε κάποια ενδεχομένως ενδιαφέρουσα δομή των δεδομένων. Δοθέντων κάποιων δεδομένων χωρίς κλάση οι αλγόριθμοι ομαδοποιήσεις κατηγοριοποιούν τα δεδομένα σε συστάδες, έτσι ώστε τα στοιχεία τα οποία ανήκουν στην ίδια ομάδα, να έχουν όμοια ή παραπλήσια χαρακτηριστικά. Η μέθοδος ομαδοποίησης που επιλέξαμε για την εκπαίδευση του RBF δικτύου μας είναι η K-means εφόσον είναι απλή στην υλοποίηση και αποτελεσματική σε απόδοση.

Ο αλγόριθμος K-means ξεκινάει με k τυχαία σημεία, τα οποία ονομάζονται κεντροειδή της συστάδας και δηλώνουν το κέντρο βάρους της συστάδας. Το k υποδηλώνει πόσες συστάδες θέλουμε ο αλγόριθμος να δημιουργήσει. Ο αλγόριθμος εκτελεί επαναληπτικά δύο βήματα. Το πρώτο βήμα αφορά την ανάθεση σε κάποια συστάδα, ενώ το δεύτερο βήμα αφορά τον επαναπροσδιορισμό και τη μετατόπιση του κεντροειδούς κάθε συστάδας. [41]

Πιο αναλυτικά, όσον αφορά στο πρώτο βήμα, δηλαδή την ανάθεση σε κάποια συστάδα, ο αλγόριθμος εξετάζει κάθε δείγμα σε σχέση με τα κεντροειδή των συστάδων. Με χρήση κάποιου μέτρου απόστασης, αναθέτει το εξεταζόμενο δείγμα στη συστάδα, της οποίας το κεντροειδές είναι το πλησιέστερο ως προς το συγκεκριμένο δείγμα. Στο δεύτερο βήμα, παίρνοντας τον μέσο όρο των δειγμάτων κάθε συστάδας, επανυπολογίζονται τα κεντροειδή της κάθε συστάδας, ώστε το κεντροειδές να είναι πιο αντιπροσωπευτικό στην πρόσφατα διαμορφωμένη συστάδα. Ο αλγόριθμος εκτελεί επαναληπτικά αυτά τα δύο βήματα, μέχρις ότου συμπληρωθεί συγκεκριμένο πλήθος επαναλήψεων ή τα κεντροειδή των συστάδων να μην μετατοπίζονται περεταίρω(δηλαδή να μην μετατοπιστούν από την τελευταία επανάληψη).

1. Αρχικοποίησε τυχαία τα k κεντροειδη των συστάδων $\mu_1, \mu_2,..,\mu_K$

2. Επαναληπτική διαδικασία
    i. Εξέτασε κάθε δείγμα και ανέθεσε το στη συστάδα με το πλησιέστερο κεντροειδές$\left(min\left|x^{(i)} - \mu_\kappa\right|^2\right)$
    ii. Επανυπολόγισε τα κεντροειδή υπολογίζοντας το μέσο όρο των δειγμάτων της συστάδας

3.5. Εκπαίδευση Δυο Σταδίων του Υβριδικού Δικτύου RBF

Για την εκπαίδευση του υβριδικού RBF δεν θα χρησιμοποιηθεί ο αλγόριθμος backpropagation αλλά η μέθοδος εκπαίδευσης δυο σταδίων. Στο πρώτο στάδιο εφαρμόζεται ο αλγόριθμος K-means για την εκπαίδευση του κρυφού επιπέδου με μη επιβλεπόμενο τρόπο, ώστε να βρεθούν τα κέντρα των συναρτήσεων Gauss. Επιπλέον στο στάδιο αυτό υπολογίζεται και η διακύμανση σ. Στο δεύτερο στάδιο εφαρμόζουμε εκπαίδευση με επίβλεψη για την εύρεση των βαρών w_j των διασυνδέσεων μεταξύ του κρυφού επιπέδου και του επιπέδου εξόδου. Πιο συγκεκριμένα για την ενημέρωση των βαρών υπολογίζονται οι μερικές παραγωγοί της συνάρτησης σφάλματος(συνάρτηση κόστους) ως προς τα βάρη.

$$E = \frac{1}{2}\sum_{i=1}^{n}\left(F_{x_i} - y_i\right)^2 \qquad \text{Εξ. 3.8}$$

$$\frac{\partial \mathrm{E}}{\partial w_j} = \left(F_{x_i} - y_i\right)\varphi_{j(x,x_j)} \qquad \text{Εξ. 3.9}$$

$$w_j = w_j + \eta\left(F_{x_i} - t_i\right)\varphi_{j(x,x_j)} \qquad \text{Εξ. 3.10}$$

ΚΕΦΑΛΑΙΟ 4.
ΠΡΟΣΟΜΟΙΩΣΗ ΔΙΚΤΥΩΝ ΚΑΙ ΑΞΙΟΛΟΓΗΣΗ ΔΕΔΟΜΕΝΩΝ

4.1. Ανάλυση Προγραμματιστικού Μοντέλου που Εκπονήθηκε
4.2. Ο Ρόλος των Παραμέτρων στο Δίκτυο MLP
4.3. Απόδοση του MLP στις 4 Συναρτήσεις που Επιλέξαμε
4.4. Απόδοση του RBF στις 4 Συναρτήσεις που Επιλέξαμε
4.5. Σύγκριση του Υβριδικού Δικτύου με το Εκτεταμένο MLP
4.6. Προσομοίωση του Υβριδικού Δικτύου για αριθμό νευρώνων RBF = 30
4.7. Διαφοροποίηση των Δεδομένων Εκπαίδευσης με τα Δεδομένα Ελέγχου

4.1. Ανάλυση του Προγραμματιστικού Μοντέλου που Εκπονήθηκε

Σε αυτό το κεφάλαιο θα ασχοληθούμε με τα αποτελέσματα από την προσομοίωση των μεθόδων που παρουσιάστηκαν. Επικεντρωθήκαμε στην παρούσα διατριβή στο πόσο καλά ή όχι συγκλίνει το υβριδικό δίκτυο σε σχέση με το εκτεταμένο αναλύοντας παράλληλα και τον χρόνο εκτέλεσης που χρειάζεται το καθένα. Οι μετρικές που χρησιμοποιήσαμε πέραν του χρόνου εκτέλεσης ήταν η σύγκλιση σε σχέση με την αρχική συνάρτηση καθώς και το μέσο τετραγωνικό σφάλμα σε εκατοστιαία μορφή, το οποίο, για χάριν ευκολίας, από εδώ και στο εξής θα ονομάζεται απλώς σφάλμα, ενώ η διαφορά μεταξύ πραγματικής τιμής και τιμής που έχει προβλεφθεί θα ονομάζεται πραγματικό σφάλμα

Για να κατασκευάσουμε τον κώδικα προσομοίωσης χρησιμοποιήσαμε το εργαλείο MATLAB στην έκδοση 2016B. Ο λόγος που επιλέχθηκε είναι η απλότητα έκφρασης και δημιουργίας γραφικών παραστάσεων σε συνδυασμό με την προηγούμενη εμπειρία σε αυτό. Ο υπολογιστής που έγιναν όλα τα πειράματα ήταν πάντοτε ο ίδιος με χαρακτηριστικά: Μνήμη Ram 4 GB και επεξεργαστής 5 γενιάς i5 7200U στα 2.50 GHz . Για να δημιουργήσουμε ακολουθήσαμε την εξής τακτική. Πρώτα δημιουργήθηκε ο αλγόριθμος για το δίκτυο MLP ο οποίος εκπαιδεύτηκε σε

μία δοκιμαστική συνάρτηση μέχρις ότου διαμορφώθηκε κατά το δοκούν για την περίστασή μας και ακολούθως εφαρμόστηκε και σε άλλες συναρτήσεις

Οι συναρτήσεις τις οποίες δοκιμάσαμε είναι οι εξής τέσσερις:

Α) y=sin(x/10) + (x/50).^2 με πεδίο ορισμού [1,100] και βήμα αύξησης την μονάδα με τη χρήση της εντολής X=[1:100].

Β) y=sin(2πx) με πεδίο ορισμού το [0,0.5] χωρισμένο σε 100 ισοδύναμα σημεία με τη χρήση της εντολής X=linspace(start,end,number_of_elements) όπου στην προκειμένη περίπτωση ήταν X=linspace(0,0.5,100).

Γ) y=2x^2+e^(π/x)*sin(2πx) με πεδίο ορισμού το [1,3] χωρισμένο σε 100 ισοδύναμα σημεία όπως και στην προηγούμενη περίπτωση.

Δ) y=x*sin(x^2) με πεδίο ορισμού το [-4,4] χωρισμένο ομοίως όπως και στις προηγούμενες δύο περιπτώσεις.

Βλέπουμε πως εξετάστηκαν συναρτήσεις που εμπεριέχουν ημιτονοειδή, εκθετική και υψωμένη στο τετράγωνο μορφή του X ούτως ώστε να προσδίδουν αρκετή δυσκολία στην προσέγγισή τους

Η διαδικασία κατασκευής του μοντέλου είχε ως εξής

- Αρχικοποίηση του X ανά περίπτωση
- Υπολογισμός του πραγματικού Y ανά περίπτωση
- Αρχικοποίηση των παραμέτρων για το δίκτυο MLP στην απλή του μορφή
- Κλήση της συνάρτησης για το MLP με έξοδο το y που προσομοίωσε ο αλγόριθμος και το σφάλμα που παρείχε
- Αρχικοποίηση των παραμέτρων για το δίκτυο MLP ως μέλος αυτή τη φορά του υβριδικού δικτύου
- Κλήση της συνάρτησης για το MLP με έξοδο το y που προσομοίωσε ο αλγόριθμος και το σφάλμα που παρείχε
- Αρχικοποίηση του RBF δικτύου με είσοδο την διαφορά της εξόδου του MLP από το πραγματικό Y με έξοδο το y που προσομοίωσε ο αλγόριθμος και το σφάλμα ως προς την πραγματική τιμή.
- Σύγκριση των δύο αλγορίθμων

Εδώ θα πρέπει να σημειωθεί ότι ο αλγόριθμος που δημιουργήθηκε για το MLP είναι ένας και κάθε φορά ανάλογα αν είναι η περίπτωση του εκτεταμένου δικτύου ή του υβριδικού καλείται με διαφορετικές παραμέτρους. Από εκεί και πέρα για να

φτάσουμε στην τελική μορφή του μοντέλου προσπαθήσαμε να ελέγξουμε το ρόλο των παραμέτρων του κάθε δικτύου με εκτέλεση πειραμάτων για διαφορετικούς συντελεστές κάθε φορά μέχρι να βρούμε εμπειρικά την ιδανική ρύθμιση των παραμέτρων

Οι παράμετροι για το MLP δίκτυο πέραν των εισόδων X και πραγματικού Y είναι

- Αριθμός εποχών
- Αριθμός νευρώνων
- Ρυθμός εκμάθησης σε είσοδο και έξοδο

εντός του δικτύου προσαρμόζονται κάθε φορά και τα βάρη με βάση τους τύπους που αναφέρθηκαν στο προηγούμενο κεφάλαιο

Οι παράμετροι για το RBF δίκτυο είναι (πέραν των εισόδων των X και Y) οι εξής

- Αριθμός νευρώνων
- Βοηθητική παράμετρος σ(τυπική απόκλιση) η οποία όσο αυξάνεται παράγει και πιο ομαλά αποτελέσματα (ιδανική εμπειρική τιμή γύρω στο 10)

Ο αλγόριθμος για το RBF καλεί και την kmeans για να βρεθούν τα κέντρα των συναρτήσεων Gauss των κρυφών κόμβων του δικτύου RBF

Για την ανάλυση και αξιολόγηση του μοντέλου επιλέξαμε την εξής δομή. Στην επόμενη παράγραφο αναλύουμε την επίδραση των παραμέτρων του MLP στην αποδοτικότητα του μοντέλου με βάση μία συνάρτηση (την Δ εν προκειμένω) με σκοπό να επιλέξουμε τις κατάλληλες τιμές για την χρήση του δικτύου μετά για όλες τις συναρτήσεις. Δεν παρουσιάζεται παρόμοια τακτική για το RBF δίκτυο λόγω της απλότητάς του στη ρύθμιση των αρχικών παραμέτρων μιας και ο αριθμός νευρώνων ακολουθεί παραπλήσια διακύμανση στο αποτέλεσμα με τον αριθμό νευρώνων για το RBF και η παράμετρος σ(τυπική απόκλιση) δεν έχει ιδιαίτερα μεγάλες διακυμάνσεις ώστε να χρήζει αναφοράς σε ξεχωριστό κεφάλαιο.

4.2. Ο Ρόλος των Παραμέτρων στο Δίκτυο MLP

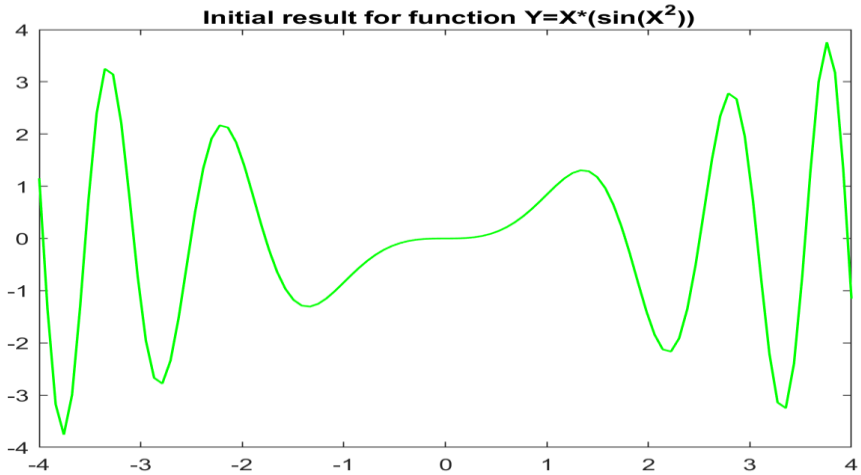
Σε αυτή την παράγραφο παρουσιάζονται τα αποτελέσματα από την ανάλυση των παραμέτρων του MLP και πόσο αυτές επηρεάζουν το τελικό αποτέλεσμα του αλγορίθμου αυτού. Εδώ θα πρέπει να σημειωθεί ότι κάθε φορά που αλλάζαμε μία παράμετρο οι άλλες παρέμεναν σταθερές σε μία τιμή τέτοια ώστε οι μεταβολές

στην παράμετρο που επιλέξαμε για μελέτη να είναι ευδιάκριτες. Εμπειρικά αυτές οι τιμές ήταν

- Αριθμός εποχών ίσος με 5000
- Αριθμός νευρώνων 48
- Ρυθμός εκμάθησης εισόδου 0.1 και εξόδου 0.01
- Τυχαία αρχικοποίηση βαρών για καλύτερη γενίκευση του δικτύου.

Παρατηρήσαμε ότι για να πετύχουμε μέγιστη ακρίβεια αδιαφορώντας για το υπολογιστικό κόστος θα πρέπει να πάρουμε αριθμό εποχών κοντά στο 20.000, αριθμό νευρώνων μεγαλύτερο του 64 και αρχικοποίηση ρυθμού εκμάθησης όπως αναφέρθηκε παραπάνω. Για λόγους ισορροπίας μεταξύ υπολογιστικού κόστους και ακρίβειας, αλλά και να γίνουν πιο εμφανείς οι διαφορές των υπολοίπων παραμέτρων, εμπειρικά αποφασίσαμε να χρησιμοποιήσουμε τις παραπάνω παραμέτρους όπως αυτές ορίστηκαν στην προηγούμενη παράγραφο του συγκεκριμένου υποκεφαλαίου

Η πραγματική συνάρτηση που θα έπρεπε να προσεγγίσει ο αλγόριθμος είναι, όπως προαναφέρθηκε, η Δ του προηγούμενου υποκεφαλαίου και της οποίας η πραγματική μορφή είναι η παρακάτω



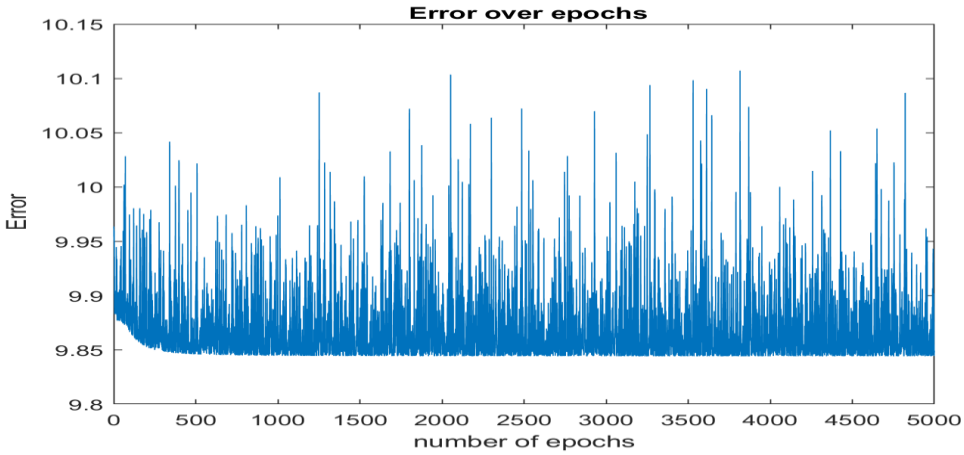
Σχήμα 4.1 Γραφική Συνάρτησης $y=x\sin(x^2)$

4.2.1. Ρόλος του Αριθμού Νευρώνων

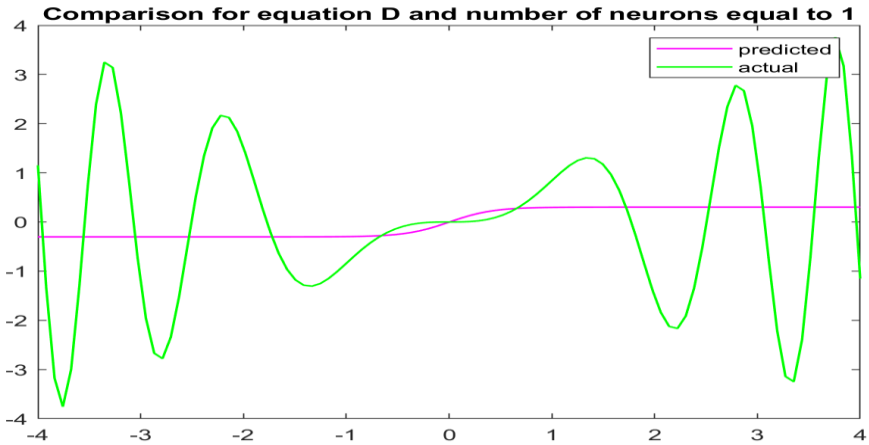
Σε αυτή την περίπτωση εργαστήκαμε κρατώντας σταθερές τις υπόλοιπες παραμέτρους και μεταβάλλοντας τον αριθμό των νευρώνων ξεκινώντας από το 1

και διπλασιάζοντάς το μέχρι και το 64. Παρατηρήσαμε ότι ανεβάζοντας τον αριθμό η απόδοση του MLP αυξανόταν λίγο στην αρχή, ραγδαία μετέπειτα και στο τέλος παρατηρήσαμε μικρή βελτίωση όσον αφορά την απόδοση του αλγορίθμου. Πιο συγκεκριμένα, εκτελέσαμε τον αλγόριθμο 7 φορές και τα αποτελέσματα που προέκυψαν ήταν τα εξής

N=1

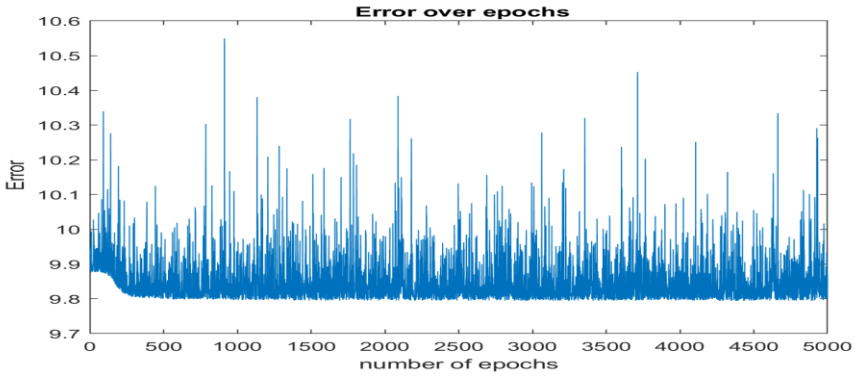


Σχήμα 4.2 Σφάλμα για MLP Δίκτυο με 1 Νευρώνα

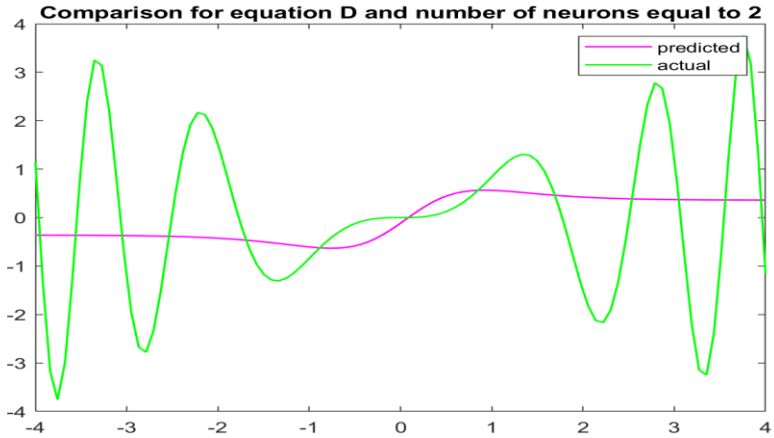


Σχήμα 4.3 Προσέγγιση Συνάρτησης $y=x\sin(x^2)$ με MLP με1 Νευρώνα

N=2

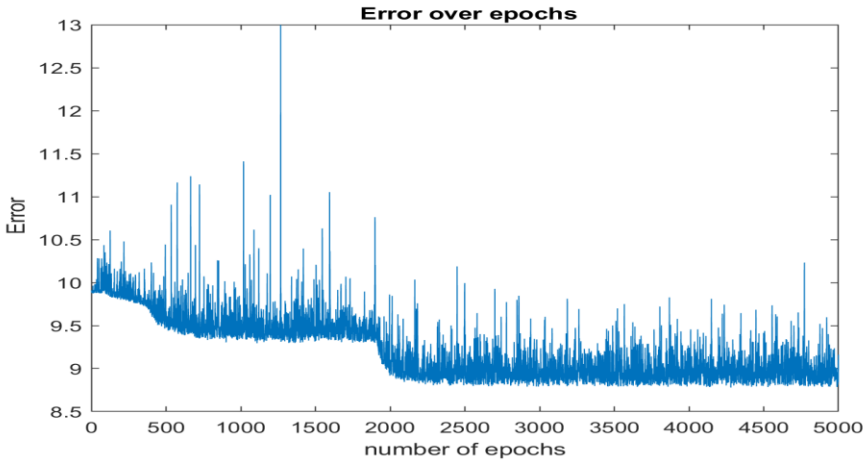


Σχήμα 4.4 Σφάλμα για MLP Δίκτυο με 2 Νευρώνες

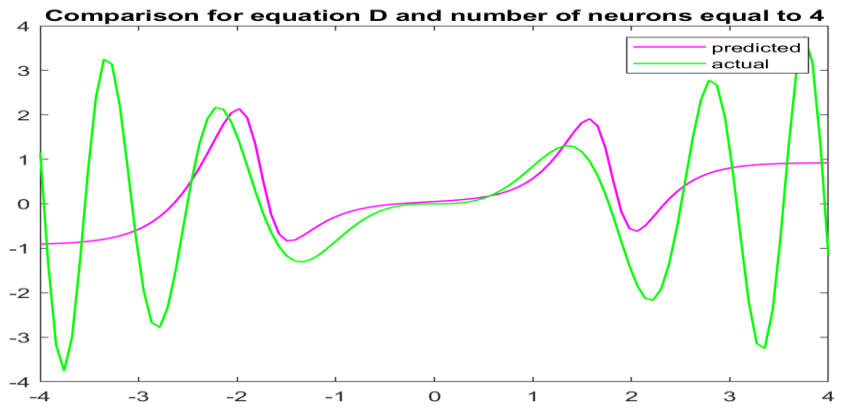


Σχήμα 4.5 Προσέγγιση Συνάρτησης $y=x\sin(x^2)$ με MLP με 2 Νευρώνες

N=4

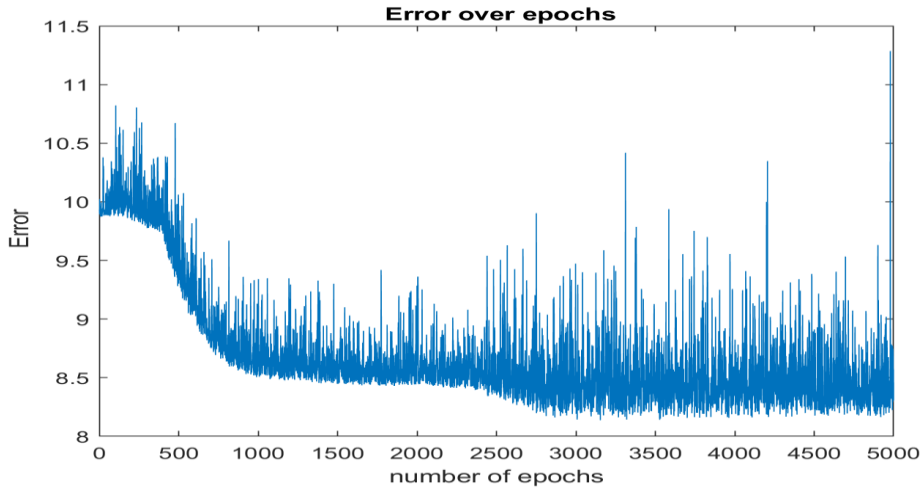


Σχήμα 4.6 Σφάλμα για MLP Δίκτυο με 4 Νευρώνες

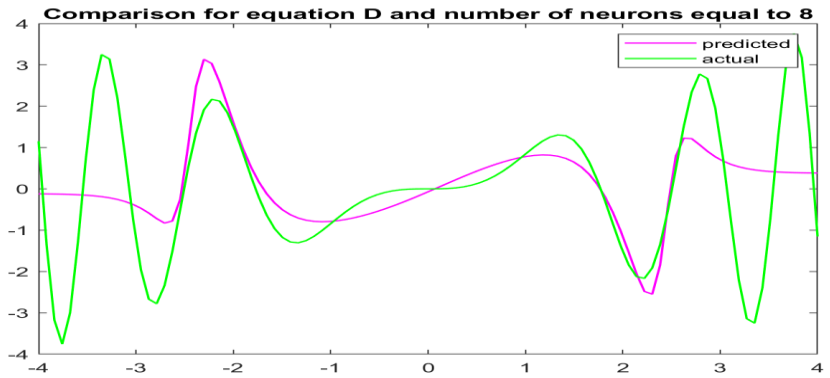


Σχήμα 4.7 Προσέγγιση Συνάρτησης y=xsin(x^2) με MLP με 4 Νευρώνες

N=8

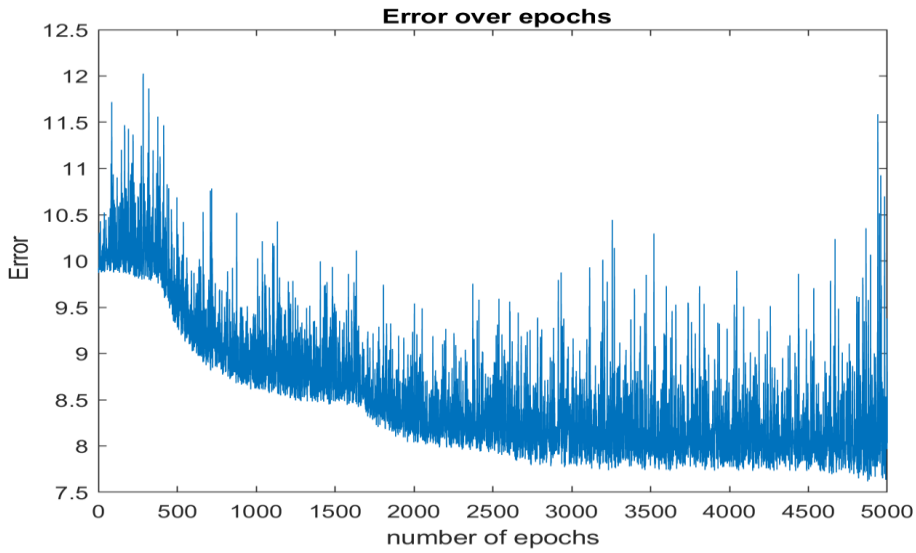


Σχήμα 4.8 Σφάλμα για MLP Δίκτυο με 8 Νευρώνες

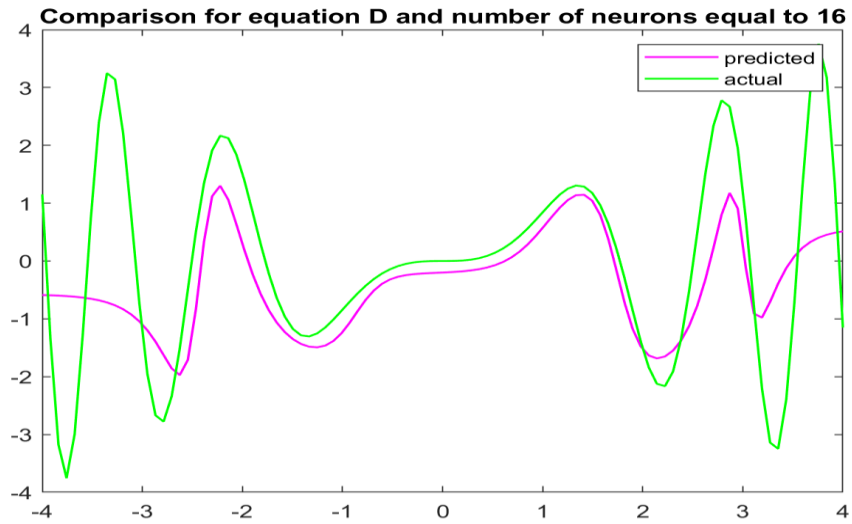


Σχήμα 4.9 Προσέγγιση Συνάρτησης y=xsin(x^2) με MLP με 8 Νευρώνες

N=16

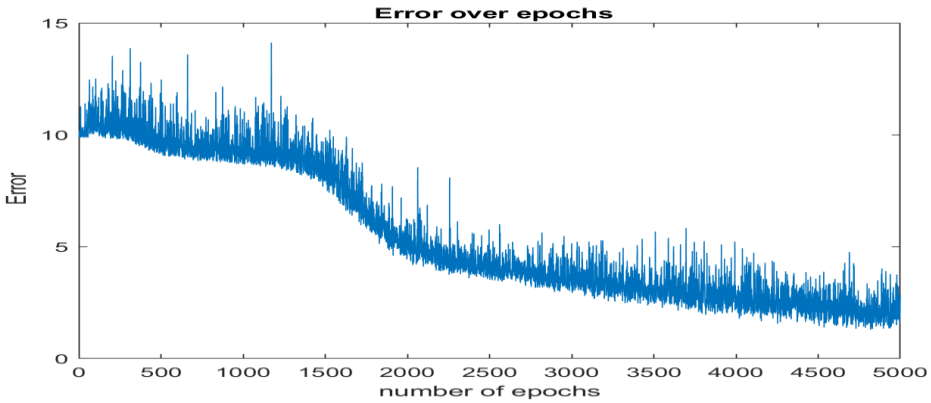


Σχήμα 4.10 Σφάλμα για MLP Δίκτυο με 16 Νευρώνες

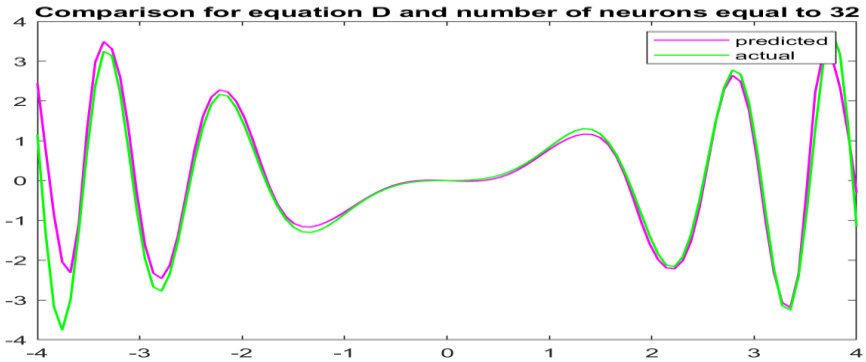


Σχήμα 4.11 Προσέγγιση Συνάρτησης $y=x\sin(x^2)$ με MLP με 16 Νευρώνες

N=32

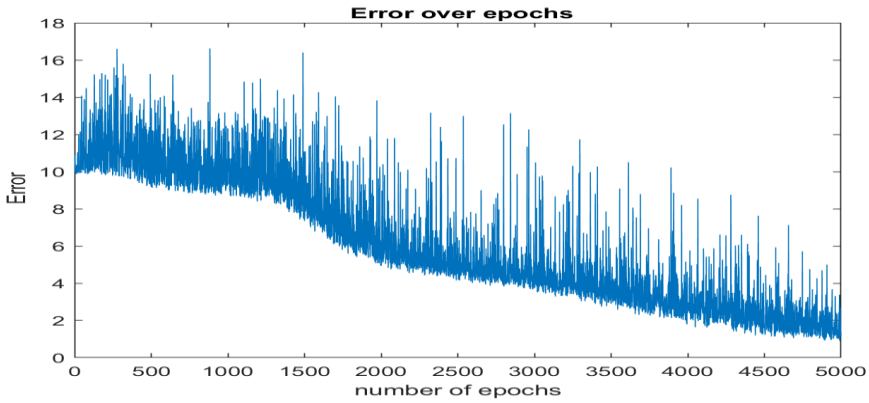


Σχήμα 4.12 Σφάλμα για MLP Δίκτυο με 32 Νευρώνες

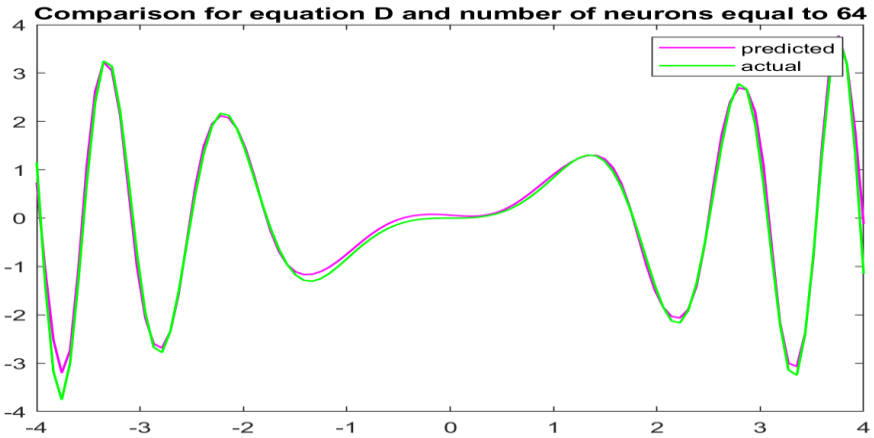


Σχήμα 4.13 Προσέγγιση Συνάρτησης $y=x\sin(x^2)$ με MLP με 32 Νευρώνες

N=64



Σχήμα 4.14 Σφάλμα για MLP Δίκτυο με 64 Νευρώνες



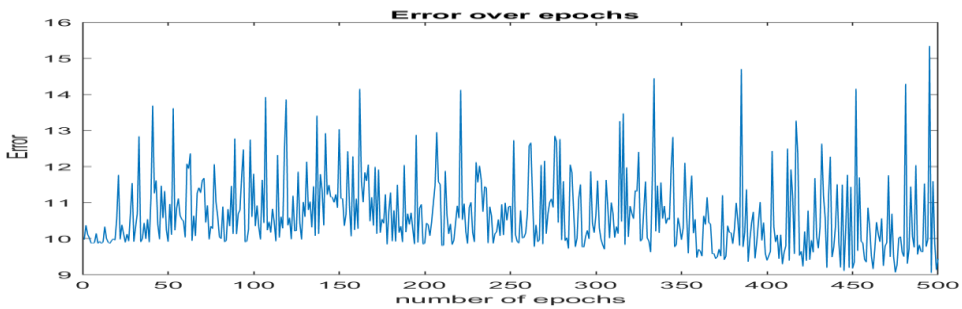
Σχήμα 4.15 Προσέγγιση Συνάρτησης $y=x\sin(x^2)$ με MLP με 64 Νευρώνες

Παρατηρούμε λοιπόν ότι μέχρι και τους 16 νευρώνες πετυχαίνουμε μια αρκετά ικανοποιητική ακρίβεια με το σφάλμα να πέφτει έως 3, ενώ αν θέλουμε να πέσουμε σε τιμή κοντά στο 1 θα πρέπει να πάρουμε αριθμό νευρώνων μεγαλύτερο του 64.

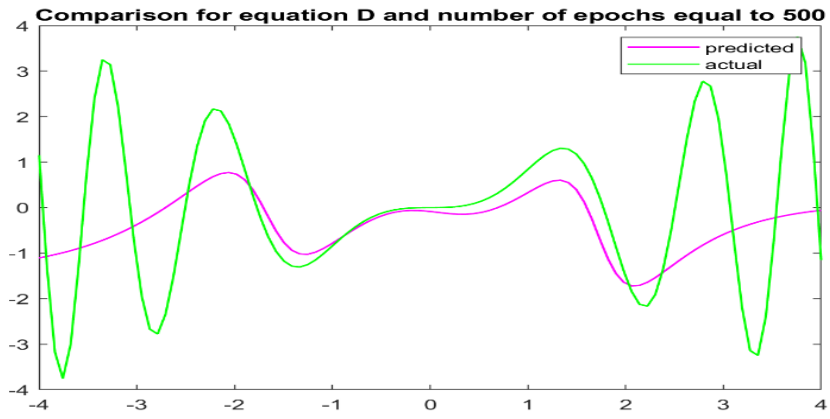
4.2.2. Ο Ρόλος του Αριθμού των Εποχών

Ο αριθμός των εποχών, επειδή είναι στην ουσία επανάληψη (loop) στο προγραμματιστικό κομμάτι αποτέλεσε και τον μοναδικό παράγοντα που έπαιξε ρόλο στο χρόνο εκτέλεσης του αλγορίθμου. Οι επαναλήψεις που εκτελέσαμε ήταν για 500, 1.000, 3.000, 5.000 και 10.000 εποχές διότι παρατηρήσαμε ότι μετά από αυτό τον αριθμό ο αλγόριθμος δεν παρουσίαζε μεγάλες μεταβολές στο αποτέλεσμα σύγκλισης και στο σφάλμα που προέκυψε. Παρακάτω παρουσιάζουμε τα αποτελέσματα από την εκτέλεση του αλγορίθμου

Εποχές = 500

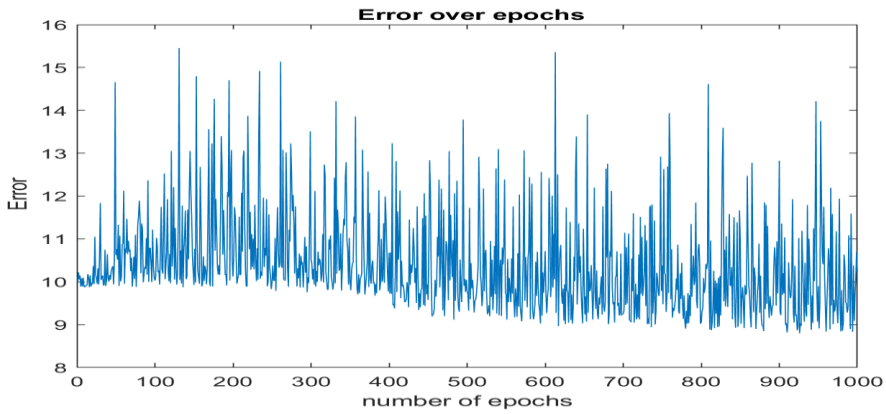


Σχήμα 4.16 Σφάλμα για MLP Δίκτυο για 500 Εποχές

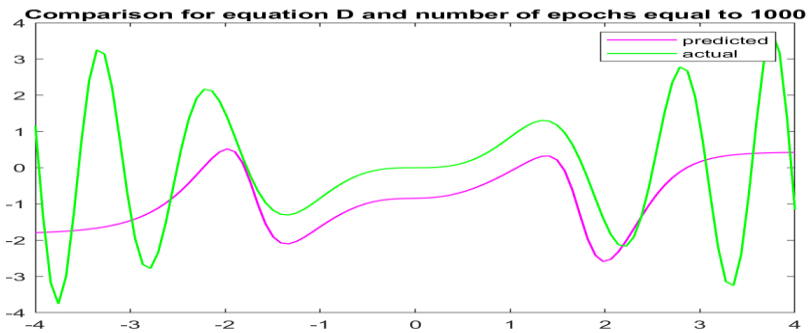


Σχήμα 4.17 Προσέγγιση Συνάρτησης y=xsin(x^2) με MLP με 500 Εποχές

Εποχές=1000

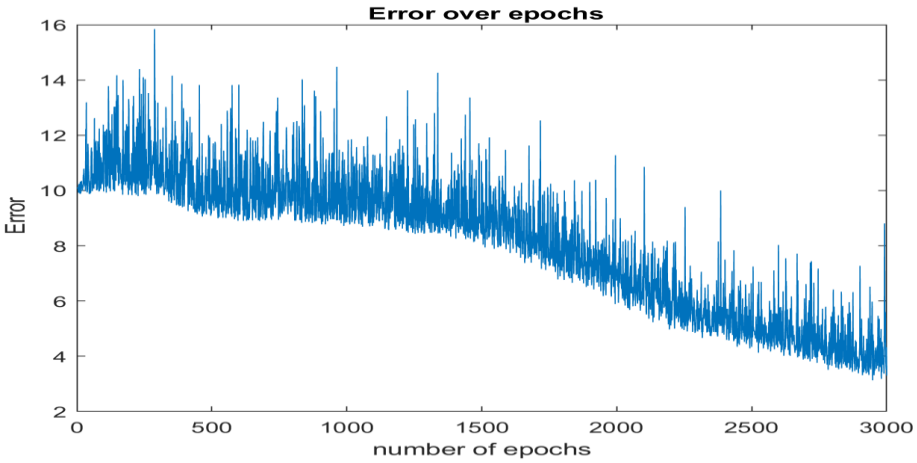


Σχήμα 4.18 Σφάλμα για MLP Δίκτυο για 1000 Εποχές

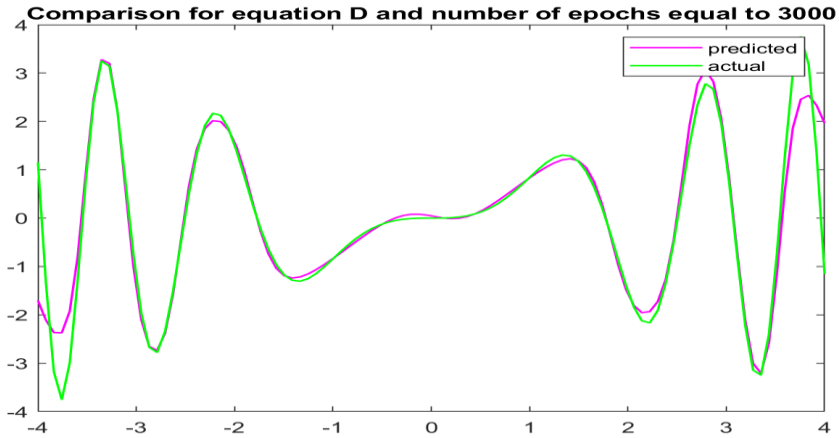


Σχήμα 4.19 Προσέγγιση Συνάρτησης y=xsin(x^2) με MLP με 1000 Εποχές

Εποχές=3000

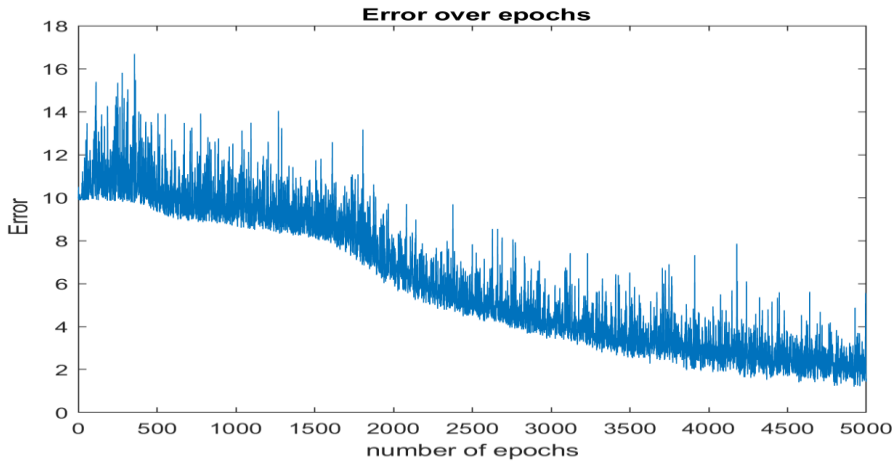


Σχήμα 4.20 Σφάλμα για MLP Δίκτυο για 3000 Εποχές

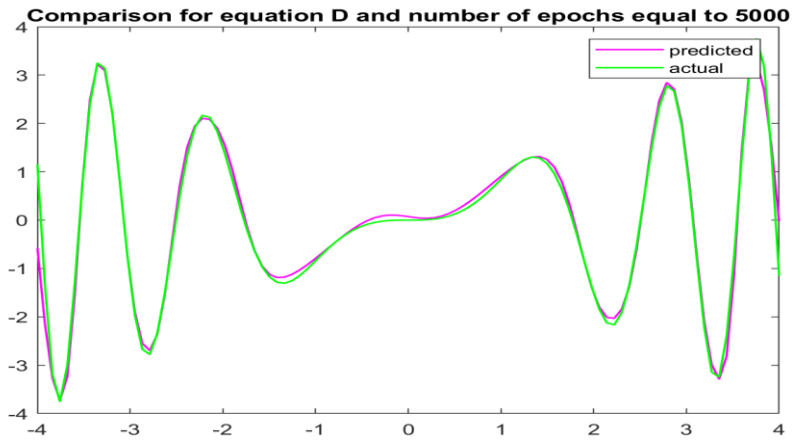


Σχήμα 4.21 Προσέγγιση Συνάρτησης $y=x\sin(x^2)$ με MLP με 3000 Εποχές

Εποχές=5000

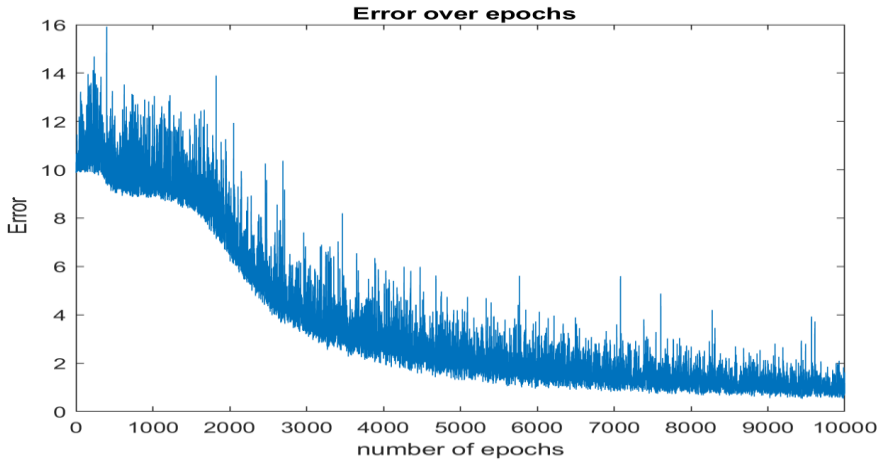


Σχήμα 4.22 Σφάλμα για MLP Δίκτυο για 5000 Εποχές

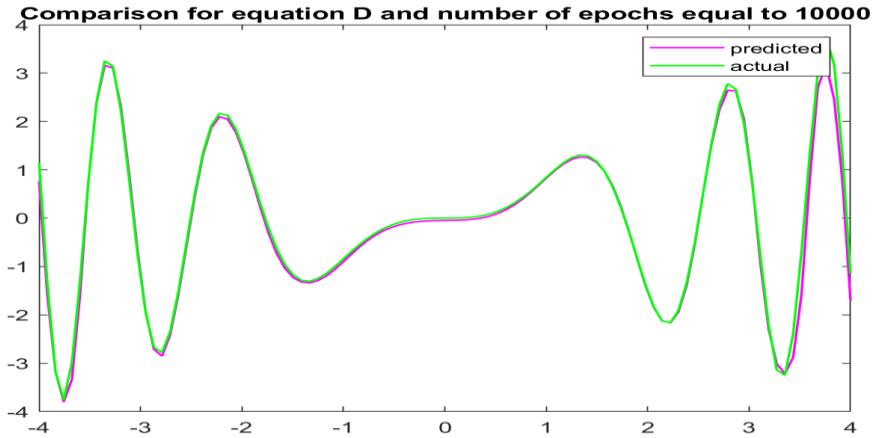


Σχήμα 4.23 Προσέγγιση Συνάρτησης $y=x\sin(x^2)$ με MLP με 5000 Εποχές

Εποχές=10000



Σχήμα 4.24 Σφάλμα για MLP Δίκτυο για 10000 Εποχές



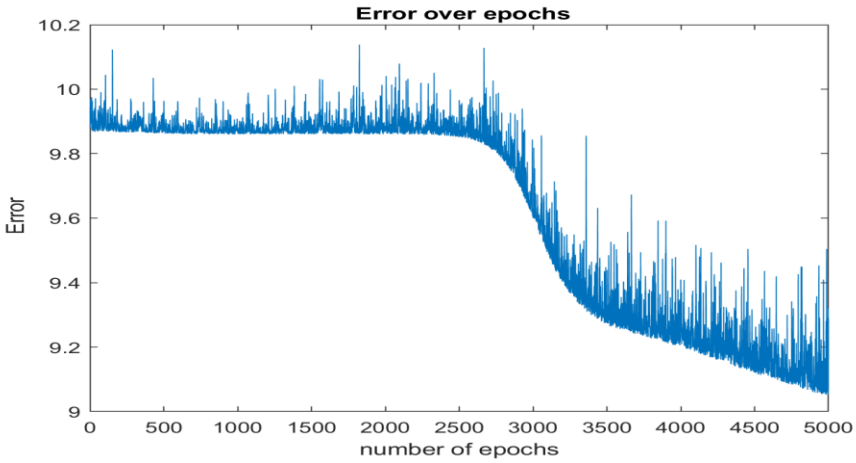
Σχήμα 4.25 Προσέγγιση Συνάρτησης $y=x\sin(x^2)$ με MLP με 10000 Εποχές

Παρατηρούμε, λοιπόν, ότι και εδώ όπως και πριν όσο αυξάνεται ο αριθμός εποχών, τόσο βελτιώνεται η απόκριση του αλγορίθμου καθώς επίσης μειώνεται και το σφάλμα. Από τις 3.000 εποχές και έπειτα όμως δεν παρατηρούμε ραγδαία αλλαγή, όμως το σφάλμα παραμένει μεγαλύτερο του 1.

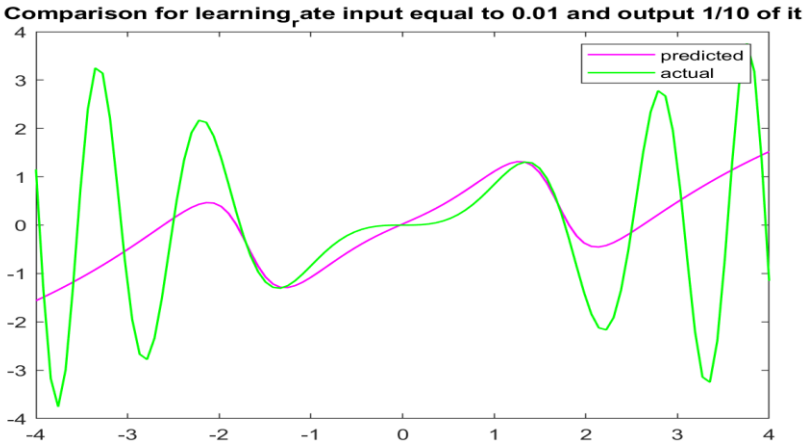
4.2.3. Ο Ρόλος του Ρυθμού Εκμάθησης

Εδώ τα πράγματα δεν ακολουθούν την ίδια λογική με επάνω. Παρατηρήσαμε ότι ιδανική τιμή είναι το 0.1 για το ρυθμό εισόδου και 0.01 για το ρυθμό εξόδου. Επίσης, παρατηρήσαμε ότι τιμές με αρνητικό πρόσημο ή τιμές μεγαλύτερες του 0.5 δίνανε άπειρη τιμή στο σφάλμα όπως επίσης ότι η ιδανική περίπτωση ήταν το σφάλμα εξόδου να είναι το 1/10 του σφάλματος εισόδου. Έτσι, οι προσομοιώσεις που κάναμε και που παρουσιάζονται παρακάτω ήταν τρεις και ήταν για ρυθμό εισόδου 0.01, 0.1, 0.3 με το ρυθμό εξόδου να είναι το 1/10 αυτού σε κάθε περίπτωση.

Ρυθμός εισόδου 0.01 και εξόδου 0.001

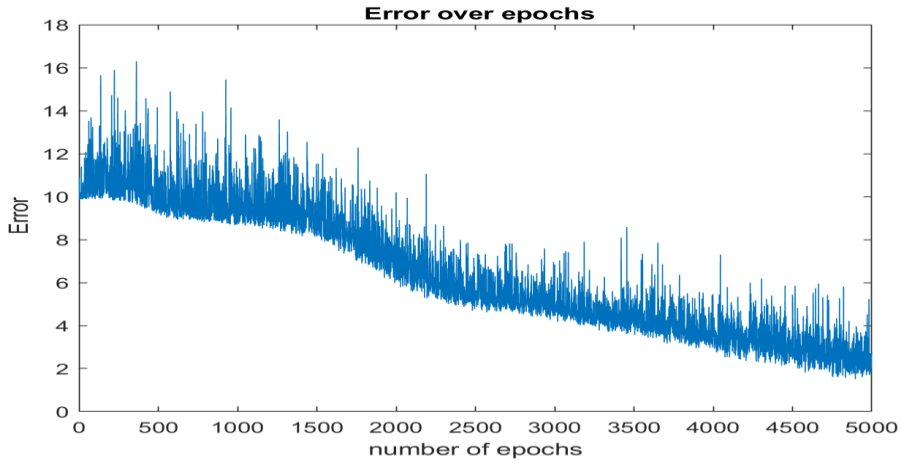


Σχήμα 4.26 Σφάλμα για MLP Δίκτυο με Learning Rate 0.01 και 0.001

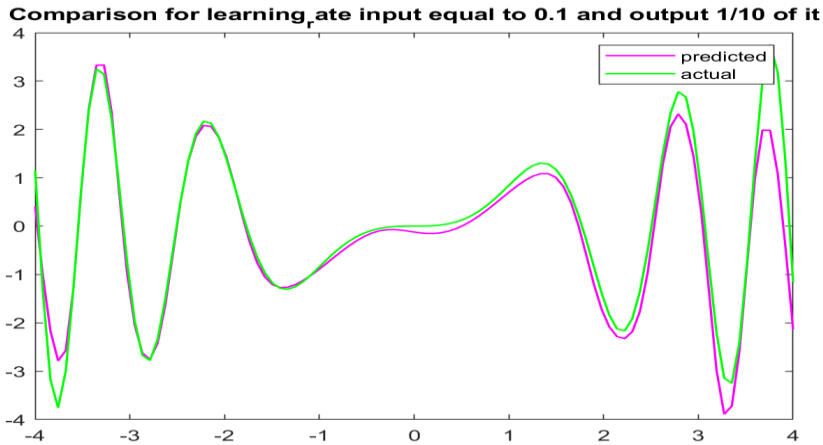


Σχήμα 4.27 Προσέγγιση Συνάρτησης $y=x\sin(x^2)$ με MLP με Learning Rate 0.01 και 0.001

Ρυθμός εισόδου 0.1 και εξόδου 0.01

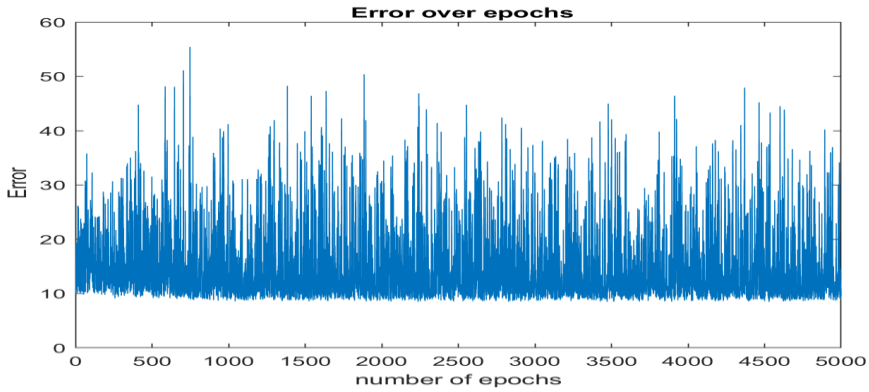


Σχήμα 4.28 Σφάλμα για MLP Δίκτυο με Learning Rate 0.1 και 0.01

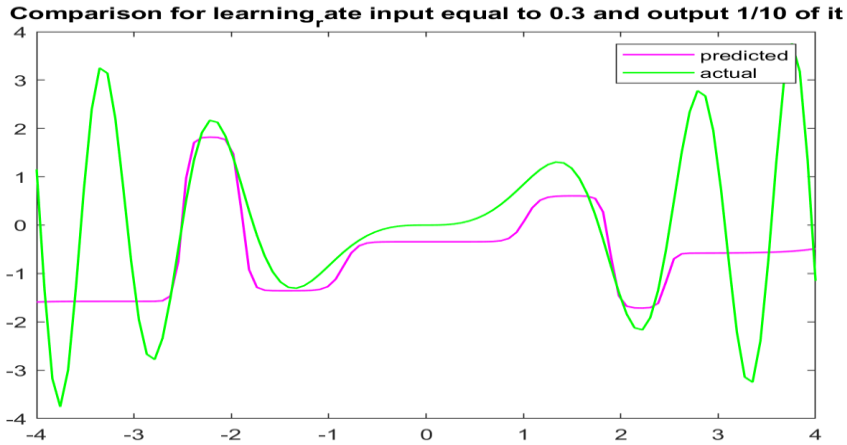


Σχήμα 4.29 Προσέγγιση Συνάρτησης $y=x\sin(x^2)$ με MLP με Learning Rate 0.1 και 0.01

Ρυθμός εισόδου 0.3 και εξόδου 0.03



Σχήμα 4.30 Σφάλμα για MLP Δίκτυο με Learning Rate 0.3 και 0.03



Σχήμα 4.31 Προσέγγιση Συνάρτησης $y=x\sin(x^2)$ με MLP με Learning Rate 0.3 και 0.03

Παρατηρούμε, λοιπόν, μονάχα οι τιμές 0.1 και 0.01 αντίστοιχα για είσοδο και έξοδο δίνουν τιμές ικανές να προσεγγίσουν τη συνάρτηση. Αυτό σημαίνει ότι μόνο αυτές τις τιμές θα χρησιμοποιήσουμε από εδώ και πέρα για οποιαδήποτε προσομοίωση κατά τη διάρκεια της παρούσας διατριβής.

4.2.4. Ο Ρόλος της Αρχικοποίησης των Βαρών

Επειδή θέλαμε να πετύχουμε όσο το δυνατό μεγαλύτερη δυνατότητα γενίκευσης του μοντέλου επιλέξαμε τυχαία τιμή κάθε φορά στα βάρη συναρτώμενη από τον

αριθμό των νευρώνων και μέσα στο διάστημα [-0.5,0.5]. Η δημιουργία έγινε με βάση τη συνάρτηση randn του MATLAB.

4.2.5. Διακύμανση του Χρόνου Προσομοίωσης και του Σφάλματος Ανάλογα με την Παράμετρο που Μεταβάλλονταν σε Κάθε Περίπτωση

Όπως αναφέρθηκε παραπάνω, ο χρόνος προσομοίωσης καθορίστηκε από τον αριθμό των εποχών και όχι από άλλες παραμέτρους. Έτσι στους παρακάτω πίνακες παρουσιάζονται οι χρονικές μεταβολές καθώς και οι μεταβολές της ελάχιστης τιμής του σφάλματος για κάθε μία από τις παραπάνω τρεις περιπτώσεις όπου μεταβάλλαμε τιμές σε κάποια παράμετρο

- Για μεταβαλλόμενο τον αριθμό των εποχών οι τιμές που πήραμε ήταν οι εξής:

Πίνακας 4.1: Τιμές για Ελάχιστο Σφάλμα Σύμφωνα με τις Εποχές

Αριθμός εποχών	Ελάχιστο σφάλμα
500	9.04
1,000	8.74
3,000	3.35
5,000	1.29
10,000	0.55

Παρατηρήσαμε ότι, όσο ο αριθμός των εποχών μεγάλωνε το σφάλμα μειωνόταν.

Πίνακας 4.2: Τιμές για Ελάχιστο Σφάλμα Σύμφωνα με τους Νευρώνες

Αριθμός νευρώνων	Ελάχιστο σφάλμα
1	9.84
2	9.79
4	8.77
8	8.06
16	2.83
32	1.22
64	1.11

Παρατηρούμε ότι το σφάλμα ελαττώνεται αργά μέχρι τους 8 νευρώνες και από εκεί και έπειτα ελαττώνεται ραγδαία με κάθε αύξηση αριθμού νευρώνων

• Για μεταβαλλόμενο τον ρυθμό εκμάθησης εισόδου και εξόδου οι τιμές που πήραμε ήταν οι εξής:

Πίνακας 4.3: Τιμές για Ελάχιστο Σφάλμα Σύμφωνα με το Ρυθμό μάθησης

Ρυθμός εκμάθησης ΕΙΣ/ΕΞ	Ελάχιστο σφάλμα
0.01/0.001	9.18
0.1/0.01	1.14
0.3/0.03	8.46

Παρατηρούμε ότι μόνο για το δεύτερο ζευγάρι το σφάλμα ήταν μικρό. Στις άλλες περιπτώσεις το σφάλμα δεν ήταν σε ικανοποιητικό αριθμό

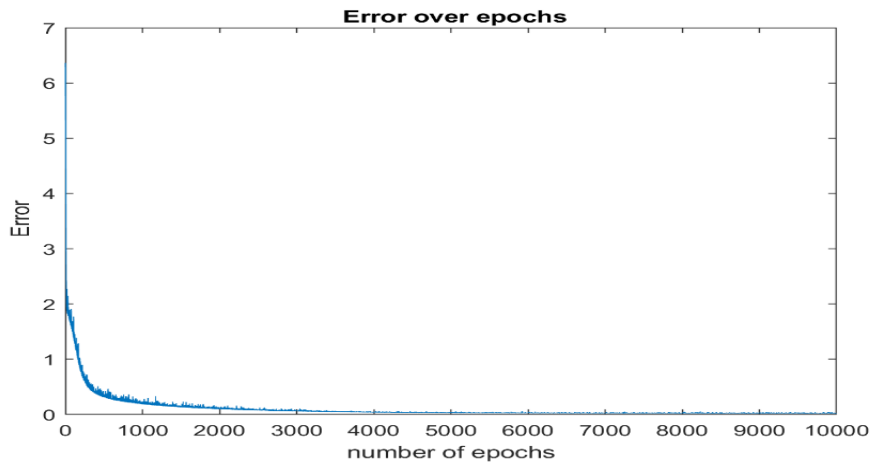
4.3. Απόδοση του MLP στις 4 Συναρτήσεις που Επιλέξαμε

Σε αυτό το υποκεφάλαιο, θα παρουσιάσουμε την απόδοση του απλού MLP για κάθε μία από τις τέσσερις συναρτήσεις που επιλέξαμε να μελετήσουμε κατά τη διάρκεια της παρούσας διατριβής. Οι παράμετροι για το δίκτυο ορίστηκαν με έμφαση στη λεπτομέρεια του αποτελέσματος και την ικανότητα του αλγορίθμου. Έτσι διαλέξαμε τις παραμέτρους ως εξής:

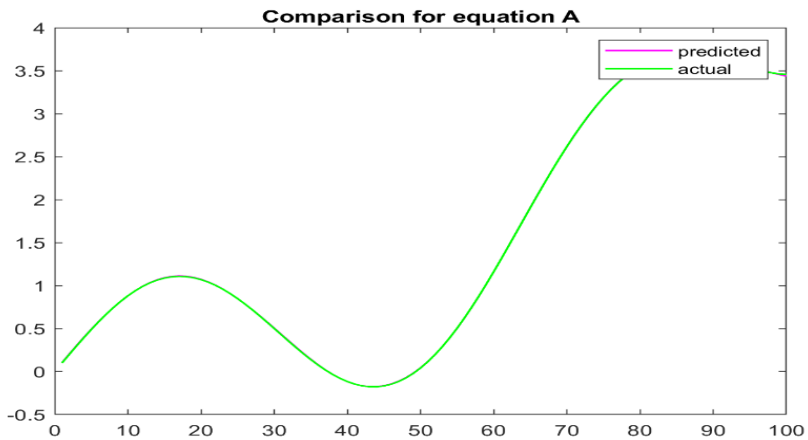
- Αριθμός εποχών ίσος με 10,000
- Αριθμός νευρώνων 64
- Ρυθμός εκμάθησης εισόδου 0.1 και εξόδου 0.01
- Τυχαία αρχικοποίηση βαρών για καλύτερη γενίκευση του δικτύου.

Γενικά, παρατηρήσαμε ότι και στις τέσσερις περιπτώσεις ο αλγόριθμος συγκλίνει εξαιρετικά και ότι θα μπορούσαμε να μικρύνουμε τον αριθμό των δύο πρώτων παραμέτρων ούτως ώστε να αποκομίσουμε οφέλη σε θέματα χρόνου προσομοίωσης και υπολογιστικού κόστους. Ακολούθως, παρουσιάζονται τα αποτελέσματα σε κάθε περίπτωση.

4.3.1. Συνάρτηση A



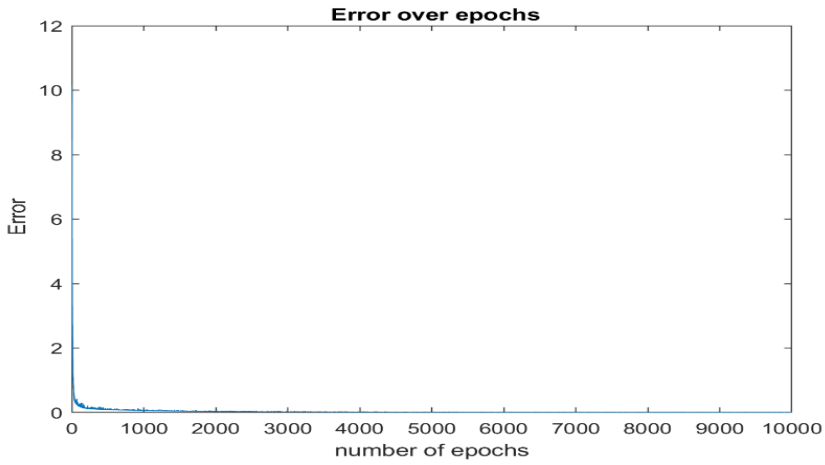
Σχήμα 4.32 Σφάλμα Συνάρτησης y=sin(x/10) + $(x/50)^2$



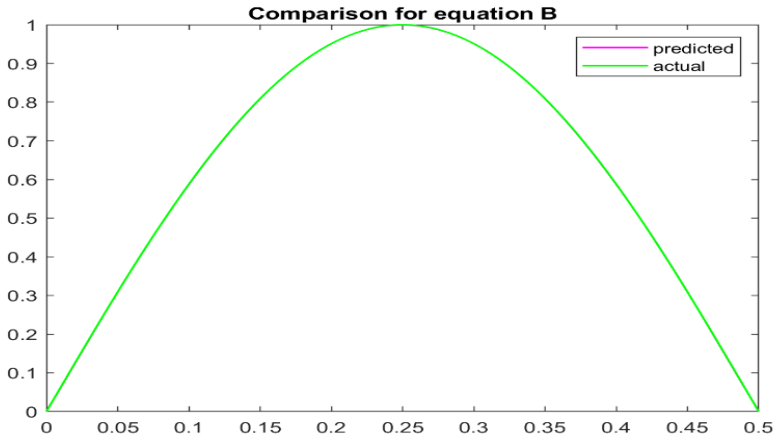
Σχήμα 4.33 Προσέγγιση Συνάρτησης y=sin(x/10) + $(x/50)^2$

Σε αυτή την περίπτωση το σφάλμα ήταν 0.027. Ο αλγόριθμος συγκλίνει γρήγορα σε τιμές κάτω του 0.5 από πολύ νωρίς ενώ βλέπουμε ότι η προσεγγιστική του ικανότητα είναι πολύ καλή.

4.3.2. Συνάρτηση B



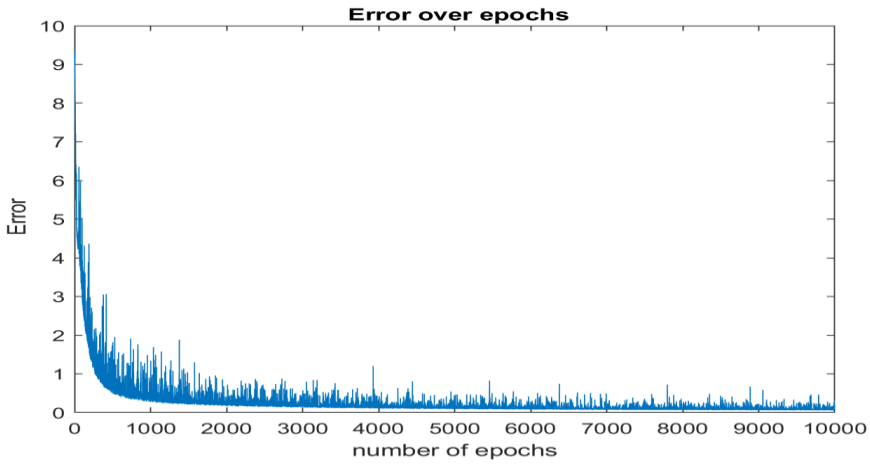
Σχήμα 4.34 Σφάλμα Συνάρτησης y=sin(2πx)



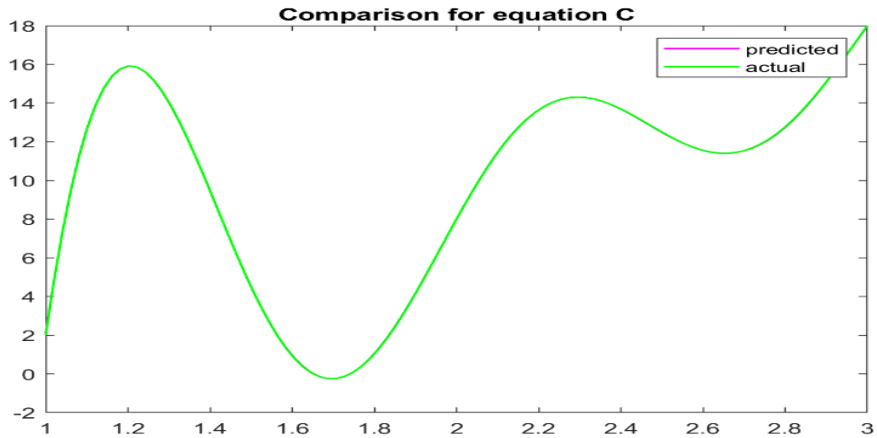
Σχήμα 4.35 Προσέγγιση Συνάρτησης y=sin(2πx)

Σε αυτή την περίπτωση το σφάλμα ήταν 0.012 και βλέπουμε εδώ ότι η συνάρτηση αυτή ήταν η λιγότερο περίπλοκη από όλες. Έτσι, αναμενόμενα, το σφάλμα είναι το μικρότερο από όλες τις περιπτώσεις και βλέπουμε ότι ο αλγόριθμος σύγκλινε σε μικρό σφάλμα από πάρα πολύ νωρίς και η ικανότητά του ήταν η καλύτερη από όλες. Πρακτικά ο συγκεκριμένος αλγόριθμος δε χρειαζόταν πάνω από 700-1000 εποχές για να έχει ένα άκρως επιτυχημένο αποτέλεσμα, αν και πάντα το αποτέλεσμα κρίνεται από το ελάχιστο σφάλμα που θέλει να φτάσει ο χρήστης σε κάθε περίπτωση.

4.3.3. Συνάρτηση Γ



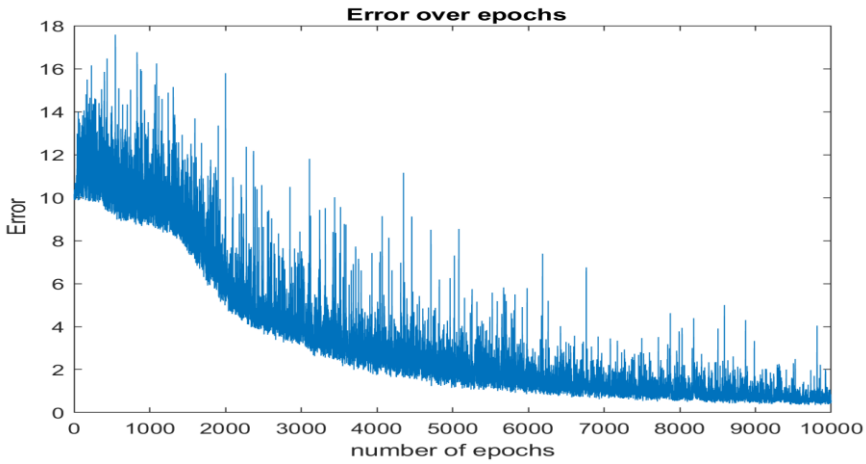
Σχήμα 4.36 Σφάλμα Συνάρτησης $y=2x^2+e^{(\pi/x)}*\sin(2\pi x)$



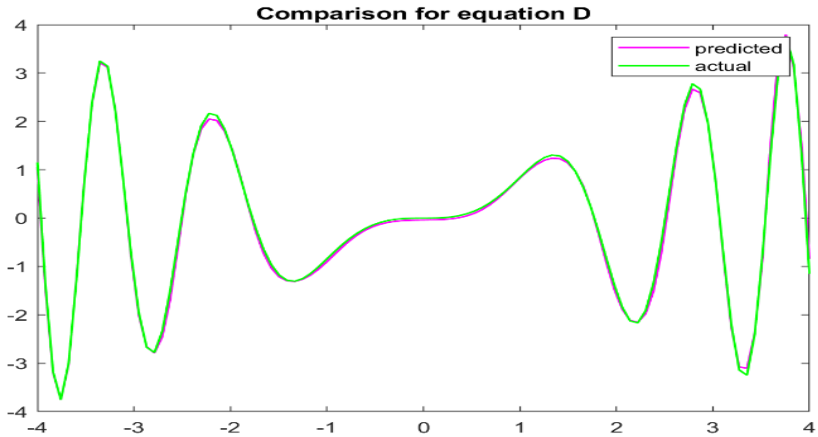
Σχήμα 4.37 Προσέγγιση Συνάρτησης $y=2x^2+e^{(\pi/x)}*\sin(2\pi x)$

Σε αυτή την περίπτωση το σφάλμα ήταν 0.025 και βλέπουμε εδώ ότι λόγω της παρουσίας εκθετικού και τετραγώνου στο X, το σφάλμα άργησε να συγκλίνει. Επίσης, διακρίνουμε μεγάλες ταλαντώσεις στην τιμή του σφάλματος όσο αυτό σύγκλινε αλλά, εν τέλει, παρατηρήσαμε πολύ καλή ικανότητα στην συνάρτηση

4.3.4. Συνάρτηση Δ



Σχήμα 4.38 Σφάλμα Συνάρτησης y=x*sin(x^2)



Σχήμα 4.39 Προσέγγιση Συνάρτησης y=x*sin(x^2)

Σε αυτή την περίπτωση είχαμε την πιο σύνθετη συνάρτηση. Αν και το σφάλμα ήταν 0.2389, που σημαίνει αρκετά μεγαλύτερο από τις υπόλοιπες συναρτήσεις, παρατηρήσαμε ότι το τελικό αποτέλεσμα πρόβλεψης για τη συνάρτηση είναι πάρα πολύ κοντά στο πραγματικό. Επίσης, παρατηρήσαμε ότι το σφάλμα είχε μεγάλες διακυμάνσεις σε κάθε εποχή.

4.4. Απόδοση του RBF στις 4 Συναρτήσεις που Επιλέξαμε

Σε αυτό το υποκεφάλαιο, θα παρουσιάσουμε την απόδοση του RBF χωριστά για κάθε μία από τις τέσσερις συναρτήσεις που επιλέξαμε να μελετήσουμε κατά τη διάρκεια της παρούσας διατριβής. Οι παράμετροι για το δίκτυο ορίστηκαν με έμφαση στη λεπτομέρεια του αποτελέσματος και την ικανότητα του αλγορίθμου. Ο αριθμός νευρώνων παρέμεινε σταθερό και επιλέξαμε να μεταβάλλουμε την παράμετρο τυπική απόκλιση σ το οποίο είναι το μέτρο του πλάτους της Gauss συνάρτησης. Έτσι πραγματοποιήσαμε 4 προσομοιώσεις για κάθε μία από τις 4 συναρτήσεις. Οι τιμές του σ που διαλέξαμε ήταν 0.5,1,2,5.

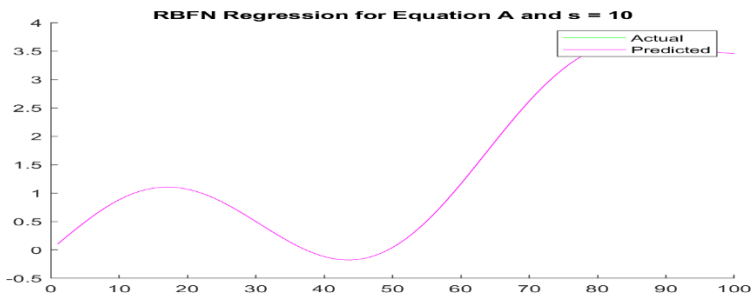
Γενικά, παρατηρήσαμε ότι και στις τέσσερις περιπτώσεις ο αλγόριθμος συγκλίνει εξαιρετικά για όσο μικρότερο αριθμό της παραμέτρου, κάτι που παρατηρήθηκε και για τιμές μικρότερες του 0.5 οι οποίες δεν παρουσιάζονται στην παρούσα διατριβή για το λόγο ότι δεν υπήρχε ευδιάκριτη τιμή στην γενικευτική ικανότητα του αλγορίθμου.

4.4.1. Συνάρτηση A

Για τη συγκεκριμένη συνάρτηση, επειδή το βήμα της είναι ανά μία μονάδα και όχι ανά εκατοστά της μονάδας όπως στις υπόλοιπες το σ εικοσαπλασιάστηκε κι αυτό διότι οι τιμές που αποφασίστηκαν για τα άλλα έδιναν μηδαμινό σφάλμα και δεν παρατηρήθηκε κάποια αξιοσημείωτη αλλαγή. Έτσι τρέξαμε τις παρακάτω προσομοιώσεις:

Για σ=10

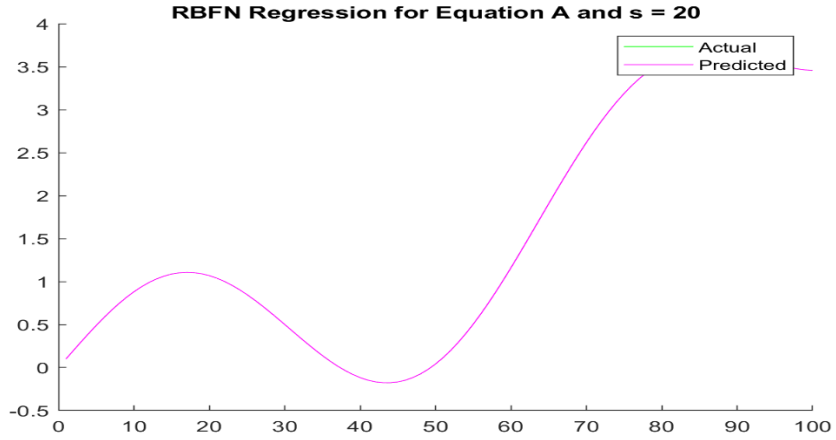
Εδώ παρατηρούμε ότι πρακτικά το RBF πέφτει ακριβώς πάνω στην καμπύλη που καλείται να προσεγγίσει πράγμα που φανερώνει την πολύ καλή γενικευτική ικανότητα του αλγορίθμου για τη συγκεκριμένη τιμή της παραμέτρου αυτής



Σχήμα 4.40 Προσέγγιση Συνάρτησης $y=\sin(x/10) + (x/50)^2$ για σ=10

Για σ=20

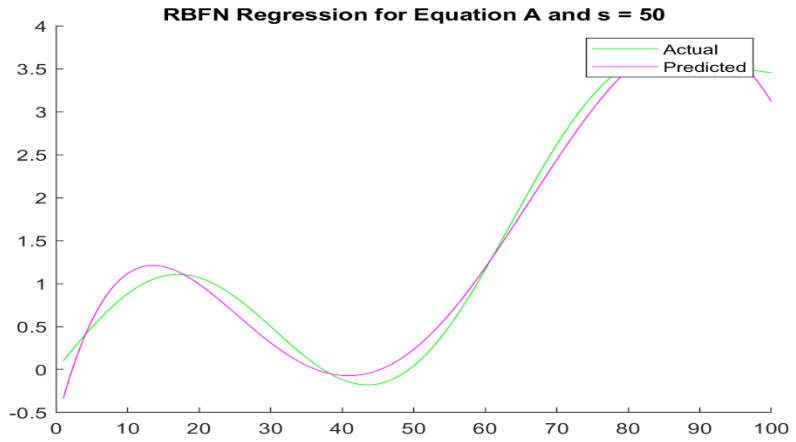
Εδώ παρατηρούμε εξακολουθεί να παραμένει σχεδόν όμοιο με την πραγματική συνάρτηση αλλά στις κορυφές από τις καμπύλες (σημεία τοπικών ακρότατων) παρατηρείται μία μικρή διαφοροποίηση



Σχήμα 4.41 Προσέγγιση Συνάρτησης Συνάρτησης y=sin(x/10) + $(x/50)^2$ για σ=20

Για σ=50

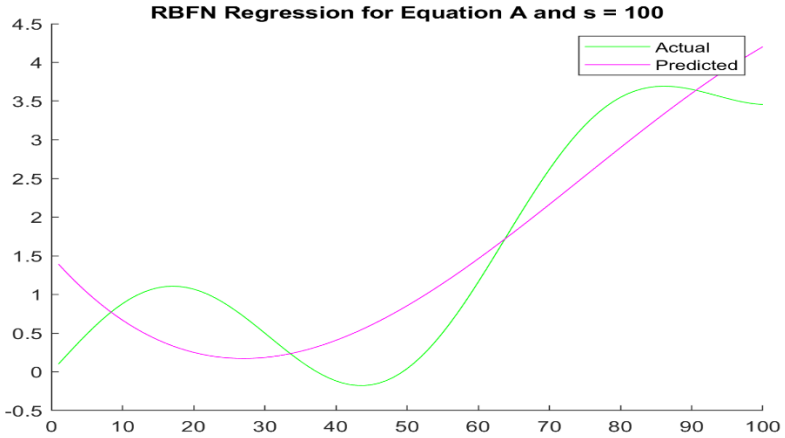
Εδώ βλέπουμε το σφάλμα να έχει αρχίσει να μεγαλώνει και η εκτιμώμενη καμπύλη να διαφοροποιείται αρκετά από την πραγματική αλλά να συνεχίζει να ακολουθεί το μοτίβο της



Σχήμα 4.42 Προσέγγιση Συνάρτησης Συνάρτησης y=sin(x/10) + $(x/50)^2$ για σ=50

Για σ=100

Εδώ βλέπουμε την εκτιμώμενη συνάρτηση να έχει τελείως απομακρυνθεί από την πραγματική που σημαίνει ότι για την συγκεκριμένη τιμή της παραμέτρου το RBF δεν κάνει καθόλου καλή δουλειά.

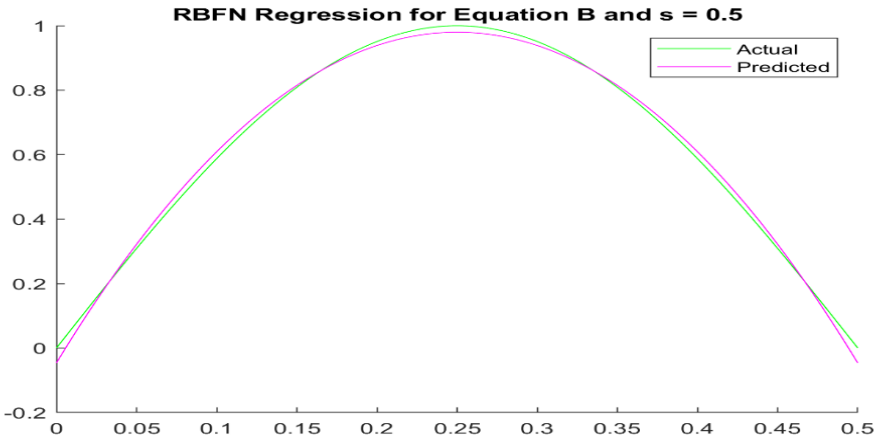


Σχήμα 4.43 Προσέγγιση Συνάρτησης Συνάρτησης y=sin(x/10) + $(x/50)^2$ σ=100

4.4.2. Συνάρτηση B

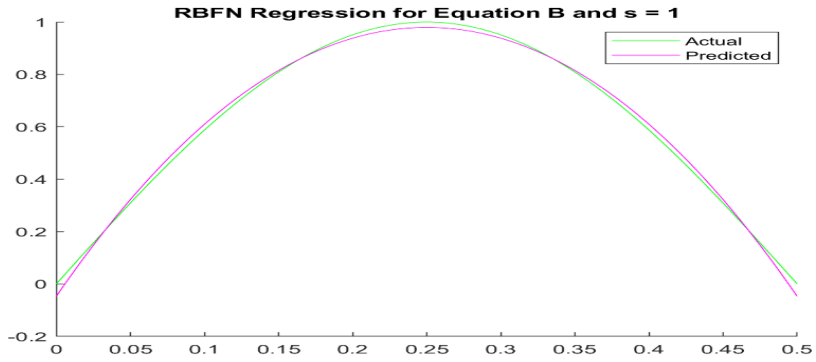
Εδώ παρατηρήσαμε ότι για τις πρώτες τρεις τιμές το RBF έβγαλε σχετικά καλά αποτελέσματα ενώ για τιμή ίση με 5 δεν κατάφερε να την προσεγγίσει καν. Αναλυτικά, τα παρακάτω διαγράμματα δείχνουν το πως το RBF προσέγγισε την πραγματική συνάρτηση για κάθε τιμή του σ

Για σ=0.5



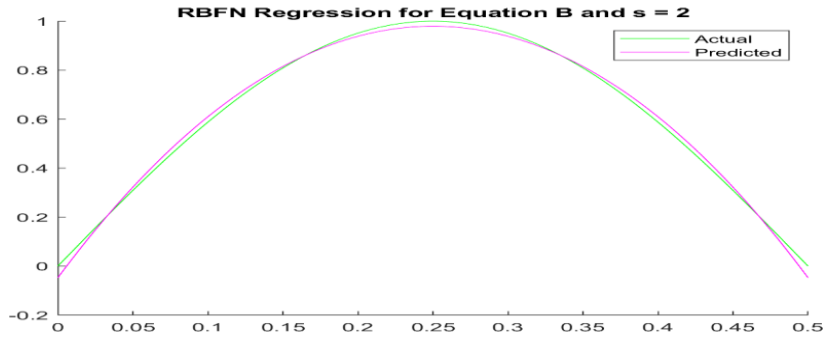
Σχήμα 4.44 Προσέγγιση Συνάρτησης y=sin(2πx) για σ=0.5

Για σ=1



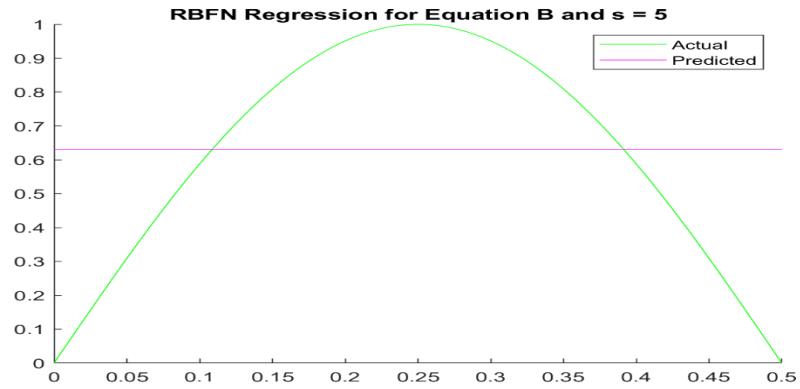
Σχήμα 4.45 Προσέγγιση Συνάρτησης y=sin(2πx) για σ=1

Για σ=2



Σχήμα 4.46 Προσέγγιση Συνάρτησης y=sin(2πx) για σ=2

Για σ=5

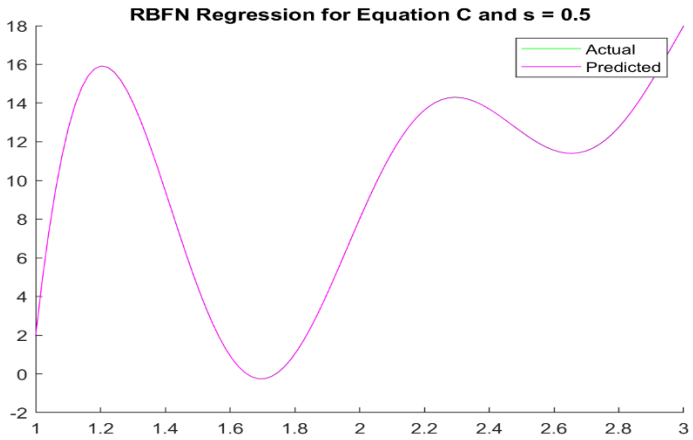


Σχήμα 4.47 Προσέγγιση Συνάρτησης y=sin(2πx) για σ=5

4.4.3. Συνάρτηση Γ

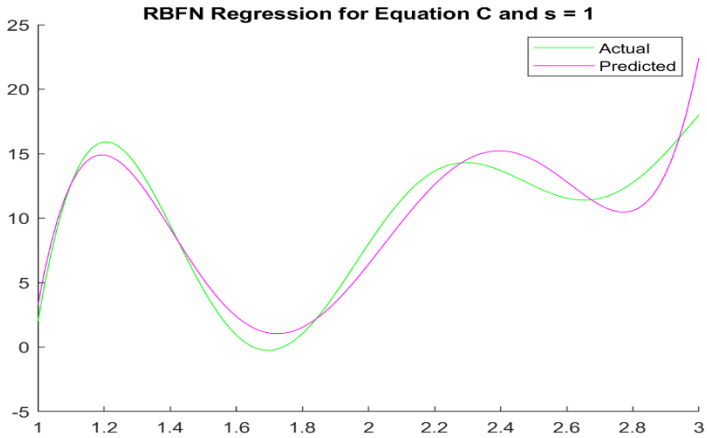
Εδώ, παρατηρούμε ότι στην πρώτη περίπτωση έγινε πολύ καλή προσέγγιση της πραγματικής συνάρτησης ενώ για τιμές του σ μεγαλύτερες της μονάδας το RBF απέτυχε ουσιαστικά να προσεγγίσει έστω την μορφή της καμπύλης της συνάρτησης

Για σ=0.5



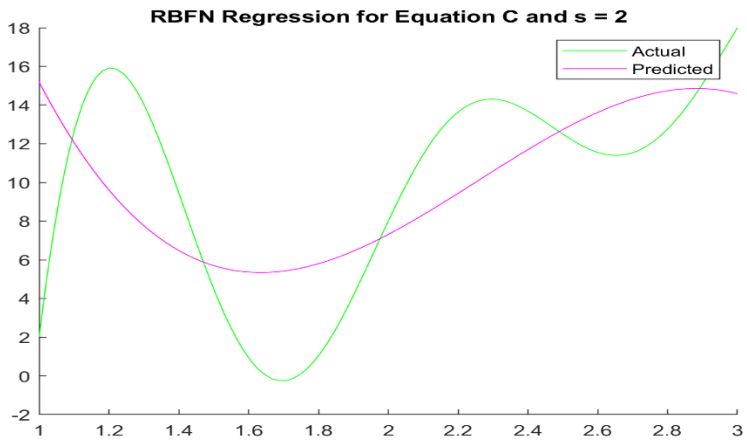
Σχήμα 4.48 Προσέγγιση Συνάρτησης $y=2x^2+e^{(\pi/x)}*\sin(2\pi x)$ για σ=0.5

Για σ=1



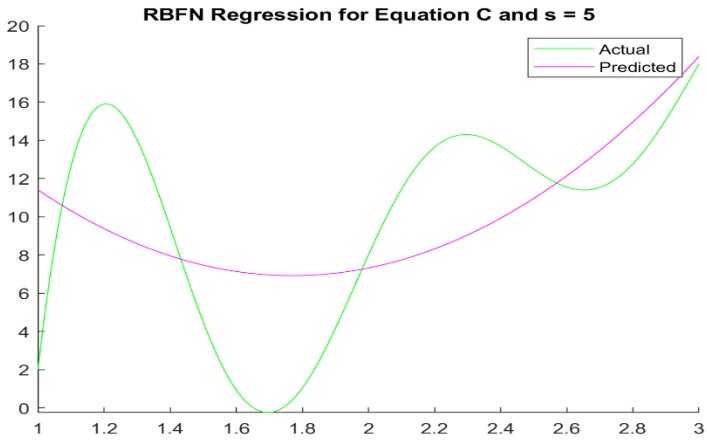
Σχήμα 4.49 Προσέγγιση Συνάρτησης $y=2x^2+e^{(\pi/x)}*\sin(2\pi x)$ για σ=1

Για σ=2



Σχήμα 4.50 Προσέγγιση Συνάρτησης $y=2x^2+e^{(\pi/x)}*\sin(2\pi x)$ για σ=2

Για σ=5

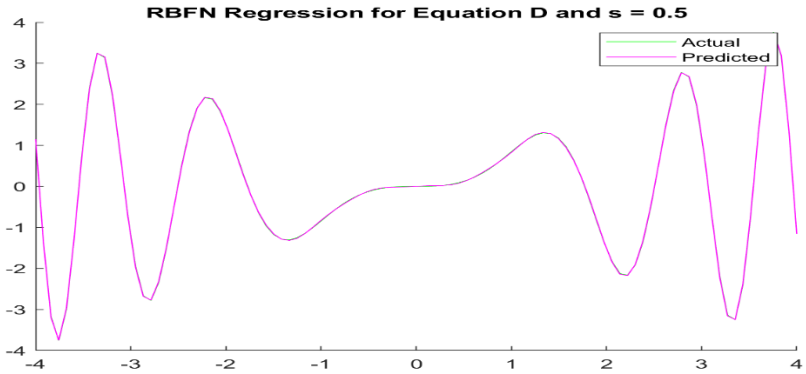


Σχήμα 4.51 Προσέγγιση Συνάρτησης $y=2x^2+e^{(\pi/x)}*\sin(2\pi x)$ για σ=5

4.4.4. Συνάρτηση Δ

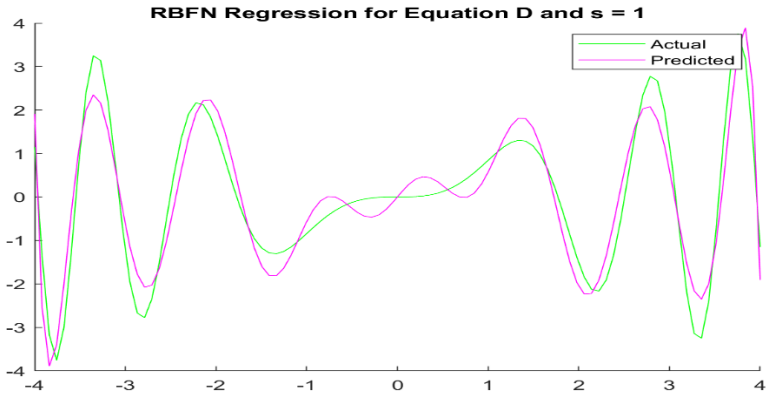
Εδώ παρατηρούμε ότι ισχύουν ακριβώς τα ίδια όπως και για τη συνάρτηση Γ παρόλο που εδώ η συνάρτηση ήταν πιο πολύπλοκη

Για σ=0.5



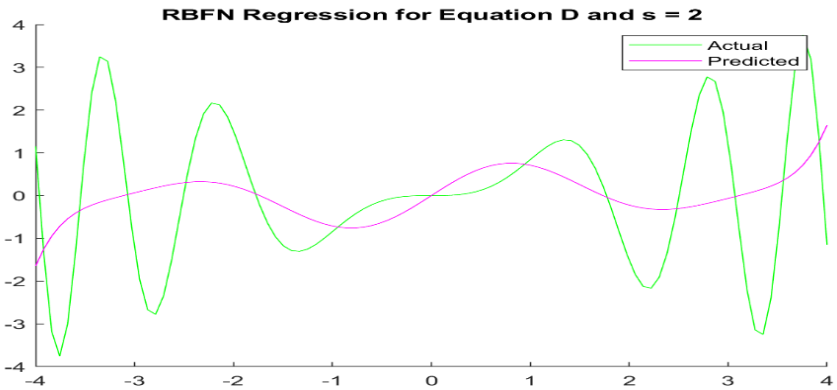
Σχήμα 4.52 Προσέγγιση Συνάρτησης y=x*sin(x^2) για σ=0.5

Για σ=1



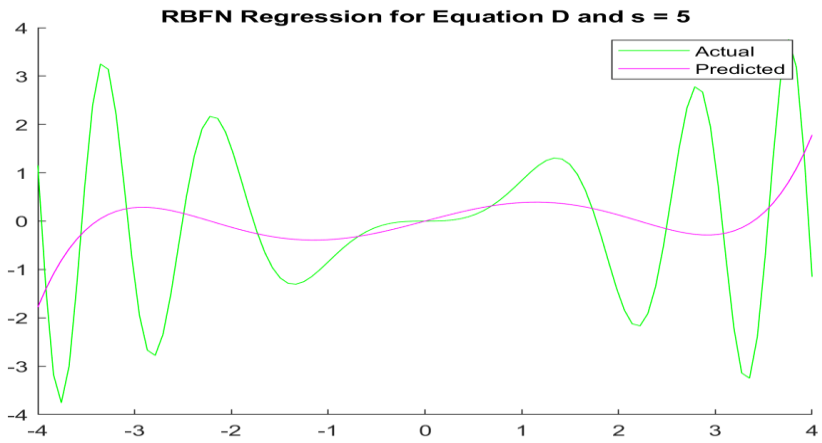
Σχήμα 4.53 Προσέγγιση Συνάρτησης y=x*sin(x^2) για σ=1

Για σ=2



Σχήμα 4.54 Προσέγγιση Συνάρτησης y=x*sin(x^2) για σ=2

Για σ=5



Σχήμα 4.55 Προσέγγιση Συνάρτησης $y=x*\sin(x^2)$ για σ=5

4.4.5. Παρουσίαση Μέσου Τετραγωνικού Σφάλματος για τις Παραπάνω Περιπτώσεις

Στον παρακάτω πίνακα παρατηρούμε το μέγεθος του σφάλματος για τις διάφορες τιμές της παραμέτρου που εξετάσαμε. Σε αυτό το σημείο αξίζει να σημειωθεί ότι στην πρώτη γραμμή μέσα σε παρενθέσεις βρίσκεται η τιμή του σ για την συνάρτηση Α

Πίνακας 4.4: Τιμές για Σφάλμα για Διάφορες Τιμές της Μεταβλητής σ

Παράμετρος σ (Α) /Συνάρτηση	0.5 (10)	1 (20)	2 (50)	5 (100)
Συνάρτηση Α	1.28^{E}-11	2.2^{E}-08	0.01	0.16
Συνάρτηση Β	1.51^{E}-05	1.59^{E}-04	1.6^{E}-03	4.89
Συνάρτηση Γ	4.5^{E}-05	0.88	7.52	8.68
Συνάρτηση Δ	6.55^{E}-05	0.14	1.17	12.01

4.5. Σύγκριση του Υβριδικού Δικτύου με το Εκτεταμένο MLP

Όπως αναφέρθηκε και στην αρχή της παρούσας διατριβής, σκοπός αυτής είναι να παρουσιαστεί μια εναλλακτική υβριδική μέθοδος προσέγγισης συνάρτησης χρησιμοποιώντας, αντί ενός εκτεταμένου MLP δικτύου, ένα πιο απλό MLP σε σειρά με ένα RBF. Πρακτικά για εμάς αυτό σημαίνει ότι η έξοδος του απλού MLP θα γίνεται είσοδος στο RBF. Για την ακρίβεια όχι η συνάρτηση που προσέγγισε το

απλό MLP αλλά το απόλυτο σφάλμα σημείο προς σημείο. Έτσι, σκοπός του RBF είναι να προσεγγίσει το σφάλμα καλώς, ούτως ώστε προσθέτοντας την τιμή του RBF στην τιμή που προσέγγισε το απλό MLP να πάρουμε παραπλήσιο ή και μικρότερο τελικό σφάλμα για το υβριδικό δίκτυο σε σχέση με το σφάλμα του εκτεταμένου MLP. Αυτό θα σημαίνει τελικώς, πως το υβριδικό σύστημα που σχεδιάσαμε είναι ισάξιο σε αποτέλεσμα με το εκτεταμένο MLP αλλά ταυτόχρονα είναι πιο απλό.

Για να πραγματοποιήσουμε αυτή τη σύγκριση θα χρησιμοποιήσουμε το μοντέλο μας σε κάθε μία από τις 4 συναρτήσεις που προηγήθηκαν σε ανάλυση. Για να πραγματοποιηθεί αυτό δημιουργήθηκε μια κεντρική συνάρτηση (main) στην οποία καθορίζονται οι παράμετροι ανάλογα με τη συνάρτηση που θέλουμε να προσεγγίσουμε και τις ρυθμίσεις του MLP και του RBF όπως αναφέραμε σε προηγούμενα υποκεφάλαια του παρόντος κεφαλαίου. Σε αυτό το σημείο, θα πρέπει να τονιστεί πως δε δημιουργήθηκαν δύο διαφορετικές συναρτήσεις για τις δύο διαφορετικές περιπτώσεις του MLP, αλλά δημιουργήθηκε μία συνάρτηση η οποία καλείται με διαφορετικές παραμέτρους. Οι παράμετροι που αποφασίσαμε να τρέξουμε τα πειράματα για να συγκρίνουμε το υβριδικό με το εκτεταμένο MLP είναι οι εξής:

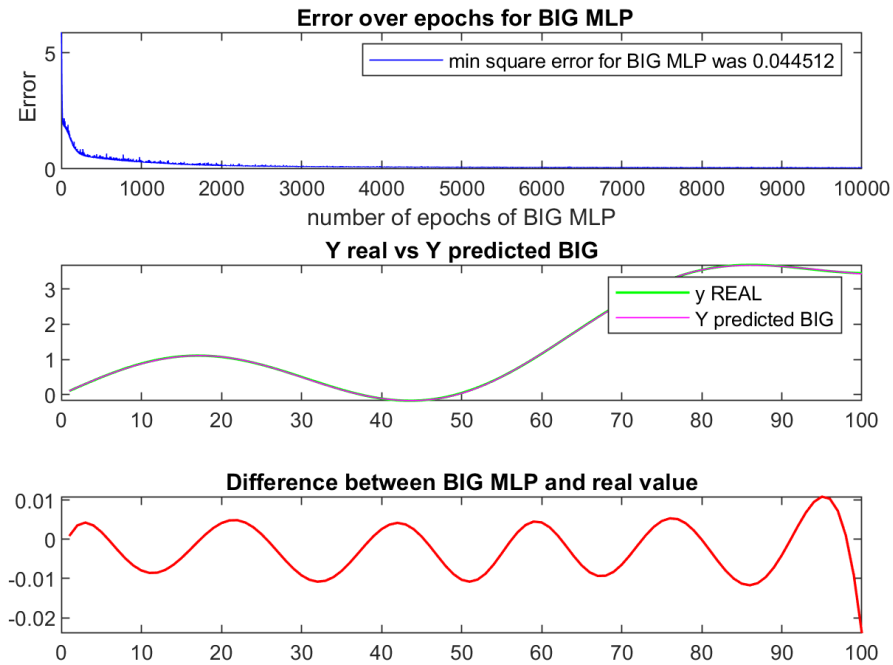
- Για το εκτεταμένο MLP επιλέξαμε τις καλύτερες των περιπτώσεων από τις προηγούμενες αναλύσεις:
  - Αριθμός νευρώνων ίσος με 64
  - Αριθμός εποχών ίσος με 10,000
  - Ρυθμός εκμάθησης εισόδου και εξόδου ίσος με 0.1 και 0.01 αντίστοιχα
  - Αρχικοποίηση βαρών τυχαία
- Για το απλό MLP επιλέξαμε μία μέτρια περίπτωση από τις προηγούμενες αναλύσεις:
  - Αριθμός νευρώνων ίσος με 24
  - Αριθμός εποχών ίσος με 5,000
  - Αριθμός εκμάθησης εισόδου και εξόδου ίσος με 0.1 και 0.01 αντίστοιχα διότι κανένα άλλο ζευγάρι δεν έδινε αρκετή σύγκλιση όπως παρουσιάστηκε παραπάνω
  - Αρχικοποίηση βαρών τυχαία
- Για το RBF δίκτυο οι ρυθμίσεις που επιλέξαμε στις παραμέτρους ήταν

- Αριθμός νευρώνων ίσος με 60
- Μεταβλητή σ ίση με 10 για την συνάρτηση A, 0.1 για την συνάρτηση B και 0.2 για τις υπόλοιπες

Συνολικά τα αποτελέσματα που πήραμε ακολουθούν αμέσως παρακάτω.

4.5.1. Συνάρτηση A

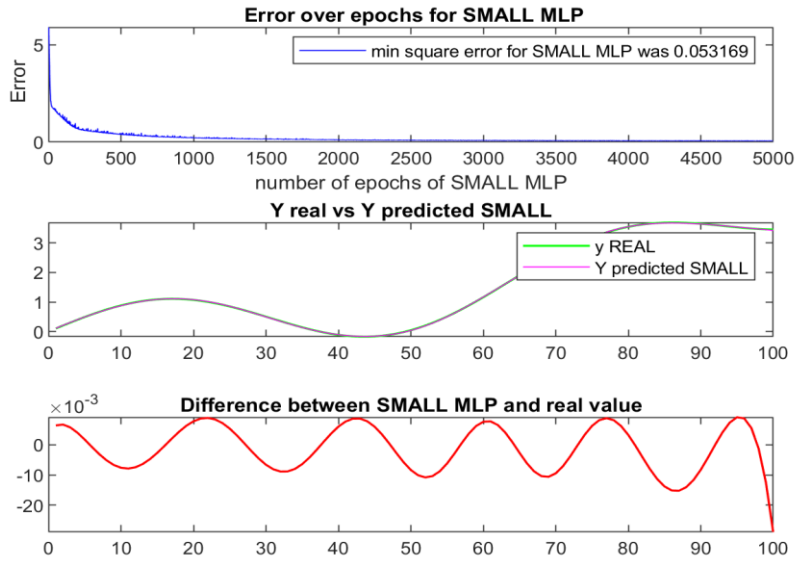
Για το εκτεταμένο MLP



Σχήμα 4.56 Προσέγγιση Συνάρτησης $y=\sin(x/10) + (x/50)^2$ για Εκτεταμένο MLP

Εδώ παρατηρούμε ότι, όπως και πριν το εκτεταμένο MLP δίνει πολύ καλή προσέγγιση. Στο παραπάνω σχήμα βλέπουμε πέραν του σφάλματος και της προσέγγισης της συνάρτησης και μία τρίτη γραφική παράσταση με κόκκινο χρώμα. Αυτή η παράσταση είναι για την πραγματική διαφορά κάθε σημείου που προσέγγισε το δίκτυο σε σχέση με την πραγματική τιμή της συνάρτησης.

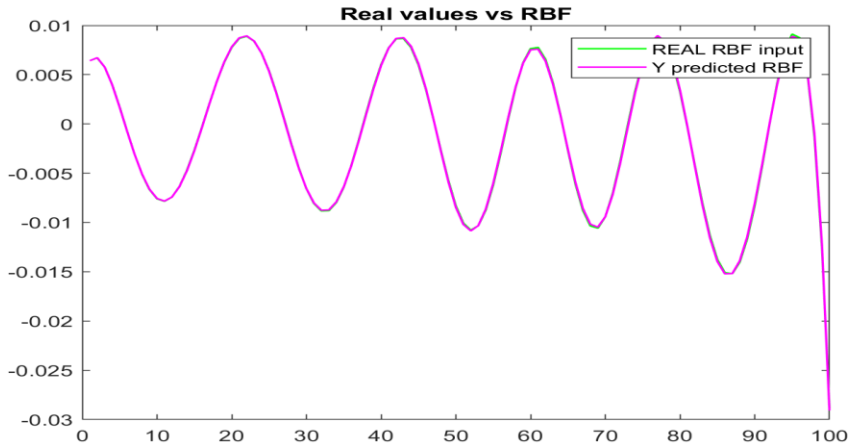
Για το απλό MLP ως πρώτο μέρος του υβριδικού



Σχήμα 4.57 Προσέγγιση Συνάρτησης $y=\sin(x/10) + (x/50)^2$ για Απλό MLP

Εδώ βλέπουμε ακριβώς τις ίδιες γραφικές παραστάσεις για το απλό MLP που είναι και το πρώτο τμήμα του υβριδικού δικτύου. Παρατηρούμε ότι η κόκκινη γραφική παράσταση που δείχνει τις τιμές απόλυτης διαφοράς έχει μεγαλύτερες διακυμάνσεις. Τέλος, αξίζει να επισημανθεί ότι αυτή η διαφορά είναι και η είσοδος για το RBF δίκτυο που βρίσκεται σε σειρά με το απλό MLP ως μέρος του υβριδικού δικτύου.

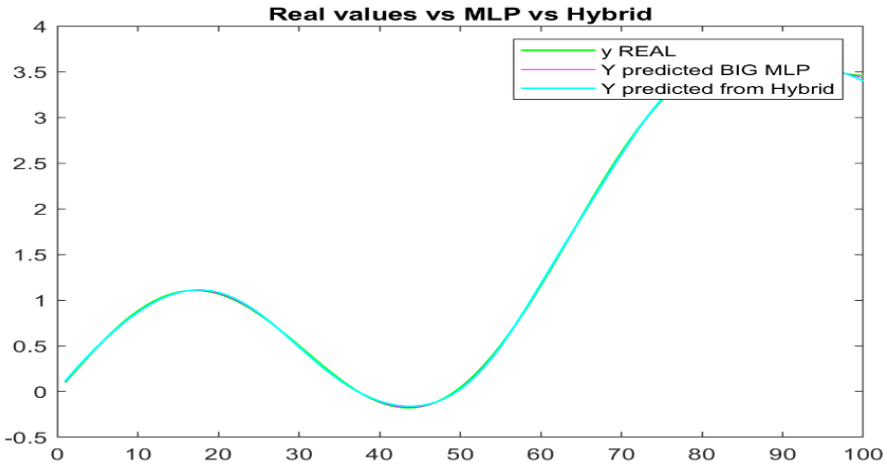
Για το RBF ως δεύτερο μέρος του υβριδικού



Σχήμα 4.58 Προσέγγιση Συνάρτησης $y=\sin(x/10) + (x/50)^2$ με RBF

Βλέπουμε εδώ ότι πρακτικά το RBF δίκτυο ως μέρος του υβριδικού δικτύου προσεγγίζει την τιμή της εισόδου του (που ήταν το απόλυτο σφάλμα του απλού MLP) με μεγάλη ακρίβεια και ελάχιστο σφάλμα κάτι που υποδεικνύει την καλή λειτουργία του υβριδικού δικτύου.

Σύγκριση μεταξύ πραγματικού, εκτεταμένου MLP και υβριδικού μοντέλου



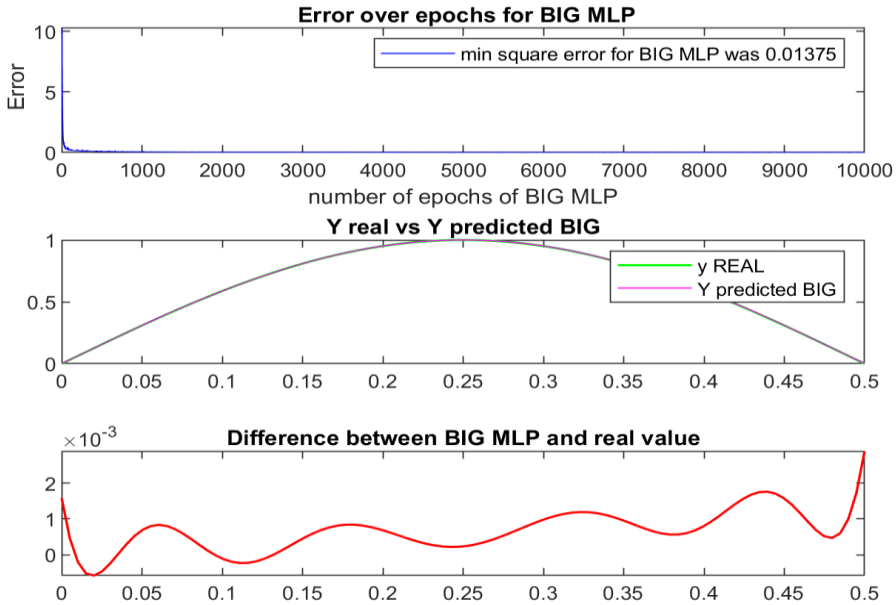
Σχήμα 4.59 Σύγκριση Μεταξύ Πραγματικού, Εκτεταμένου MLP και Υβριδικού Μοντέλου

Σε αυτό το γράφημα βλέπουμε τρεις συναρτήσεις. Την πραγματική, την προσέγγιση του εκτεταμένου MLP και την προσέγγιση του υβριδικού δικτύου. Βλέπουμε λοιπόν ότι και οι τρεις είναι πάρα πολύ κοντά μεταξύ τους.

4.5.2. Συνάρτηση B

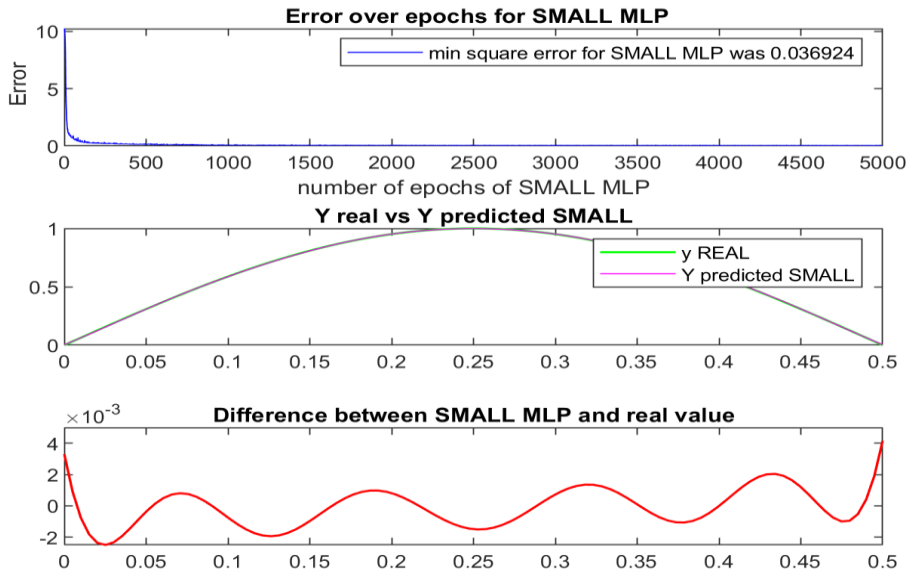
Παρόμοια αποτελέσματα βρήκαμε και για αυτή τη συνάρτηση σε σύγκριση του εκτεταμένου MLP με το υβριδικό δίκτυο. Και εδώ παρατηρήσαμε ότι τα αποτελέσματα ήταν πάρα πολύ κοντά.

Για το εκτεταμένο MLP



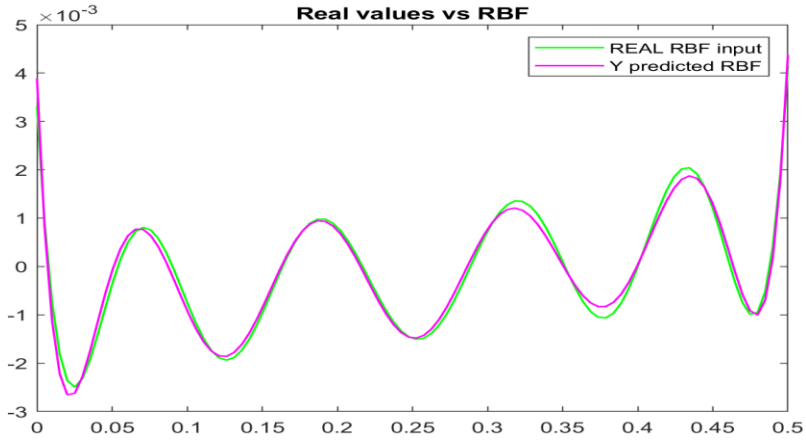
Σχήμα 4.60 Προσέγγιση Συνάρτησης y=sin(2πx) για Εκτεταμένο MLP

Για το απλό MLP ως πρώτο μέρος του υβριδικού



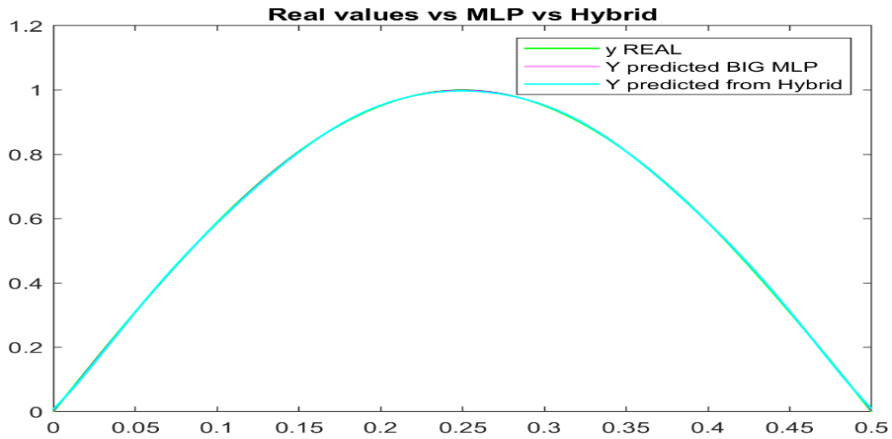
Σχήμα 4.61 Προσέγγιση Συνάρτησης y=sin(2πx) για Απλό MLP

Για το RBF ως δεύτερο μέρος του υβριδικού



Σχήμα 4.62 Προσέγγιση Συνάρτησης y=sin(2πx) για RBF

Σύγκριση μεταξύ πραγματικού, εκτεταμένου MLP και υβριδικού μοντέλου

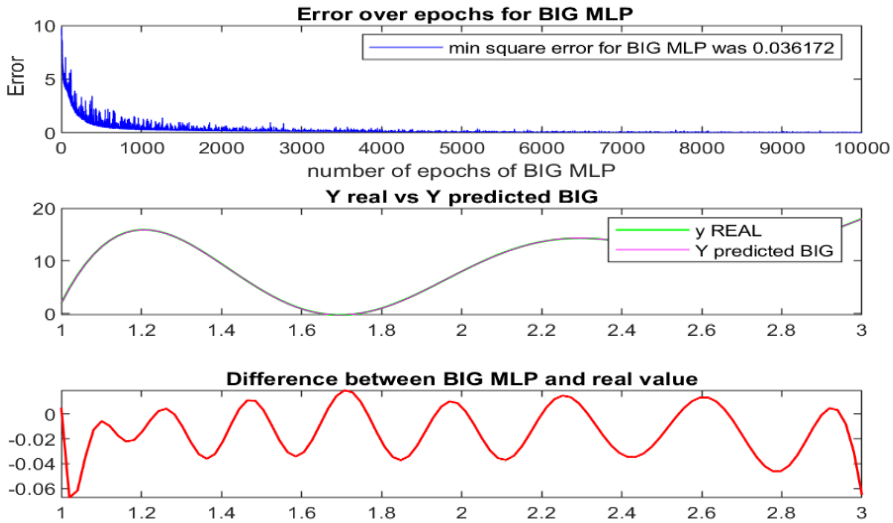


Σχήμα 4.63 Σύγκριση Μεταξύ Πραγματικού, Εκτεταμένου MLP και Υβριδικού Μοντέλου

4.5.3. Συνάρτηση Γ

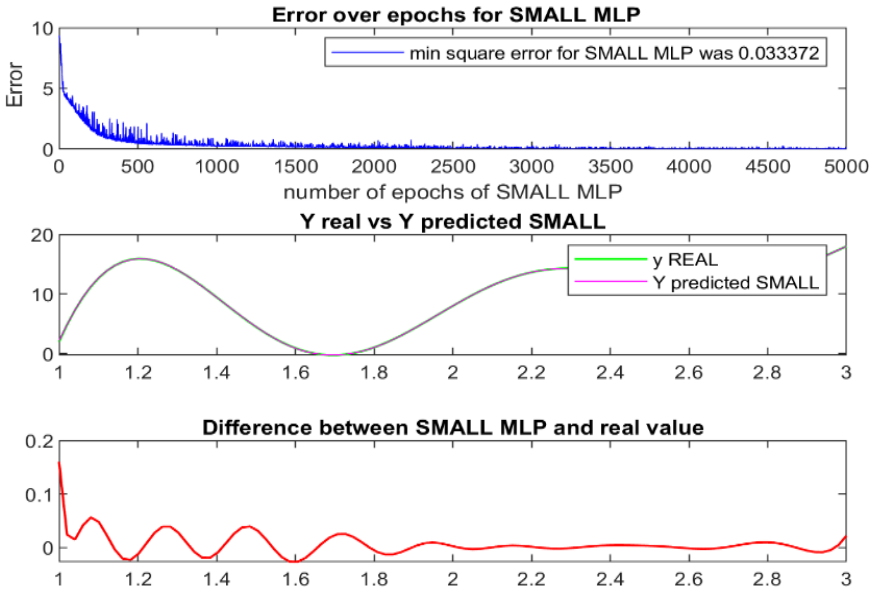
Παρόμοια αποτελέσματα σε σχέση με τις άλλες δύο συναρτήσεις βλέπουμε και σε αυτή τη συνάρτηση που δείχνει την γενικευτική ικανότητα του υβριδικού δικτύου ανεξαρτήτως συνάρτησης προς προσέγγισης. Πιο αναλυτικά, τα αποτελέσματα που πήραμε ήταν τα ακόλουθα:

Για το εκτεταμένο MLP



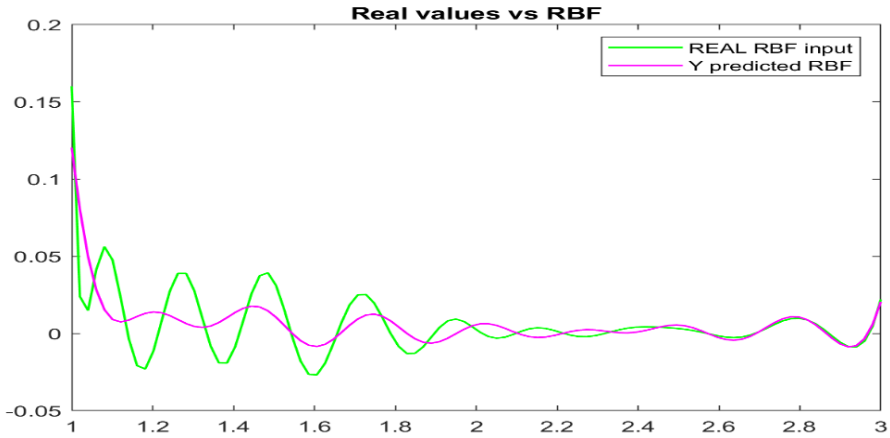
Σχήμα 4.64 Προσέγγιση Συνάρτησης $y=2x^2+e^{(\pi/x)}*\sin(2\pi x)$για Εκτεταμένο MLP

Για το απλό MLP ως πρώτο μέρος του υβριδικού



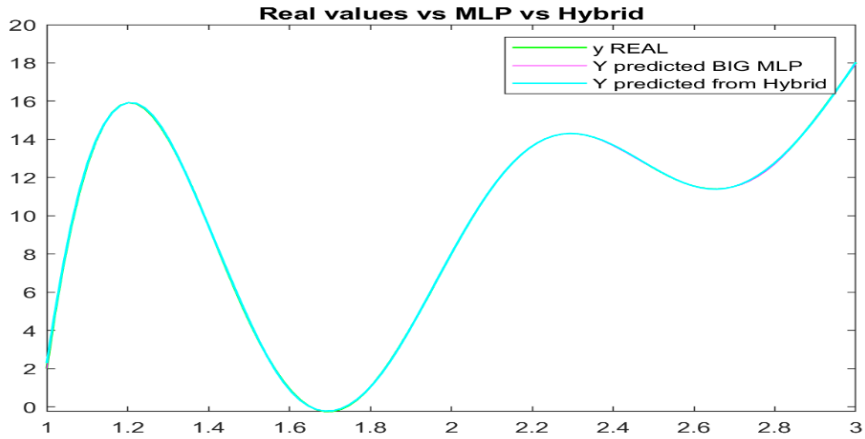
Σχήμα 4.65 Προσέγγιση Συνάρτησης $y=2x^2+e^{(\pi/x)}*\sin(2\pi x)$για Απλό MLP

Για το RBF ως δεύτερο μέρος του υβριδικού



Σχήμα 4.66 Προσέγγιση Συνάρτησης $y=2x^2+e^{(\pi/x)}*\sin(2\pi x)$για RBF

Σύγκριση μεταξύ πραγματικού, εκτεταμένου MLP και υβριδικού μοντέλου

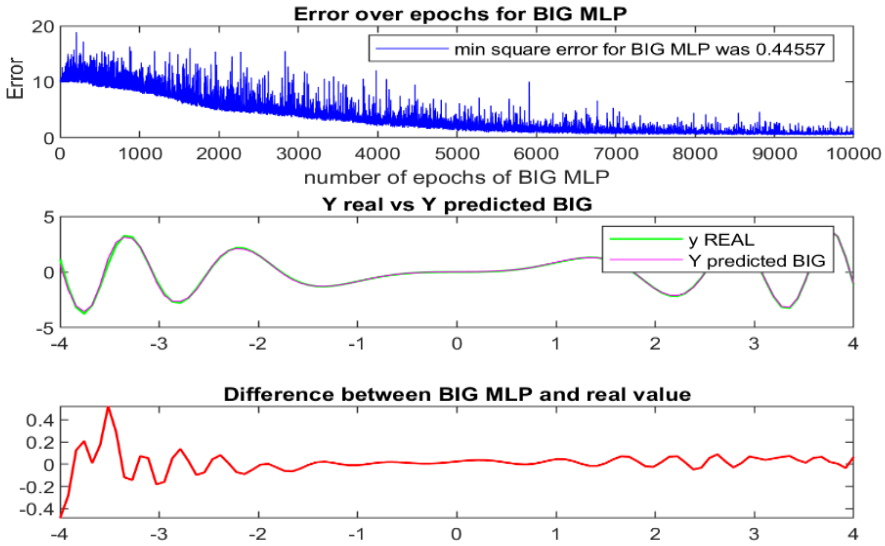


Σχήμα 4.67 Προσέγγιση Σύγκριση Μεταξύ Πραγματικού, Εκτεταμένου MLP και Υβριδικού μοντέλου

4.5.4. Συνάρτηση Δ

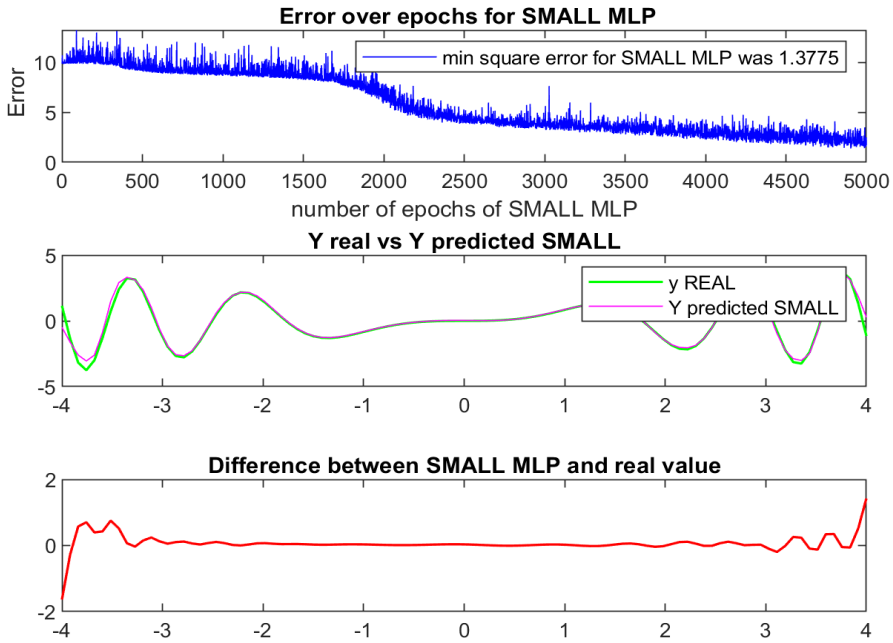
Στη συνάρτηση αυτή, που είναι και η πιο δύσκολη από αυτές που εξετάστηκαν, παρατηρούμε ότι το υβριδικό δίκτυο συμπεριφέρεται πάρα πολύ καλά αλλά όχι άριστα όπως στις προηγούμενες συναρτήσεις. Πιο αναλυτικά τα αποτελέσματα που πήραμε ήταν τα ακόλουθα:

Για το εκτεταμένο MLP



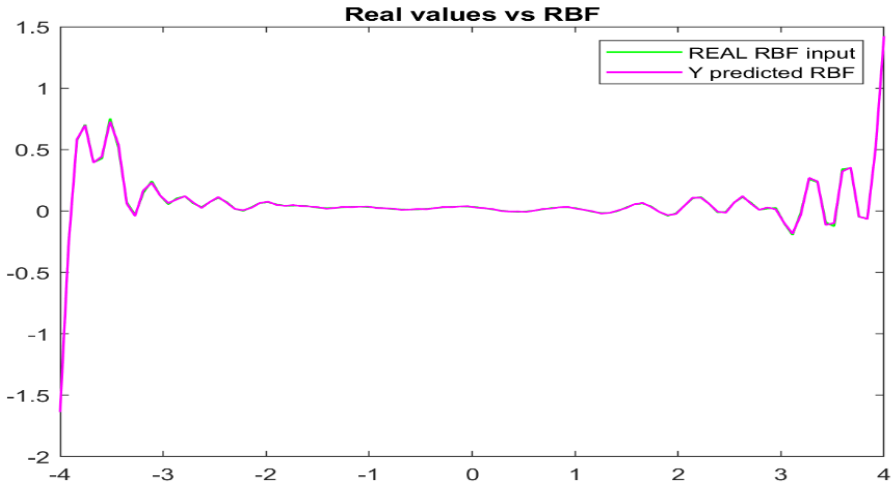
Σχήμα 4.68 Προσέγγιση Συνάρτησης $y=x*\sin(x^2)$ για Εκτεταμένο MLP

Για το απλό MLP ως πρώτο μέρος του υβριδικού



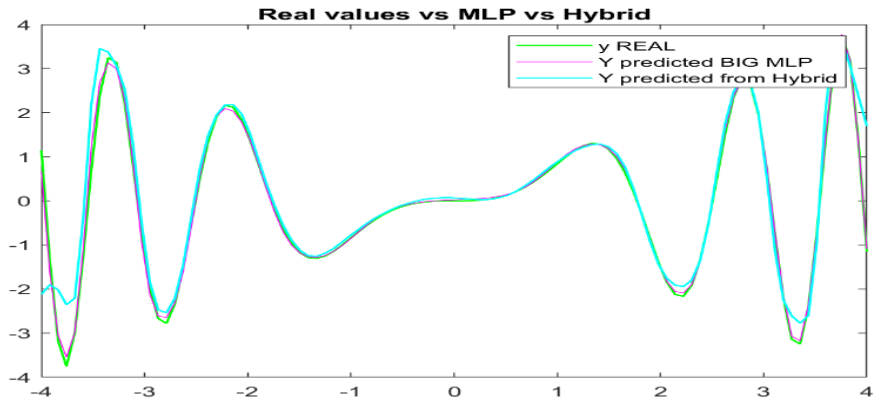
Σχήμα 4.69 Προσέγγιση Συνάρτησης y=x*sin(x2) για Απλό MLP

Για το RBF ως δεύτερο μέρος του υβριδικού



Σχήμα 4.70 Προσέγγιση Συνάρτησης $y=x*\sin(x^2)$ για RBF

Σύγκριση μεταξύ πραγματικού, εκτεταμένου MLP και υβριδικού μοντέλου



Σχήμα 4.71 Σύγκριση Μεταξύ Πραγματικού, Εκτεταμένου MLP και Υβριδικού Μοντέλου

4.5.5. Σχολιασμός Αποτελεσμάτων

Συνολικά παρατηρήσαμε ότι σε όλες τις περιπτώσεις, το υβριδικό δίκτυο απέδωσε μικρότερο σφάλμα από το εκτεταμένο MLP δείχνοντας επίσης άρτια γενικευτική ικανότητα. Αυτό δείχνει ότι η μέθοδος αυτή λειτουργεί ανεξαρτήτως συνάρτησης που επιλέξαμε. Το σφάλμα σε κάθε συνάρτηση διαμορφώθηκε όπως δείχνει ο πίνακας :

Πίνακας 4.5: Συγκεντρωτικός Πίνακας Σφαλμάτων για νευρώνες RBF = 60

Σφάλμα	ΕΚΤΕΤΑΜΕΝΟ MLP	Απλό MLP	RBF	ΥΒΡΙΔΙΚΟ
Συνάρτηση Α	2.70^{E}-02	0.047	5.68^{E}-09	1.8^{E}-04
Συνάρτηση Β	1.50^{E}-02	0.036	1.31^{E}-08	6.96^{E}-06
Συνάρτηση Γ	4.20^{E}-02	0.045	0.0001	0.012
Συνάρτηση Δ	0.463	1.721	1.14^{E}-05	0.353

Φυσικά, ανάλογα τη συνάρτηση θα πρέπει να προσαρμόζονται και οι παράμετροι των δικτύων αλλά σε κάθε περίπτωση το υβριδικό μοντέλο είναι σε θέση να προβλέψει αρκετά καλά το μοντέλο που θα του δοθεί

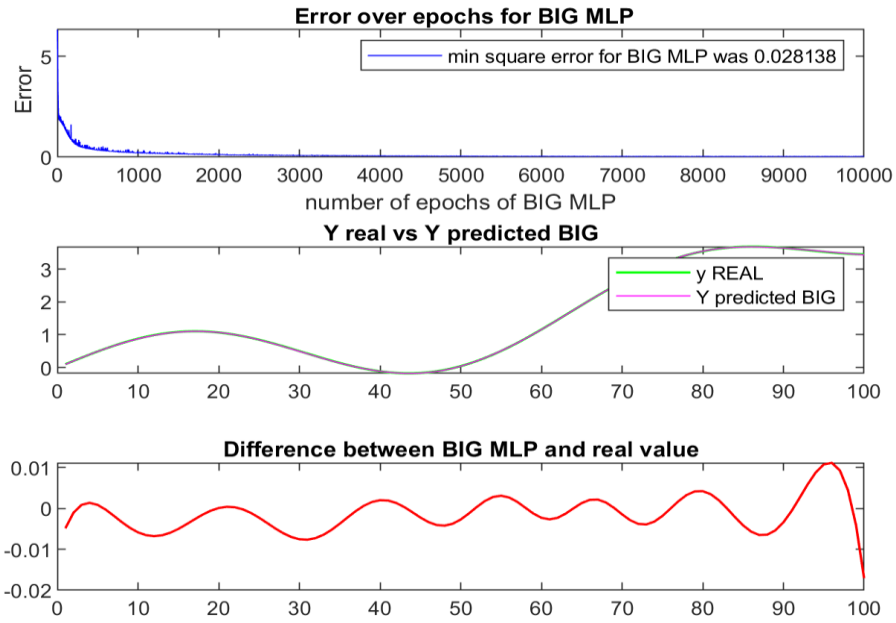
4.6. Προσομοίωση του Υβριδικού Δικτύου για αριθμό νευρώνων RBF = 30

Αρχικά, για να τεστάρουμε το υβριδικό δίκτυο ως προς το πολύπλοκο MLP, πραγματοποιήσαμε για τις διάφορες συναρτήσεις, προσομοιώσεις στις οποίες ο αριθμός νευρώνων για το RBF ήταν υψηλός και ίσος με 60. Όμως, ο σκοπός μας είναι μετά από την επιτυχή προσομοίωση του υβριδικού δικτύου στο προηγούμενο υπόκεφάλαιο, να το προσομοιώσουμε με μικρότερο αριθμό νευρώνων. Έτσι, σε αυτό το υποκεφάλαιο παρουσιάζουμε τα αποτελέσματα για αριθμό νευρώνων του RBF ίσο με 30.

4.6.1. Συνάρτηση Α

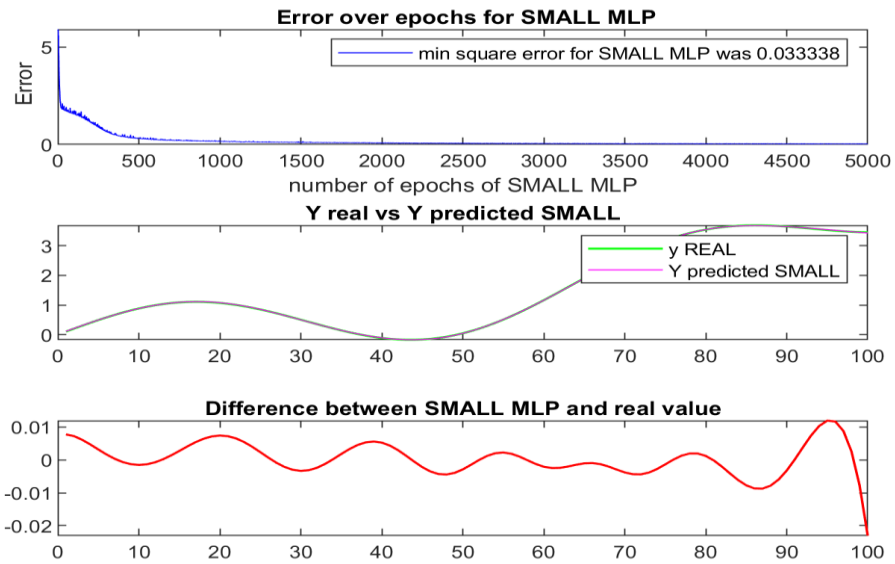
Για τη συνάρτηση αυτή δεν παρατηρήσαμε μεγάλες μεταβολές στις γραφικές παραστάσεις σε σχέση με πριν.

Για το εκτεταμένο MLP



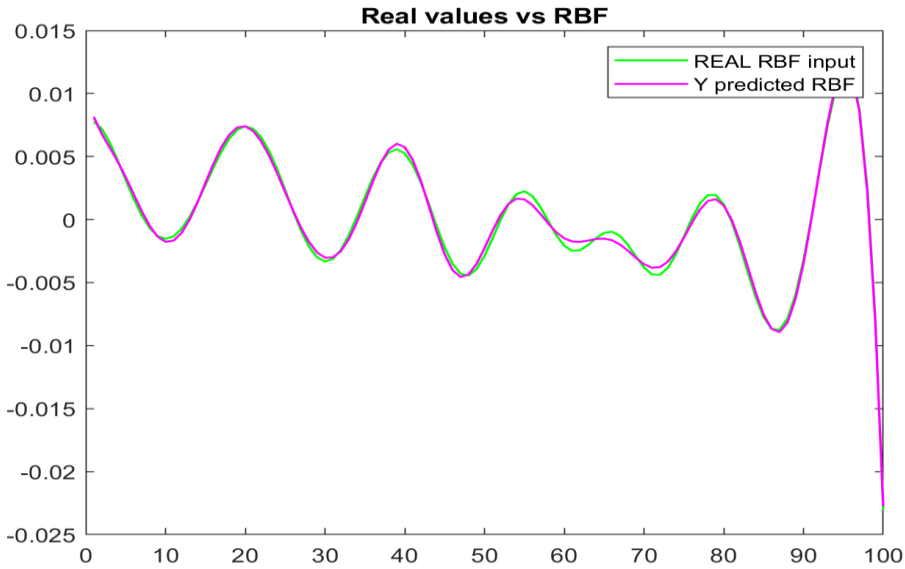
Σχήμα 4.72 Προσέγγιση Συνάρτησης y=sin(x/10) + $(x/50)^2$ με Χρήση Συνόλου Ελέγχου για Εκτεταμένο MLP

Για το απλό MLP ως πρώτο μέρος του υβριδικού



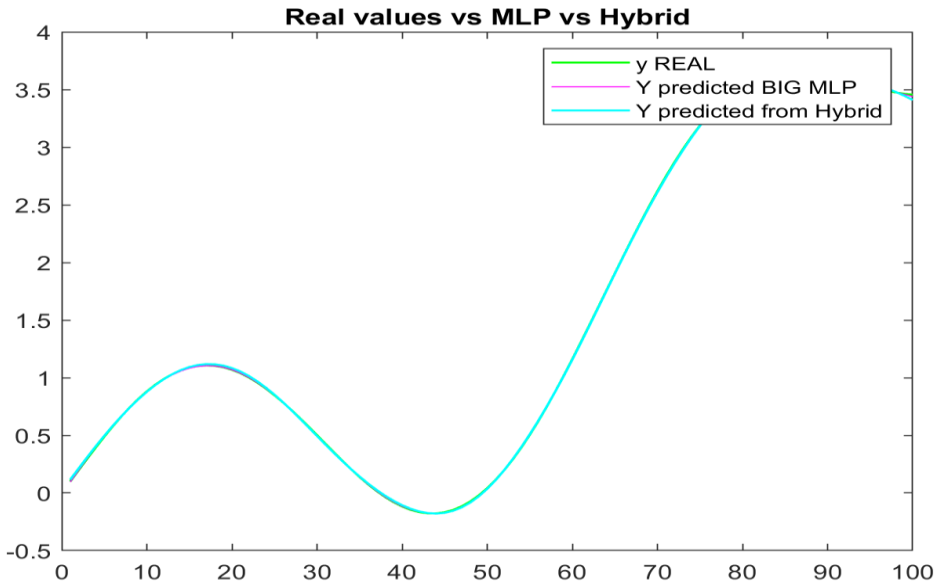
Σχήμα 4.73 Προσέγγιση Συνάρτησης y=sin(x/10) + $(x/50)^2$ με Χρήση Συνόλου Ελέγχου για Απλό MLP

Για το RBF ως δεύτερο μέρος του υβριδικού



Σχήμα 4.74 Προσέγγιση Συνάρτησης Σφάλματος της $y=\sin(x/10) + (x/50)^2$ με Χρήση Συνόλου Ελέγχου για RBF

Σύγκριση μεταξύ πραγματικού, εκτεταμένου MLP και υβριδικού μοντέλου

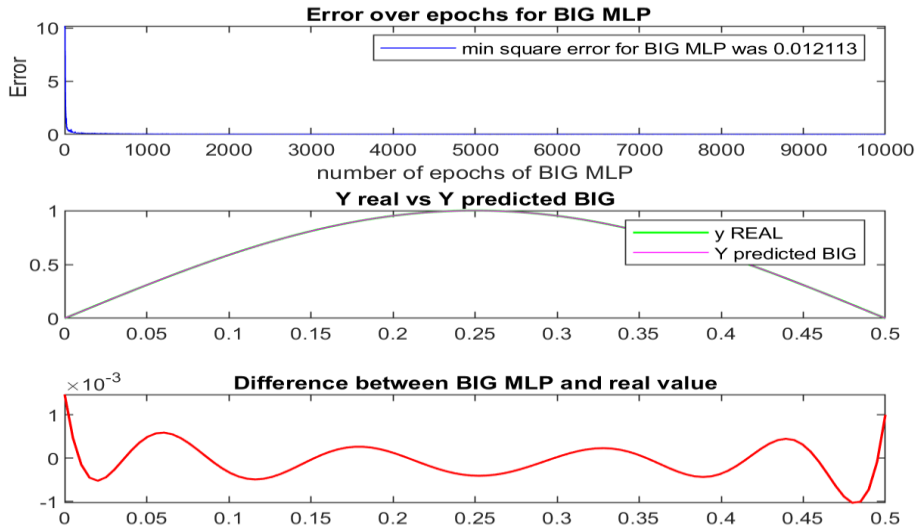


Σχήμα 4.75 Σύγκριση Μεταξύ Πραγματικού, Εκτεταμένου MLP και Υβριδικού Μοντέλου με Χρήση Συνόλου Ελέγχου της Συνάρτησης Α

4.6.2. Συνάρτηση Β

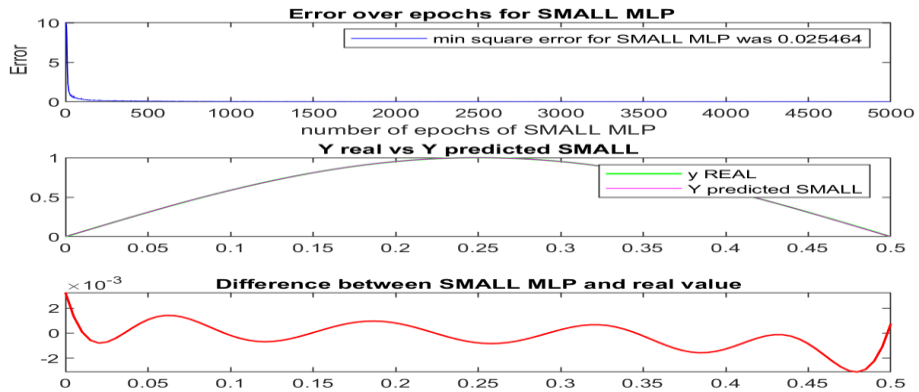
Παρόμοια αποτελέσματα βρήκαμε και για αυτή τη συνάρτηση σε σύγκριση με την προηγούμενη περίπτωση μόνο που εδώ είδαμε το RBF να δυσκολεύεται κάπως να προσεγγίσει τέλεια τη συνάρτηση.

Για το εκτεταμένο MLP



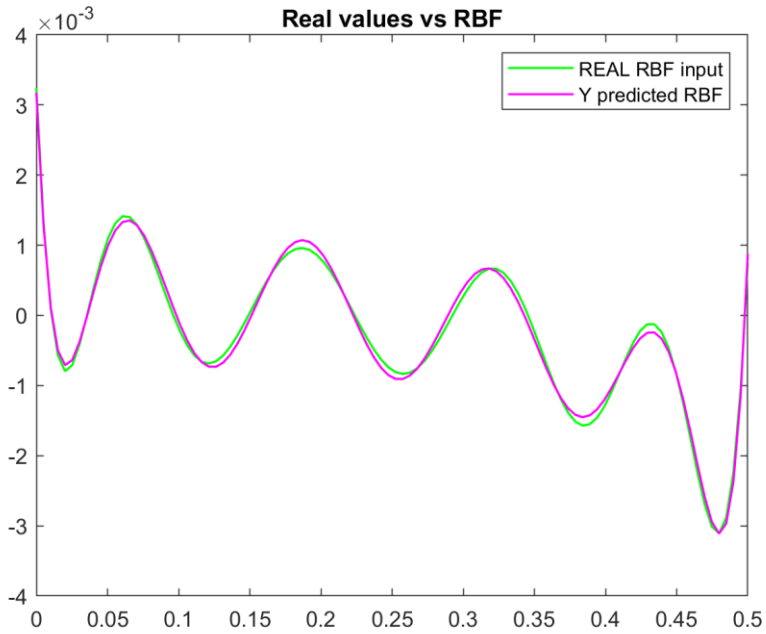
Σχήμα 4.76 Προσέγγιση Συνάρτησης y=sin(2πx) με Χρήση Συνόλου Ελέγχου για Εκτεταμένο MLP

Για το απλό MLP ως πρώτο μέρος του υβριδικού



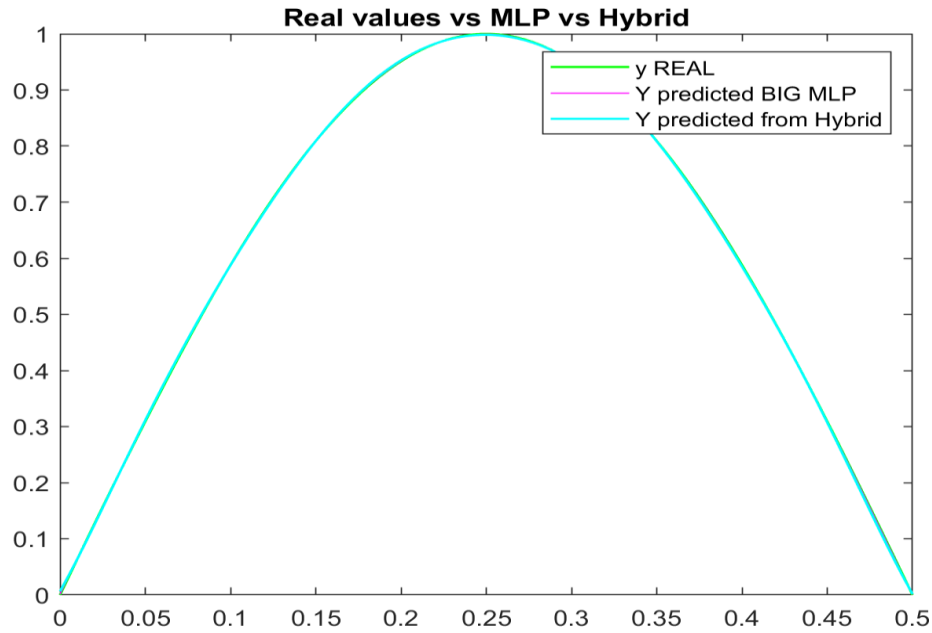
Σχήμα 4.77 Προσέγγιση Συνάρτησης y=sin(2πx) με Χρήση Συνόλου Ελέγχου για Απλό MLP

Για το RBF ως δεύτερο μέρος του υβριδικού



Σχήμα 4.78 Προσέγγιση Συνάρτησης Σφάλματος της y=sin(2πx) με Χρήση Συνόλου Ελέγχου RBF

Σύγκριση μεταξύ πραγματικού, εκτεταμένου MLP και υβριδικού μοντέλου

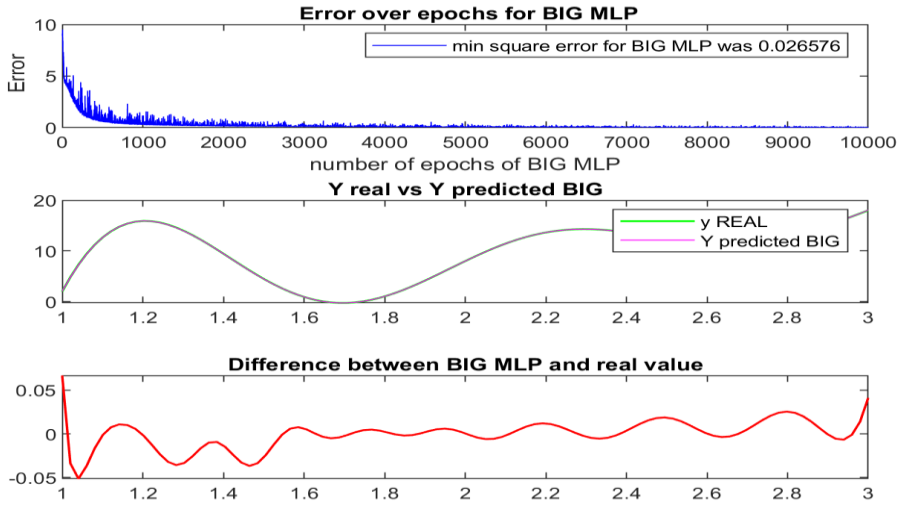


Σχήμα 4.79 Σύγκριση Μεταξύ Πραγματικού, Εκτεταμένου MLP και Υβριδικού Μοντέλου με Χρήση Συνόλου Ελέγχου της Συνάρτησης της B

4.6.3. Συνάρτηση Γ

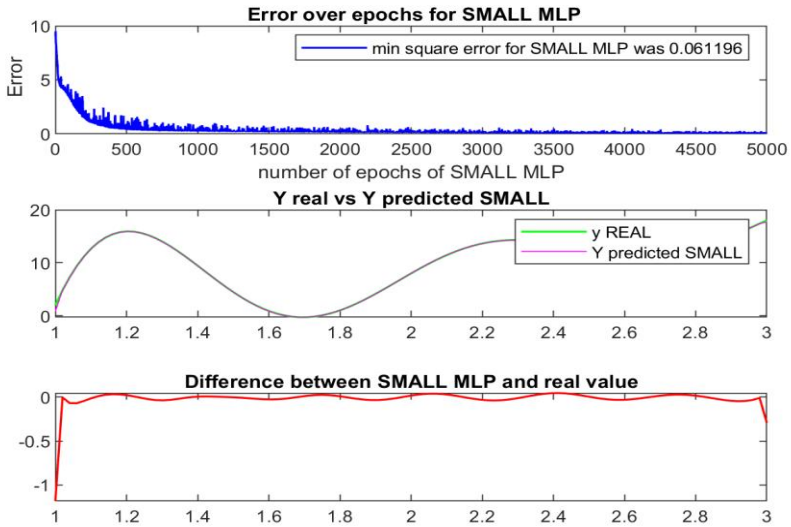
Παρόμοια αποτελέσματα σε σχέση με τις άλλες δύο συναρτήσεις βλέπουμε και σε αυτή τη συνάρτηση που δείχνει την γενικευτική ικανότητα του RBF να είναι μικρότερη σε σχέση με την προηγούμενη περίπτωση.

Για το εκτεταμένο MLP



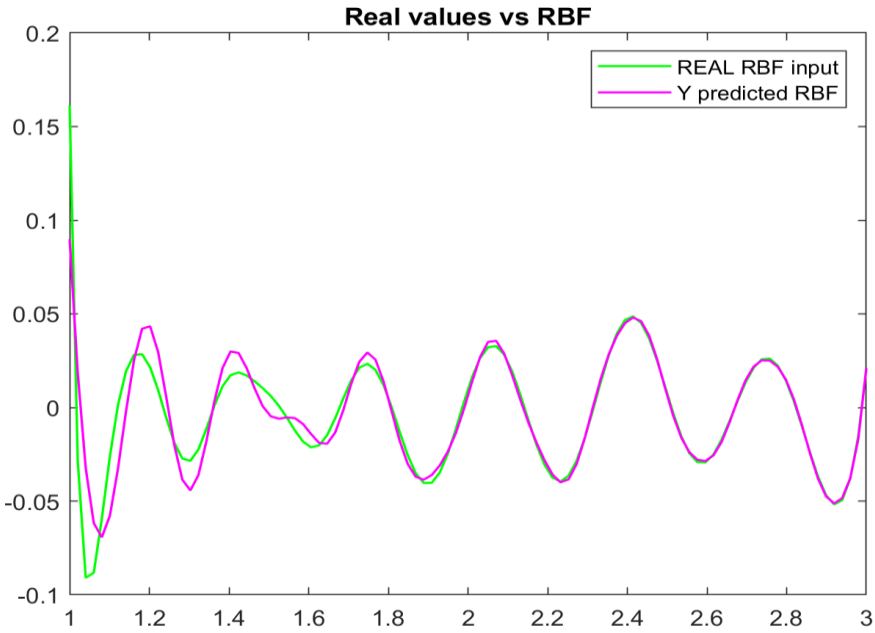
Σχήμα 4.80 Προσέγγιση Συνάρτησης $y=2x^2+e^{(\pi/x)}*\sin(2\pi x)$ με Χρήση Συνόλου Ελέγχου για Εκτεταμένο MLP

Για το απλό MLP ως πρώτο μέρος του υβριδικού

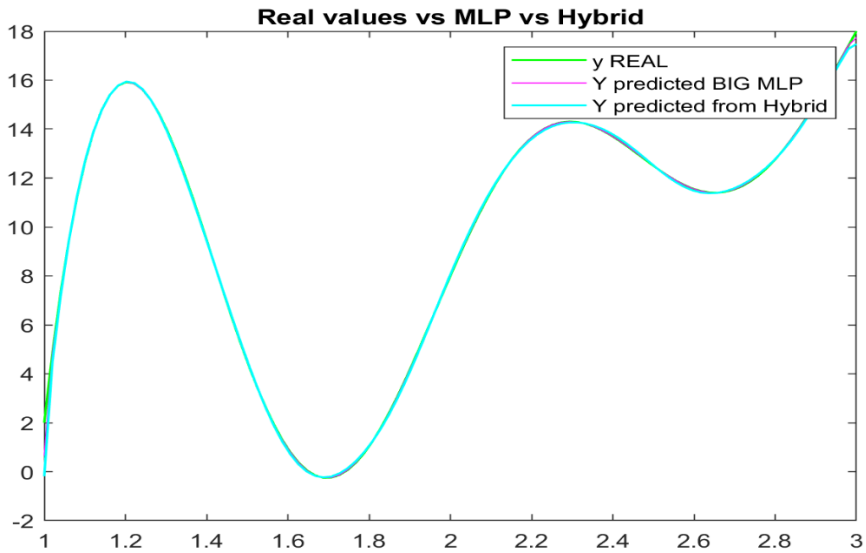


Σχήμα 4.81 Προσέγγιση Συνάρτησης $y=2x^2+e^{(\pi/x)}*\sin(2\pi x)$ με Χρήση Συνόλου Ελέγχου για Απλό MLP

Για το RBF ως δεύτερο μέρος του υβριδικού



Σχήμα 4.82 Προσέγγιση Συνάρτησης Σφάλματος της $y=2x^2+e^{(\pi/x)}*\sin(2\pi x)$ με Χρήση Συνόλου Ελέγχου με RBF

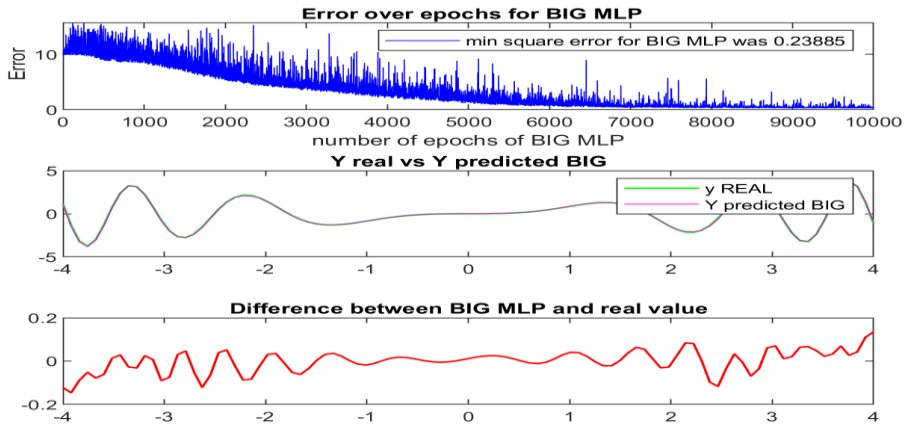


Σχήμα 4.83 Σύγκριση Μεταξύ Πραγματικού, Εκτεταμένου MLP και Υβριδικού Μοντέλου με Χρήση Συνόλου Ελέγχου της Συνάρτησης Γ

4.6.4. Συνάρτηση Δ

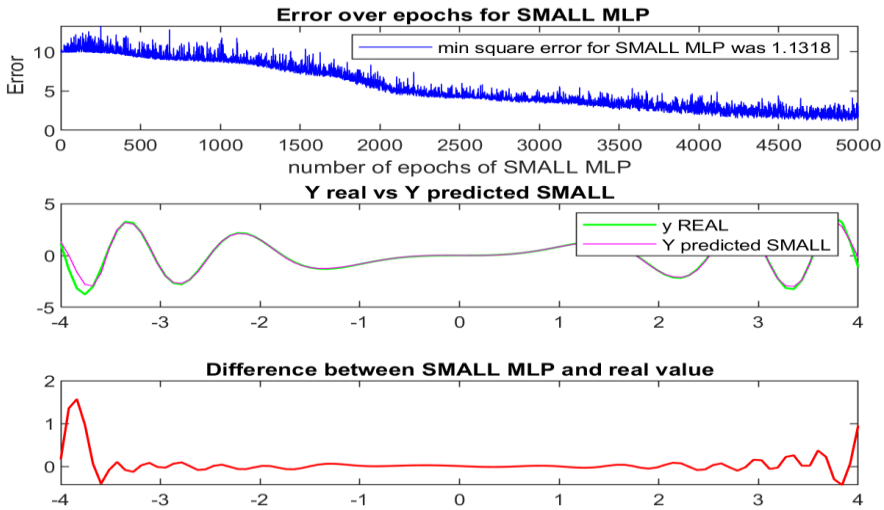
Στη συνάρτηση αυτή, παρατηρήσαμε ότι που είναι και η πιο δύσκολη από αυτές που εξετάστηκαν, παρατηρούμε ότι το RBF δυσκολεύτηκε να ακολουθήσει το ρυθμό που είχαν οι καμπύλες εισόδου. Πιο αναλυτικά τα αποτελέσματα που πήραμε ήταν τα ακόλουθα:

Για το εκτεταμένο MLP



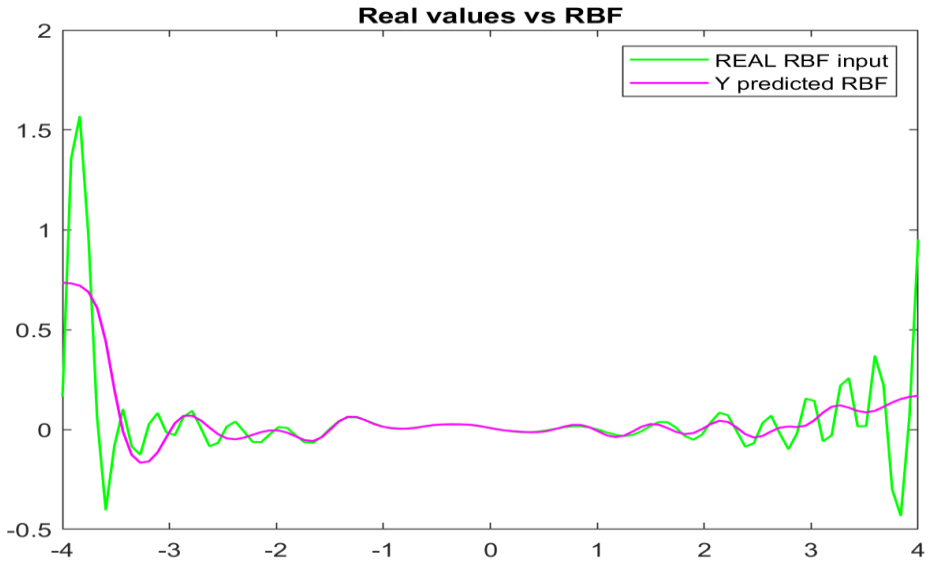
Σχήμα 4.84 Προσέγγιση Συνάρτησης y=x*sin(x^2) με Χρήση Συνόλου Ελέγχου για Εκτεταμένο MLP

Για το απλό MLP ως πρώτο μέρος του υβριδικού



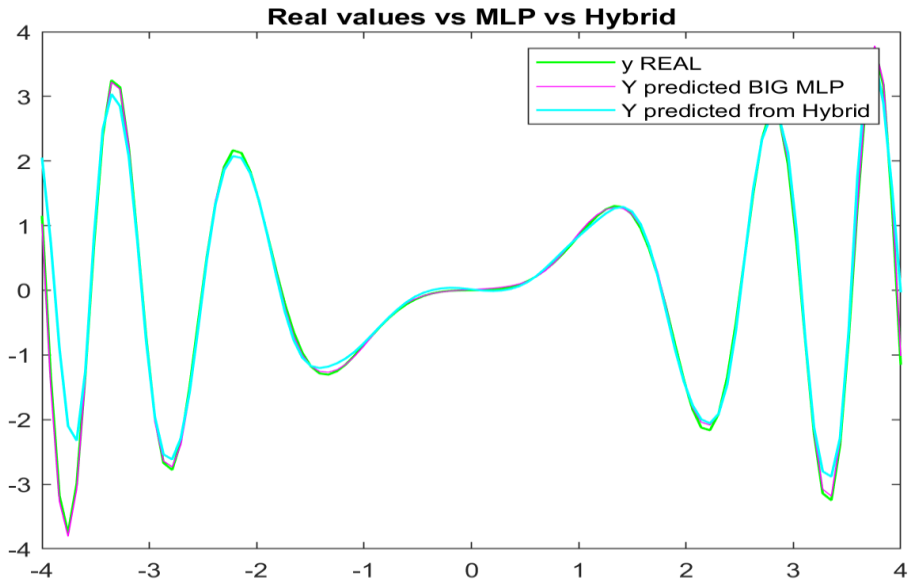
Σχήμα 4.85 Προσέγγιση Συνάρτησης y=x*sin(x^2) με Χρήση Συνόλου Ελέγχου για Απλό MLP

Για το RBF ως δεύτερο μέρος του υβριδικού



Σχήμα 4.86 Προσέγγιση Συνάρτησης Σφάλματος της $y=x*\sin(x^2)$ με Χρήση Συνόλου Ελέγχου RBF

Σύγκριση μεταξύ πραγματικού, εκτεταμένου MLP και υβριδικού μοντέλου



Σχήμα 4.87 Σύγκριση Μεταξύ Πραγματικού, Εκτεταμένου MLP και Υβριδικού Μοντέλου με Χρήση Συνόλου Ελέγχου της Συνάρτησης Δ

4.6.5. Σχολιασμός Αποτελεσμάτων

Συνολικά παρατηρήσαμε ότι εν γένει το σφάλμα αυξήθηκε αλλά δεν ήταν τέτοια η τάξη μεγέθους της αύξησής του ώστε να παρατηρήσουμε κάποια μεγάλη διαφορά Το σφάλμα σε κάθε συνάρτηση σε αυτή την περίπτωση διαμορφώθηκε όπως δείχνει

Πίνακας 4.6: Συγκεντρωτικός Πίνακας Σφαλμάτων για νευρώνες RBF = 30

Σφάλμα	Εκτεταμένο MLP	Απλό MLP	RBF	ΥΒΡΙΔΙΚΟ
Συνάρτηση Α	2.80E-02	0.033	1.26E-07	9.82E-04
Συνάρτηση Β	1.20E-02	0.025	2.83E-06	4.30E-05
Συνάρτηση Γ	2.60E-02	0.061	0.0001	0.036
Συνάρτηση Δ	0.238	1.131	4.10E-02	0.161

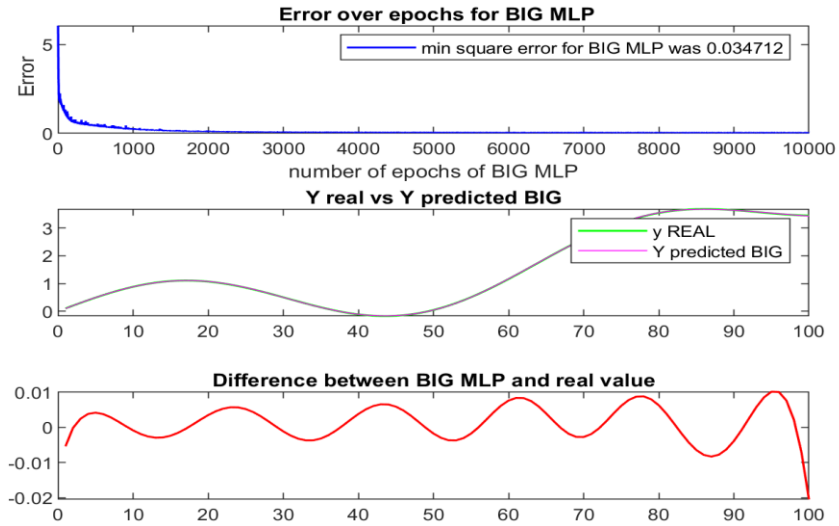
4.7. Διαφοροποίηση των Δεδομένων Εκπαίδευσης με τα Δεδομένα Ελέγχου

Αρχικά, για να τεστάρουμε το υβριδικό δίκτυο ως προς το πολύπλοκο MLP, πραγματοποιήσαμε για τις διάφορες συναρτήσεις, προσομοιώσεις στις οποίες τα δεδομένα εκπαίδευσης (train) και τα δεδομένα ελέγχου (test) ήταν το ίδιο διάνυσμα. Σε αυτή την παράγραφο παρουσιάζουμε τα αποτελέσματα των προσομοιώσεων όταν τα δεδομένα ελέγχου (test) είχαν διαφοροποιηθεί από το διάνυσμα εκπαίδευσης (train). Για την άμεση σύγκριση των αποτελεσμάτων τρέξαμε ακριβώς τις ίδιες προσομοιώσεις με τις ίδιες παραμετροποιήσεις πέραν από το γεγονός πως το διάνυσμα ελέγχου test είχε μετατοπιστεί δεξιά κατά μία μικρή τιμή κάθε φορά ανάλογη των δεδομένων εισόδου η οποία ήταν στο μέσο των δεδομένων. Δηλαδή, αν τα δεδομένα εισόδου ήταν 1,2,3,… τότε το καινούριο διάνυσμα ήταν 1.5,2.5,3.5… κ.ο.κ. Παρατηρήσαμε λοιπόν πως ναι μεν το σφάλμα αυξήθηκε αλλά το αποτέλεσμα συνέχισε να είναι ικανοποιητικό. Τα αποτελέσματα για κάθε συνάρτηση φαίνονται παρακάτω.

4.7.1. Συνάρτηση Α

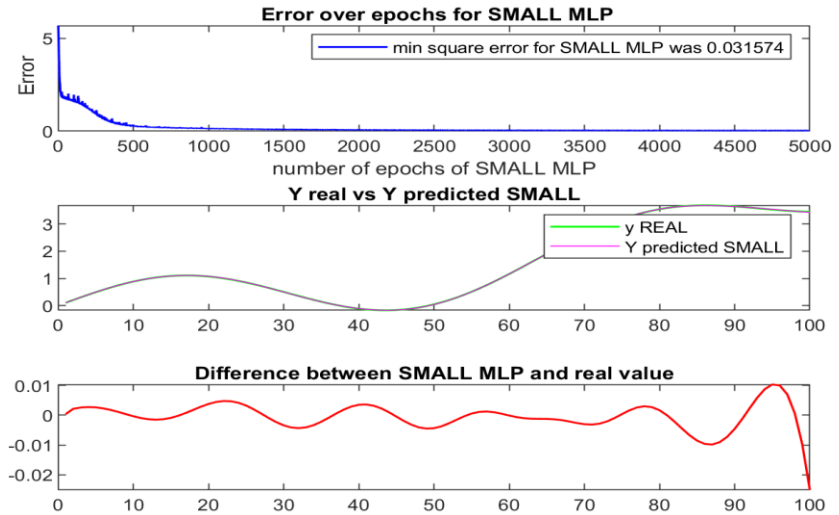
Για τη συνάρτηση αυτή δεν παρατηρήσαμε μεγάλες μεταβολές στις γραφικές παραστάσεις σε σχέση με πριν.

Για το εκτεταμένο MLP



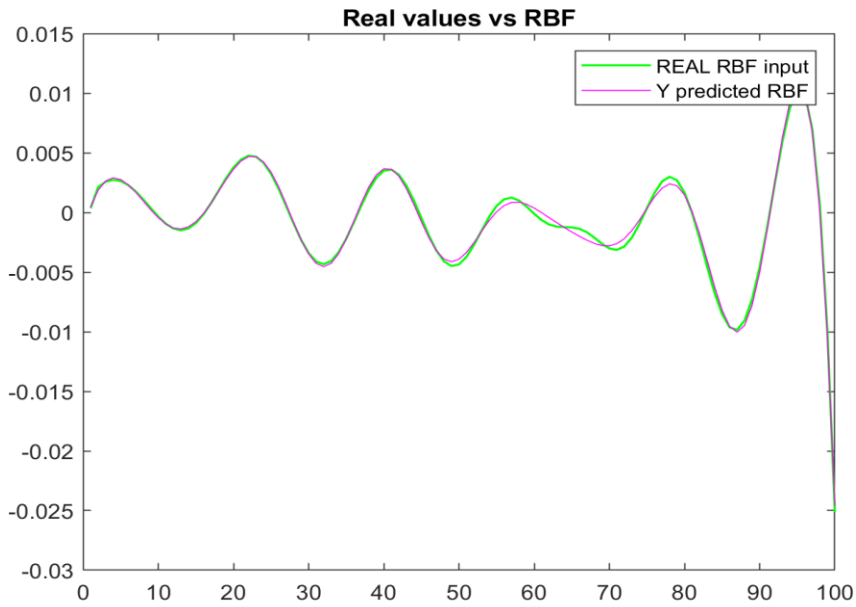
Σχήμα 4.88 Προσέγγιση Συνάρτησης $y=\sin(x/10) + (x/50)^2$ με Χρήση Συνόλου Ελέγχου για Εκτεταμένο MLP

Για το απλό MLP ως πρώτο μέρος του υβριδικού



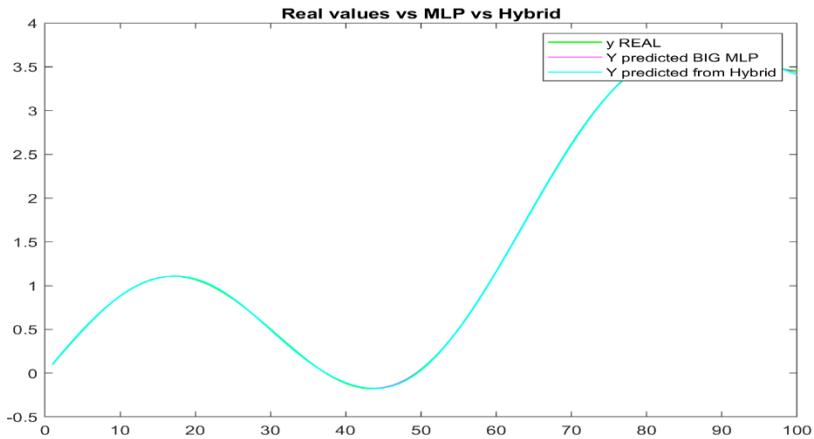
Σχήμα 4.89 Προσέγγιση Συνάρτησης y=sin(x/10) + $(x/50)^2$ με Χρήση Συνόλου Ελέγχου για Απλό MLP

Για το RBF ως δεύτερο μέρος του υβριδικού



Σχήμα 4.90 Προσέγγιση Συνάρτησης Σφάλματος της y=sin(x/10) + $(x/50)^2$ με Χρήση Συνόλου Ελέγχου για RBF

Σύγκριση μεταξύ πραγματικού, εκτεταμένου MLP και υβριδικού μοντέλου

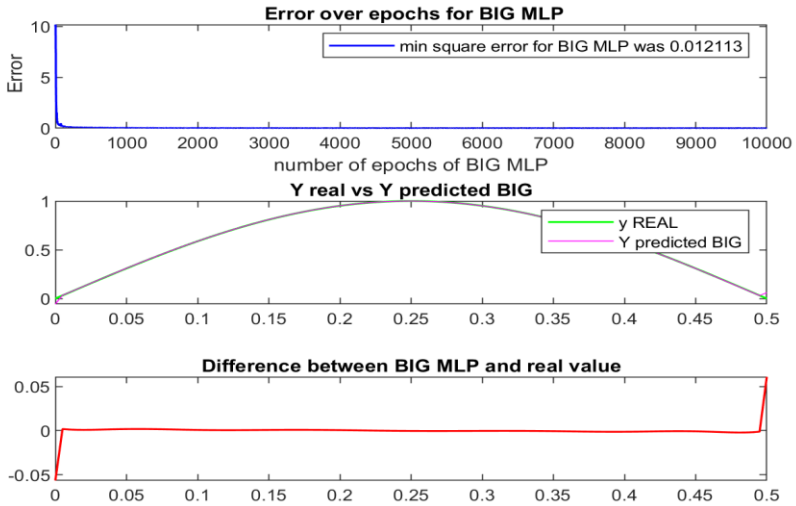


Σχήμα 4.91 Σύγκριση Μεταξύ Πραγματικού, Εκτεταμένου MLP και Υβριδικού Μοντέλου με Χρήση Συνόλου Ελέγχου της Συνάρτησης Α

4.7.2. Συνάρτηση Β

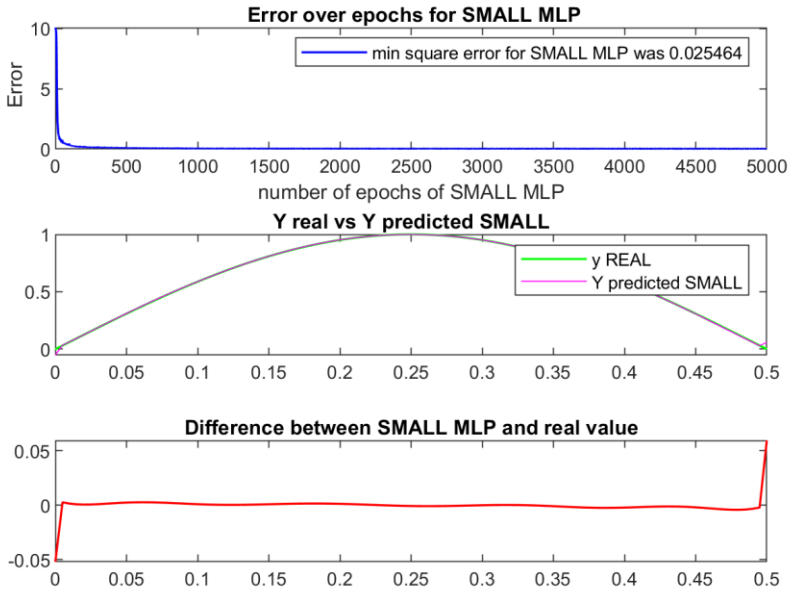
Παρόμοια αποτελέσματα βρήκαμε και για αυτή τη συνάρτηση σε σύγκριση με την προηγούμενη περίπτωση μόνο που εδώ στις άκρες της περιοχής προσομοίωσης είδαμε το σφάλμα το απλού MLP να είναι αρχικά πιο απότομο και μετέπειτα να πέφτει

Για το εκτεταμένο MLP



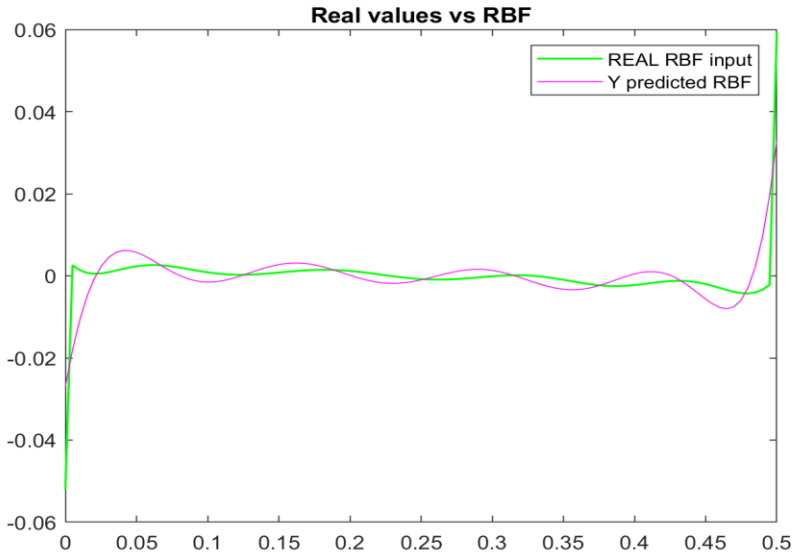
Σχήμα 4.92 Προσέγγιση Συνάρτησης y=sin(2πx) με Χρήση Συνόλου Ελέγχου για Εκτεταμένο MLP

Για το απλό MLP ως πρώτο μέρος του υβριδικού



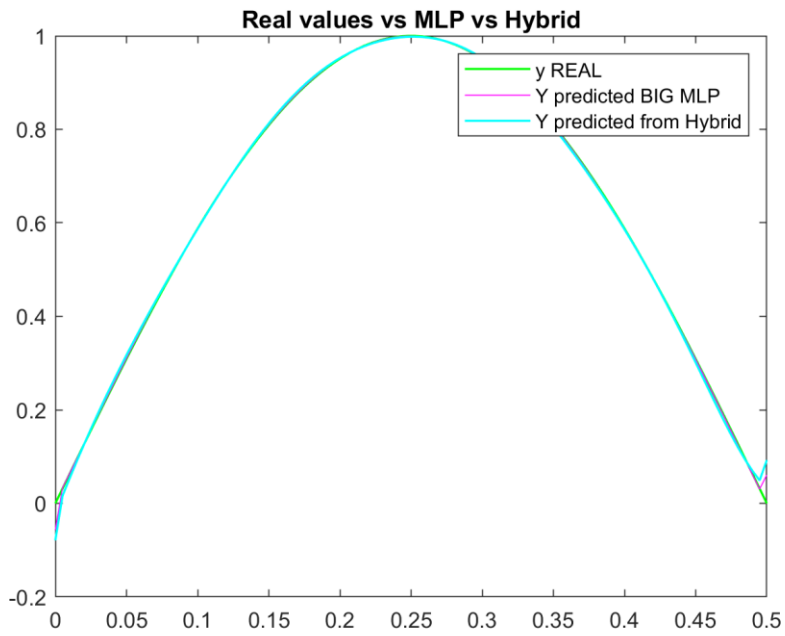
Σχήμα 4.93 Προσέγγιση Συνάρτησης y=sin(2πx) με Χρήση Συνόλου Ελέγχου για Απλό MLP

Για το RBF ως δεύτερο μέρος του υβριδικού



Σχήμα 4.94 Προσέγγιση Συνάρτησης Σφάλματος της y=sin(2πx) με Χρήση Συνόλου Ελέγχου RBF

Σύγκριση μεταξύ πραγματικού, εκτεταμένου MLP και υβριδικού μοντέλου

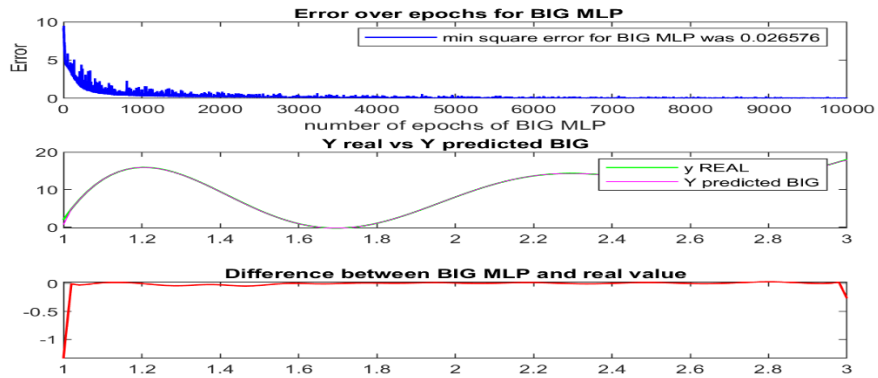


Σχήμα 4.95 Σύγκριση Μεταξύ Πραγματικού, Εκτεταμένου MLP και Υβριδικού Μοντέλου με Χρήση Συνόλου Ελέγχου της Συνάρτησης της Β

4.7.3. Συνάρτηση Γ

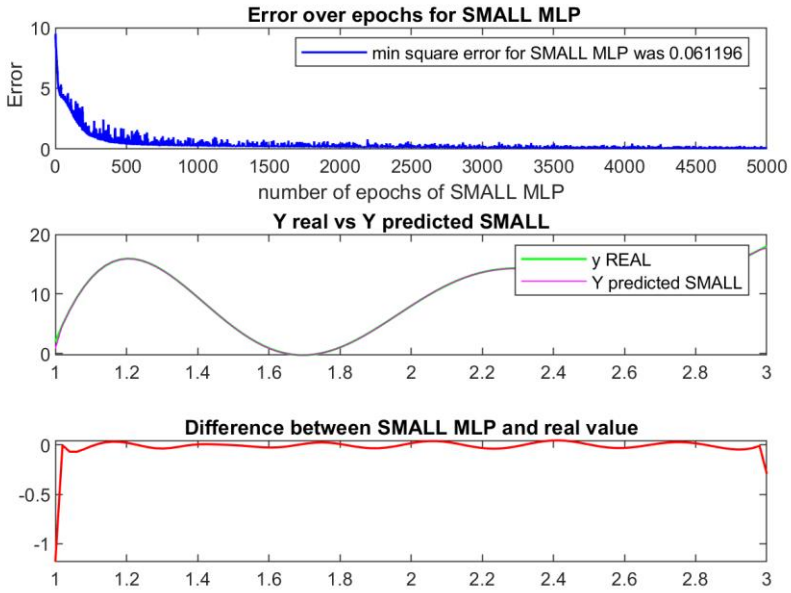
Παρόμοια αποτελέσματα σε σχέση με τις άλλες δύο συναρτήσεις βλέπουμε και σε αυτή τη συνάρτηση που δείχνει την γενικευτική ικανότητα του υβριδικού δικτύου ανεξαρτήτως συνάρτησης προς προσέγγισης. Πιο αναλυτικά, τα αποτελέσματα που πήραμε ήταν τα ακόλουθα:

Για το εκτεταμένο MLP



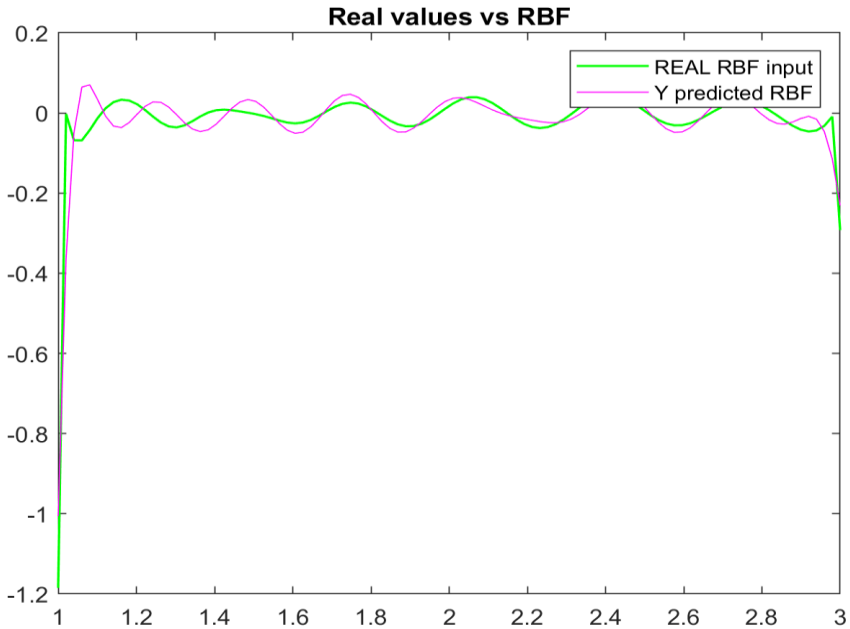
Σχήμα 4.96 Προσέγγιση Συνάρτησης $y=2x^2+e^{(\pi/x)}*\sin(2\pi x)$ με Χρήση Συνόλου Ελέγχου για Εκτεταμένο MLP

Για το απλό MLP ως πρώτο μέρος του υβριδικού

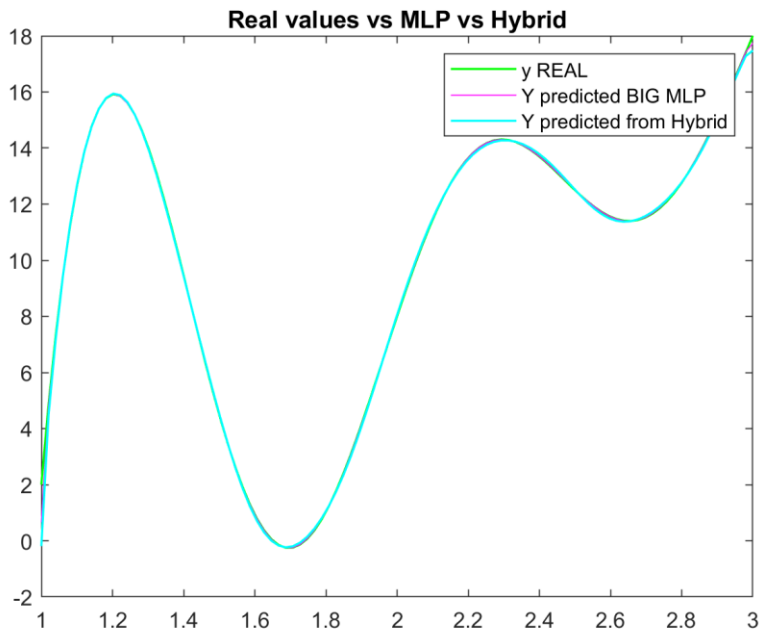


Σχήμα 4.97 Προσέγγιση Συνάρτησης $y=2x^2+e^{(\pi/x)}*\sin(2\pi x)$ με Χρήση Συνόλου Ελέγχου για Απλό MLP

Για το RBF ως δεύτερο μέρος του υβριδικού



Σχήμα 4.98 Προσέγγιση Συνάρτησης Σφάλματος της $y=2x^2+e^{(\pi/x)}*\sin(2\pi x)$ με Χρήση Συνόλου Ελέγχου με RBF

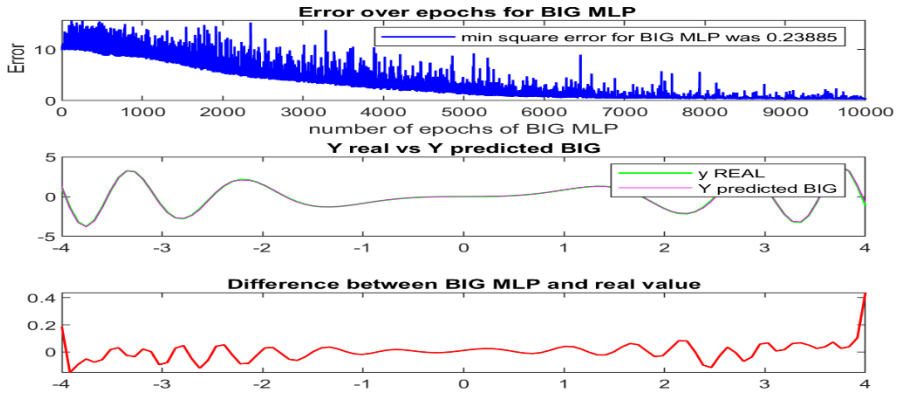


Σχήμα 4.99 Σύγκριση Μεταξύ Πραγματικού, Εκτεταμένου MLP και Υβριδικού Μοντέλου με Χρήση Συνόλου Ελέγχου της Συνάρτησης Γ

4.7.4. Συνάρτηση Δ

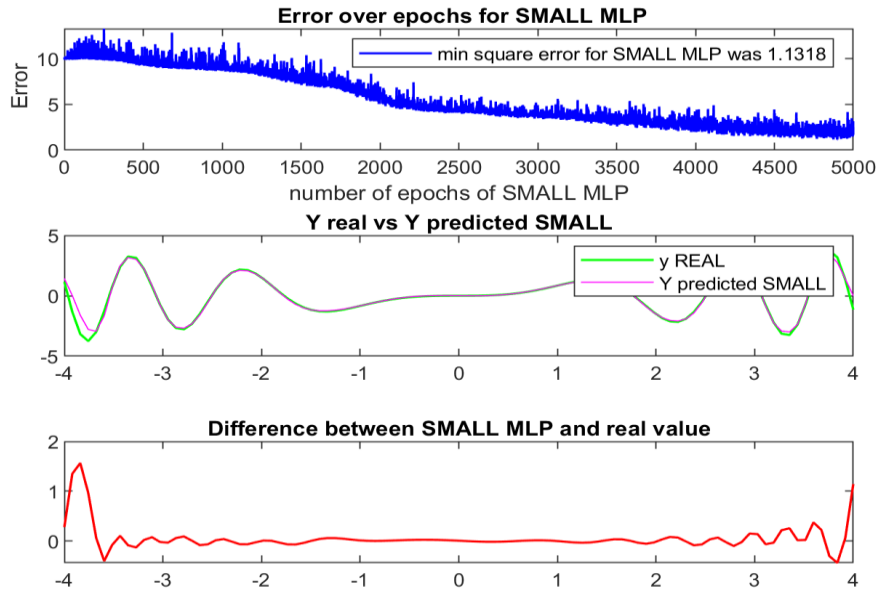
Στη συνάρτηση αυτή, παρατηρούμε ότι το υβριδικό δίκτυο συμπεριφέρεται πάρα πολύ καλά αλλά όχι άριστα όπως στις προηγούμενες συναρτήσεις. Πιο αναλυτικά τα αποτελέσματα που πήραμε ήταν τα ακόλουθα:

Για το εκτεταμένο MLP



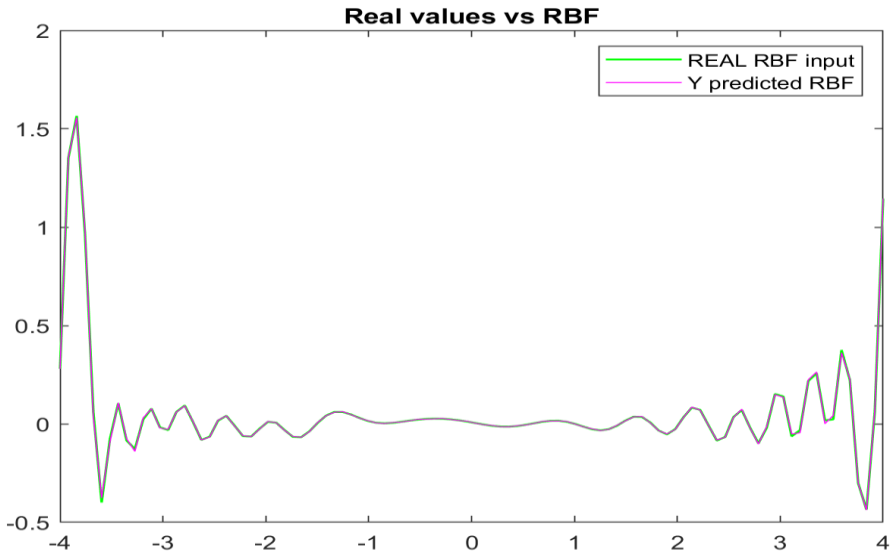
Σχήμα 4.100 Προσέγγιση Συνάρτησης $y=x*\sin(x^2)$ με Χρήση Συνόλου Ελέγχου για Εκτεταμένο MLP

Για το απλό MLP ως πρώτο μέρος του υβριδικού



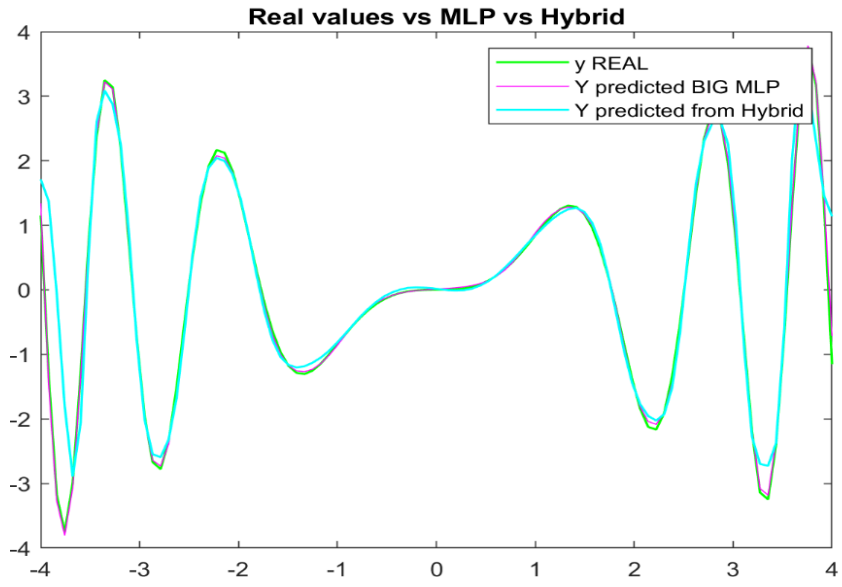
Σχήμα 4.101 Προσέγγιση Συνάρτησης y=x*sin(x^2) με Χρήση Συνόλου Ελέγχου για Απλό MLP

Για το RBF ως δεύτερο μέρος του υβριδικού



Σχήμα 4.102 Προσέγγιση Συνάρτησης Σφάλματος της y=x*sin(x^2) με Χρήση Συνόλου Ελέγχου RBF

Σύγκριση μεταξύ πραγματικού, εκτεταμένου MLP και υβριδικού μοντέλου



Σχήμα 4.103 Σύγκριση Μεταξύ Πραγματικού, Εκτεταμένου MLP και Υβριδικού Μοντέλου με Χρήση Συνόλου Ελέγχου της Συνάρτησης Δ

4.7.5. Σχολιασμός Αποτελεσμάτων

Συνολικά παρατηρήσαμε ότι εν γένει το σφάλμα αυξήθηκε αλλά δεν ήταν τέτοια η τάξη μεγέθους της αύξησής του ώστε να παρατηρήσουμε κάποια μεγάλη διαφορά Το σφάλμα σε κάθε συνάρτηση σε αυτή την περίπτωση διαμορφώθηκε όπως δείχνει

Πίνακας 4.7: Συγκεντρωτικός Πίνακας Σφαλμάτων για train ≠ test

Σφάλμα	Εκτεταμένο MLP	Απλό MLP	RBF	ΥΒΡΙΔΙΚΟ
Συνάρτηση Α	0.034	0.031	9.81^{E}-08	8.27^{E}-05
Συνάρτηση Β	0.012	0.025	2.99^{E}-08	0.0001
Συνάρτηση Γ	0.026	0.061	0.002	0.054
Συνάρτηση Δ	0.238	1.131	3.55^{E}-05	0.302

ΚΕΦΑΛΑΙΟ 5.
ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΠΡΟΤΑΣΕΙΣ

5.1. Συμπεράσματα
5.2. Προτάσεις

5.1. Συμπεράσματα

Η παρούσα διατριβή αποτελεί μια πρότυπη εργασία η οποία ασχολείται με την δημιουργία ενός σύνθετου υβριδικού δικτύου αποτελούμενου από δύο διαφορετικού τύπου νευρωνικά δίκτυα, ένα MLP σε σειρά με ένα RBF. Σκοπός της διατριβής είναι να παρουσιάσει μια εναλλακτική μέθοδος προσέγγισης συναρτήσεων σε σύγκριση με ένα πιο εκτεταμένο MLP.

Συνολικά, διεξήχθη μελέτη για την συμπεριφορά, λειτουργία και προσομοίωση τέτοιων δικτύων με σκοπό την κατανόηση της υπάρχουσας βιβλιογραφίας γύρω από αυτά. Έπειτα, δημιουργήθηκαν επιμέρους συναρτήσεις για την υλοποίηση του MLP και του RBF οι οποίες καλούνται μέσα από μία κεντρική συνάρτηση που ελέγχει όλες τις παραμέτρους τους.

Η αξιολόγηση έγινε σε τέσσερις διαφορετικές συναρτήσεις με ποικίλη μορφή αλλά κύριο συστατικό την ύπαρξη ημιτονοειδούς συνάρτησης για την παραγωγή τοπικών ελαχίστων και μεγίστων. Ακολούθως, παρουσιάστηκαν τα αποτελέσματα αρχικά του MLP δικτύου για μία συνάρτηση από αυτές σε συνάρτηση με κάθε μία από τις παραμέτρους που προσδιορίζουν την αποτελεσματικότητά του. Ομοίως παρουσιάστηκαν και τα αποτελέσματα για το RBF δίκτυο. Επιπλέον, παρουσιάστηκε η αποτελεσματικότητά του για κάθε μία από τις συναρτήσεις που επιλέχθηκαν . Τέλος παρουσιάστηκε η αποτελεσματικότητά του υβριδικού δικτύου σε σύγκριση με το πιο εκτεταμένο MLP. Το κύριο κριτήριο αξιολόγησης ήταν η ικανότητα γενίκευσης των δικτύων που σχεδιάσαμε. Έτσι, παρατηρήσαμε ότι τα αποτελέσματα που πήραμε ήταν αρκετά ενθαρρυντικά με το υβριδικό δίκτυο να παρέχει πολύ ομαλή και ακριβή προσέγγιση στο αποτέλεσμα.

5.2. Προτάσεις

Μελλοντικός στόχος είναι η διεξαγωγή περισσότερων και αναλυτικότερων προσομοιώσεων με σκοπό την καλύτερη προσαρμογή των δικτύων για την απόκτηση μέγιστης γενικευτικής ικανότητας ανεξαρτήτως εισόδου δεδομένων. Επίσης, σκοπός μας είναι η εφαρμογή του μοντέλου σε πραγματικά προβλήματα και δεδομένα εισόδου για την προσαρμογή του και την αξιολόγηση της ικανότητάς του σε αυτό το κομμάτι

ΒΙΒΛΙΟΓΡΑΦΙΑ

[1] Γ. Α. Ροβιθάκης, Τεχνικές Βελτιστοποίησεις, Τζιολα, 2007.

[2] K. E. P. &. M. N. Vrahatis, Practile Swarm optimazation and Intelligence : advances and applications, IGI Global, 2010.

[3] S. McLoone, M. D. Brown, G. Irwin and G. Lightbody, "A hybrid linear/nonlinear training algorithm for feedforward neural networks," *IEEE Transactions on Neural Networks*, pp. 669-684, 1998.

[4] H. Landahl, W. S. McCulloch and W. Pitts, "A statistical consequence of the logical calculus of nervous nets," *Bulletin of Mathematical Biology*, pp. 135-137, 1943.

[5] Κ. Διαμαντάρας, Τεχνητά Νευρωνικά Δίκτυα, Κλειδάριθμος, 2007.

[6] S. Haykin, Νευρωνικά Δίκτυα και Μηχανική Μάθηση, Αθήνα: Παπασωτηρίου, 2010.

[7] I. Isa, Z. Saad, S. Omar, . M. Osman, K. Ahmad and H. M. Sakim, "Suitable MLP network activation functions for breast cancer and thyroid disease detection," in *Computational Intelligence, Modelling and Simulation (CIMSiM), 2010 Second International Conference on*, 2010.

[8] E. El-Sebakhy, Artificial Neural Networks, Probabilistic Networks, Support Vector Machines, Adaptive-Neuro Fuzzy Systems, and Functional Networks, Elsevier Science, Saudi Arabia, 2006.

[9] L.-M. Fu, Neural networks in computer intelligence, Tata McGraw-Hill Education, 2003.

[10] J. E. Dayhoff, Neural network architectures: an introduction, New York: Van Nostrand Reinhold Co, 1990.

[11] V. Cherkassky and H. Lari-Najafi, "Data representation for diagnostic neural networks," *IEEE Expert}*, pp. 43-53, 1992.

[12] F. Schwenker, H. A. Kestler and G. Palm, "Three learning phases for radial-basis-function networks," *Neural Networks*, pp. 439-458, 2001.

[13] Y. Wu, H. Wang, B. Zhang and D. K-L, "Using radial basis function networks for function approximation and classification," *ISRN Applied Mathematics,* 2012.

[14] J. Gasteiger and J. Zupan, "Neural networks in chemistry," *Angewandte Chemie International Edition in English,* pp. 503--527, 1993.

[15] R. P. Lippmann, "Pattern classification using neural networks," *IEEE communications magazine,* pp. 47-50, 1989.

[16] A. Λύκας, "http://www.cs.uoi.gr/~arly/courses/nn/slides/K4.pdf," Απρίλιος 2018. [Online]. Available: http://www.cs.uoi.gr/~arly/courses/nn/slides/K4.pdf. [Accessed 2018].

[17] . D. Svozil, . V. Kvasnicka and J. Pospichal, "Introduction to multi-layer feed-forward neural networks," *Chemometrics and intelligent laboratory systems,* pp. 43-62, 1997.

[18] D. E. Rumelhart, G. E. Hinton and R. J. Williams, "Learning internal representations by error propagation," California Univ San Diego La Jolla Inst for Cognitive Science, 1985.

[19] J. E. Dayhoff and J. M. DeLeo, "Artificial neural networks: opening the black box," *Cancer: Interdisciplinary International Journal of the American Cancer Society,* pp. 1615-1635, 2001.

[20] C. M. Bishop, Pattern Recognition and Machine Learning, Springer Science+Business Media,LLC, 2006.

[21] R. Reed and R. J. MarksII, Neural smithing: supervised learning in feedforward artificial neural networks, Mit Press, 1999.

[22] E. Barnard and L. Wessels, "Extrapolation and interpolation in neural network classifiers," *IEEE Control Systems Magazine,* pp. 50-53, 1992.

[23] I. V. Tetko, D. J. Livingstone and A. I. Luik, "Neural network studies. 1. Comparison of overfitting and overtraining," *Journal of chemical information and computer sciences,* pp. 826-833, 1995.

[24] M. Stone, "Cross-validatory choice and assessment of statistical predictions," *Journal of the royal statistical society. Series B (Methodological),* pp. 111-147, 1974.

[25] Z. Zainuddin and P. Ong, "Function Approximation Using Artificial Neural Networks," *WSEAS Transactions on Mathematics,* pp. 333-338, 2008.

[26] G. Li, H. Alnuweiri, Y. Wu and H. Li, "Acceleration of back propagation through initial weight pre-training with delta rule," in *Neural Networks, 1993., IEEE International Conference on*, 1993.

[27] Y. Lee, S.-H. Oh and M. W. Kim, "The effect of initial weights on premature saturation in back-propagation learning," 1991.

[28] M. Hassoun, Fundamentals of artificial neural networks, MIT press, 1995.

[29] Y. Lecun, L. Bottou, G. Orr and K.-R. Muller, Neural Networks: Trick of the trade, Springer, 1998.

[30] S. Walczak and N. Cerpa, "Heuristic principles for the design of artificial neural networks," *Information and software technology,* pp. 107--117, 1999.

[31] R. A. Jacobs, "Increased rates of convergence through learning rate adaptation," *Neural networks,* pp. 295--307, 1988.

[32] J. Bilski, "The backpropagation learning with logarithmic transfer function," in *Proc. 5th Conf. on Neural Networks and Soft Computing*, 2000.

[33] S.-W. Lee and C. Moraga, "A Cosine-Modulated Gaussian activation function for Hyper-Hill neural networks," in *Signal Processing, 1996., 3rd International Conference on*, 1996.

[34] M. Rosen-Zvi, M. Biehl and I. Kanter, "Learnability of periodic activation functions: General results," *Physical Review E,* 1998.

[35] L. R. Medsker and J. Liebowitz, Design and Developement of Expert Systems and Neutral Networks, Prentice Hall PTR, 1993.

[36] L.-M. Fu, Neural networks in computer intelligence, Tata McGraw-Hill Education, 2003.

[37] I. A. a. H. M. Basheer, "Artificial neural networks: fundamentals, computing, design, and application," *Journal of microbiological methods*, pp. 3--31, 2000.

[38] M. Hung, M. Hu, M. Shanker and B. Patuwo, "Estimating posterior probabilities in classification problems with neural networks," *International Journal of Computational Intelligence and Organizations*, pp. 49-60, 1996.

[39] K. Hornik, M. Stinchcombe and H. White, "Multilayer feedforward networks are universal approximators," *Neural networks*, pp. 359-366, 1989.

[40] S. Chen, S. Billings and P. Grant, "Recursive hybrid algorithm for non-linear system identification using radial basis function networks," *International Journal of Control*, pp. 1051-1070, 1992.

[41] "Ελληνικά Ακαδημαϊκά Ηλεκτρονικά Συγγράμματα και Βοηθήματα," [Online]. Available: https://repository.kallipos.gr. [Accessed 6 2018].

ΣΥΝΤΟΜΟ ΒΙΟΓΡΑΦΙΚΟ

Η Ερίσα Σκέντα γεννήθηκε στο Τεπελένι της Αλβανίας το 1988. Το 1995 μετακόμισε μόνιμα με την οικογένεια της στα Ιωάννινα. Το 2012 αποφοίτησε από το τμήμα Τεχνολογίες Πληροφορικής και Τηλεπικοινωνίων του Τεχνολογικού Εκπαιδευτικού Ιδρύματος Ηπείρου. Το 2016 έγινε δεκτή ως μεταπτυχιακή φοιτήτρια του τμήματος Μηχανικών Η/Υ και Πληροφορικής του Πανεπιστήμιου Ιωάννινων. Από το Φεβρουάριο του 2014 έως και σήμερα εργάζεται στην εταιρεία Business Solution ως web developer. Γνωρίζει άπταιστα Αλβανικά, Ελληνικά και Αγγλικά.