



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΜΗΧΑΝΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΤΟΜΕΑΣ ΒΙΟΜΗΧΑΝΙΚΗΣ ΔΙΟΙΚΗΣΗΣ ΚΑΙ
ΕΠΙΧΕΙΡΗΣΙΑΚΗΣ ΕΡΕΥΝΑΣ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**Πειραματική διερεύνηση αλγορίθμων για βελτιστοποίηση
της απόδοσης της πρόγνωσης χρονοσειρών με τη χρήση της
μεθόδου των τεχνητών νευρωνικών δικτύων**

Τσαλαβούτης Α. Βασίλειος

Φοιτητής 10^{ου} εξαμήνου

Επιβλέπων: Τόλης Αθανάσιος, Επίκουρος Καθηγητής Ε.Μ.Π

Αθήνα, Σεπτέμβριος 2014

Αφιερώνεται
στην οικογένεια μου, για την
στήριξη τους όλα αυτά τα χρόνια

Ευχαριστίες

Καταρχήν θα ήθελα να ευχαριστήσω και να εκφράσω την ευγνωμοσύνη μου στον κύριο Τόλη. Τον ευχαριστώ για την ευκαιρία που μου έδωσε να ασχοληθώ με το συγκεκριμένο θέμα καθώς και για την εμπιστοσύνη του. Επιπλέον θα ήθελα να τον ευχαριστήσω για τις συζητήσεις που κάναμε καθ' όλη τη διάρκεια εκπόνησης του συγκεκριμένου θέματος και οι οποίες πραγματικά ήταν πολύ εποικοδομητικές για μένα.

Έπειτα θα ήθελα να ευχαριστήσω τους συμφοιτητές μου γιατί ο καθένας από αυτούς συνέβαλε με το δικό του τρόπο ώστε από αυτά τα χρόνια που πέρασα ως φοιτητής των Μηχανολόγων να μου μείνουν οι καλύτερες αναμνήσεις. Ειδικά θα ήθελα να ευχαριστήσω τους φίλους μου Χ. , Γ., Φ. και Β. γιατί μαζί τους μοιράστηκα πολλές καλές στιγμές τα τελευταία πέντε χρόνια.

Για το τέλος κράτησα το μεγαλύτερο ευχαριστώ το οποίο το οφείλω στην οικογένεια μου. Στην αδερφή μου που αυτά τα πέντε χρόνια ως συγκάτοικός μου με βοήθησε να προσαρμοστώ και αποτέλεσε για μένα ένα παράδειγμα ανθρώπου με επιμονή και θέληση να πετύχει τους στόχους του. Στους γονείς μου γιατί όλα αυτά τα χρόνια με στηρίζουν χωρίς να ζητούν αντάλλαγμα και προσπαθούν ώστε να μπορώ εγώ να συνεχίζω απρόσκοπτα τις σπουδές μου και γιατί με βοήθησαν να γίνω ο άνθρωπος που είμαι.

Περιεχόμενα

Έποψη διπλωματικής εργασίας.....	11
----------------------------------	----

ΚΕΦΑΛΑΙΟ 1

ΕΙΣΑΓΩΓΗ ΣΤΑ ΤΕΧΝΗΤΑ ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

1.1	Εισαγωγή- Ορισμός των Τεχνητών Νευρωνικών Δικτύων	14
1.2	Πηγή έμπνευσης- Τα βασικά χαρακτηριστικά του βιολογικού νευρώνα	15
1.3	Ιστορική αναδρομή.....	17
1.4	Εφαρμογές των τεχνητών νευρωνικών δικτύων.....	19

ΚΕΦΑΛΑΙΟ 2

ΤΑ ΤΕΧΝΗΤΑ ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

2.1	Βασικά δομικά στοιχεία τεχνητών νευρωνικών δικτύων	22
2.1.1	Ο απλός νευρώνας (απλής εισόδου-απλής εξόδου).....	22
2.1.2	Απλός νευρώνας με διανυσματική είσοδο	23
2.2	Είδη συναρτήσεων ενεργοποίησης.....	24
2.3	Βασικές αρχιτεκτονικές νευρωνικών δικτύων	27
2.3.1	Τροφοδοτούμενα προς τα εμπρός δίκτυα απλού στρώματος.....	27
2.3.2	Τροφοδοτούμενα προς τα εμπρός δίκτυα πολλαπλών στρωμάτων	28
2.3.3	Αναδρομικά δίκτυα	30

ΚΕΦΑΛΑΙΟ 3

ΕΚΠΑΙΔΕΥΣΗ ΤΕΧΝΗΤΩΝ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ

3.1	Εισαγωγή στην έννοια της εκπαίδευσης ενός νευρωνικού δικτύου.....	34
3.2	Κανόνες εκμάθησης των δικτύων	35
3.2.1	Εκμάθηση Σφάλματος- Διόρθωσης	35
3.3	Πρότυπα εκπαίδευσης.....	37
3.3.1	Εκπαίδευση με επίβλεψη.....	37
3.4	Αλγόριθμοι εκπαίδευσης νευρωνικών δικτύων	40
3.4.1	Το πρόβλημα της εκχώρησης πίστωσης.....	40
3.4.2	Ο αλγόριθμος της οπίσθιας διάδοσης του σφάλματος.....	41
3.4.3	Ο αλγόριθμος της απότομης καθόδου.....	44
3.4.4	Παρουσίαση εναλλακτικών αλγορίθμων εκπαίδευσης δικτύων	45
3.5	Γενίκευση δικτύου.....	51
3.5.1	Το πρόβλημα του υπερβολικής προσαρμογής.....	52

ΚΕΦΑΛΑΙΟ 4

ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ ΚΑΙ ΠΡΟΒΛΕΨΗ ΧΡΟΝΟ-ΣΕΙΡΩΝ

4.1	Εισαγωγή.....	58
4.2	Βιβλιογραφική επισκόπηση I: Νευρωνικά Δίκτυα και πρόβλεψη χρονοσειρών 60	
4.3	Βιβλιογραφική επισκόπηση II: Νευρωνικά δίκτυα και ενέργεια	64
4.4	Στόχος της παρούσας διπλωματικής.....	69

ΚΕΦΑΛΑΙΟ 5

ΠΑΡΟΥΣΙΑΣΗ ΤΗΣ ΠΕΙΡΑΜΑΤΙΚΗΣ ΔΙΑΔΙΚΑΣΙΑΣ

5.1	Παρουσίαση μαθηματικού μοντέλου	72
5.1.1	Η λογική οργάνωση των πειραμάτων	72
5.2	Παρουσίαση των αλγορίθμων του πειράματος.....	77
5.2.1	Αλγόριθμος 1: Ομαδική πρόβλεψη.....	78
5.2.2	Αλγόριθμος 2: Ανεξάρτητης εισόδου- εξαρτημένης εξόδου	78
5.2.3	Αλγόριθμος 3: Προηγούμενης εξόδου- επόμενης εξόδου.....	80
5.2.4	Αλγόριθμος 4: Χρήσης του σετ αποτελεσμάτων προσομοίωσης.....	83
5.2.5	Αλγόριθμος 5: Λαμβάνοντας υπόψη την εποχικότητα	84
5.3	Σύνοψη πειραματικής διαδικασίας	88

ΚΕΦΑΛΑΙΟ 6

ΤΟ ΥΠΟΛΟΓΙΣΤΙΚΟ ΜΟΝΤΕΛΟ

6.1	Παρουσίαση των χρησιμοποιούμενων δεδομένων.....	90
6.2	Παρουσίαση υπολογιστικής διαδικασίας.....	92

ΚΕΦΑΛΑΙΟ 7

ΠΑΡΟΥΣΙΑΣΗ ΚΑΙ ΑΝΑΛΥΣΗ ΤΩΝ ΑΠΟΤΕΛΕΣΜΑΤΩΝ

7.1	Εισαγωγή-Περιγραφή του τρόπου παρουσίασης των αποτελεσμάτων	100
7.2	Παρουσίαση των αποτελεσμάτων	101
7.2.1	Ένα κρυφό στρώμα νευρώνων	101
7.2.2	Δύο κρυφά στρώματα νευρώνων.....	106
7.3	Μελέτη περίπτωσης με πραγματικά δεδομένα.....	115

ΚΕΦΑΛΑΙΟ 8

ΣΥΜΠΕΡΑΣΜΑΤΑ

8.1	Συμπεράσματα- Επίλογος διπλωματικής.....	120
-----	--	-----

ΠΑΡΑΡΤΗΜΑ

Παράρτημα Α: Σετ δεδομένων εκπαίδευσης	124
Παράρτημα Β: Κώδικες.....	126
Βιβλιογραφία και διαδικτυακοί σύνδεσμοι	129

Έποψη διπλωματικής εργασίας

Τα νευρωνικά δίκτυα αποτελούν ένα πολλά υποσχόμενο εργαλείο στο πεδίο της πρόβλεψης χρονοσειρών. Τα ιδιαίτερα χαρακτηριστικά που παρουσιάζουν σε σχέση με τις κλασικές μεθόδους προβλέψεων, και ιδιαίτερα η ικανότητά τους να ανιχνεύουν και μη γραμμικές σχέσεις μεταξύ των δεδομένων τα καθιστούν ένα πολύ ενδιαφέρον αντικείμενο προς μελέτη.

Στα πλαίσια της παρούσας διπλωματικής μελετούνται τα νευρωνικά δίκτυα και η απόδοσή τους στο πεδίο των προβλέψεων. Πιο συγκεκριμένα μελετάται και πραγματοποιείται σύγκριση της συμπεριφοράς πέντε αλγορίθμων που στόχο έχουν τη βελτιστοποίηση της απόδοσης του δικτύου κατά την πρόβλεψη χρονοσειρών. Στόχος είναι να διαπιστωθεί αν κάποιος ή κάποιοι από αυτούς τους αλγορίθμους επιτυγχάνουν στην πράξη να βελτιώσουν την απόδοση των νευρωνικών δικτύων κατά την διαδικασία της πρόβλεψης, χωρίς να μεταβάλλουν την επιρροή του δικτύου ή να επιδρούν στη διαδικασία της εκπαίδευσης, σε σχέση με τον κλασικό τρόπο εκπαίδευσης των νευρωνικών δικτύων.

Για την μελέτη του συγκεκριμένου ζητήματος πραγματοποιήθηκε μία σειρά από υπολογιστικά πειράματα. Σε αυτά εξετάστηκαν δίκτυα ενός και δύο κρυφών στρωμάτων νευρώνων και μελετήθηκε η απόδοσή τους σε κάθε ένα από τους πέντε αλγορίθμους, για το ίδιο σετ δεδομένων μίας σύνθετης χρονοσειράς και για τους ίδιους αριθμούς νευρώνων στα κρυφά στρώματα των δικτύων. Η μελέτη και η σύγκριση των αποτελεσμάτων πραγματοποιήθηκε με τη χρήση δύο συναρτήσεων απόδοσης (MAE, MSE) αποκλειστικά για τα σημεία στα οποία τα δίκτυα πραγματοποίησαν προβλέψεις. Από τα αποτελέσματα αυτών των υπολογιστικών πειραμάτων προκύπτουν τα δύο δίκτυα (ένα για κάθε περίπτωση αριθμού κρυφών στρωμάτων) τα οποία προπονήθηκαν από τον αλγόριθμο με τα καλύτερα αποτελέσματα και τα οποία εφαρμόζονται σε μία χρονοσειρά δεδομένων του πραγματικού κόσμου για να ελεγχθεί και εκεί η απόδοσή τους.

Συμπληρωματικά στην παραπάνω μελέτη πραγματοποιείται και μία δεύτερη, κατά την οποία εξετάζεται η επίδραση του αριθμού των νευρώνων στα κρυφά στρώματα καθώς και τους μεγέθους του σετ δεδομένων εκπαίδευσης στην απόδοση του δικτύου.

Για την εξέταση της επίδρασης των συγκεκριμένων παραγόντων παρέχονται στον πρώτο από τους πέντε αλγορίθμους, που στην ουσία αποτελεί τον κλασικό τρόπο εκπαίδευσης των νευρωνικών δικτύων, δύο επιπρόσθετα σετ δεδομένων τα οποία καλύπτουν το ίδιο εύρος τιμών με το αρχικό σετ δεδομένων αλλά περιλαμβάνουν περισσότερα σημεία. Έτσι με τον ίδιο τρόπο όπως και για τους υπόλοιπους αλγορίθμους πραγματοποιούνται τα υπολογιστικά πειράματα και καταγράφονται τα αποτελέσματα με τη χρήση των ίδιων συναρτήσεων απόδοσης. Βέβαια τα αποτελέσματα που προκύπτουν από τα δίκτυα με τα περισσότερα σημεία δεν συγκρίνονται με εκείνα που προκύπτουν για τους υπόλοιπους αλγορίθμους για το αρχικό σετ δεδομένων

Μετά την εκτέλεση των υπολογιστικών πειραμάτων και την εξέταση των αποτελεσμάτων προκύπτουν κάποια ενδιαφέροντα αποτελέσματα. Ο δεύτερος από τους αλγόριθμους παράγει τα καλύτερα αποτελέσματα για το αρχικό σετ δεδομένων, δηλαδή παράγει περισσότερα δίκτυα που πραγματοποιούν καλύτερη πρόβλεψη και τα δίκτυα με τις ακριβέστερες προβλέψεις. Επίσης για τη συμπληρωματική μελέτη φαίνεται από τα αποτελέσματα ότι για δίκτυα ενός κρυφού στρώματος όσο αυξάνεται ο αριθμός του σετ δεδομένων αυξάνεται και η απόδοση των δικτύων κατά την πρόβλεψη. Αυτό το συμπέρασμα δεν ισχύει και για τα δίκτυα των δύο κρυφών στρωμάτων τα οποία δεν φαίνεται να επηρεάζονται τόσο άμεσα από τον αριθμό των σημείων στο σετ δεδομένων εκπαίδευσης.

Παρακάτω περιγράφεται συνοπτικά το περιεχόμενο κάθε κεφαλαίου ώστε ο αναγνώστης να έχει μία γενική άποψη για τη δομή της παρούσας διπλωματικής.

Στο πρώτο κεφάλαιο παρουσιάζεται συνοπτικά η πηγή έμπνευσης των τεχνητών νευρωνικών δικτύων, ο νευρώνας του κεντρικού νευρικού συστήματος των έμβιων οργανισμών, καθώς και τα πεδία στα οποία εφαρμόζονται τα τεχνητά νευρωνικά δίκτυα.

Στο δεύτερο κεφάλαιο περιγράφεται το βασικό δομικό στοιχείο των τεχνητών νευρωνικών δικτύων, ο νευρώνας, και εξετάζονται μερικές βασικές αρχιτεκτονικές τεχνητών νευρωνικών δικτύων.

Στο τρίτο κεφάλαιο επιχειρείται μία περιγραφή της διαδικασίας της εκπαίδευσης των νευρωνικών δικτύων. Γίνεται αναφορά στα πρότυπα και στους κανόνες εκπαίδευσης και γίνεται προσπάθεια να περιγραφούν με απλά λόγια μερικοί από τους πιο βασικούς αλγόριθμους εκπαίδευσης.

Το κεφάλαιο τέσσερα περιλαμβάνει μία περίληψη των πιο βασικών, σύμφωνα με τη γνώμη του συγγραφέα της διπλωματικής, άρθρων και διατριβών που αφορούν το θέμα των νευρωνικών δικτύων και της πρόβλεψης χρονοσειρών.

Στο κεφάλαιο πέντε γίνεται μία παρουσίαση του μαθηματικού μοντέλου των πειραμάτων. Πιο συγκεκριμένα γίνεται μία περιγραφή των παραμέτρων που θα μεταβάλλονται κατά τη διάρκεια του πειράματος και της λογικής οργάνωσης των πειραμάτων. Τέλος αναλύονται οι πέντε αλγόριθμοι που θα εξεταστούν στη πλαίσια της διπλωματικής.

Στο κεφάλαιο έξι παρουσιάζονται τα σετ δεδομένων που χρησιμοποιούνται στην πειραματική διαδικασία καθώς και το υπολογιστικό μοντέλο, δηλαδή τα βήματα που πραγματοποιήθηκαν κατά την εκτέλεση των πειραμάτων και την αξιολόγηση των αποτελεσμάτων

Στο κεφάλαιο επτά πραγματοποιείται η παρουσίαση των αποτελεσμάτων και μία τεχνική ανάλυσή τους.

Τέλος στο κεφάλαιο οχτώ περιλαμβάνονται τα συμπεράσματα που προκύπτουν από τα αποτελέσματα των πειραμάτων.

ΚΕΦΑΛΑΙΟ 1

ΕΙΣΑΓΩΓΗ ΣΤΑ ΤΕΧΝΗΤΑ ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

1.1 Εισαγωγή- Ορισμός των Τεχνητών Νευρωνικών Δικτύων

Τα Τεχνητά Νευρωνικά Δίκτυα είναι, όπως αναφέρει και το όνομά τους, υπολογιστικά δίκτυα τα οποία προσπαθούν χονδρικά να προσομοιάσουν τον τρόπο που λειτουργούν οι νευρώνες του κεντρικού νευρικού συστήματος των έμβιων οργανισμών (είτε πρόκειται για ανθρώπινο οργανισμό είτε όχι).

Λόγω της γνώσης που δανείζονται από τον τομέα της νευροφυσιολογίας για τους βιολογικούς νευρώνες και τα δίκτυα που αυτοί σχηματίζουν, διαφέρουν από τις συμβατικές υπολογιστικές μηχανές που χρησιμοποιούνται για να επιταχύνουν την υπολογιστική διαδικασία αντικαθιστώντας τον ανθρώπινο εγκέφαλο. Σε αντίθεση με αυτές κατά τη δημιουργία ενός τεχνητού νευρωνικού δικτύου δίνεται ιδιαίτερη σημασία στην οργάνωση των υπολογιστικών μονάδων (νευρώνες) καθώς και στις συνδέσεις μεταξύ τους.

Γιατί όμως να χρησιμοποιηθούν και να αντιμετωπιστούν τα νευρωνικά δίκτυα με διαφορετικό τρόπο από ένα απλό υπολογιστικό αλγόριθμο προσομοίωσης; Η απάντηση σε αυτό το ερώτημα έχει να κάνει με το γεγονός ότι τα νευρωνικά δίκτυα λόγω της δομής τους επιτρέπουν με τη χρήση πολύ απλών υπολογισμών (κατά βάση πολλαπλασιασμών και προσθέσεων) την επίλυση περίπλοκων μαθηματικών προβλημάτων, τα οποία μπορεί να μην είναι πλήρως ορισμένα καθώς και μη γραμμικών ή στοχαστικών προβλημάτων. Συνάμα ο τρόπος της οργάνωσής τους επιτρέπει τη χρήση τους σε ένα μεγάλο εύρος προβλημάτων. Αντίθετα ένας συμβατικός υπολογιστικός αλγόριθμος χρησιμοποιεί σεντ περίπλοκων εξισώσεων που μπορούν όμως να εφαρμοστούν μόνο στο συγκεκριμένο πρόβλημα που του έχει δοθεί.

Άλλο ένα πλεονέκτημα των νευρωνικών δικτύων σε σχέση με τους συμβατικές υπολογιστικές μηχανές είναι ο υψηλός βαθμός παραλληλίας κατά την εκτέλεση των διαδικασιών. Στον ανθρώπινο νευρικό σύστημα ενός ενηλίκου χιλιάδες νευρώνες πεθαίνουν κάθε χρόνο. Παρόλα αυτά η λειτουργία του εγκεφάλου παραμένει ανεπηρέαστη, λόγω της υψηλού βαθμού παραλληλίας που παρατηρείται στα βιολογικά νευρωνικά δίκτυα. Το ίδιο χαρακτηριστικό μπορεί να παρατηρηθεί και στα τεχνητά νευρωνικά δίκτυα.

Αν και πρόκειται για μία σχετικά καινούργια ιδέα στο χώρο των υπολογιστών καθώς μεγάλη ανάπτυξη γνώρισαν από τη δεκαετία του 80 και έπειτα, έχουν αναπτυχθεί ήδη αρκετά λογισμικά για τη δημιουργία και τη διαχείριση των νευρωνικών δικτύων. Ένα από τα κυριότερα αποτελεί και το τμήμα του toolbox του προγράμματος Matlab της εταιρείας Mathworks το οποίο και χρησιμοποιείται σε αυτή τη διπλωματική.

Συνοψίζοντας μπορεί να ειπωθεί ότι ο ενθουσιασμός που επικρατεί γύρω από τα τεχνητά νευρωνικά δίκτυα έχει να κάνει με το γεγονός ότι με μία χονδρική αναπαράσταση της λειτουργίας του ανθρώπινου νευρικού συστήματος, επιτρέπει με σχετικά απλό προγραμματισμό την επίλυση περίπλοκων προβλημάτων. Αυτό

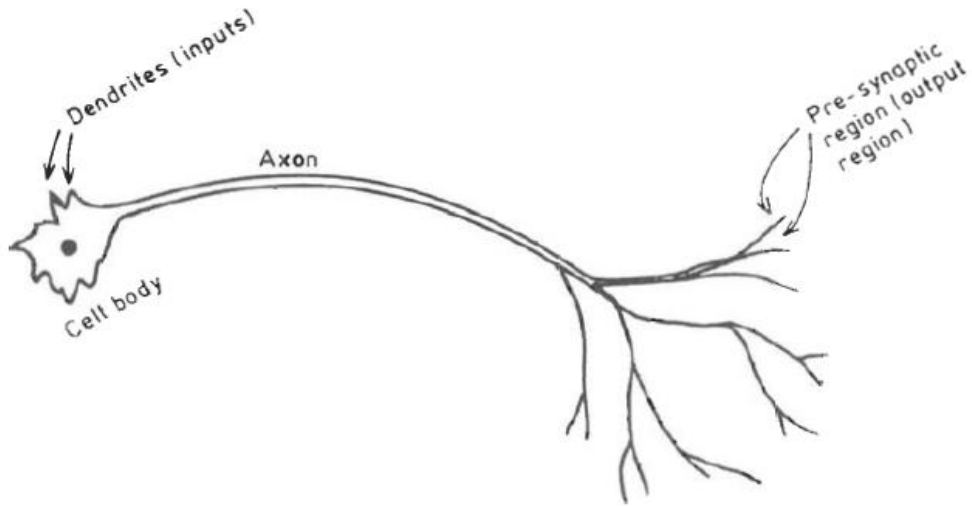
επιτυγχάνεται με μία αυτοοργάνωση του δικτύου με τέτοιο τρόπο ώστε πολλές φορές να μην απαιτείται επαναπρογραμματισμός του δικτύου για την επίλυση κάποιου άλλου προβλήματος, καθώς ο 'εκπαιδευτής' του δικτύου του επιτρέπει να προσαρμοστεί στο πρόβλημα χωρίς να είναι απαραίτητη η αναλυτική περιγραφή του.

1.2 Πηγή έμπνευσης- Τα βασικά χαρακτηριστικά του βιολογικού νευρώνα

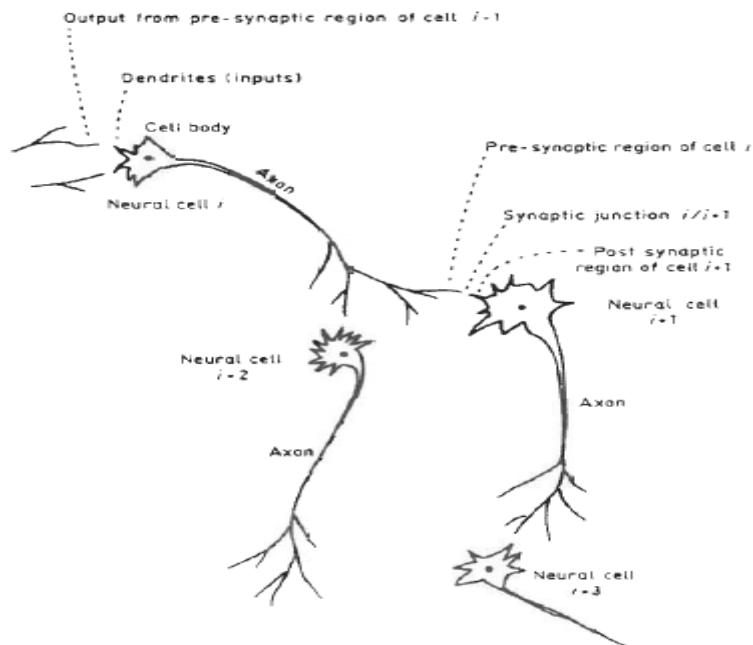
Όπως αναφέρθηκε και στην εισαγωγή πηγή έμπνευσης για τη δημιουργία των τεχνητών νευρωνικών δικτύων αποτελεί η δομή του κεντρικού νευρικού συστήματος των έμβιων οργανισμών.

Το νευρικό σύστημα ενός έμβιου οργανισμού αποτελείται από νευρώνες οι οποίοι συνδέονται μεταξύ τους. Στο κυρίως κυτταρικό σώμα του νευρώνα (cell body) περιλαμβάνεται και ο πυρήνας του και είναι εκεί που λαμβάνουν χώρα οι περισσότερες 'υπολογιστικές' διαδικασίες. Επίσης το περιβάλλουν μία η περισσότερες απολήξεις, οι λεγόμενοι «δενδρίτες» που συλλέγουν τα σήματα που στέλνονται στο κύτταρο. Το σήμα μεταδίδεται από τον ένα νευρώνα στον άλλο ως ηλεκτρικοί παλμοί διαμέσου του άξονα του νευρώνα μέσω μίας ηλεκτροχημικής διαδικασίας που δεν θα αναλυθεί περαιτέρω στα πλαίσια αυτής της διπλωματικής. Αυτή η διαδικασία μεταφοράς ξεκινάει από το κυτταρικό σώμα, διασχίζει τον άξονα και έπειτα μέσα από συναπτικές απολήξεις στο τέλος του άξονα μέσω ενός πολύ στενού συναπτικού διαστήματος καταλήγει στους δενδρίτες και το κυρίως σώμα του επόμενου νευρώνα με μία μέση ταχύτητα ίση με 3 m/sec.

Στην εικόνας που ακολουθούν φαίνεται η δομή ενός βιολογικού νευρώνα, οι συνδέσεις μεταξύ διαφορετικών νευρώνων καθώς και το συναπτικό κενό μεταξύ των απολήξεων του νευρώνα που στέλνει το σήμα και των δενδριτών του νευρώνα που το λαμβάνει.

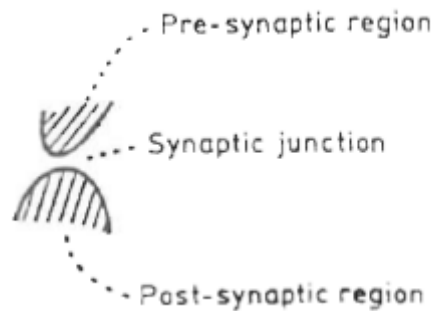


Εικόνα 1.1: Η δομή ενός βιολογικού νευρώνα



Εικόνα 1.2: Συνδέσεις μεταξύ των νευρώνων

Εδώ αξίζει να αναφερθεί ότι δεν είναι όλες οι συνάψεις ίσες μεταξύ τους. Μερικές έχουν μεγαλύτερη προτεραιότητα από κάποιες άλλες. Επίσης μερικές είναι διεγερτικού χαρακτήρα (δηλαδή μεταφέρουν το σήμα) και άλλες είναι ανασταλτικού χαρακτήρα (απαγορεύουν τη διέλευση του σήματος). Ο τρόπος με τον οποίο συνδέονται οι νευρώνες και επηρεάζουν με διαφορετικό «βάρος» τη μεταφορά του σήματος αποτελεί βασικό παράγοντα στη διαμόρφωση και των τεχνητών νευρωνικών δικτύων.



Εικόνα 1.3: Το συναπτικό διάκενο μεταξύ των νευρώνων

1.3 Ιστορική αναδρομή

Τα νευρωνικά δίκτυα άρχισαν να αναπτύσσονται ιδιαίτερα από τις αρχές τις δεκαετίας του 80. Παρόλα αυτά αυτό το επιστημονικό πεδίο είχε κατοχυρωθεί πριν ακόμα από τη δημιουργία των υπολογιστών.

Το πρώτο τεχνητό νευρωνικό δίκτυο δημιουργήθηκε από τον νευροφυσιολόγο Warren McCulloch το 1943. Αλλά η διαθέσιμη τεχνολογία εκείνη την εποχή δεν του επέτρεψε να πραγματοποιήσει κάποια περαιτέρω ανάπτυξη στο εύρημα του.

Γενικά η ιστορία των νευρωνικών δικτύων μπορεί να διαχωριστεί στις εξής περιόδους:

- **Πρώτες προσπάθειες:** Υπήρχαν κάποιες αρχικές αναπαραστάσεις χρησιμοποιώντας απλή λογική. Οι McCulloch και Pits δημιούργησαν μοντέλα νευρωνικών δικτύων βασιζόμενοι στις γνώσεις τους στην νευρολογία. Το δίκτυό τους βασιζόταν σε απλούς νευρώνες που θεωρούνταν δυαδικές συσκευές με σταθερές τιμές κατώτατων ορίων. Μία ακόμα προσπάθεια έγινε με τη χρήση προσομοίωσης σε υπολογιστή. Αυτό πραγματοποιήθηκε από δύο ομάδες τους Farley και Clark το 1954 και τους Haibit και Duda το 1956. Η πρώτη ομάδα διατηρούσε επαφές με επιστήμονες της νευρολογίας στο πανεπιστήμιο McGill, τους οποίους και συμβουλευόταν. Αυτή η αλληλεπίδραση καθιέρωσε μία διεπιστημονική τάση στον τομέα των νευρωνικών δικτύων που συνεχίζεται μέχρι σήμερα
- **Ελπιδοφόρας και Αναπτυσσόμενη Τεχνολογία:** Σε αυτή την περίοδο ψυχολόγοι και μηχανικοί συνέβαλλαν παράλληλα στην ανάπτυξη των δικτύων. Ο Rosenblatt προκαλεί τη στρέψη του ενδιαφέροντος προς αυτό το πεδίο με την ανακάλυψη του Αισθητήρα (Perceptron) το 1958. Ο Αισθητήρας είχε τρία στρώματα νευρώνων και μπορούσε να συνδέσει τη μία δοθείσα είσοδο με μία τυχαία μονάδα εξόδου.

Έπειτα ακολούθησαν και άλλες τέτοιες ανακαλύψεις όπως το ADALINE από τους Widrow και Hoff, το οποίο ήταν μία αναλογικού τύπου ηλεκτρονική συσκευή της οποίας η μέθοδος εκμάθησης διέφερε από τον Αισθητήρα, καθώς χρησιμοποιούσε τον κανόνα του ελάχιστου μέσου τετραγωνικού σφάλματος (LMS rule).

- **Η περίοδος της δυσφήμισης** : Κατά την οποία μετά από την έκδοση ενός βιβλίου από τους Minsky και Papert στο οποίο τόνιζαν τους περιορισμούς ενός Αισθητήρα απλού στρώματος μειώθηκε δραματικά η χρηματοδότηση για την έρευνα στα νευρωνικά δίκτυα και προκλήθηκε προκατάληψη απέναντι σε αυτόν τον τομέα και τις δυνατότητες των δικτύων.
- **Καινοτομίας** : Παρά την μείωση της χρηματοδότησης, αρκετή ερευνητές συνέχισαν να ασχολούνται με τη δημιουργία υπολογιστικών μεθόδων που να βασίζονται σε δίκτυα νευρώνων για προβλήματα όπως η αναγνώριση προτύπων. Μεταξύ των επιστημόνων που συνέβαλλαν στην ανάπτυξη του κλάδου ήταν και οι Grossberg και Gail οι οποίοι ανακάλυψαν τα ART (Adaptive Resonance Theory) δίκτυα βασιζόμενοι σε βιολογικά μοντέλα. Παράλληλα οι Anderson και Kohonen δημιούργησαν παρόμοια δίκτυα. Ο Klopff το 1972 ανέπτυξε μία μέθοδο εκμάθησης στα τεχνητά νευρωνικά δίκτυα, βασιζόμενος στον σε μια βασική αρχή εκμάθησης των νευρώνων την ετερόσταση. Ο Werbos το 1974 δημιούργησε και χρησιμοποίησε τη μέθοδο εκμάθησης back-propagation (ανάδρασης) που πέρασαν αρκετά χρόνια πριν αρχίσει και χρησιμοποιείται ευρέως. Υπήρξαν και αρκετοί άλλοι επιστήμονες οι οποίοι συνέβαλλαν στην ανάπτυξη του κλάδου όπως ήταν οι Amari και Fukushima.
- **Η επανάκαμψη και το σήμερα** : Η πρόοδος στο τέλος του 1970 και στις αρχές του 1980 ήταν σημαντική για την επανεμφάνιση του ενδιαφέροντος για τον τομέα των νευρωνικών δικτύων. Αρκετά βιβλία εκδόθηκαν και συνέδρια πραγματοποιήθηκαν και η ανταπόκριση σε αυτά ήταν αρκετά θετική. Παράλληλα πανεπιστημιακά προγράμματα και μαθήματα άρχισαν να πραγματοποιούνται στα περισσότερα μεγάλα Πανεπιστήμια. Σήμερα η πρόοδος που έχει σημειωθεί σε αυτόν τον κλάδο είναι αρκετή ώστε να προσεγγίσει αρκετά ικανοποιητική χρηματοδότηση. Η πρόοδος πέρα από τις συνηθισμένες εφαρμογές φαίνεται να είναι εφικτή και η έρευνα σε διάφορους τομείς εφαρμογής των νευρωνικών δικτύων αναπτύσσεται με γοργούς ρυθμούς. Ξεκάθαρα η σημερινή εποχή αποτελεί μία μεταβατική περίοδο για την τεχνολογία τη σχετική με τα νευρωνικά δίκτυα.

1.4 Εφαρμογές των τεχνητών νευρωνικών δικτύων.

Οι βασικές λειτουργίες τις οποίες μπορεί να επιτελέσει ένα τεχνητό νευρωνικό δίκτυο είναι τρεις:

1. Προσέγγιση συναρτήσεων, στην οποία με βάση τα δεδομένα που δίνονται το νευρωνικό δίκτυο έχει τη δυνατότητα να προσεγγίσει τη σχέση που συνδέει τις εισόδους με του εξόδους. Μάλιστα έχει αποδειχθεί ότι τα τεχνητά νευρωνικά δίκτυα με ένα κρυφό στρώμα μπορούν να προσεγγίσουν οποιαδήποτε συνεχή συνάρτηση με οποιαδήποτε επιθυμητή ακρίβεια (Hornik et al., 1989 , Hornik, 1991)[3],[4]. Είναι η βασική λειτουργία πάνω στην οποία βασίζεται και η δυνατότητα των νευρωνικών δικτύων για πρόβλεψη χρονοσειρών. Ένα χαρακτηριστικό παράδειγμα είναι η χρήση νευρωνικών δικτύων για την πρόβλεψη της τουριστικής κίνησης στις Βαλεαρίδες νήσοι (Palmer et al., 2006)[5].
2. Αναγνώριση προτύπων, κατά την οποία με βάση τα στοιχεία που δίνονται για ένα πρότυπο το δίκτυο μπορεί να προχωρήσει στην αναγνώριση και τον χαρακτηρισμό του. Σε αυτή την περίπτωση κατά την προπόνηση του δικτύου παρέχονται σε αυτό πρότυπα με τα χαρακτηριστικά τους και το σωστό διαχωρισμό τους και το δίκτυο μέσα από τη διαδικασία της εκμάθησης μπορεί να καταστεί ικανό να αναγνωρίζει άγνωστα σε αυτό πρότυπα βάσει κάποιων χαρακτηριστικών τους. Παράδειγμα για αυτό αποτελεί η χρήση των νευρωνικών δικτύων για την αναγνώριση των πνευμόνων που πάσχουν από καρκίνο(Kuruvilla and Gunavathi, 2013) [6].
3. Διαχωρισμός δεδομένων, κατά την οποία το δίκτυο διαχωρίζει σετ δεδομένων που του δίνονται βάση κάποιας ομοιότητας. Έτσι το δίκτυο προπονείται με ένα σωστό σετ δεδομένων ώστε μέσω της διαδικασίας εκμάθησης να έχει τη δυνατότητα να διαχωρίσει σε ομάδες οποιοδήποτε σετ δεδομένων του δοθεί στη συνέχεια βοηθώντας έτσι και στη αναγνώριση δεδομένων που δημιουργούν κάποια υποομάδα με ιδιαίτερα χαρακτηριστικά. Χαρακτηριστικό παράδειγμα αποτελεί ο διαχωρισμός δειγμάτων φυτών σε ομάδες βάση του μήκους των πετάλων τους ή του πλάτους του ανθού τους.

Έτσι με βάση αυτές τις τρεις βασικές λειτουργίες που επιτελούν τα νευρωνικά δίκτυα οι εφαρμογές τους στη σημερινή εποχή εμπίπτουν σε πολλές κατηγορίες. Οι σημαντικότερες από αυτές είναι:

- Οικονομικά: Εκτίμηση τιμής ακινήτων, συμβουλευτική για δανεισμό, αξιολόγηση τιμής εταιρικών ομολόγων, παρακολούθηση δραστηριότητας πιστωτικών καρτών, πρόβλεψη της τιμής του συναλλάγματος

- Ιατρική και Επιστήμες Υγείας : Συμβολή στις ιατρικές διαγνώσεις, εκτίμηση κόστους νοσηλείας, ανάλυση και αναγνώριση καρκινικών όγκων, εκτίμηση της απαραίτητης διαμονής ασθενούς στο νοσοκομείο.
- Βιομηχανία : Πρόβλεψη παραγωγής, πρόβλεψη πωλήσεων, διοίκηση ποιότητας και πρόβλεψη βιομηχανικών διεργασιών όπως είναι οι εκπομπές αερίων από τους κλιβάνους αντικαθιστώντας τον ακριβό εξοπλισμό που χρησιμοποιούνταν προηγουμένως για αυτή την διεργασία.
- Ενέργεια : Πρόβλεψη του παραγόμενου ηλεκτρικού φορτίου, πρόβλεψη της ενεργειακής ζήτησης, εκτίμηση του βραχυπρόθεσμου και μακροπρόθεσμου απαραίτητου ενεργειακού φορτίου, πρόβλεψη τιμών πετρελαίου και φυσικού αερίου.
- Τηλεπικοινωνίες : Αυτόματες υπηρεσίες πληροφοριών, υπηρεσίες μετάφρασης της ομιλίας σε πραγματικό χρόνο, συστήματα επεξεργασίας πληρωμών πελατών, συμπίεση εικόνων και δεδομένων, αναγνώριση φωνής.
- Αεροδιαστημική : Υψηλής απόδοσης συστήματα αυτομάτου πιλότου, συστήματα αναπαράστασης διαδρομής πτήσης, συστήματα ελέγχου του αεροσκάφους, εντοπισμός προβλημάτων σε μέρη του αεροσκάφους.

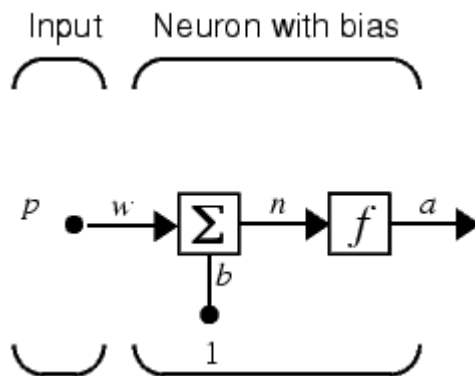
ΚΕΦΑΛΑΙΟ 2

ΤΑ ΤΕΧΝΗΤΑ ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

2.1 Βασικά δομικά στοιχεία τεχνητών νευρωνικών δικτύων

2.1.1 Ο απλός νευρώνας (απλής εισόδου-απλής εξόδου)

Το βασικό δομικό στοιχείο των νευρωνικών δικτύων είναι ο νευρώνας (neuron). Στην παρακάτω εικόνα φαίνεται η δομή ενός τέτοιου νευρώνα απλής εισόδου- απλής εξόδου [6].



Εικόνα 2.1: Δομή νευρώνα απλής εισόδου- απλής εξόδου

Το p αποτελεί την είσοδο του νευρώνα (στην συγκεκριμένη περίπτωση μονόμετρο μέγεθος) η οποία και πολλαπλασιάζεται με το συναπτικό βάρος w . Τα βάρη w ανάλογα με την τιμή τους καθορίζουν τη σημασία της συγκεκριμένης εισόδου του νευρώνα για το αποτέλεσμα που αυτός θα παράγει. Η τιμή του βάρους μπορεί να είναι θετική ή αρνητική σε αντιστοιχία με την επιταχυντική ή επιβραδυντική λειτουργία της σύναψης.

Το γινόμενο $w p$ εισέρχεται στον αθροιστή (Σ) όπου και αθροίζεται με το γινόμενο μίας άλλης εισόδου, η οποία έχει πάντα ως τιμή τη μονάδα, επί το αντίστοιχο βάρος της b . Η τελευταία αυτή είσοδος στον αθροιστή ονομάζεται πόλωση (bias). Η πόλωση έχει την ιδιότητα να αυξάνει ή να μειώνει την είσοδο στη συνάρτηση μεταφοράς όταν είναι θετική ή αρνητική αντίστοιχα.[8] Το άθροισμα που προκύπτει από τον αθροιστή και το οποίο ονομάζεται και τοπικό πεδίο του νευρώνα είναι το εξής:

$$n = w \times p + b$$

Αυτό το άθροισμα αποτελεί την είσοδο στην συνάρτηση ενεργοποίησης f η οποία είναι και εκείνη που τελικά δίνει την έξοδο του δικτύου. Οι συναρτήσεις ενεργοποίησης και τα είδη τους θα αναλυθούν παρακάτω. Οπότε το αποτέλεσμα στην έξοδο του νευρωνικού δικτύου είναι της μορφής,

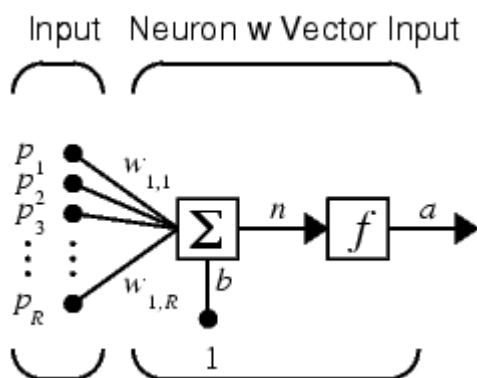
$$a = f(n) = f(w \times p + b)$$

Εδώ αξίζει να σημειώσουμε ότι οι παράμετροι b , w είναι μεγέθη τα οποία μπορούν να ρυθμιστούν. Έτσι η βασική ιδέα στα τεχνητά νευρωνικά δίκτυα είναι αυτές οι δύο παράμετροι να διαμορφώνονται ώστε το δίκτυο να παράγει τα ζητούμενα

αποτελέσματα. Αυτή η ρύθμιση των συγκεκριμένων παραμέτρων μπορεί να πραγματοποιηθεί μέσα από τη διαδικασία της εκπαίδευσης του δικτύου.

2.1.2 Απλός νευρώνας με διανυσματική είσοδο

Ένας νευρώνας μπορεί να δέχεται παραπάνω από μία εισόδους οι οποίες να παρέχονται σε αυτόν με τη μορφή διανύσματος \mathbf{p} . Η δομή του φαίνεται στην παρακάτω εικόνα [6] :



Εικόνα 2.2: Νευρώνας με είσοδο διάνυσμα

Η λειτουργία του νευρώνα με διανυσματική είσοδο είναι όμοια με τη λειτουργία του νευρώνα απλής εισόδου με τη διαφορά ότι πλέον το \mathbf{p} συμβολίζει το διάνυσμα (πίνακας στήλη) που δέχεται στην είσοδό του ο νευρώνας και το \mathbf{w} το διάνυσμα (πίνακας γραμμή) των συναπτικών βαρών. Οπότε ο πολλαπλασιασμός των τιμών της εισόδου p_1, p_2, \dots, p_R με τις τιμές των αντίστοιχων βαρών $w_{1,1}, w_{1,2}, \dots, w_{1,R}$ μας δίνει το γινόμενο $\mathbf{w}\mathbf{p}$ που είναι η μία είσοδος στον αθροιστή.

Όσον αφορά το συμβολισμό που χρησιμοποιείται για να περιγράψει τα συναπτικά βάρη ο πρώτος δείκτης δηλώνει το στρώμα στο οποίο ανήκει ο νευρώνας που καταλήγει το βάρος ενώ ο δεύτερος δείκτης δηλώνει την είσοδο που συνδέεται με το νευρώνα μέσω του συγκεκριμένου συναπτικού βάρους.

Ομοίως με τον νευρώνα απλής εισόδου στον αθροιστή το γινόμενο $\mathbf{w}\mathbf{p}$ αθροίζεται με την πόλωση, που είναι το γινόμενο μιας εισόδου ίση με τη μονάδα με ένα προσαρμοζόμενο βάρος b . Το αποτέλεσμα στην έξοδο του αθροιστή περιγράφεται από την παρακάτω εξίσωση:

$$n = \mathbf{w} \times \mathbf{p} + b = (w_{1,1} \times p_1 + w_{1,2} \times p_2 + \dots + w_{1,R} \times p_R) + b$$

Έπειτα το τοπικό πεδίο το νευρώνα εισέρχεται στη συνάρτηση ενεργοποίησης από όπου και υπολογίζεται η τελική έξοδος του δικτύου ως εξής:

$$\alpha = f(n) = f(\mathbf{w} \times \mathbf{p} + b)$$

Μία επιπλέον παρατήρηση είναι ότι ένας νευρώνας μπορεί να έχει και περισσότερες της μίας εξόδους. Διευκρινίζεται πάντως ότι η μοναδικότητα της εξόδου του νευρώνα έχει να κάνει με την τιμή της εξόδου και όχι με το πόσες εξοδοι υπάρχουν στο δεξιό μέρος του νευρώνα. [8]. Οπότε ένας νευρώνας μπορεί να έχει πολλές εξόδους που να έχουν την ίδια ή διαφορετικές τιμές.

2.2 Είδη συναρτήσεων ενεργοποίησης

Η συνάρτηση ενεργοποίησης, που συμβολίζεται με $f(n)$, προσδιορίζει την έξοδο του δικτύου με τη χρήση του τοπικού πεδίου n του νευρώνα. [8] Η σημασία της συνάρτησης ενεργοποίησης έγκειται και στο γεγονός ότι εισάγει το στοιχείο της μη γραμμικότητας στο δίκτυο πράγμα πολύ σημαντικό ώστε το δίκτυο να επιτελέσει χωρίς δυσκολίες τη λειτουργία της προσέγγισης συναρτήσεων. Οι σημαντικότερες συναρτήσεις ενεργοποίησης είναι οι εξής.

1. Η Συνάρτηση κατωφλίου ή Βηματική συνάρτηση:

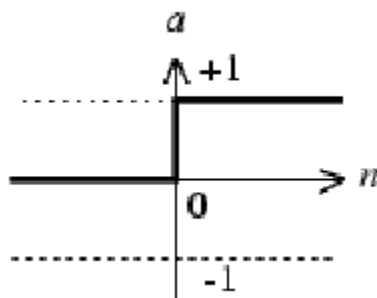
Ο τύπος που την περιγράφει είναι ο ακόλουθος:

$$a = f(n) = \begin{cases} 0, & n < 0 \\ 1, & n \geq 0 \end{cases}$$

με n να είναι το τοπικό πεδίο του νευρώνα.

Η λειτουργία της είναι απλή. Όταν το τοπικό πεδίο έχει αρνητική τιμή τότε η συνάρτηση ενεργοποίησης δίνει την τιμή 0 στην έξοδο του νευρώνα. Αντίθετα όταν το τοπικό πεδίο είναι θετικό ή μηδέν τότε η συνάρτηση ενεργοποίησης δίνει τιμή ίση με τη μονάδα στην έξοδο του νευρώνα.

Η συγκεκριμένη συνάρτηση χρησιμοποιήθηκε στον πρώτο νευρώνα από τους McCulloch και Pitts. Συνήθως χρησιμοποιείται στους Αισθητήρες, που είναι μία από τις πρώτες αρχιτεκτονικές νευρωνικών δικτύων. Η γραφική της απεικόνιση [7] φαίνεται στην παρακάτω εικόνα.

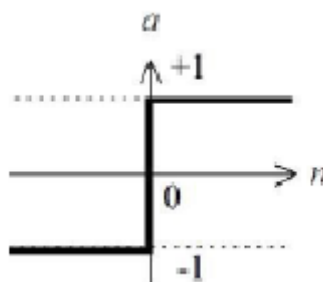


Εικόνα 2.3: Η βηματική συνάρτηση ενεργοποίησης

Παραλλαγή της συγκεκριμένης αποτελεί και η συνάρτηση κατωφλίου με όρια τις τιμές -1 και +1 που περιγράφεται από τον τύπο

$$a = f(n) = \begin{cases} -1, & n < 0 \\ 1, & n \geq 0 \end{cases}$$

και η γραφική της παράσταση [7] είναι η εξής:



Εικόνα 2.4: Η Βηματική συνάρτηση ενεργοποίησης [1, -1]

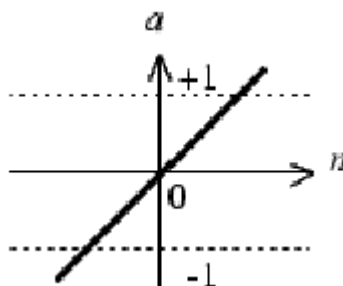
2. Η Γραμμική συνάρτηση:

Η συγκεκριμένη συνάρτηση ενεργοποίησης περιγράφεται από τον παρακάτω τύπο:

$$a = f(n) = n$$

Όπως φαίνεται και από τον τύπο της, η συγκεκριμένη συνάρτηση δίνει ως έξοδο του νευρώνα την τιμή του τοπικού του πεδίου. Χρησιμοποιείται συνήθως ως συνάρτηση ενεργοποίησης στα περισσότερα δίκτυα που επιτελούν προσέγγιση συναρτήσεων και πρόγνωση.

Η γραφική της απεικόνιση [7] φαίνεται στη συνέχεια.



Εικόνα 2.5: Γραμμική συνάρτηση ενεργοποίησης

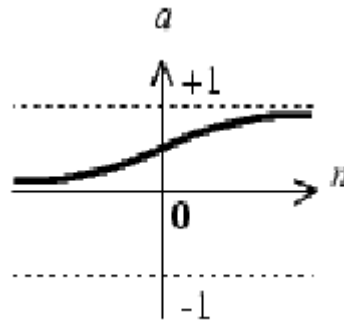
3. Η Σιγμοειδής συνάρτηση:

Η Σιγμοειδής συνάρτηση είναι μακρά η πιο ευρέως χρησιμοποιούμενη από τις συναρτήσεις ενεργοποίησης. Η γραφική της απεικόνισης είναι μορφής S, είναι μία αύξουσα διαφορίσιμη συνάρτηση. Μία έκφραση της σιγμοειδούς συνάρτησης είναι η λογιστική (logistic) συνάρτηση ενεργοποίησης της οποίας το εύρος των αποτελεσμάτων βρίσκεται στην περιοχή τιμών [1,0]. Ο τύπος της λογιστικής συνάρτησης είναι:

$$a = f(n) = \frac{1}{1 + e^{-n}}$$

Όπως φαίνεται και από τον τύπο της η συγκεκριμένη συνάρτηση ανάλογα με την τιμή του τοπικού πεδίου του νευρώνα δίνει στην έξοδο της τιμές που

βρίσκονται μέσα στα όρια $[1,0]$. Η γραφική της απεικόνιση [7] φαίνεται παρακάτω:

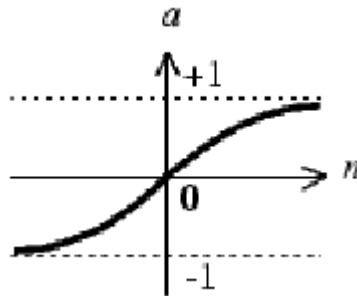


Εικόνα 2.6: Η λογιστική σιγμοειδής συνάρτηση

Μία άλλη έκφραση της σιγμοειδούς συνάρτησης είναι και η εφαπτομενική σιγμοειδής (tan-sigmoid) συνάρτηση ενεργοποίησης. Εκφράζεται από τον τύπο:

$$a = f(n) = \frac{e^n - e^{-n}}{e^n + e^{-n}}$$

και η οποία έχει την εξής γραφική αναπαράσταση [7]:



Εικόνα 2.7: Η εφαπτομενική σιγμοειδής συνάρτηση

2.3 Βασικές αρχιτεκτονικές νευρωνικών δικτύων

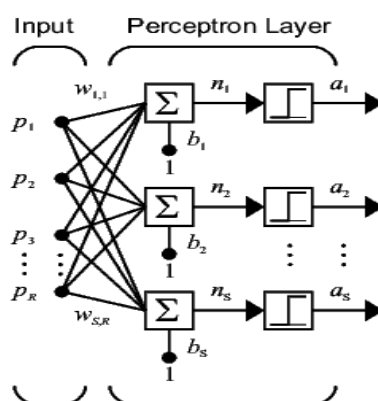
Συνήθως ο τρόπος που δομείται ένα νευρωνικό δίκτυο συνδέεται πολύ στενά με τον κανόνα που θα χρησιμοποιηθεί μετέπειτα για την εκπαίδευση του δικτύου. Έτσι συνήθως οι ονομασίες που παίρνουν τα δίκτυα προέρχονται και από τον τρόπο με τον οποίο εκπαιδεύονται. Παρόλα αυτά θα μπορούσαμε να προχωρήσουμε σε ένα βασικό διαχωρισμό ορισμένων αρχιτεκτονικών δικτύων με κριτήρια κυρίως τον τρόπο ροής της πληροφορίας καθώς και των αριθμό των χρησιμοποιούμενων στρωμάτων (layers). Εδώ να σημειώσουμε ότι στρώμα ενός νευρωνικού δικτύου είναι ένα σύνολο S νευρώνων που τοποθετούνται και λειτουργούν παράλληλα και ανεξάρτητα ο ένας από τον άλλο, δεχόμενοι το σήμα R εισόδων (με R όχι απαραίτητα ίσο του S). Με βάση λοιπόν τα παραπάνω κριτήρια οι βασικότερες αρχιτεκτονικές δικτύων ακολουθούν παρακάτω.[8]

2.3.1 Τροφοδοτούμενα προς τα εμπρός δίκτυα απλού στρώματος

Στην απλούστερη μορφή ενός νευρωνικού δικτύου έχουμε ένα στρώμα νευρώνων, το οποίο ονομάζεται και στρώμα εξόδου (output layer), που τροφοδοτείται από ένα στρώμα κόμβων εισόδου. Εδώ να επισημάνουμε ότι στην πλειοψηφία των βιβλιογραφικών αναφορών το στρώμα των κόμβων εισόδου δεν υπολογίζεται ως στρώμα του νευρωνικού δικτύου, λογική η οποία θα ακολουθηθεί και στην παρούσα διπλωματική.

Το σήμα μεταφέρεται από το επίπεδο των κόμβων εισόδου στο στρώμα εξόδου, αλλά όχι αντίστροφα. Με άλλα λόγια, το δίκτυο είναι αυστηρώς τροφοδοτούμενο προς τα εμπρός. [8]

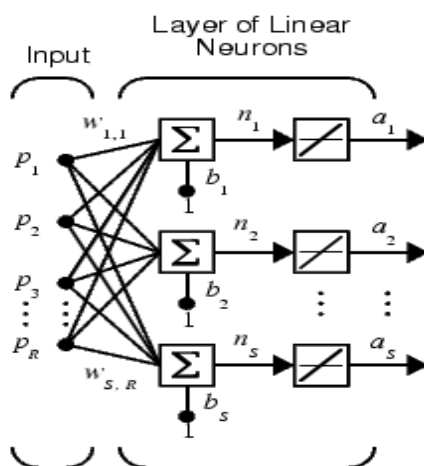
Όπως αναφέρει και ο Bishop στο βιβλίο του: «Δίκτυα απλού στρώματος με βηματικές συναρτήσεις ενεργοποίησης μελετήθηκαν από τον Rosenblatt, ο οποίος τα ονόμασε Αισθητήρες (Perceptrons). Ο Rosenblatt επίσης δημιούργησε πραγματικές εφαρμογές τους, που πραγματοποιούσαν εκμάθηση χρησιμοποιώντας τον αλγόριθμο εκμάθησης του Αισθητήρα (Perceptron Learning Rule). Αυτά τα δίκτυα εφαρμόστηκαν κυρίως σε προβλήματα ταξινόμησης, στα οποία οι είσοδοι ήταν δυαδικές εικόνες απλών σχημάτων».[9] Μία απεικόνιση ενός τέτοιου δικτύου με R εισόδους και S νευρώνες φαίνεται στην εικόνα 2.8.[7]



Εικόνα 2.8: Ο Αισθητήρας (Perceptron)

Εδώ να επισημάνουμε ότι το σήμα που χρησιμοποιείται μετά τον αθροιστή στην συνάρτηση ενεργοποίησης συμβολίζει τη βηματική συνάρτηση.

Άλλο ένα είδος δικτύων τέτοιας μορφής είναι και τα Γραμμικά Φίλτρα (Linear Filters), τα οποία είναι δίκτυα ενός στρώματος που έχουν ως συνάρτηση ενεργοποίησης τη Γραμμική Συνάρτηση. Στην εικόνα 2.9 [7] φαίνεται ένα γραμμικό φίλτρο με R εισόδους και S νευρώνες. Ομοίως με προηγουμένως το σύμβολο που χρησιμοποιείται μετά το αθροιστή είναι για τη Γραμμική συνάρτηση ενεργοποίησης.



Εικόνα 2.9: Το Γραμμικό Φίλτρο (Linear Filter)

Για να παρακαμφθούν οι αδυναμίες των δικτύων ενός στρώματος να αντιμετωπίσουν κάποιες περιπτώσεις οι δημιουργοί τους προχώρησαν και στην ανάπτυξη δικτύων πολλών στρωμάτων που θα περιγραφούν παρακάτω.

2.3.2 Τροφοδοτούμενα προς τα εμπρός δίκτυα πολλαπλών στρωμάτων

Αυτή η κατηγορία δικτύων περιλαμβάνει περισσότερα από ένα κρυφά στρώματα (hidden layers) υπολογιστικών νευρώνων, οι οποίοι ονομάζονται και κρυφοί νευρώνες (hidden neurons). Κρυφά στρώματα είναι εκείνα τα οποία βρίσκονται μεταξύ του επιπέδου των κόμβων εισόδου καθώς και του στρώματος εξόδου.

Με την προσθήκη επιπλέον στρωμάτων νευρώνων το δίκτυο μπορεί να οδηγηθεί στην πραγματοποίηση πιο σύνθετων υπολογισμών και άρα στην εξαγωγή υψηλότερης τάξης αποτελεσμάτων. Η ικανότητα αυτή των δικτύων είναι ιδιαίτερα χρήσιμη όταν οι κόμβοι στο επίπεδο εισόδου είναι πολλοί.

Μάλιστα δίκτυα τα οποία έχουν δύο στρώματα νευρώνων (1 κρυφό στρώμα και ένα στρώμα εξόδου) έχουν τη δυνατότητα να προσεγγίσουν μία οποιαδήποτε συνεχής συνάρτηση σε οποιαδήποτε επιθυμητή ακρίβεια. (Hornik et al., 1989 ; Hornik, 1991).

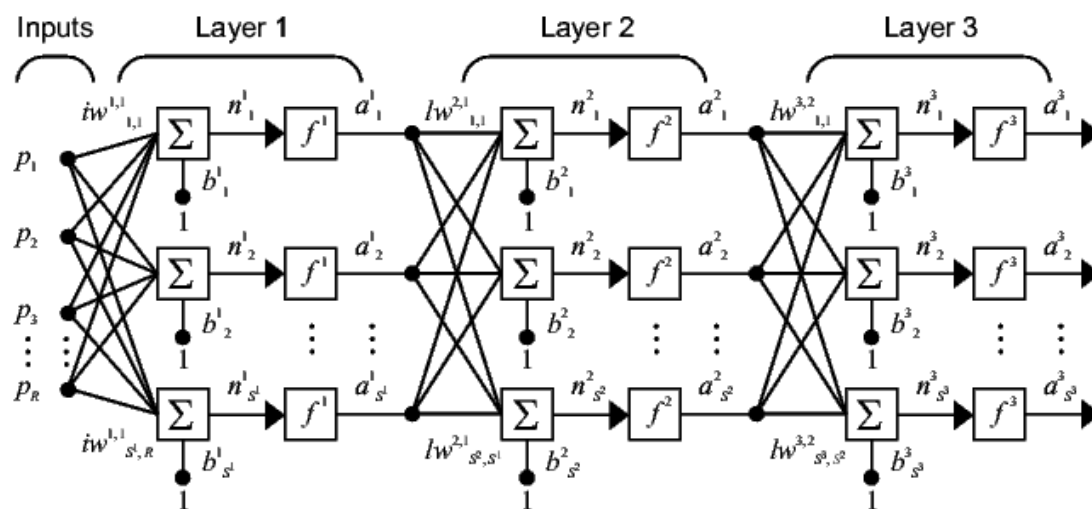
Η λειτουργία ενός δικτύου πολλαπλών στρωμάτων έχει ως εξής. Αρχικά το επίπεδο με τους κόμβους εισόδου τροφοδοτεί το πρώτο κρυφό στρώμα με το διάνυσμα εισόδου, ενεργοποιώντας το. Έπειτα πραγματοποιούνται οι υπολογισμοί στους νευρώνες του πρώτου κρυφού στρώματος και παράγονται σήματα εξόδου. Αυτά τα

σήματα εξόδου χρησιμοποιούνται ως εισοδοί στο δεύτερο κρυφό στρώμα νευρώνων που με τη σειρά του θα τροφοδοτήσει το επόμενο στρώμα. Η διαδικασία συνεχίζεται με τον ίδιο τρόπο μέχρι το σήμα να φτάσει στο στρώμα εξόδου, από το οποίο εξάγεται η απόκριση του δικτύου στο σήμα εισόδου που του δόθηκε μέσω των κόμβων στο επίπεδο εισόδου.

Αυτά τα δίκτυα καλούνται Πολυστρωματικοί Αισθητήρες (Multilayer Perceptron), όπως αναφέρει και ο Bishop [9], ακόμα και αν οι συναρτήσεις ενεργοποίησής τους είναι Σιγμοειδής. Για τέτοια δίκτυα με διαφορίσιμες συναρτήσεις ενεργοποίησης έχει ανακαλυφθεί η μέθοδος της οπίσθιας διάδοσης του σφάλματος (error back-propagation) που χρησιμοποιείται στη διαδικασία εκπαίδευσης του δικτύου.

Ένα τέτοιο δίκτυο πολλαπλών στρωμάτων καλείται πλήρως συνδεδεμένο (fully connected) όταν κάθε νευρώνας ενός κρυφού στρώματος είναι συνδεδεμένος με όλους τους νευρώνες του επόμενου στρώματος. Αν δεν ισχύει κάτι τέτοιο τότε το δίκτυο καλείται μερικώς συνδεδεμένο (partially connected).

Στο παρακάτω σχήμα φαίνεται η δομή ενός δικτύου τριών στρωμάτων (δύο κρυφά και ένα εξόδου). Το δίκτυο αυτό έχει R εισόδους και είναι πλήρως συνδεδεμένο καθώς κάθε νευρώνας των κρυφών στρωμάτων είναι συνδεδεμένος με όλους τους νευρώνες του επόμενου στρώματος

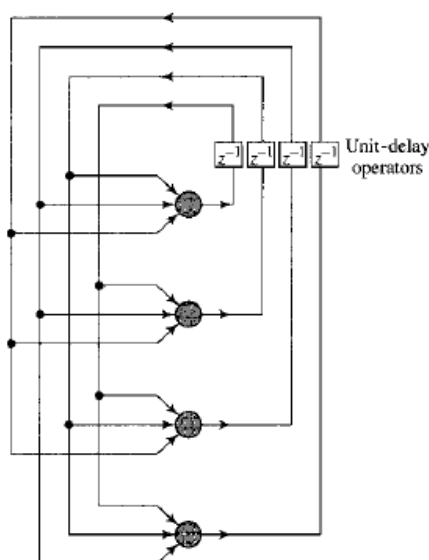


Εικόνα 2.10: Τροφοδοτούμενο προς τα εμπρός δίκτυο πολλαπλών στρωμάτων

Όσον αφορά τώρα την επεξήγηση των δεικτών που χρησιμοποιούνται στα iw, lw που είναι οι πίνακες των βαρών που συνδέουν το επίπεδο εισόδου με το πρώτο κρυφό στρώμα και τα κρυφά στρώματα μεταξύ τους αντίστοιχα. Από τους δύο κάτω δείκτες ο πρώτος δηλώνει το νευρώνα προορισμού και ο δεύτερος το νευρώνα (ή τον κόμβο) εισόδου προέλευσης. Αντίστοιχα από τους δύο πάνω δείκτες ο πρώτος δηλώνει το στρώμα προορισμού και ο δεύτερος το στρώμα προέλευσης (για το επίπεδο των κόμβων εισόδου χρησιμοποιείται ο δείκτης i).

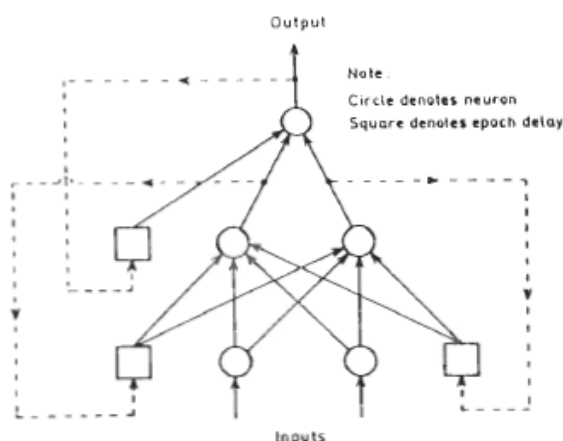
2.3.3 Αναδρομικά δίκτυα

Ένα αναδρομικό δίκτυο διαφέρει από ένα τροφοδοτούμενο προς τα εμπρός δίκτυο στο γεγονός ότι το πρώτο έχει τουλάχιστον ένα βρόγχο ανάδρασης. [8] Ως παράδειγμα για ένα τέτοιο δίκτυο θα μπορούσε να αναφερθεί ένα δίκτυο ενός στρώματος στο οποίο η έξοδος κάθε νευρώνα τροφοδοτεί την είσοδο όλων των υπόλοιπων νευρώνων στο ίδιο στρώμα όπως φαίνεται και στην παρακάτω εικόνα στην οποία ο κάθε νευρώνας συμβολίζεται με ένα κύκλο:[8]



Εικόνα 2.11: Δίκτυο απλού στρώματος με ανάδραση

Ομοίως με προηγούμενως υπάρχουν τα πλήρως αναδρομικά δίκτυα τα οποία έχουν την ιδιότητα κάθε στρώμα να ανατροφοδοτεί το αμέσως προηγούμενο από αυτό στρώμα όπως φαίνεται και στην παρακάτω εικόνα: [1]



Εικόνα 2.12: Πλήρως αναδρομικό δίκτυο

Η παρουσία των βρόγχων ανάδρασης έχει μία σημαντική επίδραση στη διαδικασία εκπαίδευσης του δικτύου και παραπέρα στην απόδοσή του. Οι βρόγχοι ανάδρασης περιλαμβάνουν την χρησιμοποίηση στοιχείων καθυστέρησης (delay elements) που στο σχήμα έχουν σημειωθεί με z^{-1} που έχουν ως αποτέλεσμα μία μη γραμμική

δυναμική συμπεριφορά του δικτύου με δεδομένο ότι οι συναρτήσεις ενεργοποίησης που χρησιμοποιούνται είναι μη γραμμικές.

Δεν θα επικεντρωθούμε περαιτέρω στην ανάλυση της λειτουργίας δικτύων τέτοιας μορφής καθώς δεν γίνεται χρήση τους στα πλαίσια αυτής της διπλωματικής. Για μια πιο αναλυτική περιγραφή αυτών των δικτύων ο αναγνώστης μπορεί να ανατρέξει στην πρώτη επιλογή στις βιβλιογραφικές παραπομπές.

ΚΕΦΑΛΑΙΟ 3

ΕΚΠΑΙΔΕΥΣΗ ΤΕΧΝΗΤΩΝ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ

3.1 Εισαγωγή στην έννοια της εκπαίδευσης ενός νευρωνικού δικτύου

Η εκπαίδευση είναι μία πολύ σημαντική διαδικασία για τα νευρωνικά δίκτυα, καθώς δηλώνει την ικανότητα του δικτύου να μαθαίνει από το περιβάλλον του και του δίνει τη δυνατότητα σταδιακά να βελτιώσει την απόδοσή του. Η διαδικασία της εκπαίδευσης σχετίζεται άμεσα με ένα μετρούμενο μέγεθος του δικτύου το οποίο μεταβάλλεται με το χρόνο και το οποίο επηρεάζει τις μεταβολές που πραγματοποιούνται στα συναπτικά βάρη καθώς και στις πολώσεις του δικτύου. Ιδανικά η απόδοση του δικτύου θα βελτιώνεται και εκείνο θα αποκτά μεγαλύτερη γνώση του περιβάλλοντος μετά από κάθε επανάληψη στην διαδικασία εκμάθησης.

Ειδικότερα για τα τροφοδοτούμενα προς τα εμπρός δίκτυα με πολλαπλά στρώματα νευρώνων (multilayer feed-forward neural networks) με τα οποία και θα ασχοληθούμε στα πλαίσια της συγκεκριμένης διπλωματικής το πρόβλημα της εκπαίδευσης ενός δικτύου μετατρέπεται σε πρόβλημα ελαχιστοποίησης της τιμής μίας συνάρτησης σφάλματος (error function) ή όπως αποκαλείται αλλιώς μίας συνάρτησης απόδοσης (performance function). Αυτή είναι μία συνάρτηση των ρυθμιζόμενων παραμέτρων του δικτύου (δηλαδή των βαρών και των πολώσεων) τα οποία και μεταβάλλονται για να επιτευχθεί ο στόχος της ελαχιστοποίησης.

Ένας ορισμός της εκπαίδευσης δίνεται από το Simon Haykin στο βιβλίο του [8] ο οποίος περιγράφει την εκπαίδευση ως « *μια διαδικασία με την οποία οι ελεύθερες παράμετροι του δικτύου μεταβάλλονται μέσω μιας διαδικασίας διέγερσης από το περιβάλλον στο οποίο το δίκτυο είναι ενσωματωμένο. Το είδος της εκμάθησης καθορίζεται από τον τρόπο με τον οποίο γίνεται η μεταβολή των παραμέτρων.*»

Για την πραγματοποίηση της εκπαίδευσης απαραίτητος είναι ένας αλγόριθμος εκπαίδευσης ο οποίος αποτελείται από ένα σετ οδηγιών που υλοποιούν τον τρόπο με τον οποίο θα εκπαιδευτεί το δίκτυο. Υπάρχει ποικιλία τέτοιων αλγορίθμων οι οποίοι διαφέρουν ο ένας από τον άλλο στον τρόπο με τον οποίο πραγματοποιούνται οι προσαρμογές των συναπτικών βαρών του δικτύου.

Επίσης επισημαίνεται ότι οι περισσότεροι από τους αλγορίθμους εκπαίδευσης που χρησιμοποιούνται σήμερα έχουν καλή απόδοση σε πληθώρα προβλημάτων. Παρόλα αυτά σε διάφορα προβλήματα υπάρχουν κάποιοι αλγόριθμοι που παρουσιάζουν καλύτερα αποτελέσματα. Αυτός είναι και ο λόγος για τον οποίο δεν μπορεί να προταθεί ένας μοναδικός αλγόριθμος εκπαίδευσης ο οποίος θα έχει καθολική ισχύ.

Στις παραγράφους του συγκεκριμένου κεφαλαίου θα γίνει μία συνοπτική περιγραφή του βασικότερου από τους κανόνες εκμάθησης και θα περιγραφεί το πρότυπο της εκμάθησης με επίβλεψη. Τέλος θα αναλυθούν οι δύο αλγόριθμοι εκπαίδευσης δικτύου που ανακαλύφθηκαν πρώτοι οι οποίοι είναι ο αλγόριθμος της οπίσθιας διάδοσης του σφάλματος και ο αλγόριθμος της απότομης καθόδου (gradient/steepest descent) και θα γίνει μία σύντομη περιγραφή νέων πιο αποδοτικών αλγορίθμων εκπαίδευσης των δικτύων.

3.2 Κανόνες εκμάθησης των δικτύων

Οι κανόνες εκπαίδευσης θέτουν τα βασικά πλαίσια για τον τρόπο με τον οποίο ο αλγόριθμος εκπαίδευσης θα αλληλεπιδράσει με το δίκτυο ώστε να πραγματοποιηθεί η εκπαίδευσή του. Ο βασικότερος από αυτούς είναι η εκμάθηση σφάλματος διόρθωσης η οποία χρησιμοποιείται κατά κόρον στα τροφοδοτούμενα προς τα εμπρός πολλαπλών στρωμάτων δίκτυα. Παρακάτω επιχειρείται μία συνοπτική περιγραφή της

3.2.1 Εκμάθηση Σφάλματος- Διόρθωσης

Για να πραγματοποιηθεί ο συγκεκριμένος τρόπος εκμάθησης πρέπει στο δίκτυο να έχει δοθεί κατά τη διαδικασία της εκπαίδευσης ένα σετ εισόδων το οποίο θα συνοδεύεται από τις επιθυμητές από το δίκτυο εξόδους-στόχους, τους οποίους συμβολίζουμε με d . Παρέχοντας στο δίκτυο αυτές τις εισόδους στο επίπεδο των κόμβων εισόδων του και μετά την εκτέλεση των υπολογισμών λαμβάνουμε από το στρώμα εξόδου του δικτύου την απόκριση του σε αυτό το σήμα εισόδου, την οποία και συμβολίζουμε με y . Τότε το σφάλμα εξόδου του δικτύου υπολογίζεται ως εξής:

$$e(ts) = d(ts) - y(ts)$$

Όπου το σύμβολο ts χρησιμοποιείται για να δηλώσει το διακριτό χρόνο ή το στάδιο της επαναληπτικής διαδικασίας στο οποίο βρίσκεται η εκμάθηση.

Με βάση αυτό το παραγόμενο σφάλμα ενεργοποιείται ένας μηχανισμός ο οποίος πραγματοποιεί σταδιακά μεταβολές στα συναπτικά βάρη και στις πολώσεις του δικτύου. Αυτές οι μεταβολές έχουν ως στόχο η απόκριση του δικτύου να προσεγγίσει σταδιακά την επιθυμητή έξοδο-στόχο. Αυτός ο σκοπός πραγματοποιείται με την ελαχιστοποίηση μίας συνάρτησης κόστους ή δείκτη απόδοσης, η οποία εκφράζεται με τον τύπο:

$$E(ts) = \frac{1}{2} \times e(ts)^2$$

Το E είναι η τιμή της ενέργειας του σφάλματος μετά την ts επανάληψη της διαδικασίας. Αυτή η σταδιακή ρύθμιση των βαρών γίνεται μέχρι το σύστημα να φτάσει σε μία σταθερή κατάσταση κατά την οποία οι μεταβολές στα βάρη του γίνονται πολύ μικρές (στην ουσία τα βάρη σταθεροποιούνται). Σε αυτό το σημείο η διαδικασία εκπαίδευσης του δικτύου τερματίζεται.

Οι μεταβολές που πραγματοποιούνται στα βάρη με σκοπό την ελαχιστοποίηση της συνάρτησης κόστους διέπονται από τον λεγόμενο κανόνα δέλτα (delta rule). Για την επεξήγησή του θα θεωρήσουμε δίκτυο ενός νευρώνα k το οποίο ενεργοποιείται από διάνυσμα p . Η μεταβολή Δw_{kj} στο βάρος w_{kj} λόγω της απόκρισης του νευρώνα k εξαιτίας της ενεργοποίησης του από το στοιχείο p_j του διανύσματος p τη χρονική στιγμή ts είναι:

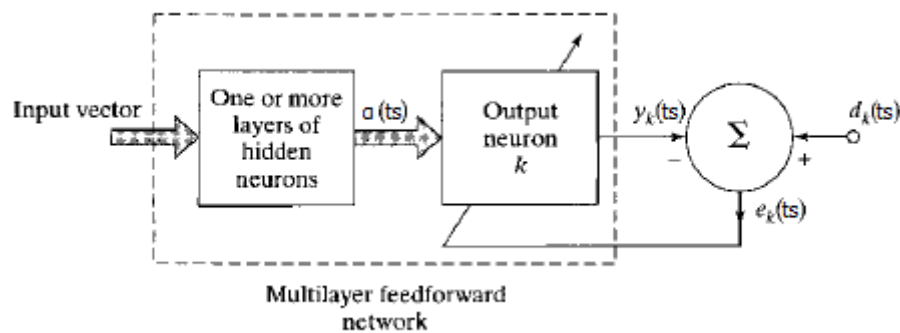
$$\Delta w_{kj}(ts) = \eta \times e_k(ts) \times p_j(ts)$$

Όπου το η είναι μία μεταβλητή η οποία δηλώνει το ρυθμό εκμάθησης καθώς προχωράμε από τη μία επανάληψη της διαδικασίας στην επόμενη.

Επισημαίνεται εδώ ότι ο τρόπος που περιγράφηκε ο κανόνας δέλτα προϋποθέτει ότι το σφάλμα στην έξοδο του δικτύου μπορεί να μετρηθεί. Με λίγα λόγια ότι το αποτέλεσμα που εξάγεται από το νευρώνα k είναι διαθέσιμο (δηλαδή δεν είναι κρυφός νευρώνας). Όπως θα περιγραφεί και παρακάτω υπάρχει και διαθέσιμος τρόπος υπολογισμού της μεταβολής στους κρυφούς νευρώνες. Είναι προφανές ότι μετά τον προσδιορισμό της μεταβολής το νέο βάρος θα είναι:

$$w_{kj}(ts + 1) = w_{kj}(ts) + \Delta w_{kj}(ts)$$

Από αυτά που περιγράφηκαν γίνεται προφανές ότι η εκπαίδευση σφάλματος-διόρθωσης είναι ένα παράδειγμα συστήματος ανάδρασης κλειστού βρόγχου όπως φαίνεται και στην εικόνα 16 [8]. Στην εικόνα παρατηρούμε το σύστημα κατά την εκτέλεση της επανάληψης ts όπου η έξοδος από τον κρυφό νευρώνα εισέρχεται στον νευρώνα εξόδου. Εκείνος παράγει το σήμα εξόδου y_k το οποίο αφαιρείται από την έξοδο στόχο d_k για να υπολογιστεί το σφάλμα e_k το οποίο επηρεάζει τον νευρώνα k μεταβάλλοντας το συναπτικό του βάρος με τον κανόνα δέλτα



Εικόνα 3.1: Το σύστημα κλειστού βρόγχου της εκμάθησης σφάλματος-διόρθωσης

Η ευστάθεια αυτού του συστήματος εξαρτάται άμεσα από την παράμετρο που ορίσαμε πριν ως ρυθμό εκμάθησης η . Μία μικρή τιμή του η μπορεί να καθυστερήσει πολύ τη διαδικασία και να οδηγήσει τον αλγόριθμο σε πολύ αργή σύγκλιση. Αντίθετα πολύ υψηλές τιμές του η μπορούν να προκαλέσουν αστάθεια και ταλαντώσεις των αποτελεσμάτων του αλγορίθμου με αποτέλεσμα τη μη επίτευξη σύγκλισης. Για αυτό απαιτείται ένας προσεκτικός προσδιορισμός της τιμής του η που θα χρησιμοποιηθεί καθώς ο ρόλος του είναι πολύ σημαντικός για την απόδοση του αλγορίθμου.

3.3 Πρότυπα εκπαίδευσης

Πολύ σημαντικό ρόλο κατά τη διαδικασία της εκπαίδευσης ενός δικτύου διαδραματίζει ο τρόπος με τον οποίο το δίκτυο συνδέεται με το περιβάλλον του. Σε αυτή την περίπτωση έχουμε να κάνουμε με τα πρότυπα εκμάθησης τα οποία αναφέρονται σε μοντέλα του περιβάλλοντος βάσει των οποίων το δίκτυο λειτουργεί. Το βασικότερο εξ αυτών και αυτό που χρησιμοποιείται στα πλαίσια της συγκεκριμένης διπλωματικής είναι η εκπαίδευση με επίβλεψη η οποία και θα περιγραφεί συνοπτικά παρακάτω.

3.3.1 Εκπαίδευση με επίβλεψη

Αυτή η μέθοδος εναλλακτικά ονομάζεται και εκμάθηση με δάσκαλο. Στην πράξη μπορούμε να πούμε ότι ο δάσκαλος έχει γνώση του περιβάλλοντος, η οποία γνώση είναι στην ουσία ένα σετ από σωστά παραδείγματα εισόδων εξόδων που είναι επιθυμητό να παράγει το δίκτυο μετά τη διαδικασία εκπαίδευσής του. Το περιβάλλον είναι άγνωστο στο δίκτυο που μας ενδιαφέρει να εκπαιδεύσουμε.

Η διαδικασία αυτή ξεκινάει παρέχοντας ένα διάνυσμα εισόδου από το περιβάλλον και στο δάσκαλο και στο υπό εκπαίδευση δίκτυο. Τότε ο δάσκαλος είναι ικανός να παρέχει στο δίκτυο την επιθυμητή απόκριση σε αυτό το διάνυσμα εισόδου. Αυτή θα είναι και η βέλτιστη λύση την οποία θέλουμε ιδανικά να παράγει το δίκτυο μετά την εκπαίδευση του. Οι παράμετροι του υπό εκπαίδευση δικτύου προσαρμόζονται υπό την επίδραση του διανύσματος εκπαίδευσης και του σφάλματος εξόδου που παράγεται. Όπως αναφέρθηκε και προηγουμένως το σφάλμα εξόδου εξαρτάται από την επιθυμητή έξοδο και από αυτή που πραγματικά παράγει το δίκτυο κατά τα στάδια της εκπαίδευσής του. Αυτή η διαδικασία συνεχίζεται σταδιακά μέχρι το δίκτυο να καταφέρει να έχει τα επιθυμητά αποτελέσματα, δηλαδή να «μιμηθεί» το δάσκαλο. Όταν γίνει αυτό τότε η εκπαίδευση του δικτύου έχει ολοκληρωθεί.

Το πρότυπο εκπαίδευσης με επίβλεψη που μόλις περιγράφηκε συμπληρώνει τη διαδικασία εκμάθησης σφάλματος- διόρθωσης. Όπως αναφέρθηκε και εκεί πρόκειται για ένα κλειστό βρόγχο με ανάδραση μόνο που οι άγνωστες τιμές που παρέχονται στο δίκτυο μετά την εκπαίδευσή του δεν περιλαμβάνονται στο περιβάλλον.

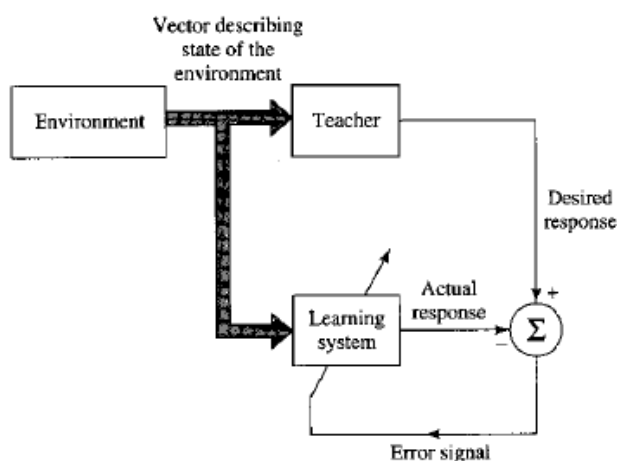
Ως ένα μέτρο μέτρησης της απόδοσης του δικτύου μπορεί να χρησιμοποιηθεί το μέσο τετραγωνικό σφάλμα το οποίο υπολογίζεται ως ο μέσος όρος του αθροίσματος των σφαλμάτων που προκύπτουν μετά από μία εποχή εκπαίδευσης, κατά την οποία παρέχονται στο δίκτυο όλα τα πρότυπα εκπαίδευσης που περιλαμβάνονται στα διανύσματα εισόδου εξόδου που χρησιμοποιούνται ως δάσκαλος του δικτύου.

Το μέσο τετραγωνικό σφάλμα μπορεί να σχεδιαστεί ως συνάρτηση των ελευθέρων παραμέτρων του δικτύου (βάρη) δίνοντας έτσι την επιφάνεια σφάλματος, που παρέχει την τιμή του σφάλματος με συντεταγμένες τα βάρη του δικτύου. Είναι προφανές ότι σε περίπτωση που τα βάρη του δικτύου είναι περισσότερα από δύο δεν είναι δυνατόν να σχεδιαστεί αυτή η επιφάνεια.

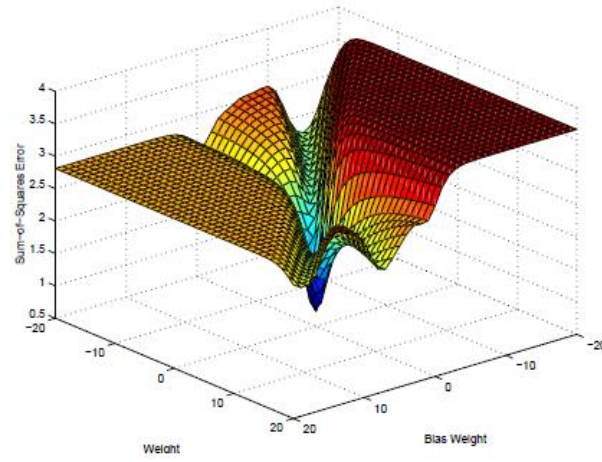
Για κάθε διάνυσμα εισόδου που παρέχεται στο δίκτυο και του οποίου έχουμε την απόκριση του δασκάλου, υπάρχει ένα αντίστοιχο σημείο στην επιφάνεια σφάλματος που δημιουργείται λόγω της απόκρισης του δικτύου. Για να είναι επιτυχής η διαδικασία της εκπαίδευσης δηλαδή το δίκτυο να μαθαίνει από το δάσκαλο θα πρέπει το σημείο αυτό που αναπαριστά τη λειτουργία του δικτύου να κινείται προς περιοχές με πιο χαμηλό σφάλμα μέχρι να φτάσει σε κάποιο τοπικό ελάχιστο. Το ιδανικό θα ήταν το σημείο λειτουργίας να καταφέρει να φτάσει στο ολικό ελάχιστο της επιφάνειας του σφάλματος αλλά λόγω των πολλών κοιλάδων που εκείνη διαθέτει συνήθως «παγιδεύει» το σημείο λειτουργίας σε περιοχές με τοπικά ελάχιστα, με αποτέλεσμα πολλές φορές το δίκτυο να μην έχει την επιθυμητή απόκριση μετά από τη διαδικασία της εκπαίδευσής του.

Αυτή η διαδικασία εύρεσης της διαδρομής προς το χαμηλότερο σημείο γίνεται συνήθως με τον υπολογισμό της κλίσης της επιφάνειας του σφάλματος στο σημείο λειτουργίας. Μάλιστα, στην περίπτωση της εκπαίδευσης με επίβλεψη, το σύστημα ή ο αλγόριθμος εκπαίδευσης μπορεί να χρησιμοποιήσει μία στιγμιαία τιμή του διανύσματος της κλίσης. Αυτό θα έχει ως αποτέλεσμα τη μετακίνηση του σημείου λειτουργίας με τη μορφή του «τυχαίου περιπάτου» το οποίο μπορεί να οδηγήσει στην εύρεση κάποιου τοπικού ελαχίστου ή του ολικού ελαχίστου της επιφάνειας του σφάλματος.

Στην εικόνα 11 [8] περιγράφεται ο βρόγχος εκπαίδευσης του δικτύου με επίβλεψη όπου βλέπουμε πως η έξοδος που παρέχεται από το δάσκαλο επηρεάζει τη συμπεριφορά του δικτύου. Τέλος στην εικόνα 12 [10] υπάρχει μία οπτική περιγραφή της επιφάνειας σφάλματος για ένα δίκτυο με δύο ελεύθερες παραμέτρους.



Εικόνα 3.2: Βρόγχος εκπαίδευσης με επίβλεψη



Εικόνα 3.3: Η επιφάνεια σφάλματος για δίκτυο δύο βαρών στο οποίο έχουν δοθεί για εκπαίδευση δέκα σετ δεδομένων από πραγματικά παραδείγματα

3.4 Αλγόριθμοι εκπαίδευσης νευρωνικών δικτύων

Όπως αναφέρθηκε και σε προηγούμενη παράγραφο του συγκεκριμένου κεφαλαίου, ένας ορισμός της έννοιας του αλγορίθμου εκπαίδευσης σύμφωνα με τον S. Haykin [8] είναι ο εξής:

« Αλγόριθμος εκπαίδευσης είναι ένα προκαθορισμένο σύνολο καλά ορισμένων εντολών που επιλύουν το πρόβλημα της εκπαίδευσης του δικτύου. »

Όπως φαίνεται και από τον ορισμό ο αλγόριθμος εκπαίδευσης παρέχει τις οδηγίες και ενεργοποιεί τη διαδικασία για να εκπαιδευτεί ένα δίκτυο. Δεν υπάρχει κάποιος αλγόριθμος που να έχει καθολική ισχύ και να μπορεί να εφαρμοστεί σε κάθε μορφής πρόβλημα εκπαίδευσης ενός δικτύου παρέχοντας τα βέλτιστα αποτελέσματα. Αντίθετα θα λέγαμε ότι υπάρχουν εργαλεία εκπαίδευσης τα οποία εκφράζονται μέσα από μία ευρεία γκάμα αλγορίθμων εκπαίδευσης το καθένα από τα οποία παρουσιάζει τα δικά του πλεονεκτήματα και μειονεκτήματα ανάλογα με το προς επίλυση πρόβλημα.

Στις υποπαραγράφους τις συγκεκριμένης παραγράφου θα γίνει μία παρουσίαση του προβλήματος εκχώρησης πίστωσης, καθώς και των αλγορίθμων της απότομης καθόδου (gradient descent / steepest descent algorithm) και της οπίσθιας διάδοσης του σφάλματος που αποτελούν τις αρχικές και πιο βασικές μεθόδους εκπαίδευσης ενός δικτύου. Τέλος θα γίνει μία συνοπτική περιγραφή μερικών από τους πιο σύγχρονους αλγορίθμους εκπαίδευσης.

3.4.1 Το πρόβλημα της εκχώρησης πίστωσης

Αρχικά πριν προχωρήσουμε στην ανάλυση του αλγορίθμου της οπίσθιας διάδοσης του σφάλματος (error back-propagation algorithm) θα δώσουμε μία σύντομη περιγραφή του προβλήματος της εκχώρησης πίστωσης (credit assignment) το οποίο και αποτέλεσε έναυσμα για την ανακάλυψη του συγκεκριμένου αλγορίθμου.

Το πρόβλημα της εκχώρησης πίστωσης παρουσιάζεται κατά την εφαρμογή του κανόνα εκμάθησης σφάλματος-διόρθωσης στα τροφοδοτούμενα προς τα εμπρός δίκτυα πολλαπλών στρωμάτων. Το πρόβλημα έγκειται στο γεγονός ότι μόνο στο στρώμα εξόδου ενός δικτύου μπορούμε να παρέχουμε τιμές από το σει διανυσμάτων που έχουν δοθεί για την εκπαίδευση του δικτύου και έτσι να ρυθμίσουμε τα βάρη του με τη διαδικασία σφάλματος διόρθωσης. Όμως αυτή η απευθείας ανάθεση τιμών δεν μπορεί να πραγματοποιηθεί στα κρυφά στρώματα νευρώνων ώστε να ρυθμιστούν απευθείας τα βάρη τους. Αυτό είχε ως αποτέλεσμα όταν σε ένα δίκτυο η απόκριση του στρώματος εξόδου δεν ήταν η επιθυμητή σύμφωνα με το παρεχόμενο διάνυσμα εισόδου τότε δεν υπήρχε τρόπος να προσδιοριστεί σε ποιους από τους κρυφούς νευρώνες οφείλεται αυτή η απόκλιση στην απόκριση ώστε να υπάρξει άμεση προσαρμογή των βαρών τους.

Λύση σε αυτό το πρόβλημα δόθηκε με τη χρήση της παραγώγου της συνάρτησης ενεργοποίησης και της συνάρτησης σφάλματος ως προς τις εισόδους αλλά και ως

προς τα βάρη και τις πολώσεις του δικτύου, όπως θα φανεί και στην περιγραφή των αλγορίθμων εκπαίδευσης οπίσθιας διάδοσης του σφάλματος και απότομης καθόδου.

3.4.2 Ο αλγόριθμος της οπίσθιας διάδοσης του σφάλματος

Ο αλγόριθμος της οπίσθιας διάδοσης του σφάλματος έγινε γνωστός από μία διατριβή των Rumelhart , Hinton και Williams το 1986. Ο αλγόριθμος αυτός αποτελεί ένα εργαλείο για τον υπολογισμό των παραγώγων της συνάρτησης σφάλματος και ονομάστηκε οπίσθιας διάδοσης (back-propagation) καθώς όπως θα φανεί στη συνέχεια στην ουσία πραγματοποιείται μία διάδοση του σφάλματος από το στρώμα εξόδου προς τα προηγούμενα στρώματα του δικτύου. Ο όρος οπίσθια διάδοση χρησιμοποιείται μερικές φορές και στην ονομασία των τροφοδοτούμενων προς τα εμπρός δικτύων πολλαπλών στρωμάτων (multilayer feed-forward neural networks) τα οποία και ονομάζονται εναλλακτικά δίκτυα οπίσθιας διάδοσης (back-propagation networks).

Οι περισσότεροι αλγόριθμοι περιλαμβάνουν μία επαναληπτική διαδικασία για την ελαχιστοποίηση μίας συνάρτησης σφάλματος με αλλαγές στα βάρη που πραγματοποιούνται σε μία σειρά βημάτων. Σε κάθε τέτοιο βήμα μπορούμε να διακρίνουμε δύο διαφορετικά στάδια.

Στο πρώτο στάδιο υπολογίζονται οι παράγωγοι της συνάρτησης σφάλματος ως προς τα βάρη. Η συμβολή του αλγορίθμου οπίσθιας διάδοσης σε αυτό το στάδιο έγκειται στο γεγονός ότι αποτελεί ένα ισχυρό εργαλείο υπολογισμού των συγκεκριμένων παραγώγων.

Στο επόμενο στάδιο οι παράγωγοι αυτοί χρησιμοποιούνται για τον υπολογισμό των μεταβολών που πραγματοποιούνται στα βάρη. Η πιο απλή τεχνική υπολογισμού αυτών των μεταβολών είναι ο αλγόριθμος της απότομης καθόδου που θα περιγραφεί στην επόμενη υποπαράγραφο.

Η διάκριση μεταξύ των σταδίων αυτών είναι εμφανής. Έτσι το πρώτο στάδιο που αφορά τη διάδοση του σφάλματος προς τα πίσω μέσω του δικτύου για να υπολογιστούν οι παράγωγοι μπορεί να εφαρμοστεί και σε πολλών άλλων ειδών δίκτυα καθώς και σε πληθώρα συναρτήσεων σφάλματος. Αντίστοιχα το δεύτερο στάδιο της μετατροπής των βαρών μπορεί να πραγματοποιηθεί με πληθώρα μεθόδων (μερικές από τις οποίες θα παρουσιαστούν συνοπτικά σε επόμενη υποπαράγραφο) πιο ισχυρές από τη μέθοδο της απότομης καθόδου.

Προχωράμε τώρα στην παρουσίαση του αλγορίθμου για μία γενική περίπτωση δικτύου οπίσθιας διάδοσης.

Σε ένα τέτοιο δίκτυο κάθε νευρώνας πραγματοποιεί ένα υπολογισμό του αθροίσματος του γινομένου των βαρών επί των αντίστοιχων εισόδων (το οποίο ονομάζεται και τοπικό πεδίο του νευρώνα) ως εξής:

$$n_j = \sum_i (w_{ij} \times z_i) \quad (3.1)$$

Το z_i συμβολίζει την είσοδο στο νευρώνα j που είτε στέλνεται σε αυτόν από τον κόμβο εισόδου i , είτε αποτελεί την έξοδο της συνάρτησης ενεργοποίησης του i νευρώνα. Το w_{ji} εκφράζει το συναπτικό βάρος με εκκίνηση το νευρώνα i και προορισμό των νευρώνα j . Στο άθροισμα συμπεριλαμβάνονται όλες οι εισοδοί ή οι νευρώνες που στέλνουν σήμα στον j νευρώνα. Όσον αφορά την πόλωση του νευρώνα την θεωρούμε ως είσοδο που είναι ίση με τη μονάδα και μπορούμε έτσι να την αντιμετωπίσουμε όπως και τις άλλες εισόδους και όχι με ξεχωριστό τρόπο.

Το παραπάνω άθροισμα μετατρέπεται από μία μη γραμμική συνάρτηση ενεργοποίησης g , στην έξοδο του νευρώνα j η οποία μπορεί να αποτελεί είσοδο σε νευρώνες επόμενου στρώματος. Οπότε:

$$\alpha_j = z_j = g(n_j) \quad (3.2)$$

Επισημαίνεται ότι στην περίπτωση που το j είναι κόμβος εισόδου τότε $z_j = p_j$.

Όπως αναφέρθηκε και προηγουμένως η διαδικασία της εκπαίδευσης περιλαμβάνει την ελαχιστοποίηση μίας συνάρτησης σφάλματος, με στόχο αυτή να φτάσει στην ελάχιστη τιμή της μετά από διαδοχικές μεταβολές στα βάρη του δικτύου. Εδώ θεωρούμε μία τυχαία συνάρτηση σφάλματος E που υπολογίζεται ως το άθροισμα των επιμέρους τιμών της πάνω σε όλα τα διανύσματα που δίνονται για εξάσκηση του δικτύου. Οπότε ισχύει ότι:

$$E = \sum_{ts} E^{ts} \quad (3.3)$$

Όπου το ts συμβολίζει τα διάφορα πρότυπα (patterns) που δίνονται στο δίκτυο, τα οποία περιλαμβάνουν τα διανύσματα εισόδου καθώς και τις αντίστοιχες επιθυμητές εξόδους που επιθυμούμε να μας παρέχει το δίκτυο. Επίσης υποθέτουμε ότι η E είναι διαφορίσιμη και ως προς τις μεταβλητές εξόδου του δικτύου, δηλαδή

$$E^{ts} = E^{ts}(y_1, \dots, y_c) \quad (3.4)$$

Στόχος είναι να βρεθεί μία διαδικασία για τον προσδιορισμό των παραγώγων της συνάρτησης σφάλματος ως προς τα βάρη και τις πολώσεις του δικτύου.

Υποθέτουμε ότι για κάθε διάνυσμα εισόδου από τα πρότυπα που δίνεται στο δίκτυο παίρνουμε το αντίστοιχο διάνυσμα εξόδου αφού έχουμε υπολογίσει τις αποκρίσεις από τις συναρτήσεις ενεργοποίησης των κρυφών νευρώνων με τις σχέσεις που δόθηκαν παραπάνω. Αυτή η διαδικασία ονομάζεται διάδοση προς τα εμπρός αφού το σήμα ξεκινάει από το επίπεδο εισόδου και μετά την εκτέλεση των υπολογισμών στους κρυφούς νευρώνες καταλήγει στους νευρώνες εξόδου από όπου και παράγεται το τελικό αποτέλεσμα.

Τώρα προχωράμε στον υπολογισμό της παραγώγου του E^{ts} (δηλαδή του σφάλματος από την εφαρμογή του προτύπου ts , δηλαδή της εισόδου και του αντίστοιχης εξόδου-στόχου) ως προς κάποιο τυχαίο βάρος w_{ji} . Οι εξοδοί των διαφόρων νευρώνων θα εξαρτώνται από το διάνυσμα εισόδου του ts . Έπειτα παρατηρούμε ότι το E συνδέεται

με το w_{ji} μόνο μέσω του n_j του νευρώνα j . Οπότε εφαρμόζοντας τον κανόνα της αλυσίδας έχουμε ότι:

$$\frac{\partial E^{ts}}{\partial w_{ji}} = \frac{\partial E^{ts}}{\partial n_j} \times \frac{\partial n_j}{\partial w_{ji}} \quad (3.5)$$

Και έπειτα χρησιμοποιούμε τον εξής συμβολισμό:

$$\delta_j = \frac{\partial E^n}{\partial n_j} \quad (3.6)$$

Το δ συχνά αναφέρεται και ως σφάλμα για λόγους που θα γίνουν προφανείς στη συνέχεια. Χρησιμοποιώντας την (3.1) μπορούμε να γράψουμε ότι

$$\frac{\partial n_j}{\partial w_{ji}} = z_i \quad (3.7)$$

Χρησιμοποιώντας τις (3.6), (3.7) στην (3.5) παρατηρούμε ότι

$$\frac{\partial E^{ts}}{\partial w_{ji}} = \delta_j \times z_i \quad (3.8)$$

Από την παραπάνω σχέση παρατηρούμε ότι παράγωγος που θέλουμε, μπορεί να υπολογιστεί από το γινόμενο του δ του νευρώνα στο άκρο προορισμού του συναπτικού βάρους και του z του νευρώνα στο άκρο εισόδου του συναπτικού βάρους. Άρα πλέον αυτό που απαιτείται για τον υπολογισμό της παραγώγου σε κάθε κρυφό νευρώνα και στους νευρώνες εξόδου είναι η τιμή του δ .

Για τους νευρώνες εξόδου k ο υπολογισμός είναι άμεσος. Από τον ορισμό του δ (3.6) έχουμε ότι

$$\delta = \frac{\partial E^{st}}{\partial n_k} = g'(n_j) \times \frac{\partial E^{st}}{\partial y_k} \quad (3.9)$$

Όπου χρησιμοποιήσαμε την (3.2) και στη θέση του z_j θέσαμε το y_j αφού οι υπολογισμοί μας αφορούν το τελικό στρώμα νευρώνων του δικτύου. Οι τιμές των $g'(n_j)$ και τις μερικής παραγώγου ως προς την έξοδο του δικτύου μπορούν να υπολογιστούν γνωρίζοντας τις συναρτήσεις ενεργοποίησης και σφάλματος.

Όσον αφορά τώρα τον υπολογισμό του δ για κάθε κρυφό νευρώνα χρησιμοποιούμε ξανά τον κανόνα της αλυσίδας για τις μερικές παραγώγους και έχουμε ότι:

$$\delta_j = \frac{\partial E^{st}}{\partial n_j} = \sum_k \frac{\partial E^{st}}{\partial a_k} \times \frac{\partial a_k}{\partial a_j} \quad (3.10)$$

Όπου το άθροισμα περιλαμβάνει όλους τους νευρώνες k με τους οποίους ο νευρώνας j διασυνδέεται. Αν τώρα αντικαταστήσουμε τον ορισμό του δ της (3.6) στην (3.10) και χρησιμοποιήσουμε τις σχέσεις (3.1) και (3.2) τότε θα πάρουμε την εξής σχέση:

$$\delta_j = g'(n_j) \times \sum_k w_{kj} \times \delta_k \quad (3.11)$$

Από την παραπάνω σχέση διαπιστώνουμε ότι το δ για ένα συγκεκριμένο κρυφό νευρώνα μπορεί να υπολογιστεί «διαδίδοντας» προς τα πίσω τα δ από νευρώνες που βρίσκονται σε επόμενα στρώματα του δικτύου. Εφόσον γνωρίζουμε τις τιμές των δ των μονάδων εξόδου ο υπολογισμός μπορεί να γίνει διαδοχικά για κάθε προηγούμενο στρώμα.

Συμπερασματικά μπορούμε να περιγράψουμε τη διαδικασία της οπίσθιας διάδοσης του σφάλματος σε 4 βήματα:

1. Ένα διάνυσμα εισόδου p^{ts} παρέχεται στο δίκτυο και διαδίδεται προς τα εμπρός υπολογίζοντας το αποτέλεσμα των συναρτήσεων ενεργοποίησης με τις (3.1) και (3.2) για όλους τους κρυφούς νευρώνες και τους νευρώνες του στρώματος εξόδου.
2. Υπολογίζονται τα δ για όλους τους νευρώνες εξόδου με τη σχέση (3.9)
3. Υπολογίζονται τα δ των κρυφών νευρώνων με τη χρήση των δ του αμέσως επόμενου στρώματος χρησιμοποιώντας τη σχέση (3.11)
4. Χρησιμοποιούμε τη σχέση (3.8) για τον υπολογισμό των ζητούμενων παραγώγων.

Έπειτα από την επανάληψη αυτών των βημάτων η παράγωγος του ολικού σφάλματος μπορεί να υπολογιστεί ως το άθροισμα των επιμέρους παραγώγων των συναρτήσεων των σφαλμάτων που προέκυψαν για κάθε πρότυπο που εφαρμόστηκε στο δίκτυο

3.4.3 Ο αλγόριθμος της απότομης καθόδου

Στην προηγούμενη υποπαράγραφο περιγράφηκε ένας βασικός αλγόριθμος υπολογισμού της παραγώγου της συνάρτησης σφάλματος ως προς τα βάρη του δικτύου. Αυτές οι παράγωγοι μπορούν να χρησιμοποιηθούν σε μια πληθώρα μεθόδων που βασίζονται σε αυτές τις παραγώγους για τον υπολογισμό του ελαχίστου της συνάρτησης σφάλματος. Σε αυτή την υποπαράγραφο θα γίνει η περιγραφή της πιο απλής από αυτές τις μεθόδους, η οποία είναι ο αλγόριθμος της απότομης καθόδου (gradient descent/ steepest descent).

Καταρχήν θα κάνουμε την εξής παραδοχή: όλες οι ελεύθερες παράμετροι του δικτύου (βάρη και πολώσεις) θα ομαδοποιηθούν και πλέον θα χρησιμοποιείται ο συμβολισμός w για εκφράσει αυτόν τον πίνακα με τα βάρη και τις πολώσεις (που στην ουσία είναι βάρη που έχουν ως είσοδο τη μονάδα).

Με βάση αυτόν το συμβολισμό η συνάρτηση του σφάλματος E όπως αναφέρθηκε και σε προηγούμενη υποπαράγραφο μπορεί να εκφραστεί και ως συνάρτηση του w . Υποθέτοντας ότι το E είναι μία διαφορίσιμη ως προς το w συνάρτηση ξεκινάμε από μία αρχική πρόβλεψη για την τιμές των βαρών στο w και μετά αλλάζουμε την τιμή τους μετακινούμενοι μία μικρή απόσταση στην επιφάνεια του σφάλματος, δηλαδή

στο χώρο των βαρών, προς την κατεύθυνση κατά την οποία το E μειώνεται περισσότερο. Επαναλαμβάνοντας αυτή τη διαδικασία μεταβάλλουμε σταδιακά τα βάρη του δικτύου χρησιμοποιώντας τη σχέση:

$$w_{ji}^{\tau+1} = w_{ji}^{\tau} - \eta \times \frac{\partial E}{\partial w_{ji}} \Big|_{w^{(\tau)}} \quad (3.12)$$

Όπου το η είναι η παράμετρος του ρυθμού εκμάθησης. Αυτή η διαδικασία συνεχίζεται μέχρις ότου συγκλίνει σε ένα σημείο όπου ελαχιστοποιείται η τιμή της συνάρτησης σφάλματος.

Όπως αναφέρθηκε και προηγουμένως η συνάρτηση του σφάλματος μπορεί να εκφραστεί ως άθροισμα των επιμέρους σφαλμάτων για κάθε πρότυπο που παρέχεται στο δίκτυο. Έτσι έχουμε ότι:

$$E = \sum_{ts} E^{ts} \quad (3.3)$$

Επίσης έχουμε την έκφραση της παραγώγου της συνάρτησης σφάλματος ως προς τα βάρη του δικτύου η οποία είναι:

$$\frac{\partial E^{ts}}{\partial w_{ji}} = \delta_j \times z_i \quad (3.8)$$

Τρόποι για να υπολογιστούν το δ και το z έχουν αναφερθεί στην προηγούμενη υποπαράγραφο.

Αντικαθιστώντας την (3.8) στην (3.12) παρατηρούμε ότι τελικά η μεταβολή που πραγματοποιείται στα βάρη είναι :

$$\Delta w_{ji} = -\eta \times \delta_j \times z_i \quad (3.13)$$

Για την εφαρμογή του αλγορίθμου με τη χρήση του ολικού σφάλματος (κατά τη διαδικασία της μαζικής εκπαίδευσης) για τον υπολογισμό των παραγώγων ολικού σφάλματος απαιτείται πρώτα ο υπολογισμός των παραγώγων των σφαλμάτων για κάθε πρότυπο που παρέχεται στο δίκτυο. Έπειτα οι παράγωγοι του ολικού σφάλματος υπολογίζονται ως άθροισμα των επιμέρους παραγώγων των σφαλμάτων για κάθε πρότυπο.

3.4.4 Παρουσίαση εναλλακτικών αλγορίθμων εκπαίδευσης δικτύων

Ο αλγόριθμος της απότομης καθόδου μπορεί να είναι ο πιο απλός αλγόριθμος που χρησιμοποιείται με σκοπό την ελαχιστοποίηση της συνάρτησης σφάλματος ως προς τα βάρη του δικτύου, παρουσιάζει όμως αρκετούς περιορισμούς ως προς την αποδοτικότητά του.

Ο πρώτος από αυτούς αφορά το γεγονός ότι η απόδοση του αλγορίθμου εξαρτάται από την καταλληλότητα της τιμής του ρυθμού εκμάθησης που επιλέγεται κατά τη διάρκεια της εκπαίδευσης του δικτύου. Αν η τιμή που θα επιλεγεί είναι πολύ υψηλή βάση των δεδομένων του προβλήματος και της φύσης του δικτύου, τότε υπάρχει

σοβαρή πιθανότητα ο αλγόριθμος να οδηγηθεί σε ταλαντώσεις στην επιφάνεια του σφάλματος και να υπάρξει κατάρρευση της διαδικασίας εκπαίδευσης. Αντίθετα αν η τιμή αυτή είναι πολύ μικρή τότε η σύγκλιση θα αργήσει να επιτευχθεί με αποτέλεσμα την κατανάλωση υψηλού υπολογιστικού χρόνου που θα μπορούσε να χρησιμοποιηθεί διαφορετικά. Αξίζει εδώ να τονίσουμε ότι όπως αναφέρει και ο Bishop στο βιβλίο του [9] «*η βέλτιστη τιμή του η διαφέρει καθώς προχωράει η διαδικασία για την ελαχιστοποίηση*».

Ένα δεύτερο πρόβλημα που αντιμετωπίζει ο αλγόριθμος της απότομης καθόδου έχει να κάνει με το γεγονός ότι βασίζεται αποκλειστικά στην τιμή της τοπικής κλίσης για τον εντοπισμό του ελαχίστου. Σε πολλά σημεία της επιφάνειας του σφάλματος η τοπική πιο απότομη κλίση δεν δείχνει προς την κατεύθυνση του ελαχίστου. Αυτό έχει ως αποτέλεσμα πολλές φορές ο αλγόριθμος να απαιτεί την πραγματοποίηση πολλών βημάτων έως ότου εντοπίσει το ελάχιστο, συνέπεια που καθιστά τον αλγόριθμο πολλές φορές μη αποδοτικό.

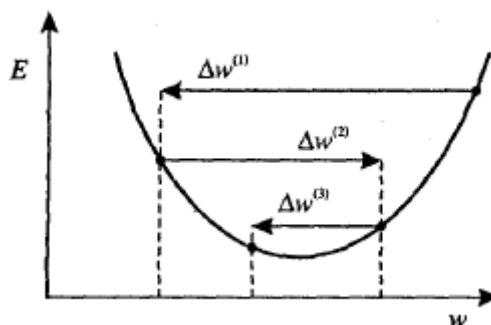
Σταδιακά αναπτύχθηκαν αλγόριθμοι εκπαίδευσης ενός νευρωνικού δικτύου που αντιμετωπίζουν τα προβλήματα που παρουσιάζει ο αλγόριθμος της απότομης καθόδου. Παρακάτω θα παρουσιαστούν τρεις από αυτούς οι οποίοι και παρουσιάζουν καλή απόδοση σε ευρεία γκάμα προβλημάτων.

1. Αλγόριθμος της απότομης καθόδου με ορμή

Στο συγκεκριμένο αλγόριθμο χρησιμοποιείται μία πολύ απλή τεχνική για την αντιμετώπιση των περιορισμών που παρουσιάζει ο αλγόριθμος της απότομης καθόδου. Η τεχνική αυτή αφορά την προσθήκη του όρου της ορμής (momentum) στον τύπο της μεταβολής των βαρών. Αποτέλεσμα αυτής της προσθήκης είναι να μειώνονται οι ταλαντώσεις στην κίνηση του σημείου λειτουργίας πάνω στην επιφάνεια του σφάλματος. Ο τύπος που δίνει τη μεταβολή στα βάρη γίνεται ως εξής:

$$\Delta w^{(\tau)} = -\eta \times \frac{\partial E}{\partial w} \Big|_{w^{(\tau)}} + \mu \times \Delta w^{(\tau-1)}$$

Όπου μ είναι η παράμετρος της ορμής. Όπως φαίνεται και από τον παραπάνω τύπο στον συγκεκριμένο αλγόριθμο για τη μεταβολή των βαρών στο παρόν βήμα λαμβάνεται υπόψη και η μεταβολή των βαρών στο αμέσως προηγούμενο βήμα. Η συνδρομή αυτού του γινομένου έγκειται στο γεγονός ότι αν η επιφάνεια σφάλματος έχει χαμηλή καμπυλότητα τότε με τη χρήση της ορμής αυξάνεται με αποτελεσματικό τρόπο ο ρυθμός εκμάθησης με συνέπεια την πιο αποδοτική εκπαίδευση του δικτύου. Αντίθετα σε περιοχές υψηλής καμπυλότητας της επιφάνειας σφάλματος, στις οποίες ο αλγόριθμος της απότομης καθόδου τείνει να ταλαντώνεται η διαδοχικές εφαρμογές του τύπου της ορμής τείνουν να μειώσουν την επιρροή του στη μεταβολή του βάρους, με αποτέλεσμα ο αποτελεσματικός ρυθμός εκμάθησης να τείνει στο η . Έτσι ο όρος αυτός συμβάλλει στην πιο εύκολη σύγκλιση του αλγορίθμου χωρίς μειώνοντας τις περαιτέρω ταλαντώσεις την επιφάνεια σφάλματος όπως φαίνεται και στην παρακάτω εικόνα [9].



Εικόνα 3.4: Η επίδραση του όρου της ορμής στην ταλαντωμένη κίνηση του σημείου στην επιφάνεια σφάλματος

Συμπερασματικά η χρήση του όρου της ορμής βελτιώνει σε μεγάλο βαθμό την απόδοση του αλγορίθμου της απότομης καθόδου. Παρόλα αυτά στο σύνολο του και αυτός ο αλγόριθμος παραμένει αναποτελεσματικός. Αυτό έχει να κάνει με το γεγονός ότι η χρήση αυτού του όρου εισάγει μία ακόμα παράμετρο, την ορμή μ , της οποίας η τιμή πρέπει να επιλεγεί από τον εκπαιδευτή του δικτύου δημιουργώντας προβλήματα αντίστοιχα με εκείνα που αφορούν την επιλογή της παραμέτρου του ρυθμού εκμάθησης η .

2. Ο αλγόριθμος Levenberg-Marquardt

Ο αλγόριθμος αυτός χρησιμοποιεί μία σχετικά απλή μέθοδο προσέγγισης μίας συνάρτησης η οποία βασίζεται στον παρακάτω τύπο:

$$(J^T J + \lambda I) \delta = J^T E$$

Όπου:

J: Είναι ο Ιακωβιανός πίνακας που περιλαμβάνει όλες τις πρώτες μερικές παραγώγους μίας συνάρτησης. Στην περίπτωση των νευρωνικών δικτύων είναι ένας πίνακας R επί W διαστάσεων, με R να είναι ο αριθμός των προτύπων που υπάρχουν στο σετ προπόνησης του δικτύου και W να είναι ο συνολικός αριθμός των βαρών και των πολώσεων του δικτύου. Τα στοιχεία του πίνακα αυτού υπολογίζονται παίρνοντας τις μερικές παραγώγους των εξόδων του δικτύου ως προς κάθε παράμετρό του.

λ : Είναι μία παράμετρος που ονομάζεται συντελεστής απόσβεσης.

δ : Είναι η μεταβλητή ως προς την οποία λύνεται η παραπάνω εξίσωση και μας υποδεικνύει το πόσο θα μεταβληθούν τα βάρη για να επιτύχουμε τη βέλτιστη λύση

E: Είναι ένα διάνυσμα στο οποίο περιλαμβάνονται όλα τα σφάλματα για κάθε διάνυσμα εισόδου που χρησιμοποιήθηκε για την προπόνηση του δικτύου.

Για να μπορεί να χρησιμοποιηθεί ο συγκεκριμένος αλγόριθμος εκπαίδευσης του δικτύου βασική προϋπόθεση αποτελεί η συνάρτηση σφάλματος ή απόδοσης υπό την

οποία πραγματοποιείται η εκπαίδευση του δικτύου να είναι εκείνη του αθροιστικού τετραγωνικού σφάλματος.

Ο συντελεστής απόσβεσης λ μεταβάλλεται μετά από κάθε επανάληψη του αλγορίθμου. Αν η μείωση του σφάλματος είναι γρήγορη τότε επιλέγονται μικρότερες τιμές του. Το αντίθετο συμβαίνει σε περίπτωση αργής μείωσης του σφάλματος. Μία τυπική τιμή εκκίνησης για το λ είναι το 0.1. Η μεταβολή στο λ πραγματοποιείται χρησιμοποιώντας ένα παράγοντα ν ο οποίος συνήθως πολλαπλασιάζεται με το λ σε περίπτωση αύξησης του λ και διαιρεί το λ στην περίπτωση που θέλουμε να μειωθεί εκείνο.

Συνοψίζοντας τα βήματα εκτέλεσης του συγκεκριμένου αλγορίθμου είναι τα εξής [11]:

1. Υπολογίζεται ο Ιακωβιανός πίνακας, με υπολογισμό των μερικών παραγώγων
2. Υπολογίζεται η κλίση του σφάλματος ως εξής: $g = J^T E$
3. Υπολογίζεται το γινόμενο πινάκων $J^T J$ το οποίο αποτελεί και πολύ καλή προσέγγιση του Hessian πίνακα.
4. Λύνεται η αρχική εξίσωση για να υπολογίσουμε το δ
5. Τα βάρη του δικτύου ανανεώνονται βάση της τιμής του δ που υπολογίστηκε στο παραπάνω βήμα
6. Πραγματοποιείται πάλι υπολογισμός των σφαλμάτων του δικτύου καθώς και του αθροίσματος των τετραγώνων τους.
7. Στην περίπτωση που το άθροισμα των τετραγώνων των σφαλμάτων δεν μειώθηκε, απορρίπτουμε τα βάρη που προέκυψαν, αυξάνουμε το λ πολλαπλασιάζοντάς το με ένα παράγοντα ν και επιστρέφουμε στο βήμα 4
8. Στην περίπτωση μείωσης του τετραγωνικού σφάλματος το λ μειώνεται διαιρούμενο με το ν , τερματίζεται το βήμα της επανάληψης και ξεκινάει η διαδικασία από την αρχή με το νέο λ και τις τιμές των βαρών που προέκυψαν από την προηγούμενη επανάληψη.

Σε γενικές γραμμές ο αλγόριθμος των Levenberg- Marquardt είναι ένας πολύ αποδοτικός αλγόριθμος που παρουσιάζει γρήγορη σύγκλιση. Ένας παράγοντας που περιορίζει την απόδοση του αλγορίθμου είναι η επιλογή των αρχικών βαρών καθώς και το γεγονός ότι δεν λαμβάνει υπόψη τις ακραίες τιμές των δεδομένων πράγμα που μπορεί να οδηγήσει σε προβλήματα κατά τη γενίκευση της λειτουργίας του δικτύου.

3. Ο Bayesian αλγόριθμος εκμάθησης

Στους περισσότερους συμβατικούς αλγόριθμους εκμάθησης για τα τροφοδοτούμενα προς τα εμπρός δίκτυα πολλαπλών στρωμάτων επιχειρείται η εύρεση ενός μοναδικού σετ βαρών, το οποίο βελτιστοποιεί την προσαρμογή του δικτύου στα δεδομένα που του παρέχονται κατά την εκπαίδευσή του[13]. Αντίθετα η θεωρία του Bayes και η εφαρμογή της στα δίκτυα βασίζεται σε μια διαφορετική οπτική της σημασίας του να μαθαίνεις από δεδομένα., στην οποία χρησιμοποιείται η έννοια της πιθανότητας για

να εκφράσει την αβεβαιότητα για τη σχέση που μαθαίνει το δίκτυο κατά την εκπαίδευσή του. Όπως αναφέρει και MacKay στη διατριβή του [12] «*οι ιδέες του τρόπου λειτουργίας των νευρωνικών δικτύων και της στατιστικής θεωρίας του Bayes μπορεί να μοιάζουν σαν δύο αταίριαστοι συγκάτοικοι. Παρόλα αυτά υπάρχει κάτι κοινό που συνδέει τα δύο παραπάνω. Και τα δύο στοχεύουν στη δημιουργία μοντέλων που ταιριάζουν καλά στα δεδομένα*».

Τα νευρωνικά δίκτυα είναι μια πιο ευέλικτη τεχνική προσέγγισης συναρτήσεων σε σχέση με τις υπόλοιπες. Αυτό γιατί περιλαμβάνει στοιχεία που εισάγουν τη μη γραμμικότητα. Αυτή η ευλυγισία των δικτύων μπορεί να τους δημιουργήσει πρόβλημα, καθώς μπορούν να παρασυρθούν από τυχαίους συσχετισμούς μεταξύ των δεδομένων με αποτέλεσμα να οδηγούνται σε λάθος αποτελέσματα. Στην αντιμετώπιση αυτού του μειονεκτήματος παίζει καταλυτικό ρόλο η θεωρία του Bayes. Χρησιμοποιώντας τη συγκεκριμένη θεωρία μπορεί κανείς να αντιληφθεί το πόσο ευέλικτο επιτρέπεται να είναι ένα δίκτυο βάσει των δεδομένων που του παρέχονται. Η αξιοποίηση του στα νευρωνικά δίκτυα εγγυάται ότι δεν θα παρασυρθεί το δίκτυο από τυχαίες σχέσεις στα δεδομένα. Αυτό συμβαίνει γιατί η θεωρία του Bayes εμπεριέχει και την αρχή του «ξυραφιού» του Occam (Occam's Razor) μία αρχή που ευνοεί περισσότερο τα απλούστερα μοντέλα αναπαράστασης της σχέσης μεταξύ των δεδομένων.

Πώς όμως συμβάλλει αυτή η θεωρία στην βελτίωση της εκπαίδευσης του δικτύου; Αυτό πραγματοποιείται σύμφωνα με τον MacKay [12] σε δύο επίπεδα. Σε αυτά αξιολογούνται πιο από τα μοντέλα H_i με παραμέτρους w , τα οποία υποθέτουμε ότι ταιριάζουν στα παρεχόμενα δεδομένα D είναι εκείνο που τα προσεγγίζει καλύτερα. Κατά την αξιολόγηση αυτών των μοντέλων σημαντικό ρόλο επιτελούν δύο ειδών πιθανότητες:

- Η εκ των προτέρων πιθανότητα (prior probability) $P(w|H_i)$ η οποία υποδεικνύει τις πιθανές τιμές που μπορεί να πάρουν οι παράμετροι του μοντέλου
- Ένα σετ δεσμευμένων πιθανοτήτων, μία για κάθε παράμετρο w που προσδιορίζουν την πρόβλεψη $P(D|w, H_i)$ που το μοντέλο πραγματοποιεί για τα δεδομένα.

Με βάση τα παραπάνω τα δύο στάδια που πραγματοποιείται αυτή η αξιολόγηση των μοντέλων είναι η εξής:

1. **Διαμόρφωσης του μοντέλου:** Σε αυτό το στάδιο υποθέτουμε ότι κάποιο από τα μοντέλα είναι το καλύτερο και προσπαθούμε να υπολογίσουμε τις παραμέτρους του (w) βάσει των δεδομένων D . Για τον υπολογισμό τους γίνεται χρήση της εκ των υστέρων πιθανότητας (posterior probability) που υπολογίζεται από τον τύπο του Bayes:

$$P(w|D, H_i) = \frac{P(D|w, H_i) \times P(w|H_i)}{P(D|H_i)}$$

Συχνά χρησιμοποιούνται και μέθοδοι απότομης καθόδου (gradient descent methods) για να υπολογιστεί το μέγιστο της εκ των υστέρων πιθανότητας, που συμβάλλει στον προσδιορισμό της πιο πιθανής τιμής των παραμέτρων w .

2. **Σύγκρισης Μοντέλων** : Σε αυτό το στάδιο τα υπό εξέταση μοντέλα συγκρίνονται και επιλέγεται αυτό με την καλύτερη απόδοση βάσει των δεδομένων. Η εκ των υστέρων πιθανότητα κάθε μοντέλου είναι ανάλογη των παρακάτω:

$$P(H_i|D) \propto P(D|H_i) \times P(H_i)$$

Η πιθανότητα $P(H_i)$ εκφράζει το πόσο πιθανό θεωρούμε ότι είναι κάθε μοντέλο να ταιριάζει στα δεδομένα πριν τα αποκτήσουμε. Συνήθως όλα τα μοντέλα έχουν αρχικά την ίδια πιθανότητα να είναι εκείνα που έχουν την καλύτερη προσέγγιση πριν ακόμα αποκτηθούν τα δεδομένα. Οπότε γίνεται προφανές ότι πλέον η επιλογή του πιο κατάλληλου μοντέλου βασίζεται στον υπολογισμό της $P(D|H_i)$.

Η παραπάνω διαδικασία είναι αποτελεσματική είτε το δίκτυο προορίζεται για προσέγγιση συναρτήσεων, είτε για αναγνώριση προτύπων είτε ακόμα και για διαχωρισμό και ομαδοποίηση των δεδομένων εισόδου.

Ένα από τα βασικά πλεονεκτήματα της χρήσης της θεωρίας του Bayes στα νευρωνικά δίκτυα αφορά και την επιλογή της κατάλληλης αρχιτεκτονικής του δικτύου. Στην επιλογή αυτή παίζει πολύ σημαντικό ρόλο η γνώση της σωστής εκ των προτέρων πιθανότητας, η οποία αναπαριστά την αβεβαιότητα για το ποιες είσοδοι στο δίκτυο είναι σωστές, για το πόσο θόρυβος περιλαμβάνεται στα δεδομένα και για άλλες παραμέτρους που αφορούν το δίκτυο. Αυτό το πλεονέκτημα έγκειται στο γεγονός ότι η χρήση της θεωρίας του Bayes επιτρέπει στον εκπαιδευτή να χρησιμοποιήσει ένας γενικής φύσεως δίκτυο το οποίο θα έχει πολλούς νευρώνες κατανεμημένους σε πολλά στρώματα. Αυτή η ιδιότητα υποστηρίζεται και από το Neal (Neal, 1993) [14], ο οποίος στη διατριβή του υποστηρίζει ότι «δεν υπάρχει στατιστικά ανάγκη να περιορίσουμε την πολυπλοκότητα ενός δικτύου όταν αυτό εκπαιδεύεται μέσω μίας καλά σχεδιασμένης μεθόδου που βασίζεται στη θεωρία του Bayes».

Προς το παρόν η θεωρία του Bayes χρησιμοποιείται κατά κύριο λόγο για την εκπαίδευση προς τα εμπρός τροφοδοτούμενων δικτύων πολλαπλών στρωμάτων αλλά η μέθοδος αυτή μπορεί να εφαρμοστεί και σε δίκτυα διαφόρων άλλων ειδών, αρκεί αυτά να μπορούν να αναπαρασταθούν με στατιστικούς όρους.

3.5 Γενίκευση δικτύου

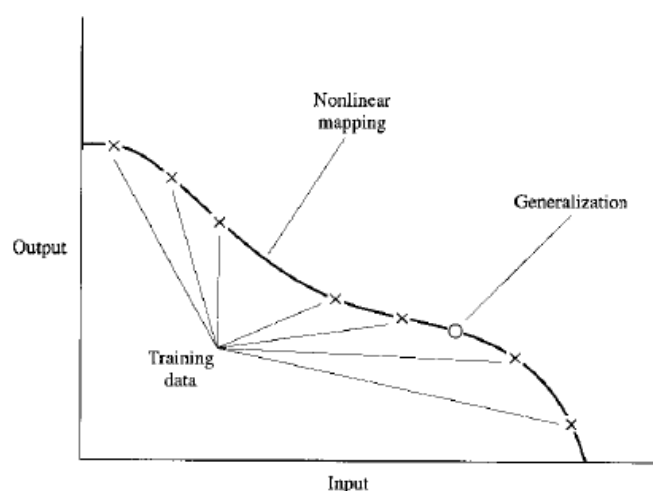
Όπως αναφέρθηκε και σε προηγούμενη υποπαράγραφο σκοπός της εκπαίδευσης ενός δικτύου είναι να μπορέσει να μάθει τη γενική σχέση που συνδέει τα διανύσματα εισόδου με τους στόχους που αντιστοιχούν στο καθένα. Η πραγματοποίηση αυτού του στόχου έχει ως αποτέλεσμα το δίκτυο να παρουσιάζει καλή γενίκευση (generalization), δηλαδή να ανταποκρίνεται σωστά σε εισόδους που δεν βρίσκονταν μέσα στα διανύσματα εκπαίδευσης του και να προχωρά σε σχετικά ακριβείς προβλέψεις για αυτές τις τιμές.

Σύμφωνα με τον W. Sarle [13] υπάρχουν τρεις απαραίτητες συνθήκες ώστε το δίκτυο να πραγματοποιεί σωστή γενίκευση. Αυτές είναι οι εξής:

1. Η πρώτη συνθήκη αφορά την πληροφορία που παρέχεται από τις εισόδους του δικτύου που τις συσχετίζει με τους αντίστοιχους στόχους στην έξοδο του δικτύου. Αυτή η πληροφορία θα πρέπει να είναι επαρκής ώστε να καταδεικνύει την μαθηματική σχέση που συνδέει τις εισόδους με τις εξόδους του δικτύου με μία επιθυμητή ακρίβεια. Τα νευρωνικά δίκτυα δεν μπορούν να μαντέψουν μία σχέση μεταξύ των εισόδων και των εξόδων. Είναι απαραίτητο να παρέχεται σε αυτά η κατάλληλη πληροφορία που θα συμβάλλει στην αναγνώριση της σχέσης που συνδέει τις εισόδους με τις εξόδους του δικτύου.
2. Η δεύτερη συνθήκη αφορά τη συνάρτηση που συνδέει τις εισόδους με τις εξόδους. Σύμφωνα με το Sarle για να παρουσιάζει το δίκτυο καλές ιδιότητες γενίκευσης είναι απαραίτητο αυτή να είναι λεία (smooth). Αυτό σημαίνει ότι σε μικρές αλλαγές στις εισόδους να αντιστοιχούν μικρές αλλαγές και στα αποτελέσματα της συνάρτησης. Συναρτήσεις οι οποίες δεν παρουσιάζουν ομαλότητα στις μεταβολές δεν μπορούν να γενικευθούν σωστά από τα νευρωνικά δίκτυα. Πολλές φορές για να επιτευχθεί η απαραίτητη ομαλότητα στη συνάρτηση επιχειρείται ένας μη γραμμικός μετασχηματισμός των εισόδων του δικτύου.
3. Η τρίτη απαραίτητη συνθήκη αφορά το σετ των δεδομένων που παρέχονται στο δίκτυο κατά τη διαδικασία της εκπαίδευσης και υποστηρίζει ότι είναι απαραίτητο αυτό το σετ να αποτελεί ένα επαρκές και αντιπροσωπευτικό σετ των περιπτώσεων για τις οποίες θέλουμε να γενικεύει το δίκτυο. Αυτή η συνθήκη σχετίζεται άμεσα με τα δύο διαφορετικά είδη γενίκευσης που παρατηρούνται, της εσωτερικής ή αλλιώς παρεμβολής (interpolation) και τις εξωτερικής ή παρέκτασης (extrapolation). Η πρώτη αφορά τις περιπτώσεις κατά τις οποίες η γενίκευση πραγματοποιείται για σημεία τα οποία έχουν γύρω τους σημεία τα οποία χρησιμοποιήθηκαν κατά την εκπαίδευση του δικτύου. Η δεύτερη αφορά σημεία τα οποία βρίσκονται εκτός του διαστήματος των σημείων εκπαίδευσης και είναι εκείνη που παράγει τα πιο αναξιόπιστα αποτελέσματα από τις δύο. Η ανάγκη για την αποφυγή περιπτώσεων στις οποίες θα πραγματοποιείται εξωτερική γενίκευση καθιστά απαραίτητη την ύπαρξη ενός επαρκώς μεγάλου σετ δεδομένων για την προπόνηση του δικτύου.

Συνοψίζοντας μπορούμε να υποστηρίξουμε ότι ένα νευρωνικό δίκτυο το οποίο έχει σχεδιαστεί για να γενικεύει καλά, τηρώντας τις τρεις παραπάνω συνθήκες, θα παράγει

σωστά αποτελέσματα ακόμα και αν οι εισόδοι που παρέχονται σε αυτό διαφέρουν αρκετά από τα δεδομένα που χρησιμοποιήθηκαν κατά την εκπαίδευσή του. Τα αποτελέσματα ενός δικτύου που γενικεύει σωστά παρατηρούνται στην εικόνα 20 [8]. Στον αντίποδα όμως όταν η ποσότητα των δεδομένων που παρέχονται στο δίκτυο για την εκπαίδευσή του είναι υπερβολική τότε το δίκτυο μπορεί να καταλήξει να «απομνημονεύσει» τα δεδομένα που χρησιμοποιήθηκαν κατά την προπόνηση, με καταστρεπτικές συνέπειες για την μετέπειτα απόδοσή του. Αυτό μπορεί να πραγματοποιηθεί όταν το δίκτυο παρασυρθεί από σχέσεις που μπορεί να υπάρχουν μέσα στα δεδομένα της εκπαίδευσής του, οι οποίες είναι τυχαίες και δεν έχουν καμία σχέση με την πραγματική μαθηματική σχέση που συνδέει τα δεδομένα εισόδου με τις εξόδους του δικτύου. Αυτό το φαινόμενο θα εξεταστεί στην παρακάτω υποπαράγραφο.



Εικόνα 3.5: Παράδειγμα καλής γενίκευσης δικτύου

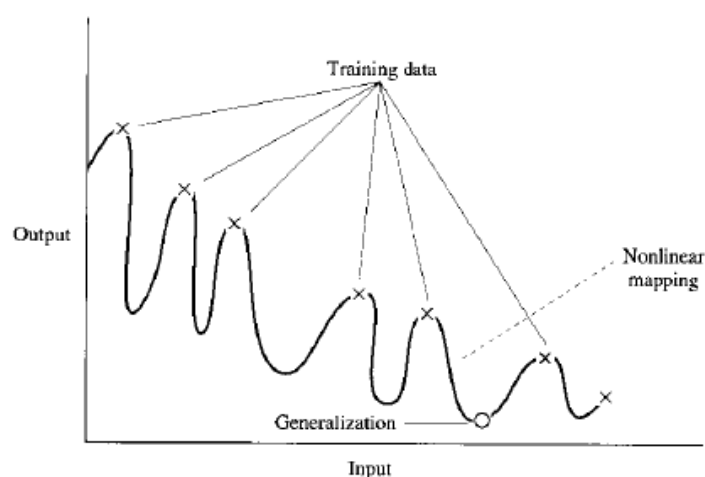
3.5.1 Το πρόβλημα του υπερβολικής προσαρμογής

Κατά τη διαδικασία της εκπαίδευσης βασικός στόχος αποτελεί το δίκτυο μετά από το στάδιο αυτό να μπορεί να προχωρά σε καλή απόκριση όταν του παρέχονται άγνωστες σε αυτό εισόδους, δηλαδή να πραγματοποιεί καλή γενίκευση. Ένα από τα βασικά προβλήματα που αποτελούν τροχοπέδη στην πραγματοποίηση του συγκεκριμένου στόχου είναι εκείνο της υπερβολικής προσαρμογή (overfitting) ή της ελλιπούς προσαρμογής (underfitting) στη σχέση που συνδέει τις εισόδους με τις εξόδους κατά την εκπαίδευση του δικτύου.

Όπως αναφέρει και ο Sarle [13] για το συγκεκριμένο πρόβλημα: « Ένα δίκτυο το οποίο δεν έχει την απαραίτητη πολυπλοκότητα όσον αφορά τη δομή του, μπορεί να αποτύχει να εντοπίσει πλήρως το σήμα σε ένα περίπλοκο σετ δεδομένων με αποτέλεσμα την ελλιπή προσαρμογή σε αυτά. Ένα δίκτυο που είναι πολύ περίπλοκο

μπορεί να προσεγγίσει και το θόρυβο και όχι μόνο το σήμα, με αποτέλεσμα την υπερβολική προσαρμογή στα παρεχόμενα δεδομένα.»

Οι συνέπειες που μπορεί να έχει αυτή η λάθος προσέγγιση του σήματος είναι σημαντικές. Ένα δίκτυο το οποίο εκπαιδεύτηκε με λάθος τρόπο, οδηγώντας το σε υπερβολική ή ελλιπή προσαρμογή στα δεδομένα, έχει ως συνέπεια η απόκριση του δικτύου να απέχει σημαντικά από τις σωστές τιμές που εκείνο θα έπρεπε να παράγει. Στην περίπτωση που αυτό το δίκτυο πραγματοποιεί προβλέψεις για ένα μέγεθος, αυτές οι προβλέψεις μπορούν να οδηγήσουν σε λανθασμένη εκτίμηση της προβλεπόμενης ποσότητας. Είναι επίσης συχνό φαινόμενο ένα τέτοιο δίκτυο να παράγει ακραίες τιμές αποπροσανατολίζοντας τελείως τον αναλυτή των αποτελεσμάτων του. Παράδειγμα απόκρισης δικτύου με υπερβολική προσαρμογή φαίνεται στην εικόνα 3.6 [8]. Η σωστή γενίκευση για το συγκεκριμένο δίκτυο υπάρχει στην εικόνα 3.5.



Εικόνα 3.6: Παράδειγμα απόκρισης δικτύου με overfitting

Υπάρχουν αρκετοί τρόποι για να αποφύγει κανείς την υπερβολική προσαρμογή του δικτύου στα δεδομένα. Ο Sarle [13], βασιζόμενος στα αποτελέσματα του Amari (Amari et al., 1997), υποστηρίζει ότι ο καλύτερος τρόπος για την αποφυγή του overfitting είναι η παροχή κατά την εκπαίδευση στο δίκτυο ενός σετ δεδομένων των οποίων οι περιπτώσεις που θα παρέχονται στο δίκτυο για εκπαίδευση να είναι τουλάχιστον 30 φορές περισσότερες από τις παραμέτρους προς ρύθμιση του δικτύου. Αυτή η πρόταση διερευνήθηκε πειραματικά και καταρρίφθηκε από τον Wang (Wang et al., 2005), ο οποίος στη διατριβή του επιχειρεί την εκπαίδευση δικτύων με σετ δεδομένων των οποίων η ποσότητα είναι από 5 έως 50 φορές μεγαλύτερο του αριθμού των ελεύθερων παραμέτρων. Μετά από την εκτέλεση των πειραμάτων συμπεραίνει ότι καθώς αυξάνεται το μέγεθος του σετ προπόνησης παρατηρείται μείωση του overfitting στα δίκτυα και καλύτερη συμπεριφορά τους κατά την απόκριση. Παρόλα αυτά συμπεραίνει ότι δεν υπάρχει μία ξεκάθαρη τιμή του δείκτη R (πηλίκο μεγέθους σετ προπόνησης προς αριθμό ελεύθερων παραμέτρων του δικτύου) μετά από την οποία το δίκτυο είναι σίγουρο ότι δεν θα παρουσιάσει overfitting.

Μάλιστα στα συγκεκριμένα πειράματα για $R=30$ συνεχίζει να παρατηρείται overfitting και στα τρία δίκτυα που εκπαιδεύονται.

Υπάρχουν και κάποιοι επιπρόσθετοι τρόποι για να αντιμετωπιστεί το πρόβλημα του overfitting στα νευρωνικά δίκτυα όταν σε αυτά παρέχεται μία σταθερή ποσότητα δεδομένων για εκπαίδευση. Ο Sarle [16] επισημαίνει ότι οι δύο σημαντικότερες από αυτές είναι η πρόωρη διακοπή (early stopping) της εκπαίδευσης, η χρήση της θεωρίας του Bayes κατά την εκπαίδευση.

Όσον αφορά την πρώτη από αυτές τις δύο μεθόδους, βασίζεται στον υπολογισμό του σφάλματος κατά τη φάση της επικύρωσης (validation). Σύμφωνα με αυτή τα διαθέσιμα προς εκπαίδευση δεδομένα χωρίζονται σε δύο σετ, ένα για εκπαίδευση του δικτύου (training set) και ένα το οποίο χρησιμοποιείται για τη ρύθμιση του σφάλματος της απόκρισής του μετά από κάθε εποχή εκπαίδευσης (validation set). Κατά τη διαδικασία της εκπαίδευσης μετά από το τέλος κάθε εποχής εκπαίδευσης το δίκτυο καλείται να αποκριθεί στις εισόδους που βρίσκονται στο validation set. Έπειτα βάση της απόκρισης του δικτύου και των αντίστοιχων στις εισόδους, εξόδων στόχων υπολογίζεται το σφάλμα επικύρωσης. Η διαδικασία διακόπτεται όταν για ένα συνεχόμενο αριθμό περιόδων (ο οποίος καθορίζεται από τον εκπαιδευτή) το σφάλμα επικύρωσης ακολουθεί ανοδική πορεία. Αυτή η μέθοδος παρουσιάζει αρκετά πλεονεκτήματα αλλά και σημαντικά μειονεκτήματα τα οποία δεν θα αναλυθούν στα πλαίσια της παρούσας διπλωματικής.

Για τη δεύτερη από αυτές τις δύο μεθόδους που είναι η κανονικοποίηση κατά Bayes (Bayesian regularization) έχουν επισημανθεί αρκετά στην υποπαράγραφο 3.4.4. Εδώ απλά να επισημάνουμε ότι σύμφωνα με τον Sarle [16] η μέθοδος αυτή είναι πολύ αποτελεσματική για την εκπαίδευση του δικτύου αλλά είναι απαιτεί αρκετό υπολογιστικό χρόνο και είναι πολύ ακριβή εκτός εάν το σετ των δεδομένων της προπόνησης είναι μικρό.

Τέλος μία ακόμα μέθοδος για την αντιμετώπιση του overfitting είναι και η κατάλληλη επιλογή του δικτύου. Η επιλογή του δικτύου έχει να κάνει όχι μόνο με τον αριθμό των βαρών αλλά και με τα μεγέθη τους. Όσο πιο πολλά βάρη υπάρχουν σχετικά με τον αριθμό των περιπτώσεων στο σετ προπόνησης τόσο πιο πιθανό είναι το δίκτυο να παρουσιάσει overfitting. Το παραπάνω συμπέρασμα σχετίζεται και με το λεγόμενο δίλημμα μεταξύ πόλωσης και διακύμανσης (bias-variance dilemma) το οποίο αναλύεται διεξοδικά από τους Geman και Bienenstock.[17]. Σύμφωνα με αυτό η πόλωση (bias) μειώνεται και η διακύμανση (variance) αυξάνεται όσο αυξάνεται η πολυπλοκότητα του δικτύου.

Στη συγκεκριμένη περίπτωση εξετάζεται η ισορροπία μεταξύ του σφάλματος που οφείλεται στην πόλωση και εκείνου που οφείλεται στη διακύμανση. Το πρώτο είναι η διαφορά ανάμεσα στην μέση τιμή των αποκρίσεων του δικτύου και την πραγματική τιμή που θα έπρεπε να έχει η απόκριση. Το δεύτερο λαμβάνεται ως η διακύμανση των αποκρίσεων των μοντέλων ως προς μία συγκεκριμένη είσοδο. Και στις δύο

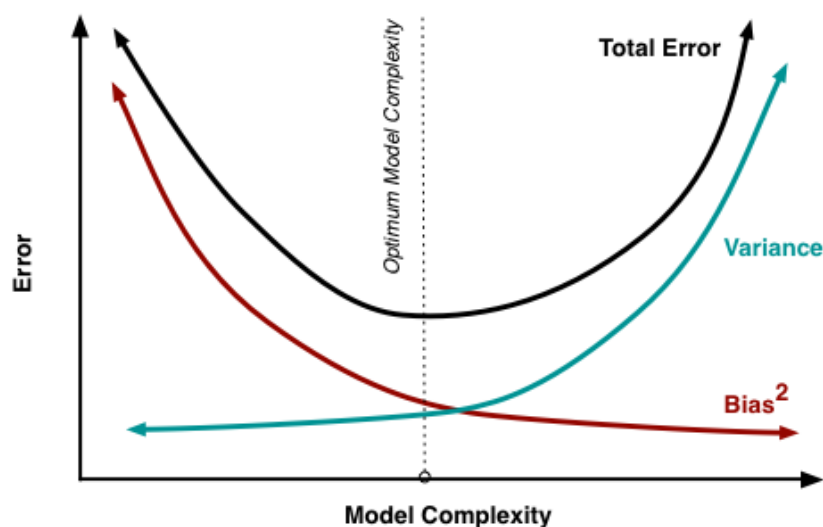
περιπτώσεις θεωρούμε ότι η διαδικασία της δημιουργίας του μοντέλου και της εκπαίδευσής του λαμβάνει χώρα πολλές φορές ώστε να μπορούμε να λάβουμε τα αποτελέσματα των πολλαπλών αποκρίσεων.

Το άθροισμα αυτών των δύο σφαλμάτων στην ουσία συνθέτει το ολικό σφάλμα κατά την απόκριση του δικτύου. Γενικά ισχύει ότι:

$$Error(x) = Bias^2 + Variance + Noise\ term$$

Ο τρίτος όρος αποτελεί το «θόρυβο» που υπάρχει στην πραγματική σχέση μεταξύ εισόδων και εξόδων και ο οποίος είναι πολύ δύσκολο να μειωθεί από οποιοδήποτε μοντέλο.

Στο παρακάτω σχήμα [18] φαίνεται η μεταβολή των δύο αυτών ποσοτήτων καθώς αυξάνεται η πολυπλοκότητα του δικτύου και πως αυτό επηρεάζει το συνολικό σφάλμα. Υπάρχει ένα σημείο στο οποίο αυτά τα δύο σφάλματα ισορροπούν και δεν υπάρχει η υπερβολή προσαρμογή. Όσο απομακρυνόμαστε από αυτό το σημείο τόσο πιο έντονο γίνεται το φαινόμενο της υπερβολικής ή ελλιπούς προσαρμογής στα δεδομένα που δόθηκαν κατά την εκπαίδευση του δικτύου.



Εικόνα 3.7: Σχέση της πόλωσης και της διακύμανση με το ολικό σφάλμα και την πολυπλοκότητα του δικτύου

ΚΕΦΑΛΑΙΟ 4

ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ ΚΑΙ ΠΡΟΒΛΕΨΗ ΧΡΟΝΟ-ΣΕΙΡΩΝ

4.1 Εισαγωγή

Όπως αναφέρθηκε και σε προηγούμενα κεφάλαια ένα νευρωνικό δίκτυο έχει τη δυνατότητα κατά την απόκρισή του να γενικεύει δηλαδή βάση των δεδομένων που του έχουν δοθεί κατά την εκπαίδευσή του να μπορεί να αποκριθεί κατάλληλα και σε εισόδους που δεν βρίσκονταν μέσα σε αυτά τα δεδομένα. Αυτό μπορεί να το κάνει και για σημεία τα οποία βρίσκονται εκτός του διαστήματος των δεδομένων της εκπαίδευσης, που όπως είπαμε προηγουμένως ονομάζεται παρέκταση (extrapolation). Λόγω κυρίως αυτής της δυνατότητας τους τα νευρωνικά δίκτυα σταδιακά καθιερώνονται ως ένα πολύ ισχυρό εργαλείο για την πρόβλεψη μελλοντικών τιμών χρονοσειρών.

Σύμφωνα μάλιστα με το Zhang (Zhang et al., 1998) [19] ο οποίος επιχειρεί μία βιβλιογραφική ανασκόπηση στο θέμα της πρόβλεψης με τη χρήση νευρωνικών δικτύων υπάρχουν αρκετά χαρακτηριστικά τους που τα καθιστούν δυνατά εργαλεία πρόβλεψης, συγκριτικά με τις παραδοσιακές μεθόδους πρόβλεψης. Αρχικά σε αντίθεση με τις παραδοσιακές μεθόδους πρόβλεψης τα νευρωνικά δίκτυα είναι μία μέθοδος που καθοδηγείται από τα δεδομένα και προσαρμόζεται από μόνη της σε αυτά, καθώς οι αρχικές υποθέσεις που πραγματοποιούνται για τα υπό μελέτη προβλήματα είναι πολύ λίγες. Με λίγα λόγια έχουν τη δυνατότητα να μάθουν από τα παραδείγματα και να κατανοήσουν τη σχέση που υπάρχει ανάμεσα στα δεδομένα ακόμα και αν αυτή είναι άγνωστη ή πολύ δύσκολο να περιγραφεί.

Το δεύτερο από αυτά τα χαρακτηριστικά, σύμφωνα με τον Zhang είναι η δυνατότητα των νευρωνικών δικτύων να γενικεύουν όπως αναφέρθηκε και στην αρχή της συγκεκριμένης παραγράφου.

Στη συνέχεια το τρίτο πολύ σημαντικό χαρακτηριστικό τους είναι η δυνατότητά τους να προσεγγίσουν οποιαδήποτε συνάρτηση, όπως έχει αναφερθεί και προηγουμένως. Οι παραδοσιακές μέθοδοι πρόβλεψης έχουν κάποιους περιορισμούς στο να εντοπίσουν τη σχέση μεταξύ των δεδομένων που τους παρέχονται εξαιτίας της πολυπλοκότητάς τους. Αντίθετα τα νευρωνικά έχουν τη δυνατότητα λόγω της πιο γενικής και ευέλικτης μορφής των συναρτήσεων που χρησιμοποιούν να εντοπίζουν, τις περισσότερες φορές, αυτή τη σχέση παρά την πολυπλοκότητα της.

Τέλος, όπως αναφέρει ο Zhang τα νευρωνικά δίκτυα είναι μη γραμμικά δίκτυα, πράγμα που τους δίνει σημαντικό πλεονέκτημα σε σχέση με τις παραδοσιακές τεχνικές οι οποίες πραγματοποιούν προβλέψεις έχοντας ως βασική υπόθεση ότι η υπό μελέτη χρονοσειρά παράγεται από γραμμικές συναρτήσεις.

Μία παρόμοια βιβλιογραφική επισκόπηση γύρω από τα νευρωνικά δίκτυα και τη συμβολή τους στην πρόβλεψη του απαιτούμενου ηλεκτρικού φορτίου αλλά για μικρό χρονικό ορίζοντα έχει πραγματοποιηθεί και από τον Hippert (Hippert et al., 2002) στην οποία γίνεται αναφορά στις διατριβές που έχουν γραφτεί το διάστημα 1991-1999 γύρω από το θέμα της χρήσης των νευρωνικών δικτύων για πρόβλεψη του

ηλεκτρικού φορτίου δίνοντας έμφαση κυρίως στον τρόπο με τον οποίο σχεδιάστηκαν και ελέγχθηκαν τα δίκτυα που χρησιμοποιήθηκαν σε αυτές.

Στη συνέχεια του συγκεκριμένου κεφαλαίου θα πραγματοποιηθεί μία βιβλιογραφική επισκόπηση ορισμένων διατριβών που έχουν πραγματοποιηθεί γύρω από τη χρήση των νευρωνικών δικτύων για πρόβλεψη χρονοσειρών η οποία θα χωριστεί σε δύο μέρη. Στο πρώτο θα γίνει αναφορά στα άρθρα τα οποία αφορούν γενικά τα νευρωνικά δίκτυα και τη συμβολή τους στην πρόβλεψη. Στο δεύτερο μέρος θα επισημανθούν τα άρθρα εκείνα τα οποία αφορούν τα νευρωνικά δίκτυα και τη χρήση τους στο πεδίο της ενέργειας και ιδιαίτερα της ηλεκτρικής.

4.2 Βιβλιογραφική επισκόπηση I: Νευρωνικά Δίκτυα και πρόβλεψη χρονοσειρών

Η χρήση των νευρωνικών δικτύων για την πρόβλεψη μελλοντικών τιμών διαφόρων χρονοσειρών είναι ένα θέμα με το οποίο έχει ασχοληθεί πληθώρα ερευνητών με συνέπεια την παραγωγή μίας πλούσιας σε περιεχόμενο και περιπτώσεις βιβλιογραφίας. Παρακάτω θα γίνει αναφορά στα σημαντικότερα, σύμφωνα με την κρίση του συγγραφέα της παρούσας διπλωματικής, άρθρα που καταπιάνονται με αυτό το ζήτημα.

Ένα από τα βασικότερα ζητήματα που αφορούν την απόδοση των νευρωνικών δικτύων κατά τη διαδικασία της πρόβλεψης είναι και οι παράγοντες που επηρεάζουν τη δυνατότητα του δικτύου να πραγματοποιεί προβλέψεις. Ένα από τα πιο σημαντικά άρθρα που έχουν γραφτεί σχετικά με αυτούς τους παράγοντες είναι εκείνο του Zhang (Zhang et al., 2001) [21]. Σε αυτό το άρθρο επιχειρείται από τους συγγραφείς να αναλυθεί μέσα από μία πειραματική διαδικασία δοκιμών η επιρροή που έχουν στην απόδοση των νευρωνικών δικτύων που χρησιμοποιούνται για πρόβλεψη τρεις πολύ σημαντικοί παράγοντες: 1) ο αριθμός των κόμβων εισόδου, 2) ο αριθμός των κρυφών νευρώνων και 3) το μέγεθος του σετ δεδομένων για την εκπαίδευση του δικτύου. Για την εκτέλεση των πειραμάτων χρησιμοποιούνται προς τα εμπρός τροφοδοτούμενα δίκτυα ενός στρώματος που εκπαιδεύονται σε 8 διαφορετικές χρονοσειρές για να πραγματοποιήσουν προβλέψεις για διάστημα μίας μελλοντικής περιόδου (one-step ahead forecasting). Για την εκτέλεση των πειραμάτων οι παραπάνω παράγοντες μεταβάλλονται στα δίκτυα ως εξής:

- Κόμβοι εισόδου: Ο αριθμός τους μεταβάλλεται από το 1 έως το 5
- Κρυφοί νευρώνες: Το πλήθος τους κυμαίνεται από 1 έως 10 νευρώνες
- Σετ δεδομένων: Εδώ αφορά σημεία της χρονοσειράς. Δίνονται διαδοχικά σετ με 100, 200 και 300 σημεία των χρονοσειρών.

Μετά από την εκτέλεση των πειραμάτων τα πιο σημαντικά από τα συμπεράσματα στα οποία καταλήγει είναι τα εξής:

1. Ο αριθμός των κόμβων εισόδου καθώς και των κρυφών νευρώνων επηρεάζουν σε σημαντικό βαθμό την δυνατότητα του δικτύου να πραγματοποιεί προβλέψεις. Από τους δύο παραπάνω παράγοντες την πιο σημαντική επίδραση έχει ο αριθμός των κόμβων εισόδου
2. Τα απλά δίκτυα μπορούν να πραγματοποιήσουν καλύτερες προβλέψεις από τα πιο σύνθετα.
3. Τα νευρωνικά δίκτυα έχουν τη δυνατότητα να αναγνωρίσουν τον κατάλληλο αριθμό παρελθοντικών εισόδων που απαιτούνται για μία αποτελεσματική πρόβλεψη.
4. Το μέγεθος του δείγματος δεν έχει τόσο σημαντική επιρροή στην απόδοση του δικτύου.

Παρόμοια διατριβή η οποία δεν συστηματοποιεί όμως με τον ίδιο τρόπο τα αποτελέσματα όπως εκείνη του Zhang είχε πραγματοποιηθεί λίγα χρόνια νωρίτερα

από τον Tang (Tang and Fishwick, 1993) [22]. Σε αυτήν κατά κύριο λόγο ερευνάται η επίδραση που έχει ο αριθμός των κρυφών νευρώνων στην απόδοση του δικτύου κατά τη διαδικασία της πρόβλεψης. Μετά τη διεξαγωγή των πειραμάτων οι συγγραφείς δεν καταλήγουν σε κάποιο ξεκάθαρο συμπέρασμα σχετικά με την επίδραση αυτού του παράγοντα στην απόδοση του δικτύου.

Ένα από τα πεδία τα οποία έχουν μελετηθεί ιδιαίτερα και αφορά την απόδοση των νευρωνικών δικτύων κατά την πρόβλεψη χρονοσειρών, είναι το κατά πόσο εκείνη επηρεάζεται από την παρουσία του στοιχείου της εποχικότητας και της τάσης στις χρονοσειρές. Διάφορες μελέτες έχουν παρουσιαστεί γύρω από αυτό το θέμα.

Μία από εκείνες η οποία εξετάζει με συστηματικό τρόπο το παραπάνω ζήτημα είναι εκείνη του Zhang (Zhang and Qi, 2005)[23]. Στη συγκεκριμένη μελέτη επιχειρούνται πειράματα σε 3 τεχνητές χρονοσειρές και 10 που περιλαμβάνουν δεδομένα από τον πραγματικό κόσμο. Τα δίκτυα που χρησιμοποιούνται είναι MLPs στα οποία μεταβάλλεται ο αριθμός των κόμβων εισόδου, οι οποίες τροφοδοτούν στην ουσία το δίκτυο με παρελθοντικές τιμές των χρονοσειρών βάση της εποχικότητας που αυτές παρουσιάζουν. Η απόδοση των δικτύων υπολογίζεται για τέσσερις διαφορετικές περιπτώσεις που αφορούν τις χρονοσειρές που τους παρέχονται κατά τη διαδικασία της εκπαίδευσης. Στην πρώτη η χρονοσειρές παρέχονται χωρίς να έχουν υποστεί κάποια προεργασία ενώ στις υπόλοιπες τρεις αφαιρούνται διαδοχικά από αυτές τα στοιχεία της τάσης, της εποχικότητας ή και τα δύο μαζί. Τέλος τα αποτελέσματα της πρόβλεψης συγκρίνονται με εκείνα που παράγονται από τη μέθοδο ARIMA που λαμβάνει υπόψη της το στοιχείο της εποχικότητας. Από τα αποτελέσματα που παράγονται ο Zhang οδηγείται στο συμπέρασμα ότι τα νευρωνικά δίκτυα δεν πραγματοποιούν καλές προβλέψεις στην περίπτωση που τα δεδομένα που τους παρέχονται δεν έχουν υποστεί κάποια προεργασία. Αντίθετα αν τα δεδομένα έχουν απαλλαγεί από το στοιχείο της εποχικότητας και της τάσης τότε παράγουν πολύ ακριβή αποτελέσματα που ξεπερνούν τα αντίστοιχα που παράγονται από τη μέθοδο ARIMA. Τέλος υποστηρίζει ότι με δεδομένα που έχουν υποστεί κατάλληλη προεργασία και χωρίς τα στοιχεία του να συνδέονται στενά με παρελθοντικά στοιχεία της ίδιας χρονοσειράς την καλύτερη απόδοση παρουσιάζουν τα πιο απλά δίκτυα.

Παρόμοια μελέτη που είχε πραγματοποιηθεί πάνω στην επιρροή του στοιχείου της εποχικότητας στην πρόβλεψη χρονοσειρών από τα νευρωνικά δίκτυα και εκπονήθηκε λίγα χρόνια νωρίτερα από τον Nelson (Nelson et al., 1999) καταλήγει σε παρόμοια αποτελέσματα καθώς υποστηρίζει ότι δίκτυα τα οποία εκπαιδεύονται με δεδομένα από τα οποία έχει αφαιρεθεί το εποχικό στοιχείο παρουσιάζουν καλύτερη απόδοση σε σχέση με εκείνα που εκπαιδεύονται με τις αρχικές χρονοσειρές. Μάλιστα στη συγκεκριμένη μελέτη αιτιολογείται και το γεγονός ότι αυτό αντιβαίνει στο συμπέρασμα της γενικής ικανότητας των δικτύων να προσεγγίζουν οποιαδήποτε συνάρτηση όπως απέδειξε ο Hornik. Η αιτία σύμφωνα με τον Nelson είναι το γεγονός ότι τα δεδομένα των χρονοσειρών είναι σχετικά περιορισμένα πράγμα που αποτρέπει τη χρήση μεγάλου αριθμού κρυφών νευρώνων στο δίκτυο, που αποτελεί βασική παραδοχή του Hornik κατά την απόδειξη της θεωρίας του.

Μία ακόμα ενδιαφέρουσα μελέτη πάνω στο συγκεκριμένο ζήτημα η οποία δεν συμερίζεται τα αποτελέσματα των δύο παραπάνω μελετών πραγματοποιήθηκε από τον Hamzacebi (Hamzacebi, 2008) [25]. Σε αυτήν ο συγγραφέας προτείνει μία νέα αρχιτεκτονική νευρωνικού δικτύου το SANN (Seasonal Artificial Neural Network), η οποία βελτιώνει την απόδοση του δικτύου κατά την πρόβλεψη χωρίς να απαιτείται κάποια προεργασία των παρεχόμενων στο δίκτυο δεδομένων για εκπαίδευση. Σε αυτή την αρχιτεκτονική οι είσοδοι και οι έξοδοι του δικτύου είναι ίσες με τον παράγοντα s , του οποίου η τιμή καθορίζεται από το διάστημα στο οποίο λαμβάνονται τιμές σε μία χρονοσειρά, π.χ. αν έχει μηνιαίες τιμές τότε το $s=12$. Τα αποτελέσματα μετά από την εκπαίδευση του δικτύου αυτού με τέσσερις χρονοσειρές που προέρχονται από τον πραγματικού κόσμο, υποδεικνύουν ότι η απόδοση κατά την πρόβλεψη βελτιώνεται σημαντικά με τη χρήση δικτύου της συγκεκριμένης αρχιτεκτονικής, αρκεί η εποχικότητα να είναι πολύ έντονη σε αυτές τις χρονοσειρές.

Η αρχιτεκτονική ενός δικτύου είναι ένας από τους παράγοντες που διαδραματίζουν σημαντικό ρόλο στην δυνατότητα του δικτύου να προβλέπει τιμές χρονοσειρών. Διάφορες αρχιτεκτονικές δικτύων έχουν προταθεί με σκοπό της βελτίωση της απόδοσής του όπως και στο τελευταίο άρθρο στο οποίο έγινε αναφορά [25].

Μία μελέτη που ασχολείται με το ζήτημα της αρχιτεκτονικής του δικτύου είναι και εκείνη του Balkin (Balkin and Ord, 2000) [26]. Στην συγκεκριμένη μελέτη επιχειρείται από τους συγγραφείς η δημιουργία μίας μεθόδου η οποία καλείται Automated ANN's και η οποία θα παρέχει μία τυποποιημένη διαδικασία για την επιλογή της καταλληλότερης αρχιτεκτονικής του δικτύου ανάλογα με τη χρονοσειρά στην οποία θα πραγματοποιηθεί η πρόβλεψη. Αυτή η διαδικασία αφορά μόνο δίκτυα προς τα εμπρός τροφοδοτούμενα με ένα στρώμα κρυφών νευρώνων και περιλαμβάνει τα εξής βήματα:

1. Μελέτη για το αν απαιτείται λογαριθμικός μετασχηματισμός στα δεδομένα της χρονοσειράς
2. Επιλογή του αριθμού παρελθοντικών τιμών που θα χρησιμοποιούνται ως είσοδοι στο δίκτυο βάση της χρονικής φύσης της χρονοσειράς
3. Εκπαίδευση του δικτύου με τον αριθμό των κρυφών νευρώνων να μεταβάλλεται από το 1 έως τον αριθμό των τιμών εισόδου.

Έπειτα τα αποτελέσματα του δικτύου που επιλέγεται συγκρίνονται με αντίστοιχα αποτελέσματα άλλων μεθόδων για δύο χρονοσειρές και οι συγγραφείς καταλήγουν στο συμπέρασμα ότι τα νευρωνικά δίκτυα αποδίδουν καλύτερα σε σχέση με τις κλασικές γραμμικές μεθόδους για μη γραμμικές χρονοσειρές με μεγάλο αριθμό δεδομένων.

Μία ακόμα μελέτη που παρουσιάζει ένα νέο μοντέλο δικτύου και εξετάζει την απόδοσή του εκπονήθηκε από τον Wong (Wong et al. , 2010) [27]. Σε αυτή οι συγγραφείς προτείνουν ένα νέο είδος δικτύου το οποίο ονομάζουν προσαρμοζόμενο δίκτυο (Adaptive Neural Network) και το οποίο στοχεύει στην αποφυγή του προβλήματος της υπερβολικής προσαρμογής (overfitting) στα

δεδομένα εκπαίδευσης. Αυτό αποτελείται στην ουσία από δύο δίκτυα MLP's ενός κρυφού στρώματος, των οποίων οι έξοδοι συνδυάζονται για να δώσουν την τελική πρόβλεψη. Το πρώτο από αυτά δέχεται ως είσοδο τις παρελθοντικές τιμές της υπό εξέταση χρονοσειράς αρχικοποιημένες από τους παράγοντες λ και u από τους οποίους ο πρώτος εκφράζει τη διαφορά στο εύρος των τιμών μεταξύ δύο ομάδων τιμών που προέρχονται από τη χρονοσειρά και ο δεύτερος αφορά την τάση που παρουσιάζεται στη χρονοσειρά. Το δεύτερο δίκτυο έχει πάντα 5 κρυφούς νευρώνες και χρησιμοποιείται για την πρόβλεψη του σχετικού σφάλματος RE μεταξύ των πραγματικών τιμών και των προβλέψεων. Η εφαρμογή του δικτύου αυτού και η σύγκρισή του με τα αποτελέσματα που παράγονται από ένα απλό νευρωνικό δίκτυο, τη μέθοδο της παλινδρόμησης (AR) και από τη μέθοδο των κ πλησιέστερων γειτονικών σημείων (AKN) δείχνουν ότι η απόδοση του συγκεκριμένου δικτύου είναι καλύτερη σε χρονοσειρές πραγματικών δεδομένων και χαοτικές χρονοσειρές.

Τέλος μία διατριβή που εξετάζει το ενδιαφέρον ζήτημα της επίδρασης της προσθήκης «θορύβου» στα δεδομένα εκμάθησης στην απόδοση του δικτύου κατά την πρόβλεψη χρονοσειρών πραγματοποιήθηκε από τον Zhang (Zhang, 2007) [28]. Σε αυτή ο συγγραφέας επιχειρεί να εκπαιδεύσει ένα δίκτυο χρησιμοποιώντας διάφορες χρονοσειρές οι οποίες παράγονται από μία βασική χρονοσειρά Y με την προσθήκη διαφορετικών επιπέδων θορύβου. Στόχος του είναι να αποδείξει ότι δίκτυα που προπονούνται με ένα τέτοιο σύνολο χρονοσειρών παρουσιάζουν καλύτερη απόδοση και μεγαλύτερη ακρίβεια στις προβλέψεις τους σε σχέση με αντίστοιχα δίκτυα που προπονούνται μόνο με την αρχική χρονοσειρά. Για να αποδείξει το παραπάνω πραγματοποιεί πειράματα πρόβλεψης χρονοσειρών χρησιμοποιώντας νευρωνικό δίκτυο και μεταβάλλοντας την ποσότητα των δεδομένων προς εκπαίδευση, το επίπεδο θορύβου στα δεδομένα καθώς και των αριθμό των επιπρόσθετων χρονοσειρών στις οποίες θα προπονηθεί το δίκτυο που δημιουργούνται με την προσθήκη ενός ανεξάρτητου σετ τυχαίων θορύβων στην αρχική χρονοσειρά. Μετά την εκτέλεση των πειραμάτων παρατηρείτε ότι τα καλύτερα αποτελέσματα εμφανίζουν τα δίκτυα τα οποία προπονήθηκαν με 5 και 10 επιπρόσθετες χρονοσειρές που δημιουργήθηκαν με την προσθήκη διαφορετικών τυχαίων σετ θορύβων στην αρχική χρονοσειρά για να δημιουργήσουν παρεμφερείς με αυτή χρονοσειρές.

Στην επόμενη παράγραφο του κεφαλαίου γίνεται αναφορά στις διατριβές που αφορούν την εφαρμογή των νευρωνικών δικτύων στην πρόβλεψη ποσών ενέργειας είτε αυτά αφορούν παραγωγή και ζήτηση ηλεκτρικής ενέργειας είτε αφορούν άλλες μορφές ενέργειας όπως η Αιολική.

4.3 Βιβλιογραφική επισκόπηση II: Νευρωνικά δίκτυα και ενέργεια

Στη βιβλιογραφία που αφορά τα νευρωνικά δίκτυα και την ικανότητά τους να μπορούν να προβλέπουν μελλοντικές τιμές χρονοσειρών σημαντικό μέρος καταλαμβάνει η σχεδίαση και δοκιμή δικτύων και η χρήσης τους για την πρόβλεψη τιμών χρονοσειρών που αφορούν την κατανάλωση και την παραγωγή ενέργειας. Παρακάτω παρουσιάζονται μερικά από τα άρθρα που έχουν γραφτεί για το συγκεκριμένο πεδίο ενασχόλησης.

Ένα από τα πρώτα άρθρα που εξετάζει τις δυνατότητες των νευρωνικών δικτύων όσον αφορά την πρόβλεψη των βραχυπρόθεσμων αναγκών σε ηλεκτρικό φορτίο είναι εκείνο του Darbellay (Darbellay and Slama, 2000) [29]. Σε αυτή τη διατριβή επιχειρείται να πραγματοποιηθεί μία στατιστική μέθοδο με την οποία θα μπορεί κανείς να εντοπίσει αν η σχέση που συνδέει τα στοιχεία μιας χρονοσειράς είναι γραμμική ή όχι για να αποφασίσει τη χρήση ή μη νευρωνικών δικτύων για τη διαδικασία της πρόβλεψης. Αυτό πραγματοποιείται με τον υπολογισμό και τη σύγκριση των μεγεθών r και ρ , εκ των οποίων ο πρώτος είναι ο συντελεστής γραμμικής συσχέτισης μεταξύ των μεγεθών μιας χρονοσειράς και ο δεύτερος συμβάλλει στην αναγνώριση μη γραμμικών στοιχείων σε χρονοσειρές. Αυτή η μέθοδος εφαρμόζεται στη χρονοσειρά που περιλαμβάνει την κατανάλωση ηλεκτρικού φορτίου στην Τσεχία τα έτη 1994-1995. Επειδή από την παραπάνω σύγκριση δεν εξήχθησαν ξεκάθαρα συμπεράσματα για τη φύση της σχέσης που συνδέει τα στοιχεία της χρονοσειράς, για την πρόβλεψη των τιμών χρησιμοποιήθηκαν νευρωνικά δίκτυα των οποίων τα αποτελέσματα συγκρίθηκαν με αντίστοιχα αποτελέσματα που παρήχθησαν από τη μέθοδο ARIMA και ARMAX προσδιορισμού γραμμικών σχέσεων. Μελετήθηκαν δύο περιπτώσεις νευρωνικών δικτύων βάση των εισόδων που αυτά δέχονται. Στην πρώτη είσοδο ήταν μόνο παλιές τιμές του φορτίου ενώ στη δεύτερη στις εισόδους συμπεριλαμβάνονταν και τιμές της θερμοκρασίας. Και στις δύο περιπτώσεις τα συγκρινόμενα μοντέλα παρουσίασαν παρόμοια απόδοση με τα γραμμικά μοντέλα να είναι ελαφρώς ακριβέστερα των νευρωνικών δικτύων

Μία από τις διατριβές που ασχολείται με τη δοκιμή διάφορων τρόπων εκπαίδευσης των δικτύων και την εφαρμογή τους στην μακροχρόνια πρόβλεψη της κατανάλωσης ενέργειας στην Ελλάδα πραγματοποιήθηκε από τον Οικονομού (L. Oikonomou, 2010) [30]. Στη συγκεκριμένη διατριβή ο συγγραφέας χρησιμοποιεί την αρχιτεκτονική του προς τα εμπρός τροφοδοτούμενου δικτύου πολλαπλών στρωμάτων για την πρόβλεψη της κατανάλωσης ενέργειας για τις περιόδους 2005-2008, 2010, 2012 και 2015. Για την επιλογή του καταλληλότερου δικτύου πραγματοποιούνται δοκιμές σε δίκτυα τα οποία διαφέρουν ως προς τον αριθμό των κρυφών νευρώνων και των κρυφών στρωμάτων, ως προς τις συναρτήσεις ενεργοποίησης και ως προς τον αλγόριθμο που χρησιμοποιείται κατά την εκπαίδευση του δικτύου. Είσοδοι του δικτύου αποτελούν 4 παράγοντες που επηρεάζουν σημαντικά σύμφωνα με το συγγραφέα την κατανάλωση ενέργειας και

αυτοί είναι οι εξής : 1) Η ετήσια θερμοκρασία του περιβάλλοντος, 2) η εγκατεστημένη ισχύς του δικτύου, 3) η ετήσια ανά κάτοικο κατανάλωση ενέργειας και 4) το ακαθάριστο εθνικό προϊόν. Οι προβλέψεις που πραγματοποιήθηκαν με το δίκτυο που τελικώς επιλέχθηκε μετά από τη διαδικασία των δοκιμών συγκρίνονται με αντίστοιχες προβλέψεις που παρήχθησαν με τη μέθοδο της γραμμικής παρεμβολής και ενός μοντέλου SVM με αρκετά ενθαρρυντικά αποτελέσματα.

Μία αντίστοιχη έρευνα πραγματοποιήθηκε και για την βραχυπρόθεσμη πρόβλεψη του ηλεκτρικού φορτίου μερικά χρόνια νωρίτερα από τον Μπακιρτζή (Bakirtzis et al., 1996) [31]. Σε αυτή τη διατριβή κατασκευάζεται ένα προς τα εμπρός τροφοδοτούμενο δίκτυο πολλαπλών στρωμάτων το οποίο χρησιμοποιείται για την πρόβλεψη του ηλεκτρικού φορτίου για κάθε ώρα της ημέρας. Ενδιαφέρουσα είναι η δομή του δικτύου που χρησιμοποιείται, καθώς αυτό περιλαμβάνει 24 εξόδους (μία για κάθε ώρα της μέρας πρόβλεψης), 24 κρυφούς νευρώνες (αυτός ο αριθμός μειώνει το χρόνο εκπαίδευσης του δικτύου χωρίς να μειώνει την απόδοσή του) και 63 εισόδους, από τις οποίες:

- οι 48 πρώτες περιλαμβάνουν ιστορικά ωριαία δεδομένα από τις τελευταίες δύο μέρες πριν την μέρα πρόβλεψης,
- οι εισοδοί 49 έως 56 αφορούν τις τιμές της υψηλότερης και της χαμηλότερης θερμοκρασίας της προηγούμενης ημέρας από την ημέρα πρόβλεψης. Επίσης περιλαμβάνουν προβλέψεις από δύο μετεωρολογικούς σταθμούς για τις τιμές αυτές τη μέρα πρόβλεψης.
- οι εισοδοί 57 έως 63 αντιστοιχούν σε μέρες της εβδομάδας η οποίες αναπαριστούνται με δυαδική μορφή.

Τα αποτελέσματα μετά την εκπαίδευση αυτού του δικτύου με 365 πρότυπα εισόδου εξόδου του προηγούμενου έτους είναι άκρως ενθαρρυντικά, αν εξαιρεθούν οι προβλέψεις που αφορούν τις μέρες των διακοπών. Για να παράγει το δίκτυο σωστές προβλέψεις για τις μέρες αυτές οι συγγραφείς προχώρησαν σε αλλαγές στις εισόδους του δικτύου αυξάνοντας της κατά ένα παράγοντα που σχετίζεται με τις διακοπές (holiday effect adjustment). Το δίκτυο που χρησιμοποιήθηκε για την ημερήσια πρόβλεψη του φορτίου, με την πραγματοποίηση κάποιων προσαρμογών χρησιμοποιήθηκε από τους συγγραφείς και για την πραγματοποίηση εβδομαδιαίων προβλέψεων και προβλέψεων που δεν περιλαμβάνουν στις εισόδους τους τις τιμές του ηλεκτρικού φορτίου για το διάστημα από τις 1 μ.μ έως τις 12 μ.μ. της προηγούμενης από της μέρα πρόβλεψης.

Μία ακόμα διατριβή που συσχετίζει τα νευρωνικά δίκτυα με την παραγωγή ενέργειας είναι και εκείνη του Singhal (Singhal and Swarup, 2011) [32]. Στη συγκεκριμένη διατριβή επιχειρείται από τους συγγραφείς η πρόβλεψη της μέσης τιμής πώλησης του παραγόμενου ηλεκτρικού ρεύματος (δηλαδή της τιμής που πωλείται η MWh όταν δεν υπάρχει συμφόρηση στη ζήτηση του ρεύματος). Η αρχιτεκτονική που χρησιμοποιείται είναι εκείνη του προς τα εμπρός τροφοδοτούμενου δικτύου πολλαπλών στρωμάτων το οποίο έχει ως εισόδους την ημέρα πρόβλεψης, την ώρα

της ημέρας για την οποία θα γίνει η πρόβλεψη, την προβλεπόμενη ζήτηση και τη μεταβολή της σε σχέση με την πρόβλεψη της προηγούμενης χρονικής στιγμής καθώς και τιμές που έλαβε η τιμή του ηλεκτρικού ρεύματος την προηγούμενη, από την ημέρα πρόβλεψης, ημέρα καθώς και αντίστοιχες τιμές των τεσσάρων προηγούμενων εβδομάδων. Το δίκτυο προπονείται με δεδομένα από τους 6 προηγούμενους μήνες και παράγει προβλέψεις που αντιστοιχούν στην τιμή του ρεύματος για κάθε ένα από τα 48 μισάωρα μιας ημέρας. Η απόδοση του δικτύου ελέγχεται με χρήση δεδομένων για ημέρες ενός συγκεκριμένου μήνα που περιλαμβάνουν περιπτώσεις κανονικής διακύμανσης των τιμών αλλά και περιπτώσεις που οι τιμές παρουσιάζουν κορυφές, είτε μικρές είτε μεγάλες. Τα αποτελέσματα δείχνουν ότι ένα τέτοιο δίκτυο έχει καλή απόδοση στην πρόβλεψη τιμών όταν αυτές παρουσιάζουν κανονική διακύμανση, αλλά αντιμετωπίζει σημαντικά προβλήματα στην πρόβλεψη των τιμών όταν αυτές παρουσιάζουν κορυφές.

Το ίδιο ζήτημα, δηλαδή την πρόβλεψη της τιμής του ηλεκτρικού ρεύματος, πραγματεύεται και η διατριβή του Ραο (Ραο, 2007)[33] όμως σε πιο μακροπρόθεσμο ορίζοντα. Ο συγγραφέας επιχειρεί την πραγματοποίηση προβλέψεων από δίκτυο το οποίο δέχεται ως εισόδους παρελθοντικές τιμές για το φορτίο της επόμενης μέρας τις οποίες τις επιλέγει βάση της συσχέτισης που παρατηρεί ότι υπάρχει στα υπάρχοντα δεδομένα των τιμών. Οι προβλέψεις αυτές που πραγματοποιούνται αφορούν την τιμή του ηλεκτρικού φορτίου για m μέρες από την τελευταία τιμή που χρησιμοποιείται ως είσοδο στο δίκτυο. Αντίθετα δηλαδή με τις προηγούμενες αντίστοιχες διατριβές στις οποίες πραγματοποιούνταν προβλέψεις είτε ενός βήματος προς τα εμπρός (π.χ. για την επόμενη ώρα) είτε ο αριθμός εξόδων του δικτύου ήταν ίσως με τα βήματα προς τα εμπρός για τα οποία θέλουμε να πραγματοποιήσουμε την πρόβλεψη (π.χ. εικοσιτέσσερις εξόδους για πρόβλεψη των τιμών του επόμενου εικοσιτετραώρου), ο συγγραφέας επιλέγει να πραγματοποιήσει προβλέψεις για μελλοντικά βήματα απευθείας με χρήση δικτύου μίας εξόδου. Τα αποτελέσματα από τα πειράματα που πραγματοποιήθηκαν συγκρίθηκαν με τα αντίστοιχα που παρήχθησαν από το γραμμικό μοντέλο AUTOREG. Από τη σύγκριση προέκυψε ότι τα αποτελέσματα που παράγονται από το προτεινόμενο μοντέλο είναι πιο ακριβή από τα αντίστοιχα που παράγονται με τη μέθοδο AUTOREG.

Μία εξίσου ενδιαφέρουσα διατριβή που εξετάζει την χρήση των νευρωνικών δικτύων στην πρόβλεψη του ηλεκτρικού φορτίου από μία διαφορετική σκοπιά εκπονήθηκε από τον Hippert (Hippert et al., 2005)[34]. Σε αυτή οι συγγραφείς εξετάζουν κατά πόσο η αύξηση του αριθμού των παραμέτρων προς βελτιστοποίηση (βάρη και πολώσεις) σε ένα νευρωνικό δίκτυο επηρεάζει την απόδοση τους κατά τη διαδικασία της πρόβλεψης φορτίου. Για το σκοπό αυτό κατασκευάζουν ορισμένα μεγάλου μεγέθους νευρωνικά δίκτυα πολλών παραμέτρων και συγκρίνουν την απόδοσή τους στην πρόβλεψη του προφίλ του ηλεκτρικού φορτίου για ένα εικοσιτετράωρο, με την απόδοση πιο συμβατικών μεθόδων πρόβλεψης που βασίζονται κυρίως στην παλινδρόμηση και σε φίλτρα εξομάλυνσης (smoothing filters) και ταξινομούνται από τους συγγραφείς σε πέντε κατηγορίες βάση του τρόπου που πραγματοποιούν την

πρόβλεψη. Από αυτές τις κατηγορίες επιλέγεται κάθε φορά η μέθοδος που παράγει τα πιο ακριβή αποτελέσματα και της οποίας οι προβλέψεις θα συγκριθούν με τις αντίστοιχες καλύτερες των άλλων μεθόδων. Όσον αφορά τα νευρωνικά δίκτυα που χρησιμοποιήθηκαν ήταν MLP's ενός κρυφού στρώματος νευρώνων, εικοσιτεσσάρων εξόδων και είχαν ως εισόδους τις ωριαίες τιμές του φορτίου της προηγούμενης ή των δύο προηγούμενων ημερών, προβλέψεις τις θερμοκρασίας και δύο μεταβλητές που καθόριζαν αν η μέρα πρόβλεψης είναι καθημερινή ή όχι. Παρότι μετά από την εκτέλεση των πειραμάτων και την εξέταση των αποτελεσμάτων παρατηρήθηκε ότι η καλύτερη απόδοση όσον αφορά τις προβλέψεις παρουσιάστηκε από το νευρωνικό δίκτυο μεγάλου μεγέθους που χρησιμοποιήθηκε, οι συγγραφείς εκφράζουν τις αμφιβολίες τους για το κατά πόσο θα πρέπει να υιοθετηθεί η χρήση τους για την πρόβλεψη φορτίου. Η αμφιβολία τους έγκειται κυρίως στο γεγονός ότι αυτά τα δίκτυα μπορούν να παρουσιάσουν πρόβλημα υπερβολικής προσαρμογής στα δεδομένα λόγω του υψηλού αριθμού παραμέτρων προς υπολογισμό σε σχέση με τα δεδομένα εκμάθησης.

Ένα άρθρο το οποίο διαφέρει από τα προηγούμενα καθώς διαπραγματεύεται το θέμα της χρήσης των νευρωνικών δικτύων σε σχέση με την εκμετάλλευση της «πράσινης ενέργειας» είναι εκείνο του Ermis (Ermis et al., 2007) [35]. Στο συγκεκριμένο άρθρο οι συγγραφείς με τη χρήση ενός προς τα εμπρός τροφοδοτούμενου νευρωνικού δικτύου ενός κρυφού στρώματος με τρεις νευρώνες επιχειρούν να πραγματοποιήσουν προβλέψεις μέχρι το έτος 2050 για την κατανάλωση άνθρακα (κάρβουνου) , πετρελαίου και φυσικού αερίου με εισόδους τις παρελθοντικές τιμές της κατανάλωσης καθώς και το έτος που αυτές πραγματοποιήθηκαν. Στη συνέχεια με χρήση αυτών των τιμών που προβλέφθηκαν στο προηγούμενο βήμα, ως εισόδους σε ένα νέο δίκτυο με κρυφό στρώμα 7 νευρώνων επιχειρείται η πρόβλεψη της ενέργειας που θα παράγεται από ανανεώσιμες πηγές μέχρι το έτος 2050. Σύμφωνα, λοιπόν με τα αποτελέσματα που προέκυψαν μετά την εκτέλεση αυτών των πειραμάτων η ποσότητα «πράσινη» ενέργειας που θα καταναλώνεται μελλοντικά αυξάνεται με πιο γρήγορους ρυθμούς σε σχέση με τις αντίστοιχες ποσότητες ενέργειας που παράγονται από τα ορυκτά καύσιμα, το πετρέλαιο και το φυσικό αέριο.

Τέλος μία αντίστοιχη με την προηγούμενη διατριβή είναι και εκείνη που πραγματοποιήθηκε από τον Yeh (Yeh et al., 2014) [36] ο οποίος χρησιμοποίησε ένα προς τα εμπρός τροφοδοτούμενο νευρωνικό δίκτυο πολλαπλών στρωμάτων σε συνδυασμό με την βελτιωμένη μέθοδο απλοποιημένης βελτιστοποίησης σωματιδίων (improved Simplified Swarm Optimization, iSSO) για τον υπολογισμό της Αιολικής ενέργειας στο αιολικό πάρκο Mai Liao της Ταϊβάν. Η μέθοδος iSSO που προτείνεται είναι βελτιωμένη σε σχέση με την αρχική μέθοδο SSO η οποία είναι ευάλωτη σε τοπικά ακρότατα. Αυτή η μέθοδος χρησιμοποιείται από το συγγραφέα για τη βελτιστοποίηση της διαδικασίας προσαρμογής των βαρών κατά την εκπαίδευση του δικτύου. Ως είσοδοι για το δίκτυο που πραγματοποίησε τις προβλέψεις χρησιμοποιήθηκαν οι τελευταίες τιμές της ταχύτητας και της ενέργειας και της ταχύτητας του ανέμου καθώς και 4 μεταβλητές που αφορούν την ταχύτητα και την

ενέργεια του ανέμου και προέκυψαν από την PCA (Principal Component Analysis) που έκανε ο συγγραφέας. Μετά από τη διαδικασία επιλογής της καλύτερης αρχιτεκτονικής για το δίκτυο που θα χρησιμοποιηθεί τα πειραματικά αποτελέσματα συγκρίνονται με τα αποτελέσματα που παράγονται από δίκτυα που έχουν εκπαιδευτεί με τους πιο γνωστούς αλγόριθμους βελτιστοποίησης συμπεριλαμβανομένων και των αλγορίθμων οπίσθιας διάδοσης σφάλματος (Back-propagation algorithm) και σμήνους σωματιδίων (Particle Swarm Optimization). Τα αποτελέσματα της σύγκρισης δείχνουν ότι η χρήση του αλγορίθμου iSSO στην εκπαίδευση του δικτύου παράγει ενθαρρυντικά αποτελέσματα ως προς την απόδοση του δικτύου και έχει το μικρότερο σφάλμα σε σχέση με τις προβλέψεις των άλλων μεθόδων.

Οι παραπάνω διατριβές αποτελούν ένα πολύ μικρό μέρος από το σύνολο της έρευνας που έχει πραγματοποιηθεί στο πεδίο των νευρωνικών δικτύων σχετικά με τις προβλέψεις ενέργειας είτε ηλεκτρικής είτε κάποιας άλλης μορφής. Η συστηματική αυτή προσπάθεια μελέτης των νευρωνικών δικτύων οφείλεται κατά κύριο λόγο στις δυνατότητες που μπορούν αυτά να παρέχουν λόγω τις ικανότητάς τους να παράγουν αποτελέσματα χρησιμοποιώντας μόνο τα δεδομένα που τους παρέχονται και χωρίς την πραγματοποίηση υποθέσεων που θα μπορούσαν να αλλοιώσουν το τελικό αποτέλεσμα.

4.4 Στόχος της παρούσας διπλωματικής

Όπως έγινε αντιληπτό και από την βιβλιογραφική επισκόπηση που πραγματοποιήθηκε παραπάνω έχει πραγματοποιηθεί μία συστηματική προσπάθεια εκμετάλλευσης των δυνατοτήτων των νευρωνικών δικτύων για την πραγματοποίηση αποδοτικών προβλέψεων μελλοντικών τιμών χρονοσειρών.

Στην βιβλιογραφία που αναφέρθηκε έγιναν προσπάθειες από τους συγγραφείς των διατριβών να προτείνουν καινούργιες αρχιτεκτονικές δικτύων και διαφορετικούς τρόπους προπόνησής τους για να επιτύχουν μεγαλύτερη ακρίβεια προβλέψεων και καλύτερη απόδοση του δικτύου. Παράλληλα ένα θέμα που απασχόλησε πολλούς από τους παραπάνω συγγραφείς είναι και το κατά πόσο η προεργασία των δεδομένων που θα δοθούν στο δίκτυο κατά τη διαδικασία της εκμάθησης είναι απαραίτητη για την καλύτερη και αποδοτικότερη εκπαίδευση του δικτύου. Παρόλα αυτά σε λίγες είναι εκείνες στις οποίες έγινε αναφορά για επεξεργασία των αποτελεσμάτων αφού το δίκτυο πάρει την τελική του μορφή μετά τη διαδικασία της εκπαίδευσης.

Στόχος της παρούσας διπλωματικής είναι η παρουσίαση και η πειραματική διερεύνηση της απόδοσης αλγορίθμων οι οποίοι κατά κύριο λόγο δεν επιδρούν στην αρχιτεκτονική του δικτύου ούτε επηρεάζουν τη διαδικασία της εκπαίδευσής του. Αντίθετα μέσα από αυτούς τους αλγορίθμους γίνεται προσπάθεια να βελτιωθεί η ακρίβεια των προβλέψεων του δικτύου με διαδικασίες που έπονται της εκπαίδευσης του δικτύου και προσπαθούν να εξάγουν αποτελέσματα από την τελική του μορφή.

Επιπρόσθετα στα πλαίσια της παρούσας διπλωματικής θα γίνει και μία διερεύνηση της επίδρασης που έχουν στην απόδοση του δικτύου κατά την πρόβλεψη παράγοντες όπως ο αριθμός των κρυφών στρωμάτων, ο αριθμός των νευρώνων σε κάθε στρώμα καθώς και ο αριθμός των σημείων στο σετ εκπαίδευσης του δικτύου.

Στα παρακάτω κεφάλαια της παρούσας διπλωματικής θα γίνει παρουσίαση του κώδικα που χρησιμοποιήθηκε καθώς και τη λογική που ακολουθήθηκε κατά την πραγματοποίηση των πειραμάτων. Επίσης θα γίνει παρουσίαση των δεδομένων που χρησιμοποιήθηκαν στα πειράματα καθώς και παρουσίαση, ανάλυση και εξαγωγή συμπερασμάτων από τα πειραματικά αποτελέσματα.

ΚΕΦΑΛΑΙΟ 5

ΠΑΡΟΥΣΙΑΣΗ ΤΗΣ ΠΕΙΡΑΜΑΤΙΚΗΣ **ΔΙΑΔΙΚΑΣΙΑΣ**

5.1 Παρουσίαση μαθηματικού μοντέλου

Στην παρούσα διπλωματική όπως αναφέρθηκε και σε προηγούμενη παράγραφο επιχειρείται η διερεύνηση αλγορίθμων που στόχο έχουν τη βελτιστοποίηση της απόδοσης του δικτύου κατά τη διαδικασία της πρόβλεψης.

Η διερεύνηση αυτή θα πραγματοποιηθεί μέσα από μία πειραματική διαδικασία κατά την οποία κάθε ένας από τους 5 αλγορίθμους θα ελεγχθεί ως προς την απόδοσή των προβλέψεών του μετά την εκπαίδευσή του σε μία συγκεκριμένη χρονοσειρά. Στις παρακάτω παραγράφους θα αναλυθεί η λογική με την οποία δομήθηκε η πειραματική διαδικασία. Επίσης θα πραγματοποιηθεί μία παρουσίαση των αλγορίθμων που θα εξεταστούν πειραματικά και των οποίων θα δοθούν παράλληλα και οι κώδικες (χωρίς να τοποθετηθούν σε κάποιο παράρτημα) για διευκόλυνση της ανάγνωσης της διπλωματικής.

5.1.1 Η λογική οργάνωση των πειραμάτων

Για τον τρόπο με τον οποίο θα οργανωθούν τα πειράματα ο συγγραφέας της παρούσας διπλωματικής βασίστηκε σε ορισμένα στοιχεία από τη διατριβή του Balestrassi (Balestrassi et al., 2009) [37]. Πιο συγκεκριμένα τα πειράματα οργανώθηκαν βάση ορισμένων από τις οδηγίες που παρέχονται από τους Coleman και Montgomery (Coleman and Montgomery, 1993) [38] για τη διεξαγωγή βιομηχανικών πειραμάτων τα οποία ο Balestrassi [37] υποστηρίζει ότι μπορούν να εφαρμοστούν και σε προβλήματα υπολογιστικών προσομοιώσεων. Έτσι λοιπόν οι βασικές οδηγίες που ακολουθήθηκαν είναι οι εξής[37]:

- Αναγνώριση και καθορισμός του προβλήματος
- Επιλογή των παραγόντων που θα μεταβάλλονται και του εύρους μεταβολής τους
- Σχεδίαση της πειραματικής διαδικασίας
- Διεξαγωγή του πειράματος
- Στατιστική ανάλυση των αποτελεσμάτων
- Εξαγωγή των συμπερασμάτων

Αναλυτικότερα για τα τρία πρώτα βήματα παραπάνω που αφορούν κατά κύριο λόγο την περιγραφή του πειράματος και των συνιστωσών του έχουμε τα εξής:

- Αναγνώριση και καθορισμός του προβλήματος: Στη συγκεκριμένη διπλωματική το πρόβλημα έγκειται στην πειραματική μελέτη 5 αλγορίθμων που αποσκοπούν στη βελτιστοποίηση της απόδοσης των νευρωνικών δικτύων κατά τη διαδικασία της πρόβλεψης. Στόχος είναι η εύρεση εκείνου ή εκείνων των αλγορίθμων από τους 5 υπό μελέτη των οποίων η εφαρμογή σε κάποιο νευρωνικό δίκτυο αυξάνει της πιθανότητες το νευρωνικό δίκτυο να καταφέρει να ανιχνεύσει τη σχέση (αν υπάρχει) που συνδέει τα στοιχεία μίας χρονοσειράς και να πραγματοποιήσει όσο το δυνατόν πιο ακριβείς προβλέψεις για τις μελλοντικές τιμές της.

Η απόδοση των συγκεκριμένων αλγορίθμων κατά τη διαδικασία της πρόβλεψης αρχικά θα μελετηθεί για ένα κοινό σετ δεδομένων μιας χρονοσειράς. Η σύγκριση μεταξύ των αλγορίθμων θα γίνει βάση συναρτήσεων απόδοσης (performance functions) αλλά και των γραφημάτων που παράγονται από την εφαρμογή τους για να διαπιστωθεί η ύπαρξη ενός ή παραπάνω αλγορίθμων που δίνουν τα καλύτερα αποτελέσματα. Έπειτα οι συγκεκριμένοι αλγόριθμοι που θα έχουν επιλεγεί από το παραπάνω βήμα θα εφαρμοστούν σε ένα ξεχωριστό σετ δεδομένων που προέρχονται από τον πραγματικό κόσμο, για να διαπιστωθεί και η απόδοσή τους σε τέτοιου είδους δεδομένα.

- **Επιλογή των παραγόντων που θα μεταβάλλονται και των ορίων μεταβολής τους:** Καταρχήν εδώ να αναφέρουμε ότι τα νευρωνικά δίκτυα τα οποία θα χρησιμοποιηθούν στις επαναληπτικές διαδικασίες είναι τα προς τα εμπρός τροφοδοτούμενα δίκτυα πολλαπλών στρωμάτων.

Δεδομένου του γεγονότος ότι οι αλγόριθμοι δεν αφορούν άμεσα την διαδικασία της εκπαίδευσης αλλά πρόκειται για αλγόριθμους οι οποίοι επιδρούν στο δίκτυο αφού έχει λάβει την τελική του μορφή μετά από τη διαδικασία εκπαίδευσης θεωρήθηκε λογικό να μην πραγματοποιηθεί σχεδόν καμία αλλαγή στις προεπιλεγμένες ρυθμίσεις που παρέχονται από το πρόγραμμα Matlab 12 με το οποίο πραγματοποιήθηκαν οι υπολογισμοί. Η μόνη αλλαγή η οποία πραγματοποιήθηκε ήταν εκείνη που αφορούσε τον αλγόριθμο εκπαίδευσης του δικτύου. Η προεπιλογή του προγράμματος ήταν ο αλγόριθμος Levenberg-Marquardt όμως επιλέχθηκε τελικά ο Bayesian αλγόριθμος εκπαίδευσης για αποφυγή του προβλήματος της υπερβολικής προσαρμογής.

Οπότε οι παράγοντες (factors) στους οποίους θα πραγματοποιηθούν οι αλλαγές αφορούν κατά κύριο λόγο τα δίκτυα που εκπαιδεύονται. Έτσι επιλέχθηκε να μεταβάλλεται ο αριθμός των κρυφών στρωμάτων νευρώνων καθώς και ο αριθμός των νευρώνων σε αυτά τα στρώματα.

Επίσης επιλέχθηκε ο αριθμός των δεδομένων εκπαίδευσης να είναι κοινώς για κάθε ένα από τους αλγόριθμους (40 σημεία) εκτός από κάποιες επιπρόσθετες δοκιμές που γίνονται στον πρώτο αλγόριθμο και οι οποίες περιλαμβάνουν δοκιμές με 40, 160 και 320 σημεία στο σετ δεδομένων για τη συμπληρωματική μελέτη που θα γίνει περί της επιρροής του μεγέθους του σετ δεδομένων στην απόδοση του δικτύου.

Παρακάτω δίνεται ένας πίνακας που συνοψίζει τα μεγέθη τα οποία θα μεταβάλλονται κατά την πειραματική διαδικασία.

Παράγοντες	Ελάχιστη τιμή	Μέγιστη τιμή
Αριθμός κρυφών στρωμάτων	1	2
Νευρώνες για δίκτυο ενός κρυφού στρώματος	1	50
Νευρώνες πρώτου κρυφού στρώματος δικτύου δύο κρυφών στρωμάτων	1	10
Νευρώνες δεύτερου κρυφού στρώματος δικτύου δύο κρυφών στρωμάτων	1	10
Σημεία εκπαίδευσης πρώτου κώδικα	40	320
Αρχικοποιήσεις βαρών δικτύου	10	10

Πίνακας 5.1: Όρια μεταβαλλόμενων παραγόντων στα πειράματα

Εδώ να αναφερθεί ότι η επιλογή των ορίων των παραγόντων για τον αριθμό των νευρώνων δεν έγινε τυχαία. Βασικό στοιχείο για την επιλογή τους αποτέλεσε ο αριθμός παραμέτρων προς βελτιστοποίηση καθώς επιδίωξη ήταν τα δίκτυα ενός και δύο κρυφών στρωμάτων να έχουν περίπου ίδιο αριθμό παραμέτρων προς βελτιστοποίηση ώστε να μπορούν να γίνουν και κάποιες συγκρίσεις μεταξύ δικτύων του ίδιου αλγορίθμου. Συγκεκριμένα ο αριθμός των παραμέτρων προς βελτιστοποίηση δικτύου ενός κρυφού στρώματος με πενήντα κρυφούς νευρώνες είναι 151. Αντίστοιχα ο αριθμός παραμέτρων δικτύου δύο κρυφών στρωμάτων με δέκα νευρώνες σε κάθε στρώμα είναι 141. Γίνεται κατανοητό ότι ο αριθμός των παραμέτρων είναι αρκετά κοντινός ο ένας στον άλλο.

Ένας επιπρόσθετος λόγος για την επιλογή των ορίων έχει να κάνει και με τον αριθμό των σημείων που περιλαμβάνονται στο σετ δεδομένων που θα δοθεί στα δίκτυα κατά τη διαδικασία εκπαίδευσης. Δεδομένο ότι αυτό έχει 40 σημεία προπόνησης έγινε προσπάθεια ο αριθμός των παραμέτρων προς βελτιστοποίηση στα δίκτυα να περιοριστεί σε πιο χαμηλούς αριθμούς για να αποφευχθεί το φαινόμενο του overfitting λόγω του μικρού αριθμού σημείων σε σχέση με τις παραμέτρους.

Τέλος σημειώνεται ότι τα όρια των παραμέτρων επιλέχθηκαν και για λόγους υπολογιστικού χρόνου καθώς ο χρόνος που απαιτείται για την εκπαίδευση ενός δικτύου αυξάνεται αρκετά καθώς αυξάνονται οι παράμετροι προς βελτιστοποίηση. Για παράδειγμα ένα δίκτυο ενός κρυφού στρώματος επτά νευρώνων χρειάζεται περίπου 25 δευτερόλεπτα για την εκπαίδευσή του ενώ ένα δίκτυο δύο κρυφών στρωμάτων με επτά νευρώνες το κάθε ένα μπορεί να χρειαστεί έως και 100 δευτερόλεπτα.

Εδώ αξίζει να αναφερθεί ότι ένα επιπρόσθετο πρόβλημα αποτελεί η επιλογή των αρχικών βαρών. Μία λάθος επιλογή αρχικών βαρών μπορεί να οδηγήσει τη διαδικασία σε τοπικό ελάχιστο με αποτέλεσμα να διακοπεί πρόωρα οδηγώντας σε λάθος συμπεράσματα. Για να αποτραπεί αυτή η εξάρτηση από την επιλογή αρχικών βαρών αποφασίσθηκε να πραγματοποιούνται δέκα αρχικοποιήσεις σε κάθε δίκτυο και να εξετάζεται το καθένα ξεχωριστά για τα αποτελέσματά του.

- **Σχεδίαση της πειραματικής διαδικασίας** : Με δεδομένα τα όρια των παραμέτρων που θα μεταβάλλονται στην πειραματική διαδικασία, αυτή θα περιλαμβάνει κατά κύριο λόγο την εκπαίδευση και την πραγματοποίηση προβλέψεων από δίκτυα τα οποία θα καλύπτουν όλους τους πιθανούς συνδυασμούς αυτών των παραμέτρων. Αυτό θα πραγματοποιηθεί για κάθε ένα από τους 5 αλγορίθμους σε σει 40 σημείων, εκτός από τον πρώτο αλγόριθμο στον οποίο όπως αναφέρθηκε τα πειράματα θα πραγματοποιηθούν και σε σει δεδομένων 160 και 320 σημείων.

Στην επόμενη σελίδα δίνεται ένα διάγραμμα το οποίο αποτελεί συνοπτική περιγραφή της υπολογιστικής διαδικασίας.

Μετά από την εκτέλεση των επαναληπτικών διαδικασιών θα προκύπτουν τα δίκτυα που θα πραγματοποιούν τις προβλέψεις για ένα διάστημα σημείων που θα περιλαμβάνει άγνωστα στο δίκτυο σημεία. Η απόδοση του δικτύου και η ακρίβεια των προβλέψεων του θα αξιολογούνται μέσω δύο συναρτήσεων απόδοσης. Η επιλογή του αριθμού των συναρτήσεων απόδοσης έγινε με στόχο την πιο αποτελεσματική κρίση των αποτελεσμάτων για την εξαγωγή συμπερασμάτων.

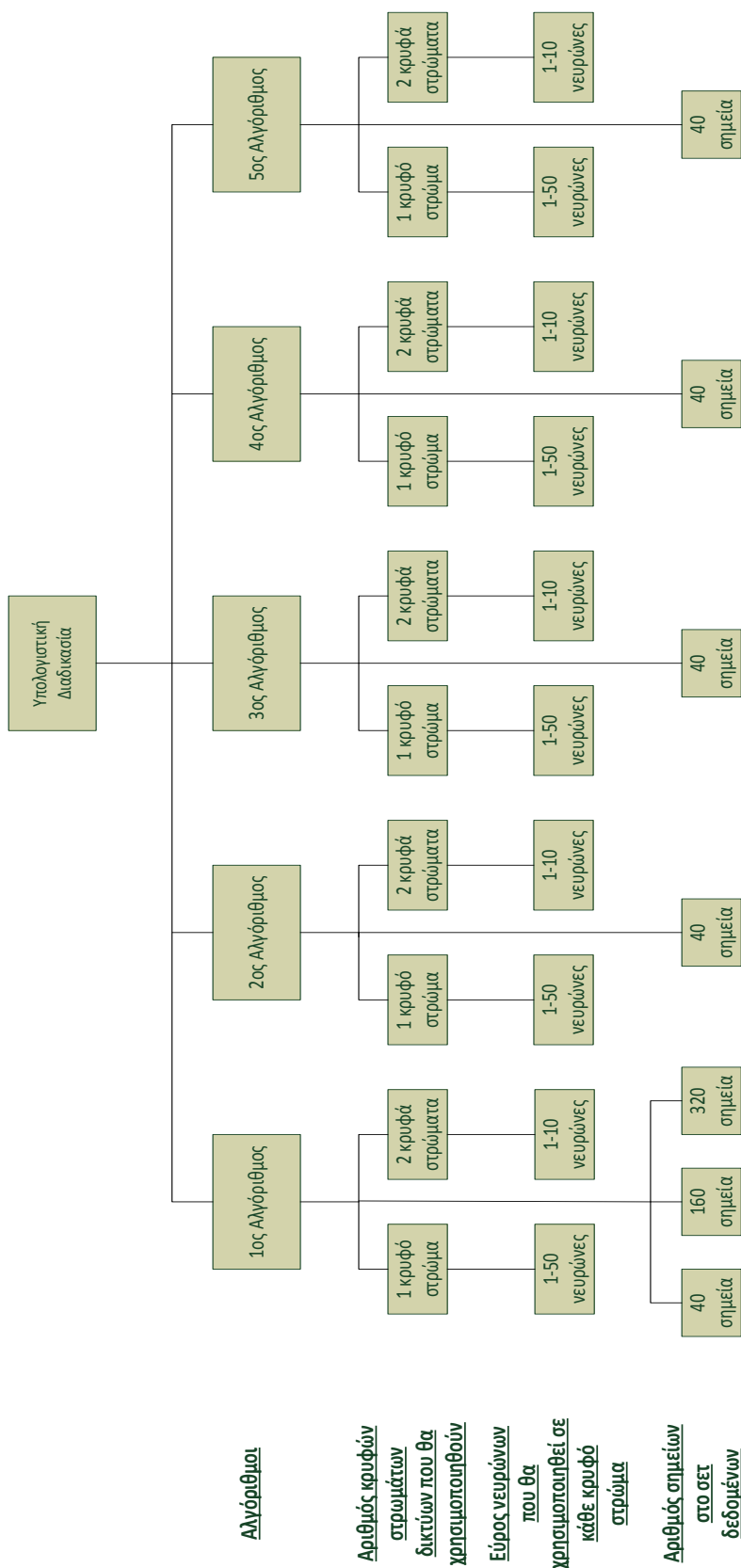
Η πρώτη από αυτές είναι του μέσου τετραγωνικού σφάλματος (Mean Square Error-MSE) της οποίας ο τύπος είναι ο εξής:

$$MSE = \frac{1}{n} \times \sum_{i=1}^n (y_i - \hat{y})^2 \quad (5.1)$$

Η δεύτερη συνάρτηση απόδοσης είναι η συνάρτηση του μέσου απόλυτου σφάλματος (Mean Absolute Error-MAE) με τύπο:

$$MAE = \frac{1}{n} \times \sum_{i=1}^n |y_i - \hat{y}| \quad (5.2)$$

Στους παραπάνω τύπους το y_i συμβολίζει την πραγματική τιμή της χρονοσειράς και το \hat{y} την πρόβλεψη του δικτύου για την τιμή της χρονοσειράς. Επίσης το παραπάνω άθροισμα πραγματοποιείται για όλα τα n δηλαδή όλα τα σημεία στα οποία το δίκτυο πραγματοποιεί πρόβλεψη. Να τονίσουμε εδώ ότι η πρόβλεψη που κάνουν τα δίκτυα είναι ενός βήματος και τα σημεία στα οποία έγινε πρόβλεψη στο προηγούμενο βήμα λαμβάνονται ως είσοδοι στο δίκτυο για να κάνει την επόμενη πρόβλεψη. Ο υπολογισμός των συναρτήσεων απόδοσης έγινε για το τελευταίο βήμα κατά το οποίο έχει πραγματοποιηθεί πρόβλεψη σε όλα τα σημεία.



Εικόνα 5.1: Περιγραφή υπολογιστικής διαδικασίας

5.2 Παρουσίαση των αλγορίθμων του πειράματος

Παρακάτω παρουσιάζονται οι αλγόριθμοι των οποίων η απόδοση θα αξιολογηθεί με την υπολογιστική διαδικασία. Τα βασικά σύμβολα που χρησιμοποιούνται σε αυτούς τους αλγορίθμους είναι τα εξής :

- **p**: Συμβολίζει τα σημεία που δίνονται στην είσοδο του δικτύου κατά την διαδικασία εκπαίδευσης του.
- **t**: Πρόκειται για το διάνυσμα που περιλαμβάνει τα σημεία στόχους που αντιστοιχούν στα σημεία εισόδου του διανύσματος **p**. Ίδιο διάνυσμα με το **t** είναι και το **tnit** που χρησιμοποιείται σε κάποιους από τους κώδικες
- **p1**: Περιλαμβάνει τα σημεία στα οποία το δίκτυο πραγματοποιεί πρόβλεψη. Στην ουσία προκύπτει από το άθροισμα του **p** και μίας μεταβλητής **FP** που καθορίζει τον αριθμό των σημείων για τα οποία το δίκτυο θα πραγματοποιήσει την πρόβλεψη.
- **i**: Είναι ο αριθμός των νευρώνων που χρησιμοποιούνται στο κρυφό στρώμα του δικτύου. Σε περίπτωση που το δίκτυο έχει δύο στρώματα το **i** εκφράζει τον αριθμό των νευρώνων στο πρώτο στρώμα ενώ για τους νευρώνες του δεύτερου στρώματος χρησιμοποιείται το σύμβολο **k**.

Παρακάτω παρατίθεται και ένας πίνακας που περιλαμβάνει τις βασικές παραμέτρους που περιλαμβάνει ένα νευρωνικό δίκτυο μετά τη δημιουργία του στο πρόγραμμα Matlab με την εξής εντολή δημιουργίας δικτύων MLP's:

```
net = newff(p,t,[i],{'trainbr'});
```

Παράμετροι	Τιμή
Αρχιτεκτονική δικτύου	MLP
Αριθμός κρυφών στρωμάτων	1
Αριθμός κρυφών νευρώνων	i
Συνάρτηση ενεργοποίησης κρυφών στρώματος	Εφαπτομενική σιγμοειδής
Συνάρτηση ενεργοποίησης στρώματος εξόδου	Γραμμική
Αλγόριθμος εκπαίδευσης δικτύου	Bayesian
Όριο Εποχών εκπαίδευσης	1000
Χρονικό όριο εκπαίδευσης	-
Συνάρτηση απόδοσης	MSE
Συνάρτηση αρχικοποίησης βαρών	initlay
Σφάλμα στόχος	0

Πίνακας 5.2: Προεπιλογές προγράμματος για τις παραμέτρους του δικτύου

Παρακάτω λοιπόν παρουσιάζονται οι αλγόριθμοι που εξετάστηκαν κατά τη παρούσα διπλωματική.

5.2.1 Αλγόριθμος 1: Ομαδική πρόβλεψη

Ο κώδικας του πρώτου από τους εξεταζόμενους αλγορίθμους παρατίθεται παρακάτω.

```
net = newff(p,t,[i],{},'trainbr');
net = train(net,p,t);

p1=p+FP;
tt=sim(net,p1);
plot (p,t); hold on; grid on; plot (p1,tt, 'r.')
```

Η λογική του παραπάνω κώδικα δεν απαιτεί κάποια ιδιαίτερη εξήγηση, διότι ο κώδικας αυτός δεν κάνει τίποτα παραπάνω από το να εφαρμόζει τις εντολές που παρέχονται από την Matlab για τη δημιουργία (newff) , την εκπαίδευση (train) και την εξαγωγή αποτελεσμάτων (sim) από το εκπαιδευμένο δίκτυο με την εφαρμογή ενός σετ σημείων στην είσοδό του. Έτσι το αρχικό δίκτυο που δημιουργείται εκπαιδεύεται με δεδομένα τα p,t και έπειτα το έτοιμο δίκτυο υπολογίζει τις εξόδους για το διάνυσμα t1 που του παρέχεται στην είσοδο. Με την τελευταία εντολή του κώδικα πραγματοποιείται η γραφική αναπαράσταση σε κοινό διάγραμμα της αρχικής χρονοσειράς καθώς και του αποτελέσματος από τον υπολογισμό του δικτύου για την περίπτωση του t1.

5.2.2 Αλγόριθμος 2: Ανεξάρτητης εισόδου- εξαρτημένης εξόδου

Ο κώδικας του δεύτερου αλγορίθμου είναι ο εξής:

```
net = newff(p,t,[i],{},'trainbr');
net = train(net,p,t);

tt=t; tnew=t;
for ITER=1:FP,
    p=p+1;
    tt=sim(net,p);

    if ITER>=2,
        tnew=[t (ITER+1:end),tnew(end-ITER+2:end),tt(end)];
    elseif ITER==1,
        tnew=[t (ITER+1:end),tt(end)];
    end
    net = newff(p,tnew,[i],{},'trainbr');
    net = train(net,p,tnew);

end
plot (pinit,t); hold on; grid on; plot (p,tnew, 'r.')
```

Στον συγκεκριμένο κώδικα αρχικά δημιουργείται το δίκτυο και προπονείται βάση των p,t . Έπειτα ακολουθεί μία επαναληπτική διαδικασία της οποίας στόχος είναι να μεταβάλλει το διάνυσμα p,tnew και σταδιακά να τα μεταφέρει προς το διάστημα στο οποίο θέλουμε να πραγματοποιηθεί η πρόβλεψη από το δίκτυο.

Πιο συγκεκριμένα, αρχικά αυξάνεται η τιμή των σημείων του διανύσματος p κατά 1 και έπειτα υπολογίζεται το αποτέλεσμα από την εφαρμογή του νέου p στην είσοδο του αρχικού δικτύου. Έπειτα υπάρχουν οι δύο εξής περιπτώσεις:

- Στην πρώτη επανάληψη της διαδικασίας ο πίνακας t_{new} θα περιλαμβάνει τα σημεία του αρχικού πίνακα t από το δεύτερο μέχρι το τελευταίο ενώ το τελευταίο σημείο θα συμπληρώνεται από το τελευταίο σημείο της προσομοίωσης του δικτύου με το καινούργιο p , δηλαδή το τελευταίο σημείο του πίνακα tt .
- Στις επόμενες επαναλήψεις ο πίνακας t_{new} αποτελείται από τα σημεία του πίνακα t τα οποία σε κάθε επανάληψη μειώνονται κατά 1, από τα τελικά σημεία του πίνακα t_{new} που προέκυψε από την προηγούμενη επαναληπτική διαδικασία και από το τελικό σημείο του πίνακα tt . Άρα σε κάθε επαναληπτική διαδικασία ο πίνακας t_{new} περιλαμβάνει όλο και περισσότερο τα σημεία που προέκυψαν μετά από προσομοίωση του καινούργιου κάθε φορά δικτύου με το νέο διάνυσμα p .

Τη δημιουργία του νέου t_{new} ακολουθεί η δημιουργία και η εκπαίδευση του δικτύου με τα νέα p και t_{new} . Αυτό γίνεται ώστε αυτό το νέο δίκτυο αφού αυξηθεί το p κατά ένα στην επόμενη επανάληψη να μπορέσει να πραγματοποιήσει νέα πρόβλεψη tt από την οποία θα πάρουμε το τελευταίο σημείο για να το προσθέσουμε στο παλιό t_{new} . Έτσι γίνεται προσπάθεια να εκμεταλλευτούμε την ιδιότητα της εσωτερικής γενίκευσης (το σημείο πρόβλεψης/γενίκευσης έχει κοντά του σημεία τα οποία ανήκουν στο σετ δεδομένων προπόνησης του δικτύου) και να επιτύχουμε υψηλής ακρίβειας προβλέψεις κάνοντας τις λεγόμενες προβλέψεις ενός βήματος (one step ahead forecasting)

Για την καλύτερη κατανόηση της λογικής της επαναληπτικής διαδικασίας παρακάτω παρατίθεται η δομή του t_{new} στις τρεις πρώτες επαναλήψεις. Με τον κάτω δείκτη επισημαίνεται κάθε φορά ο αριθμός της επανάληψης από την οποία προέκυψε το κάθε μέγεθος, ενώ ο πάνω δείκτης δηλώνει τη θέση του στοιχείου στο διάνυσμα στο οποίο ανήκει.

- Για $ITER=1$

$$p_1 = p_0 + 1$$

$$tt_1 = \text{sim}(\text{net}_0, p_1) = [tt_1^1 \quad tt_1^2 \quad \dots \quad tt_1^{\text{end}}]$$

$$t_{new_1} = [t_0(2:\text{end}), tt_1(\text{end})] = [t_0^2 \quad t_0^3 \quad \dots \quad t_0^{\text{end}} \quad tt_1^{\text{end}}]$$

$$\text{net}_1 = \text{net}_1(p_1, t_{new_1})$$

- Για $ITER=2$

$$p_2=p_1+1$$

$$tt_2=\text{sim}(\text{net}_1,p_2)= [tt_2^1 \quad tt_2^2 \dots tt_2^{\text{end}}]$$

$$tnew_2=[t_0(3:\text{end}),tnew_1(\text{end}:\text{end}), tt_2(\text{end})]=[t_0^3 \quad t_0^4 \dots t_0^{\text{end}} \quad tt_1^{\text{end}} \quad tt_2^{\text{end}}]$$

$$\text{net}_2=\text{net}_2(p_2, tnew_2)$$

- Για ITER=3

$$p_3=p_2+1$$

$$tt_3=\text{sim}(\text{net}_2,p_3)= [tt_3^1 \quad tt_3^2 \dots tt_3^{\text{end}}]$$

$$tnew_3=[t_0(4:\text{end}),tnew_2(\text{end}-1:\text{end}), tt_3(\text{end})]=[t_0^3 \quad t_0^4 \dots t_0^{\text{end}} \quad tt_1^{\text{end}} \quad tt_2^{\text{end}} \quad tt_3^{\text{end}}]$$

$$\text{net}_1=\text{net}_1(p_1, tnew_1)$$

Μετά το τέλος της επαναληπτικής διαδικασίας τα τελικά p και $tnew$ παριστάνονται γραφικά σε κοινό διάγραμμα με τα αρχικά p,t για να πραγματοποιηθεί σύγκριση.

Τέλος να επισημάνουμε ότι παρότι μεταβάλλονται τα στοιχεία του $tnew$ και του p ο συνολικός αριθμός (μέγεθος του κάθε διανύσματος) τους παραμένει ίδιος.

5.2.3 Αλγόριθμος 3: Προηγούμενης εξόδου- επόμενης εξόδου

Ο τρίτος κώδικας έχει ως εξής:

```
p=tinit(1:end-1); t=tinit(2:end);
net = newff(p,t,[i],{ },'trainbr');
net = train(net,p,t);
for ITER=1:FP,
    pnew=t; tt=sim(net,pnew);
    if ITER>=2,
        tnew=[tinit(ITER+2:end),tnew(end-ITER+2:end),tt(end)];p=t;
t=tnew;
    elseif ITER==1,
        tnew=[tinit(ITER+2:end),tt(end)] p=t; t=tnew;
    end
    net = newff(p,t,[i],{ },'trainbr');
    net = train(net,p,t);
end
plot(pinit,tinit); hold on; grid on; plot(pinit(2:end)+ITER,tnew,
'r.')
```

Στον συγκεκριμένο κώδικα επιχειρείται μία διαφορετική προσέγγιση στο θέμα της εκπαίδευσης των νευρωνικών δικτύων που χρησιμοποιούνται για πρόβλεψη χρονοσειρών. Υιοθετείται ο τρόπος με τον οποίο εκπαιδεύεται η πλειοψηφία των δικτύων που χρησιμοποιούνται, από τους συγγραφείς των περισσότερων διατριβών, για πρόβλεψη χρονοσειρών. Στη συγκεκριμένη περίπτωση επιχειρείται το δίκτυο να προπονηθεί με εισόδους παρελθοντικές τιμές της χρονοσειράς αντί για κάποια άλλη

μεταβλητή από την οποία εξαρτάται η τιμή της χρονοσειράς. Οπότε προσπαθούμε να εκπαιδεύσουμε το δίκτυο ώστε δεχόμενο στην είσοδο την τιμή της χρονοσειράς τη χρονική στιγμή $t-1$ να μας δίνει την προβλεπόμενη τιμή για τη χρονική στιγμή t .

Η δομή του τρίτου κώδικα είναι παρόμοια με του δεύτερου. Αρχικά δημιουργούνται το δίκτυο και τα διανύσματα p,t με τα οποία αυτό εκπαιδεύεται. Όπως φαίνεται και από τους τύπους τα αρχικά p,t παίρνουν τις τιμές του διανύσματος $tinit$. Το p έχει τις τιμές του $tinit$ από την αρχή μέχρι την προτελευταία τιμή του διανύσματος και το t παίρνει τις τιμές του $tinit$ από τη δεύτερη μέχρι την τελευταία. Έπειτα ακολουθεί μία επαναληπτική διαδικασία. Στην επαναληπτική διαδικασία αρχικά η μεταβλητή $pnew$ παίρνει την τιμή του t που προέκυψε από την προηγούμενη επανάληψη και έπειτα το δίκτυο (που έχει εκπαιδευτεί με τα προηγούμενα κάθε φορά p,t) με είσοδο το $pnew$ υπολογίζει το tt . Έπειτα υπάρχουν οι δύο εξής περιπτώσεις:

- Στην πρώτη επανάληψη της διαδικασίας η μεταβλητή $tnew$ σχηματίζεται από το αρχικό $tinit$ και από το διάνυσμα που δημιουργήθηκε από την προσομοίωση δηλαδή το tt . Έπειτα το p παίρνει την τιμή του αρχικού t και το νέο t παίρνει την τιμή του $tnew$ που δημιουργήθηκε στη συγκεκριμένη επανάληψη.
- Στις επόμενες επαναλήψεις σταδιακά τα στοιχεία του πίνακα $tinit$ που σχηματίζουν τον $tnew$ μειώνονται κατά ένα και τη θέση τους παίρνουν τα στοιχεία του $tnew$ από προηγούμενα βήματα, ενώ τον πίνακα συμπληρώνει και η τελευταία τιμή του πίνακα tt . Όμοια με την πρώτη επαναληπτική το p παίρνει την τιμή του t της προηγούμενης επανάληψης και το νέο t παίρνει την τιμή του νέου $tnew$.

Αυτή η επαναληπτική διαδικασία γίνεται με στόχο οι προβλέψεις να είναι ενός βήματος μπροστά και σταδιακά τα δεδομένα που λαμβάνονται για την εκπαίδευση του νέου δικτύου που δημιουργείται σε κάθε επαναληπτική, δηλαδή τα νέα p,t , να προχωρούν ώστε να καλύψουν όλο τα σημεία μέχρι το τελικό σημείο πρόβλεψης που καθορίζεται από το FP. Με αυτό τον τρόπο αποσκοπούμε ανανεώνοντας το σετ δεδομένων εκπαίδευσης του δικτύου, εκείνο να φτάσει σταδιακά μέχρι να περιλαμβάνει όλα τα σημεία πρόβλεψης μέχρι το προτελευταίο έτσι ώστε στην τελευταία επαναληπτική να γίνει η πρόβλεψη από το δίκτυο για το τελευταίο σημείο.

Για να γίνει πιο κατανοητή η λειτουργία του συγκεκριμένου κώδικα παρακάτω παρατίθενται οι τρεις πρώτες επαναλήψεις. Όπως και στον προηγούμενο αλγόριθμο, ο κάτω δείκτης συμβολίζει τον αριθμό της επανάληψης από την οποία προέκυψε το κάθε μέγεθος, ενώ ο πάνω δείκτης δηλώνει τη θέση του στοιχείου στο διάνυσμα στο οποίο ανήκει.

- Για $ITER=1$

$$pnew_{1=t_0} = tinit(2:end) = [tinit^2 \quad tinit^3 \quad \dots \quad tinit^{end}]$$

$$tt_1 = \text{sim}(\text{net}_0, pnew_1) = [tt_1^1 \quad tt_1^2 \quad \dots \quad tt_1^{end}]$$

$$tnew_1=[tinit(3:end),tt_1(end)]=[tinit^3 \ tinit^4 \ \dots \ tinit^{end} \ tt_1^{end}]$$

$$p_1=t_0=tinit(2:end)=[\ tinit^2 \ tinit^3 \ \dots \ tinit^{end} \]$$

$$t_1= tnew_1=[tinit^3 \ tinit^4 \ \dots \ tinit^{end} \ tt_1^{end}]$$

$$net_1=net_1(p_1, t_1)$$

- Για ITER=2

$$pnew_2=t_1=[tinit^2 \ tinit^3 \ \dots \ tinit^{end} \ tt_1^{end}]$$

$$tt_2=sim(net_1,pnew_2)= [\ tt_2^1 \ tt_2^2 \ \dots \ tt_2^{end} \]$$

$$tnew_2=[\ tinit(4:end), \ tnew_1(end: \ end), \ tt_2(end)]=[\ tinit^4 \ \dots \ tinit^{end} \ tt_1^{end} \ tt_2^{end} \]$$

$$p_2=t_1= tnew_1=[tinit^3 \ tinit^4 \ \dots \ tinit^{end} \ tt_1^{end}]$$

$$t_2= tnew_2=[tinit^4 \ \dots \ tinit^{end} \ tt_1^{end} \ tt_2^{end}]$$

$$net_2=net_2(p_2, t_2)$$

- Για ITER=3

$$pnew_3=t_2=[tinit^4 \ \dots \ tinit^{end} \ tt_1^{end} \ tt_2^{end}]$$

$$tt_3=sim(net_2,pnew_3)= [\ tt_3^1 \ tt_3^2 \ \dots \ tt_3^{end} \]$$

$$tnew_3=[\ tinit(5:end), \ tnew_1(end-1:end), \ tt_3(end)]=[\ tinit^5 \ \dots \ tinit^{end} \ tt_1^{end} \ tt_2^{end} \ tt_3^{end} \]$$

$$p_3=t_2= tnew_2=[tinit^4 \ \dots \ tinit^{end} \ tt_1^{end} \ tt_2^{end}]$$

$$t_3= tnew_3=[\ tinit^5 \ \dots \ tinit^{end} \ tt_1^{end} \ tt_2^{end} \ tt_3^{end} \]$$

$$net_2=net_2(p_2, t_2)$$

Όπως φαίνεται και από την περιγραφή των επαναλήψεων η πρόβλεψη που γίνεται στην n-1 επανάληψη χρησιμοποιείται ως δεδομένο προπόνησης στη n επανάληψη ώστε το δίκτυο να πραγματοποιήσει πρόβλεψη για το n σημείο.

Τέλος να σημειώσουμε ότι στις περισσότερες περιπτώσεις τις οποίες τα δίκτυα εκπαιδεύονται με παρελθοντικές τιμές της χρονοσειράς στην οποία θα πραγματοποιήσουν πρόβλεψη οι παρελθοντικές τιμές που παρέχονται ως είσοδοι στο δίκτυο είναι συνήθως περισσότερες της μίας. Επίσης οι τιμές που παρέχονται σε τέτοιες περιπτώσεις ως είσοδοι στο δίκτυο δεν είναι συνεχόμενες παρελθοντικές τιμές. Συνήθως εξετάζονται οι υπάρχουσες τιμές της χρονοσειράς και αναζητείται κάποια σύνδεση μεταξύ σημείων σε διαδοχικά διαστήματα ή κάποιο στοιχείο που να

καταδεικνύει ότι υπάρχει εποχικότητα ή τάση στην χρονοσειρά. Από αυτά τα στοιχεία καθορίζονται συνήθως ποιες παρελθοντικές τιμές θα δίνονται ως είσοδοι στο δίκτυο για να πραγματοποιήσει την ζητούμενη πρόβλεψη. Μία περίπτωση που βασίζεται στον παραπάνω τρόπο αντιμετώπισης και εκπαίδευσης του δικτύου θα μελετηθεί στον πέμπτο αλγόριθμο.

5.2.4 Αλγόριθμος 4: Χρήσης του σετ αποτελεσμάτων προσομοίωσης

Ο κώδικας του αλγορίθμου αυτού έχει ως εξής:

```
net = newff(p,t,[i],{},'trainbr');
net = train(net,p,t);

for ITER=1:FP,
    p=p+1;
    tt=sim(net,p);
    net = newff(p,tt,[i],{},'trainbr');
    net = train(net,p,tt);
end
plot (pinit,t); hold on; grid on; plot (p,tt, 'r.')
```

Στον συγκεκριμένο κώδικα επιχειρείται το δίκτυο να εκπαιδευτεί από τα αποτελέσματα που προέκυψαν από τη διαδικασία της προσομοίωσης. Κατά τη δημιουργία και την εκπαίδευση των δικτύων που δημιουργούνται στο συγκεκριμένο κώδικα, το αρχικό διάνυσμα t που περιλαμβάνει τις τιμές στόχους που αντιστοιχούν στα σημεία του διανύσματος εισόδου p , αντικαθίσταται από το διάνυσμα tt το οποίο είναι το αποτέλεσμα της προσομοίωσης που πραγματοποιεί το δίκτυο που δημιουργείται στην προηγούμενη επανάληψη με το p της τρέχουσας επανάληψης.

Έτσι στον συγκεκριμένο κώδικα δημιουργείται και εκπαιδεύεται το αρχικό νευρωνικό δίκτυο με τα αρχικά p,t και έπειτα ξεκινάει η επαναληπτική διαδικασία. Κατά την επαναληπτική διαδικασία οι τιμές των σημείων του διανύσματος p αυξάνονται κατά μία μονάδα και έπειτα πραγματοποιείται προσομοίωση της λειτουργίας του αρχικού δικτύου (ή αυτού που δημιουργήθηκε στην προηγούμενη επανάληψη) με το τρέχων διάνυσμα p . Το αποτέλεσμα αυτής της προσομοίωσης, δηλαδή το διάνυσμα tt χρησιμοποιείται ως διάνυσμα στόχος για τη δημιουργία και την εκπαίδευση ενός νέου δικτύου που έχει ως είσοδο το καινούργιο p που δημιουργήθηκε κατά τη διάρκεια της τρέχουσας επανάληψης.

Παρακάτω παρατίθενται τα τρία πρώτα βήματα της επαναληπτικής διαδικασίας για την καλύτερη κατανόηση του αλγορίθμου.

- Για $ITER=1$

$$p_1=p_0+1$$

$$tt_1=sim(net_0,p_1)=[tt_1^1 \ tt_1^2 \ \dots \ tt_1^{end}]$$

$$\text{net}_1 = \text{net}_1(p_1, \text{tt}_1)$$

- Για ITER=2
 $p_2 = p_1 + 1$

$$\text{tt}_2 = \text{sim}(\text{net}_1, p_2) = [\text{tt}_2^1 \text{tt}_2^2 \dots \text{tt}_2^{\text{end}}]$$

$$\text{net}_2 = \text{net}_2(p_2, \text{tt}_2)$$

- Για ITER=3
 $p_3 = p_2 + 1$

$$\text{tt}_3 = \text{sim}(\text{net}_2, p_3) = [\text{tt}_3^1 \text{tt}_3^2 \dots \text{tt}_3^{\text{end}}]$$

$$\text{net}_3 = \text{net}_3(p_3, \text{tt}_3)$$

Με τη διαδικασία που ακολουθεί ο παραπάνω αλγόριθμος στοχεύει στην εξέταση του κατά πόσο ένα δίκτυο μπορεί να αποδώσει ακριβείς προβλέψεις, επηρεαζόμενο από την ακρίβεια των προβλέψεων ενός άλλου δικτύου και επηρεάζοντας την ακρίβεια των προβλέψεων του δικτύου που ακολουθεί.

5.2.5 Αλγόριθμος 5: Λαμβάνοντας υπόψη την εποχικότητα

Ο κώδικας του συγκεκριμένου αλγορίθμου είναι ο εξής:

```

for per=1:s
    flag1=t(per:(per+s-1));
    flag2=t(per+s);
    tprop(:,per)=flag1;
    ttarger(per)=flag2;
end

net=newff(tprop,ttarger,[i],{},'trainbr');

for iter=1:1:FP
    net=train(net,tprop,ttarger);
    flag3=[tprop(2:end),end];ttarger(end)]
    tprop=[tprop flag3]
    tt=sim(net,tprop);
    ttarger=[ttarger,tt(end)];
    net=newff(tprop,ttarger,[i],{},'trainbr');
end

```

Όπως αναφέρθηκε και σε προηγούμενο κεφάλαιο της παρούσας διπλωματικής, έχει αποδειχτεί από τον Hornik ([3] και [4]) η ιδιότητα των προς τα εμπρός τροφοδοτούμενων δικτύων πολλαπλών στρωμάτων να μπορούν να προσεγγίσουν οποιαδήποτε συνάρτηση.

Επίσης σε μια χρονοσειρά είναι συνήθης η εμφάνιση κάποιων αναγνωρίσιμων μορφών που επαναλαμβάνονται ανά συγκεκριμένα χρονικά διαστήματα και αποτελούν δείγμα ότι η «κίνηση» του συγκεκριμένου μεγέθους που εκφράζεται μέσω της χρονοσειράς παρουσιάζει κάποια εποχικότητα

Στον συγκεκριμένο κώδικα λοιπόν επιχειρείται να δημιουργηθεί μία κατάλληλη αρχιτεκτονική νευρωνικού δικτύου με την οποία θα γίνεται προσπάθεια να αντιμετωπιστεί το πρόβλημα της πρόβλεψης τιμών χρονοσειρών ως πρόβλημα προσέγγισης συναρτήσεων, χωρίς να χρειάζεται τα δεδομένα που θα δοθούν στο δίκτυο κατά την εκπαίδευση να υποστούν κάποια προεργασία για να απαλλαγούν από το στοιχείο της εποχικότητας.

Η παράμετρος s η οποία χρησιμοποιείται στον συγκεκριμένο αλγόριθμο έχει να κάνει με τον αριθμό των χρονικών περιόδων που εμφανίζεται η εποχικότητα. Πιο συγκεκριμένα το s συμβολίζει τον αριθμό των περιόδων που μεσολαβούν μέχρι σε κάποια χρονοσειρά να εμφανιστούν δύο διαδοχικά σημεία που θα έχουν την ίδια περίπου τιμή. Επιπλέον η παράμετρος αυτή συμβάλλει στον προσδιορισμό του αριθμού εισόδων που θα υπάρχουν στο εκπαιδευόμενο δίκτυο.

Όπως και στους προηγούμενους αλγορίθμους έτσι και σε αυτόν οι προβλέψεις που πραγματοποιεί το δίκτυο είναι ενός βήματος ($t+1$). Επίσης όπως και στον τρίτο κώδικα έτσι και εδώ το δίκτυο ως εισόδους για την εκπαίδευση και την πραγματοποίηση προβλέψεων λαμβάνει παλιές τιμές της χρονοσειράς. Η βασική διαφορά στο συγκεκριμένο κώδικα είναι ότι ο αριθμός των εισόδων θα είναι ίσος με τον παράγοντα s και οι εισοδοί στο δίκτυο θα αντιστοιχούν στις τιμές της χρονοσειράς από τη χρονική στιγμή $t-s+1$ έως την στιγμή t . Αυτή είναι η πρώτη ιδιαιτερότητα του πέμπτου αλγορίθμου.

Η δεύτερη ιδιαιτερότητά του έχει να κάνει με τον τρόπο που προπονείται το δίκτυο. Ο πίνακας με τα στοιχεία που χρησιμοποιούνται ως εισοδοί στο δίκτυο είναι ο $trgop$ ο οποίος έχει διαστάσεις $(s, N-s)$ με N να είναι ο αριθμός σημείων στο σετ δεδομένων. Εδώ να διευκρινιστεί ότι οι συνδυασμοί που θα χρησιμοποιούνται στην είσοδο είναι $N-s$ και όχι $N-s+1$ γιατί το τελευταίο σημείο του πίνακα δεδομένων δεν θα χρησιμοποιείται ως είσοδο αλλά ως το τελευταίο σημείο του πίνακα στόχων πρόβλεψης. Αυτός λοιπόν ο πίνακας σε κάθε επαναληπτική διαδικασία ανανεώνεται κατά μία στήλη της οποίας το τελευταίο σημείο είναι η πρόβλεψη που έγινε στο προηγούμενη επανάληψη με τη χρήση του sim . Όμως σε κάθε επανάληψη σε αυτόν τον πίνακα προστίθεται μία στήλη χωρίς να αφαιρείται κάποια. Έτσι στην ουσία επιτυγχάνεται παράλληλα και μία αύξηση του σετ δεδομένων στο οποίο προπονείται το δίκτυο.

Πιο αναλυτικά στον κώδικα η πρώτη επαναληπτική διαδικασία χρησιμοποιείται για τη δημιουργία των αρχικών $trgop$ και $ttarget$ (πίνακας εισόδου και διάνυσμα στόχος αντίστοιχα) που θα χρησιμοποιηθούν για την εκπαίδευση του δικτύου. Στη δεύτερη επαναληπτική διαδικασία που ακολουθεί γίνονται διαδοχικά τα εξής:

1. Το δίκτυο που έχει δημιουργηθεί προπονείται με τα tprop και ttarget
2. Δημιουργείται η τελευταία στήλη που θα μπει στο tprop
3. Δημιουργείται το νέο tprop με την προσθήκη αυτής της στήλης
4. Γίνεται simulate στο δίκτυο με το καινούργιο tprop
5. Παίρνουμε το τελευταίο σημείο του simulate που είναι και αυτό στο οποίο κάθε φορά γίνεται η πρόβλεψη και το προσθέτουμε στο διάνυσμα του ttarget.
6. Δημιουργούμε νέο δίκτυο με τα καινούργια tprop και ttarget.

Για καλύτερη κατανόηση παρατίθενται οι τρεις πρώτες επαναλήψεις της δεύτερης επαναληπτικής διαδικασίας για σετ δεδομένων με N=40 και s=20.

Έστω ότι τα αρχικά tprop και ttarget με τα οποία δημιουργείται και εκπαιδεύεται το δίκτυο έχουν ως εξής:

$$tprop_0 = \begin{bmatrix} t^1 & \dots & t^{20} \\ \vdots & \ddots & \vdots \\ t^{20} & \dots & t^{39} \end{bmatrix} \quad \text{και} \quad ttarget_0 = [t^{21} \dots t^{40}]$$

- Για ITER=1

net₀=train(net₀, tprop₀, ttarget₀)

$$flag3 = [tprop_0((2:end),end); ttarget_0(end)] = \begin{bmatrix} & t^{21} \\ & \vdots \\ & t^{40} \end{bmatrix}$$

$$tprop_1 = [tprop_0 \quad flag3] = \begin{bmatrix} t^1 & \dots & t^{21} \\ \vdots & \ddots & \vdots \\ t^{20} & \dots & t^{40} \end{bmatrix}$$

$$tt_1 = \text{sim}(\text{net}_0, tprop_1) = [tt_1^{21} \dots tt_1^{41}]$$

$$ttarget_1 = [ttarget_0, tt_1(\text{end})] = [t^{21} \dots t^{40} \quad tt_1^{41}]$$

net₁=net₁(tprop₁, ttarget₁)

- Για ITER=2

net₁=train(net₁, tprop₁, ttarget₁)

$$flag3 = [tprop_1((2:end),end); ttarget_1(end)] = \begin{bmatrix} & t^{22} \\ & \vdots \\ & tt_1^{41} \end{bmatrix}$$

$$tprop_2 = [tprop_1 \quad flag3] = \begin{bmatrix} t^1 & \dots & t^{22} \\ \vdots & \ddots & \vdots \\ t^{20} & \dots & tt_1^{41} \end{bmatrix}$$

$$tt_2 = \text{sim}(\text{net}_1, tprop_2) = [tt_2^{21} \dots tt_2^{42}]$$

$$ttarget_2 = [ttarget_1, tt_2(\text{end})] = [t_{21} \dots t_{40} \quad tt_1^{41} \quad tt_2^{42}]$$

net₂=net₂(tprop₂, ttarget₂)

- Για ITER=3

net₂=train(net₂, tprop₂, ttarget₂)

$$\begin{aligned}
 \text{flag3} &= [\text{tprop}_2((2:\text{end}),\text{end});\text{ttarget}_2(\text{end})] = \begin{bmatrix} t^{23} \\ \vdots \\ tt_2^{42} \end{bmatrix} \\
 \text{tprop}_2 &= [\text{tprop}_1 \text{ flag3}] = \begin{bmatrix} t^1 & \dots & t^{23} \\ \vdots & \ddots & \vdots \\ t^{20} & \dots & tt_2^{42} \end{bmatrix} \\
 \text{tt}_3 &= \text{sim}(\text{net}_1, \text{tprop}_2) = [tt_3^{21} \dots tt_3^{43}] \\
 \text{ttarget}_2 &= [\text{ttarget}_1, \text{tt}_1(\text{end})] = [t_{21} \dots t_{40} \quad tt_1^{41} \quad tt_2^{42} \quad tt_3^{43}] \\
 \text{net}_3 &= \text{net}_3(\text{tprop}_3, \text{ttarget}_3)
 \end{aligned}$$

5.3 Σύνοψη πειραματικής διαδικασίας

Οι πέντε αλγόριθμοι που παρουσιάστηκαν παραπάνω θα εξεταστούν για τη δυνατότητα τους να βελτιστοποιούν την απόδοση του δικτύου κατά τη διαδικασία της πρόβλεψης.

Συνοπτικά η συνολική διαδικασία του πειράματος και της αξιολόγησης θα περιλαμβάνει την εκτέλεση των πειραμάτων με τη χρήση των αλγορίθμων για όλους τους συνδυασμούς των εξής παραμέτρων:

- Αριθμό κρυφών στρωμάτων νευρώνων
- Αριθμό νευρώνων σε κάθε κρυφό στρώμα
- Αριθμό σημείων στο σετ δεδομένων εκπαίδευσης (μόνο για τον πρώτο κώδικα)

Η εξέταση αυτή των αλγορίθμων θα πραγματοποιηθεί σε μία συνθετική χρονοσειρά.

Μετά την εκτέλεση των υπολογισμών τα αποτελέσματα που προκύπτουν θα αναλυθούν με χρήση αλγορίθμων που θα παρουσιαστούν στο επόμενο κεφάλαιο βάση δύο συναρτήσεων απόδοσης (MSE, MAE) στα σημεία που πραγματοποιήθηκε η πρόβλεψη. Επιπρόσθετα εξετάζονται και οι γραφικές παραστάσεις για να διαπιστωθεί η ακρίβεια των προβλέψεων καθώς σε ορισμένες περιπτώσεις η χρήση μόνο των συναρτήσεων απόδοσης δεν οδηγεί σε ασφαλή συμπεράσματα.

Βάση αυτών επιλέγεται ο αλγόριθμος ο οποίος παρουσιάζει την καλύτερη απόδοση και συγκεκριμένα το δίκτυο εκείνο που δίνει τα καλύτερα αποτελέσματα. Τέλος το συγκεκριμένο δίκτυο θα χρησιμοποιηθεί για πρόβλεψη αποτελεσμάτων μίας χρονοσειράς με δεδομένα από τον πραγματικό κόσμο και θα αξιολογηθούν τα αποτελέσματα που θα προκύψουν.

ΚΕΦΑΛΑΙΟ 6

ΤΟ ΥΠΟΛΟΓΙΣΤΙΚΟ ΜΟΝΤΕΛΟ

6.1 Παρουσίαση των χρησιμοποιούμενων δεδομένων

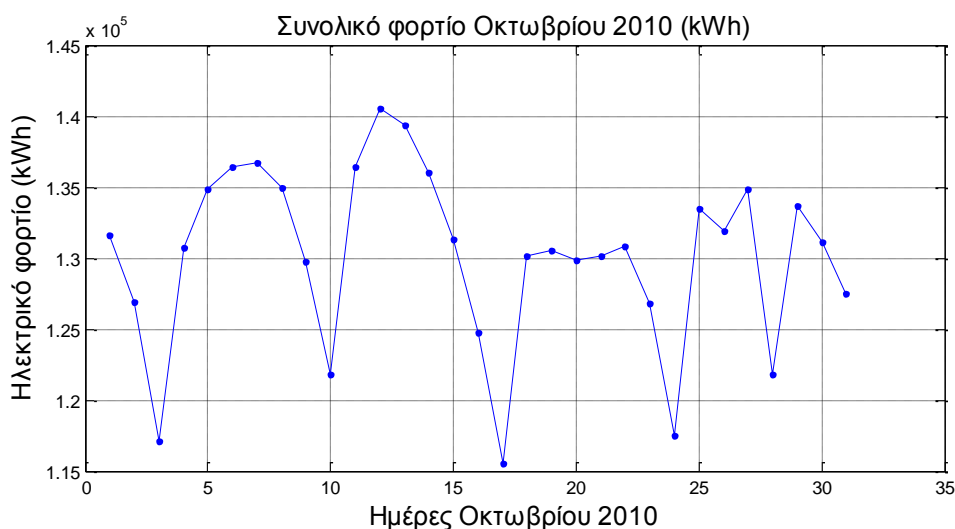
Όπως αναφέρθηκε και στο προηγούμενο κεφάλαιο οι αλγόριθμοι θα εξεταστούν σε μία τεχνητή χρονοσειρά και ο αλγόριθμος που παρουσιάζει την καλύτερη απόδοση θα εφαρμοστεί και στην πρόβλεψη σημείων χρονοσειράς που τα δεδομένα της προέρχονται από τον πραγματικό κόσμο

Έτσι στο πρώτο στάδιο τα δεδομένα που θα χρησιμοποιηθούν προέρχονται από τη συνάρτηση του ημιτόνου. Πιο συγκεκριμένα σε κάθε αλγόριθμο θα δίνονται 40 σημεία της συνάρτησης του ημιτόνου τα οποία καλύπτουν ομοιόμορφα διάστημα περίπου δύο περιόδων $[0, 1\pi, 4\pi]$. Στον πρώτο αλγόριθμο θα δοθούν επίσης 160 και 320 σημεία τα οποία είναι θα καταναμημένα ομοιόμορφα στο ίδιο διάστημα.

Το ημίτονο παρουσιάζει συγκεκριμένα χαρακτηριστικά στη γραφική του παράσταση καθώς μπορούμε να πούμε ότι εμφανίζει εποχικότητα η οποία εδώ εμφανίζεται κάθε 20 σημεία. Αυτό θα πρέπει να ληφθεί υπόψη κατά την διαδικασία της εκπαίδευσης.

Στο δεύτερο στάδιο, αφού έχουν επιλεγεί οι πιο αποδοτικοί από τους αλγορίθμους που θα έχουν εξεταστεί, τα δεδομένα που θα χρησιμοποιηθούν είναι οι ωριαίες μετρήσεις του συνολικού φορτίου (παραγόμενου και εισαγόμενου) στην Ελλάδα για το έτος 2010, τα οποία είναι διαθέσιμα στην ιστοσελίδα του Λειτουργού ΑΓΟράς Ηλεκτρικής Ενέργειας (ΛΑΓΗΕ). Αυτές οι μετρήσεις καλύπτουν τις ημερομηνίες από 01/01/2010 μέχρι και 31/12/2010 και είναι στο σύνολό τους 8760.

Από αυτές τις μετρήσεις θα χρησιμοποιηθούν μόνο εκείνες που αφορούν το μήνα Οκτωβρίου για να πραγματοποιηθούν τα πειράματα για την αξιολόγηση της απόδοσης του επιλεγμένου αλγορίθμου σε χρονοσειρά πραγματικών τιμών. Η επιλογή του μήνα αυτού έγινε γιατί είναι ένας μήνας στον οποίο δεν υπάρχουν αργίες ή διακοπές, διαστήματα στα οποία οι τιμές του συνολικού φορτίου ξεφεύγουν από τα συνήθη όριά τους. Παρακάτω δίνεται η γραφική παράσταση του ημερήσιο συνολικού φορτίου για όλες τις μέρες του συγκεκριμένου μήνα.



Εικόνα 6.1: Ημερήσιο συνολικό φορτίο Οκτωβρίου 2010 (kWh)

Τα συνολικά δεδομένα με τις ωριαίες τιμές για τις πρώτες 20 μέρες του Οκτωβρίου δίνονται στο Παράρτημα Α στο τέλος της διπλωματικής.

Στη γραφική παράσταση παρατηρούμε ότι τα ημερήσια συνολικά φορτία ρεύματος παρουσιάζουν ένα παρεμφερές προφίλ που επαναλαμβάνεται σε εβδομαδιαία βάση (η 1 Οκτωβρίου 2010 ήταν Παρασκευή) με την κατανάλωση να ακολουθεί ανοδική πορεία τις πρώτες μέρες τις εβδομάδας και καθοδική πορεία της τελευταίες. Εξαιρέση αποτελεί η τελευταία εβδομάδα στην οποία παρουσιάζεται μία απότομη πτώση του συνολικού φορτίου την Πέμπτη 28 Οκτωβρίου, η οποία έχει να κάνει με το γεγονός ότι η μέρα εκείνη είναι εθνική εορτή. Στοιχείο τάσης δεν φαίνεται να υπάρχει στα συγκεκριμένα δεδομένα.

Τέλος να επισημάνουμε ότι συνήθης διαδικασία κατά το στάδιο της εκπαίδευσης ενός δικτύου είναι να χωρίζεται το σετ δεδομένων σε δύο (ενίοτε και σε τρία) ξεχωριστά σετ. Το πρώτο (συνήθως ονομάζεται training set) από αυτά χρησιμοποιείται αποκλειστικά για την εκπαίδευση του δικτύου. Αντίθετα το δεύτερο (συνήθως ονομάζεται test set) δεν επηρεάζει καθόλου τη διαδικασία εκπαίδευσης και χρησιμοποιείται αφού εκπαιδευτεί το δίκτυο για να ελεγχθεί η απόδοσή του, καθώς αποτελεί για το δίκτυο ένα άγνωστο σετ δεδομένων. Αρκετές φορές το σετ δεδομένων χωρίζεται και σε ένα τρίτο σετ το οποίο ονομάζεται validation set. Αυτό χρησιμοποιείται για να συμβάλλει στη διακοπή του αλγορίθμου σε περίπτωση διαδοχικής αύξησης της τιμής της συνάρτησης απόδοσης για ένα αριθμό βημάτων. Εδώ επειδή χρησιμοποιούμε τον Bayesian αλγόριθμο εκπαίδευσης κάτι τέτοιο δεν είναι απαραίτητο.

Επιλέχθηκε το σετ δεδομένων εκπαίδευσης για το πρώτο στάδιο να αποτελείται και από τα 40 σημεία που του διαστήματος που καλύπτει το ημίτονο. Αυτό έγινε γιατί γνωρίζουμε την απόκριση του ημιτόνου και για τα σημεία πέραν αυτών των 40. Το test set επιλέχθηκε να περιλαμβάνει FP=15 σημεία στα οποία θα πραγματοποιείται πρόβλεψη. Να επισημάνουμε ότι για λόγους ευκολίας υπολογισμών το test set έχει και αυτό 40 σημεία εκ των οποίων τα 25 πρώτα είναι γνωστά στο δίκτυο. Αυτό έγινε για έλεγχο της απόκρισης του δικτύου για εισόδους γνωστές σε αυτό και για σύγκριση του με αντίστοιχα αποτελέσματα σε άγνωστα σε αυτό στοιχεία. Έτσι τα σετ έχουν ως εξής (με p συμβολίζεται το διάνυσμα εισόδου και με t το αντίστοιχο διάνυσμα στόχων) :

- Training set:
 $p=[1, 2, 3, \dots, 40]$
 $t = \sin(0.1 * p * \pi) = [0.3090, 0.5878, 0.8090, \dots 0]$
- Test set
 $p_1=p+15=[16, 17, 18, \dots, 55]$
 $t_1 = \sin(0.1 * p_1 * \pi) = [-0.9511, - 0.8090, - 0.5878, \dots - 1]$

Αναλυτικοί πίνακες με τα σημεία υπάρχουν στο Παράρτημα Α

6.2 Παρουσίαση υπολογιστικής διαδικασίας

Οι υπολογισμοί για την πραγματοποίηση των πειραμάτων και την αξιολόγηση των αποτελεσμάτων πραγματοποιήθηκαν σε τέσσερα βήματα εκ των οποίων στα τρία πρώτα αντιστοιχεί και ένα τμήμα υπολογιστικού κώδικα:

1. Εκπαίδευση και πραγματοποίηση πρόβλεψης από τα δίκτυα για όλους τους πιθανούς συνδυασμούς των παραμέτρων που μεταβάλλονταν.
2. Υπολογισμό της τιμής των συναρτήσεων απόδοσης στα σημεία στα οποία το δίκτυο πραγματοποίησε προβλέψεις.
3. Ταξινόμηση των αποτελεσμάτων με βάση τις τιμές που υπολογίστηκαν στο δεύτερο βήμα.
4. Παρατήρηση των γραφικών παραστάσεων που για την απόκριση του δικτύου για διόρθωση τυχόν λαθών και επιβεβαίωση των αποτελεσμάτων

Παρακάτω θα γίνει περιγραφή του κάθε βήματος και θα δοθούν τα βασικά σημεία από τους κώδικες σε κάθε βήμα που αυτοί χρησιμοποιήθηκαν.

• **Βήμα 1^ο: Εκπαίδευση και πραγματοποίηση προβλέψεων:**

Αποτελεί το βασικό βήμα της διπλωματικής καθώς σε αυτό πραγματοποιήθηκε η πλειοψηφία των απαραίτητων υπολογισμών για την εκπαίδευση και την πραγματοποίηση προβλέψεων από τα δίκτυα.

Σύμφωνα με τον Balestrassi [37] : <<...***Ο σωστός προγραμματισμός αποτρέπει την ανάγκη για την απόδοση συγκεκριμένων αριθμητικών τιμών εντός του κώδικα. Αντίθετα ο κώδικας πρέπει να διαβάζει τις τιμές των παραμέτρων, έτσι ώστε το πρόγραμμα να μπορεί να τρέχει για πολλούς συνδυασμούς των τιμών...***>>.

Ο κώδικας που δημιουργήθηκε για αυτό το βήμα, βασίστηκε στην παραπάνω αρχή και επιδιώχθηκε όσο το δυνατόν οι μεταβολές των τιμών των παραμέτρων να γίνονται αυτόματα μέσα στον κώδικα. Έτσι ο αριθμός των νευρώνων για κάθε κρυφό στρώμα μεταβάλλεται μόνος του κατά τη διάρκεια εκτέλεσης του κώδικα. Εδώ να παρατηρήσουμε ότι έχει γραφεί διαφορετικός κώδικας για τα δίκτυα δύο νευρώνων καθώς ο συγγραφέας θεώρησε καλύτερο να μην γίνεται η αλλαγή από ένα σε δύο κρυφά στρώματα αυτόματα από τον κώδικα, αλλά οι κώδικες για ένα και δύο κρυφά στρώματα είναι παρεμφερείς στις βασικές λειτουργίες που επιτελούν.

Παρακάτω παρατίθεται το βασικότερο μέρος του κώδικα ο οποίο εφαρμόστηκε για δίκτυο νευρώνων δύο κρυφών στρωμάτων με χρήση του δεύτερου από τους πέντε αλγόριθμους. Να σημειωθεί ότι οι κώδικες για τους υπόλοιπους αλγόριθμους είναι όμοιοι στις βασικές τους λειτουργίες με τον συγκεκριμένο κώδικα και έχουν υποστεί μικρές διαφοροποιήσεις για να είναι συμβατοί με κάθε αλγόριθμο. Ολόκληρος ο κώδικας βρίσκεται στο Παράρτημα Β. Εδώ να σημειώσουμε ότι τα `nnmin1`, `nnmin2`, `nnmax1`, `nnmax2` είναι ο ελάχιστος και ο μέγιστος αριθμός νευρώνων στα δύο κρυφά στρώματα και έχουν δοθεί από τον χρήστη προηγουμένως.

```

arithmos_dokimhs=0;
for i=nnmin1:1:nnmax1
    for k=nnmin2:1:nnmax2
        tout1=cell(1,10);
        net1=cell(1,10);
        fitting_error1=cell(1,10);
        fitting_mse=zeros(1,10);
        fitting_mae=zeros(1,10);
        tt1=cell(1,10);
        forecast_error1=cell(1,10);
        forecast_mse=zeros(1,10);
        forecast_mae=zeros(1,10);
        tr1=cell(1,10);
        net=newff(p,t,[i k],{'','trainbr'});
        for j=1:1:10
            arithmos_dokimhs=arithmos_dokimhs+1;
            disp('Dokimh= ')
            disp(arithmos_dokimhs)
            net=train(net,p,t);
            tsec=t; tnew=t; psec=p;
            for iter=1:1:FP
                psec=psec+1;
                tsec=sim(net,psec);
                if iter==1
                    tnew=[t(iter+1:end),tsec(end)];
                elseif iter>=2
                    tnew=[t(iter+1:end),tnew(end-...
...iter+2:end),tsec(end)];
                end
                net = newff(psec,tnew,[i],{'','trainbr'});
                [net,tr] = train(net,psec,tnew);
            end
            net1{j}=net;
            tr1{j}=tr;
            tout=sim(net,p);
            tout1{j}=tout;
            fitting_error=t-tout;
            fitting_error1{j}=fitting_error;
            fitting_mse(j)=mse(fitting_error);
            fitting_mae(j)=errperf(t,tout,'mae');
            tt=sim(net,p1);
            tt1{j}=tt;
            forecast_error=t1-tt;
            forecast_error1{j}=forecast_error;
            forecast_mse(j)=mse(forecast_error);
            forecast_mae(j)=errperf(t1,tt,'mae');
            plot(p,t); hold on; grid on; plot(p1,tt, 'r. ');hold...
... on;plot(p,tout,'g+');hold on;
            title(strcat('Dokimh=',num2str(arithmos_dokimhs),'
Ar.Arxicop=',num2str(j),' Arithmos neurons=',num2str(i),'-
',num2str(k)))

saveas(1,strcat('C:\Users\Billy\Desktop\Dok\2_code\2_hidden_layers\'
...num2str(i),'-',num2str(k),strcat('\figure',num2str(j))));
        net=init(net);
        hold off
        clear psec tsec tnew
    end
end
end
end

```

Ο κώδικας λοιπόν βασίζεται σε τέσσερις βρόγχους επανάληψης :

1. Οι πρώτοι δύο αφορούν την μεταβολή του αριθμού των νευρώνων στο πρώτο (**i**) και το δεύτερο (**k**) κρυφό στρώμα του δικτύου. Έτσι ο κώδικας τρέχει για όλους τους βασικούς συνδυασμούς των i, k ανάλογα με όρια που έχουν δοθεί από τον χρήστη. Σε αυτό το βρόγχο δημιουργείται το δίκτυο (τελευταία εντολή πριν από τον εμφωλευμένο βρόγχο j) και οι βασικοί πίνακες (matrices) και συστοιχίες (arrays) για τα δεδομένα που θα κρατηθούν για περαιτέρω επεξεργασία. Οι πίνακες και οι συστοιχίες που δημιουργούνται είναι οι εξής:
 - **fitting_error1, forecast_error1**: Είναι συστοιχίες (arrays) στις οποίες αποθηκεύονται οι πίνακες σφαλμάτων που δημιουργούνται για κάθε μία από τις 10 αρχικοποιήσεις που γίνονται για κάθε δίκτυο. Οι τύποι για τον υπολογισμό κάθε ενός από τους πίνακες των συστοιχιών αυτών θα αναλυθούν στην παρουσίαση του εμφωλευμένου βρόγχου αρχικοποιήσεων.
 - **fitting_mse/mae, forecast_mse/mae**: Οι πίνακες στους οποίους υπολογίζονται τα MSE, MAE για κάθε ένα από τα σετ δεδομένων p εκπαίδευσης (fitting) πρόβλεψης $p1$ (forecasting) και για κάθε μία από τις 10 αρχικοποιήσεις του δικτύου.
 - **tout1, tt1**: Οι συστοιχίες που αποθηκεύονται οι πίνακες με τα αποτελέσματα από την προσομοίωση του προπονημένου δικτύου με τα p και $p1$ αντίστοιχα.
 - **net1, tr1**: Συστοιχίες στις οποίες αποθηκεύεται το προπονημένο δίκτυο καθώς και τα στοιχεία που αφορούν τη διαδικασία εκπαίδευσης του, όπως αριθμός εποχών, τιμή συνάρτησης απόδοσης και άλλα.
2. Ο εμφωλευμένος βρόγχος των δέκα αρχικοποιήσεων που γίνονται στα αρχικά βάρη του δικτύου πριν από την εκπαίδευση για λόγους που αφορούν την αποτελεσματικότητα της διαδικασίας της εκπαίδευσης και έχουν να κάνουν με την επιφάνεια σφάλματος (υποπαράγραφος 3.3.1). Σε αυτό το βρόγχο εμφωλευμένος είναι ο κώδικας που έχει τον υπό εξέταση αλγόριθμο. Επίσης σε αυτό το βρόγχο κάθε ένα από τα αρχικοποιημένα δίκτυα εκπαιδεύεται και μετά την εφαρμογή του κάθε αλγορίθμου υπολογίζονται και αποθηκεύονται στους αντίστοιχους πίνακες και συστοιχίες τα εξής:
 - **tout, tt**: Τα αποτελέσματα της προσομοίωσης του εκπαιδευμένου δικτύου με τα σετ δεδομένων p και $p1$ αντίστοιχα.
 - **fitting_error, forecast_error**: Υπολογίζονται ως η διαφορά των πραγματικών τιμών-στόχων εκπαίδευσης (t) και πρόβλεψης ($t1$) από τις τιμές $tout$ και tt αντίστοιχα.
 - **fitting_mae/mse, forecast_mae/mse**: Οι τιμές των MAE και MSE οι οποίες υπολογίζονται βάση των `fitting_error` και `forecast_error`

Τέλος σε αυτό το βρόγχο δημιουργείται η γραφική παράσταση στο ίδιο διάγραμμα των προβλέψεων του δικτύου για το $p1$ και των πραγματικών τιμών του για το p και αρχικοποιείται το δίκτυο

3. Ο τελευταίος εμφωλευμένος βρόγχος είναι εκείνος που αφορά τους αλγορίθμους υπό εξέταση και είναι αυτός που μεταβάλλεται ανάλογα με τον υπό εξέταση κώδικα.

• **Βήμα 2^ο: Υπολογισμός συναρτήσεων απόδοσης στα σημεία πρόβλεψης.**

Ο συγκεκριμένος κώδικας έγινε για την επεξεργασία ορισμένων στοιχείων που θα συμβάλλουν μετά στην ταξινόμηση των αποτελεσμάτων. Πιο συγκεκριμένα όπως είδαμε και προηγουμένως το `forecast_error` υπολογίζεται για όλα τα σημεία του διαστήματος `p1` το οποίο αποτελεί το διάστημα στο οποίο ζητάμε από το δίκτυο να πραγματοποιήσει προβλέψεις. Το `p1` όμως περιέχει και στοιχεία τα οποία περιέχει και το `p` που είναι το σετ με τα δεδομένα προπόνησης του δικτύου πράγμα που μπορεί να οδηγήσει σε σύγχυση κατά την ανάγνωση των αποτελεσμάτων των συναρτήσεων απόδοσης. Για αυτό το λόγο δημιουργήθηκε αυτός ο κώδικας με τον οποίο διαλέγονται από κάθε πίνακα της συστοιχίας `forecast_error` τα σφάλματα εκείνα που αφορούν τα σημεία τα οποία δεν παρουσιάστηκαν στο δίκτυο κατά τη διαδικασία της εκπαίδευσης και βάση αυτών υπολογίζονται οι τιμές των συναρτήσεων απόδοσης MAE και MSE. Το βασικό μέρος του κώδικα του συγκεκριμένου βήματος για την περίπτωση δικτύου δύο κρυφών στρωμάτων που αφορά τον αλγόριθμο 2 παρουσιάζεται στο παράρτημα Γ μαζί με μία συνοπτική περιγραφή της λειτουργίας του.

• **Βήμα 3^ο : Ταξινόμησης των αποτελεσμάτων:**

Ο κώδικας που αντιστοιχεί στο βήμα αυτό στην ουσία κατατάσσει τα δίκτυα που υπολογίστηκαν σε κάποιες κατηγορίες βάση των μεγεθών `fitting_mae/mse`, `actual_forecast_mae/mse`. Οι τιμές αυτές που αποτελούν τα όρια των διαφόρων κατηγοριών στις οποίες χωρίζονται τα δίκτυα βάση αυτού του κώδικα επιλέχθηκαν από το συγγραφέα της διπλωματικής έπειτα από παρατήρηση των τιμών που προσεγγίζουν αυτά τα μεγέθη όταν το δίκτυο παρουσιάζει συγκεκριμένα στοιχεία κατά την απόκρισή του στα σημεία πραγματικής πρόβλεψης (`actual_forecast_mae/mse`) και στα σημεία τα οποία του δόθηκαν κατά την εκπαίδευσή του. Οι κατηγορίες στις οποίες ταξινομούνται τα δίκτυα από τον συγκεκριμένο αλγόριθμο είναι οι εξής:

- **Καλής πρόβλεψης (good forecasting):** Σε αυτή ανήκουν τα δίκτυα, η απόκριση των δικτύων είναι πολύ καλή και στο διάστημα στο οποίο υπάρχουν τα σημεία εκπαίδευσης αλλά και στο διάστημα κατά το οποίο το δίκτυο καλείται να αποκριθεί σε σημεία άγνωστα σε αυτό. Σε αυτές τις περιπτώσεις οι τιμές των τεσσάρων μεγεθών που αποτελούν τα κριτήρια της ταξινόμησης είναι χαμηλές
- **Καλής προσαρμογής (good fitting) :** Η απόκριση των συγκεκριμένων δικτύων είναι καλή στα σημεία τα οποία έχουν δοθεί στο δίκτυο και κατά τη διαδικασία της εκπαίδευσης. Αντίθετα το δίκτυο δεν καταφέρνει να αποκριθεί καλά στα σημεία πρόβλεψης. Αυτό συνήθως οφείλεται στο φαινόμενο της υπερβολικής προσαρμογής στα σημεία εκπαίδευσης. Σε αυτές τις περιπτώσεις

τα μεγέθη `fitting_mae/mse` παρουσιάζουν χαμηλές τιμές και τα μεγέθη `actual_forecast_mae` και `actual_forecast_mse` πιο υψηλές.

- **Κακής προσαρμογής (bad fitting)**: Σε αυτή την κατηγορία τοποθετούνται τα δίκτυα τα οποία κατά την απόκρισή τους δεν καταφέρνουν να προσεγγίσουν ούτε τα σημεία στα οποία έχουν εκπαιδευτεί. Αυτά είναι τα δίκτυα τα οποία απέτυχαν να ανιχνεύσουν το σήμα κατά τη διαδικασία της εκπαίδευσης. Σε αυτήν την περίπτωση οι αποκρίσεις των δικτύων παρουσιάζουν συγκεκριμένες μορφές (συνήθως ευθείες που συμπίπτουν με τον άξονα των x) και οι τιμές των μεγεθών ταξινόμησης παίρνουν τιμές που ανήκουν σε πολύ συγκεκριμένα πλαίσια.
- **Άλλες περιπτώσεις (Others)**: Υπάρχουν και περιπτώσεις δικτύων στις οποίες οι τιμές που παίρνουν τα μεγέθη βάση των οποίων γίνεται η ταξινόμηση δεν ανήκουν στα διαστήματα που έχουμε θέσει για την ταξινόμηση. Αυτά τα δίκτυα τοποθετούνται στην συγκεκριμένη κατηγορία και η απόκρισή τους εξετάζεται οπτικά μέσω της γραφικής τους παράστασης, για να γίνει η τελική ταξινόμησή τους σε κάποια από τις τρεις πρώτες κατηγορίες. Οι τιμές των μεγεθών για τα δίκτυα αυτά δεν ανήκουν σε κάποια συγκεκριμένα πλαίσια απλά αποτελούν περιπτώσεις που ξεφεύγουν από τις συνηθισμένες αποκρίσεις ή είναι πολύ κοντά στα όρια των τιμών βάση των οποίων γίνεται η ταξινόμηση.

Με τον συγκεκριμένο κώδικα πραγματοποιείται μία πρώτη ταξινόμηση των δικτύων σε αυτές τις κατηγορίες. Η ταξινόμηση αυτή καταγράφεται από τον κώδικα κατευθείαν σε ένα αρχείο excel. Να επισημανθεί ότι η τελική ταξινόμηση πραγματοποιείται μετά από εξέταση από το συγγραφέα της διπλωματικής των γραφικών παραστάσεων της απόκρισης των δικτύων. Ο κώδικας λοιπόν αυτός παίζει βοηθητικό ρόλο και διευκολύνει την τελική ταξινόμηση των δικτύων βάση της απόκρισής του. Το βασικό μέρος αυτού του κώδικα για την ταξινόμηση των αποτελεσμάτων του δεύτερου αλγορίθμου υπάρχει στο Παράρτημα Β.

• **Βήμα 4ο: Ελέγχου των γραφικών παραστάσεων:**

Στο βήμα αυτό δεν αντιστοιχεί κάποιος κώδικας. Αντίθετα κατά το βήμα αυτό ο συγγραφέας της διπλωματικής εξετάζει τις γραφικές παραστάσεις της απόκρισης των δικτύων. Αυτό κρίθηκε απαραίτητο για τρεις λόγους κυρίως. Πρώτον γιατί όπως αναφέρθηκε στο βήμα τρία υπάρχει μία κατηγορία περιπτώσεων της οποίας τα δίκτυα δεν μπορούν να ταξινομηθούν βάση των κριτηρίων που έχουμε θέσει, οπότε κρίνεται απαραίτητο να εξεταστούν οι γραφικές από τον συγγραφέα για να τοποθετηθούν αυτά τα δίκτυα σε μία από τις τρεις πρώτες κατηγορίες ταξινόμησης. Δεύτερον γιατί υπάρχουν περιπτώσεις δικτύων τα οποία ενώ οι τιμές των κριτηρίων οδηγούν τον αλγόριθμο να τα κατατάξει σε μία κατηγορία ενώ αυτά ανήκουν σε κάποια άλλη. Οπότε σε αυτήν την περίπτωση ο έλεγχος γίνεται για λόγους διόρθωσης. Αυτό το φαινόμενο παρουσιαζόταν κυρίως για μέσους αριθμούς νευρώνων (από 15 μέχρι 30 νευρώνες για δίκτυο ενός στρώματος και για συνδυασμούς με περισσότερους των τεσσάρων νευρώνων σε κάθε στρώμα για δίκτυα δύο κρυφών στρωμάτων). Τρίτον

για να επιλεγεί από το συγγραφέα το συγκεκριμένο δίκτυο για κάθε συνδυασμό νευρώνων που παρουσιάζει την καλύτερη απόδοση.

Τέλος να επισημάνουμε ότι για κάθε δίκτυο υπήρχε συγκεκριμένη κωδικοποίηση η οποία του δινόταν μέσα από τον κώδικα ταξινόμησης του έτσι όταν ένα δίκτυο ταξινομούνταν στην τελευταία κατηγορία υπήρχαν οι συντεταγμένες του (αριθμός νευρώνων πρώτου στρώματος- αριθμός νευρώνων δεύτερου στρώματος- αριθμός αρχικοποίησης) με τις οποίες αυτό το δίκτυο μπορούσε να βρεθεί και να εξεταστεί με ευκολία. Το ίδιο πραγματοποιούνταν και για τα δίκτυα τα οποία παρουσίαζαν καλή απόκριση στα σημεία πρόβλεψης.

Στο κεφάλαιο που ακολουθεί γίνεται η παρουσίαση των αποτελεσμάτων της υπολογιστικής διαδικασίας και η τεχνική ανάλυσή τους.

ΚΕΦΑΛΑΙΟ 7

ΠΑΡΟΥΣΙΑΣΗ ΚΑΙ ΑΝΑΛΥΣΗ ΤΩΝ ΑΠΟΤΕΛΕΣΜΑΤΩΝ

7.1 Εισαγωγή-Περιγραφή του τρόπου παρουσίασης των αποτελεσμάτων

Στα δύο προηγούμενα κεφάλαια επιχειρήθηκε μία περιγραφή αρχικά του πειράματος που θα πραγματοποιηθεί, των παραμέτρων που μεταβάλλονται και των βασικών αλγορίθμων που εξετάζονται. Έπειτα περιγράφηκαν τα βήματα που ακολουθήθηκαν κατά την πειραματική διαδικασία και δόθηκαν οι κώδικες στους οποίους βασίστηκε αυτή.

Στο συγκεκριμένο κεφάλαιο θα πραγματοποιηθεί η παρουσίαση των αποτελεσμάτων που προέκυψαν μετά την εκτέλεση των υπολογισμών και θα επιχειρηθεί μία τεχνική ανάλυσή τους, η οποία θα συμβάλλει στην εξαγωγή συμπερασμάτων όσον αφορά τα αποτελέσματα και την κατάληξη του πειράματος.

Προτού όμως συνεχίσουμε με την παράθεση των αποτελεσμάτων θα δώσουμε μία σύντομη περιγραφή για τον τρόπο που θα παρουσιαστούν. Τα αποτελέσματα θα παρουσιάζονται ανάλογα με τον αριθμό κρυφών στρωμάτων σε πίνακες. Οι πίνακες αυτοί θα περιλαμβάνουν για κάθε αριθμό νευρώνων ή συνδυασμό αριθμών νευρώνων στα διαφορετικά κρυφά στρώματα τις εξής γραμμές:

- **bad fitting**: όπου θα συμπληρώνεται ο αριθμός των δικτύων που δεν κατάφεραν κατά την απόκρισή τους να ανιχνεύσουν το σήμα στη συνάρτηση του ημιτόνου.
- **good fitting**: όπου θα συμπληρώνεται ο αριθμός δικτύων με καλή απόδοση στα σημεία εκπαίδευσης του δικτύου αλλά χαμηλή στα σημεία πρόβλεψης.
- **good forecasting**: Όπου συμπληρώνεται το πλήθος των δικτύων που πραγματοποιούν καλή πρόβλεψη.
- **best network**: Σε αυτή καταγράφουμε τον αριθμό της αρχικοποίησης του δικτύου το οποίο είχε την καλύτερη απόδοση από τα δίκτυα που έκαναν καλή πρόβλεψη για το συγκεκριμένο αριθμό νευρώνων. Αν κανένα δίκτυο από αυτά δεν έχει κάνει καλή πρόβλεψη τότε η θέση αυτή μένει κενή.
- **best for. mae/mse**: Οι δύο αυτές γραμμές συμπληρώνονται με τις τιμές των συναρτήσεων απόδοσης στα τα σημεία πρόβλεψης του δικτύου που παρουσίασε την καλύτερη απόδοση κατά την πρόβλεψη. Σε περίπτωση που δεν υπάρχει τέτοιο οι θέσεις παραμένουν κενές.

Το δίκτυο με την καλύτερη απόδοση για κάθε περίπτωση θα σημειώνεται με πράσινο χρώμα.

Για τα αποτελέσματα θα πραγματοποιηθεί και τεχνική ανάλυση και σύγκριση των αποτελεσμάτων μεταξύ των αλγορίθμων. Τέλος θα επιλεγούν τα καλύτερα δίκτυα για ένα και δύο κρυφά στρώματα νευρώνων τα οποία και θα δοκιμαστούν στα πραγματικά δεδομένα της χρονοσειράς του συνολικού φορτίου του Οκτωβρίου του 2010.

7.2 Παρουσίαση των αποτελεσμάτων

7.2.1 Ένα κρυφό στρώμα νευρώνων

- Αλγόριθμος 1

# Neurons	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
bad_fitting	10	10	8	10	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
good_fitting	0	0	2	0	7	6	10	10	10	10	10	10	10	10	10	10	10	10	10	10
good_forecasting	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
best_network	-	-	-	-	-	10	-	-	-	-	-	-	-	-	-	-	-	-	-	-
best_for_mse	-	-	-	-	-	0,23166	-	-	-	-	-	-	-	-	-	-	-	-	-	-
best_for_mae	-	-	-	-	-	0,37371	-	-	-	-	-	-	-	-	-	-	-	-	-	-
# Neurons	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
bad_fitting	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
good_fitting	10	10	10	9	10	10	10	10	10	10	10	9	10	9	10	10	10	10	10	10
good_forecasting	0	0	0	1	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0
best_network	-	-	-	3	-	-	-	-	-	-	-	3	-	5	-	-	-	-	-	-
best_for_mse	-	-	-	0,59158	-	-	-	-	-	-	-	0,31625	-	0,99413	-	-	-	-	-	-
best_for_mae	-	-	-	0,5723	-	-	-	-	-	-	-	0,42549	-	0,73969	-	-	-	-	-	-
# Neurons	41	42	43	44	45	46	47	48	49	50	Σύνολο	Ποσοστά								
bad_fitting	1	0	0	0	1	5	7	10	10	10	88	17,60%								
good_fitting	9	10	10	10	9	5	3	0	0	0	408	81,60%								
good_forecasting	0	0	0	0	0	0	0	0	0	0	4	0,80%								
best_network	-	-	-	-	-	-	-	-	-	-	500	100%								
best_for_mse	-	-	-	-	-	-	-	-	-	-										
best_for_mae	-	-	-	-	-	-	-	-	-	-										

Πίνακας 7.1: Αποτελέσματα δικτύου ενός κρυφού στρώματος για 40 σημεία εκπαίδευσης (Αλγόριθμος 1) (αδιάστατο)

# Neurons	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
bad_fitting	10	10	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
good_fitting	0	0	5	10	10	9	8	10	10	10	10	10	10	10	10	10	10	10	10	10
good_forecasting	0	0	0	0	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0
best_network	-	-	-	-	-	3	6	-	-	-	-	-	-	-	-	-	-	-	-	-
best_for_mse	-	-	-	-	-	0,05722	0,01154	-	-	-	-	-	-	-	-	-	-	-	-	-
best_for_mae	-	-	-	-	-	0,18431	0,08274	-	-	-	-	-	-	-	-	-	-	-	-	-
# Neurons	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
bad_fitting	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
good_fitting	10	10	10	10	9	9	10	10	10	10	10	10	10	10	10	10	10	10	10	10
good_forecasting	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
best_network	-	-	-	-	10	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-
best_for_mse	-	-	-	-	1,05413	1,08126	-	-	-	-	-	-	-	-	-	-	-	-	-	-
best_for_mae	-	-	-	-	0,77463	0,83183	-	-	-	-	-	-	-	-	-	-	-	-	-	-
# Neurons	41	42	43	44	45	46	47	48	49	50	Σύνολο	Ποσοστά								
bad_fitting	0	0	0	0	0	0	0	0	0	0	25	5,00%								
good_fitting	10	10	10	10	10	10	10	10	10	10	470	94,00%								
good_forecasting	0	0	0	0	0	0	0	0	0	0	5	1,00%								
best_network	-	-	-	-	-	-	-	-	-	-	500	100%								
best_for_mse	-	-	-	-	-	-	-	-	-	-										
best_for_mae	-	-	-	-	-	-	-	-	-	-										

Πίνακας 7.2: Αποτελέσματα δικτύου ενός κρυφού στρώματος για 160 σημεία εκπαίδευσης (Αλγόριθμος 1) (αδιάστατο)

ΚΕΦΑΛΑΙΟ 7. ΠΑΡΟΥΣΙΑΣΗ ΚΑΙ ΑΝΑΛΥΣΗ ΤΩΝ ΑΠΟΤΕΛΕΣΜΑΤΩΝ

# Neurons	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
bad fitting	10	10	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
good fitting	0	0	7	10	10	10	9	9	10	10	10	10	10	10	10	10	10	10	10	10
good forecasting	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
best_network	-	-	-	-	-	-	4	4	-	-	-	-	-	-	-	-	-	-	-	-
best_mse	-	-	-	-	-	-	0,98483	0,24509	-	-	-	-	-	-	-	-	-	-	-	-
best_mae	-	-	-	-	-	-	0,74464	0,36051	-	-	-	-	-	-	-	-	-	-	-	-
# Neurons	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
bad fitting	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
good fitting	9	9	9	8	10	9	10	10	10	10	10	10	10	10	10	10	10	10	10	10
good forecasting	1	1	1	2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
best_network	9	3	10	6	-	6	-	-	-	-	-	-	-	-	-	-	-	-	-	-
best_mse	0,30766	0,09911	0,29737	0,00722	-	0,13586	-	-	-	-	-	-	-	-	-	-	-	-	-	-
best_mae	0,37928	0,23718	0,37964	0,06032	-	0,29689	-	-	-	-	-	-	-	-	-	-	-	-	-	-
# Neurons	41	42	43	44	45	46	47	48	49	50	Σύνολο	Ποσοστά								
bad fitting	0	0	0	0	0	0	0	0	0	0	23	4,60%								
good fitting	10	10	10	10	10	10	10	10	10	10	469	93,80%								
good forecasting	0	0	0	0	0	0	0	0	0	0	8	1,60%								
best_network	-	-	-	-	-	-	-	-	-	-	500	100%								
best_mse	-	-	-	-	-	-	-	-	-	-										
best_mae	-	-	-	-	-	-	-	-	-	-										

Πίνακας 7.3: Αποτελέσματα δικτύου ενός κρυφού στρώματος για 320 σημεία εκπαίδευσης (Αλγόριθμος 1) (αδιάστατο)

• Αλγόριθμος 2

# Neurons	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
bad fitting	10	10	8	10	6	5	0	1	0	0	1	0	0	0	0	0	0	1	0	0
good fitting	0	0	2	0	4	5	10	8	10	10	9	10	9	8	9	9	10	9	8	9
good forecasting	0	0	0	0	0	0	0	1	0	0	0	0	1	2	1	1	0	0	2	1
best_network	-	-	-	-	-	-	-	2	-	-	-	-	1	1	8	6	-	-	4	1
best_for_mse	-	-	-	-	-	-	-	0,30376	-	-	-	-	0,15076	0,24476	0,09135	0,31171	-	-	0,01387	0,08038
best_for_mae	-	-	-	-	-	-	-	0,36634	-	-	-	-	0,24087	0,34528	0,21392	0,43101	-	-	0,09397	0,20124
# Neurons	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
bad fitting	0	0	0	1	0	0	0	1	1	1	0	0	1	0	0	0	0	0	0	0
good fitting	10	10	10	9	8	9	10	9	8	9	9	10	8	10	9	9	10	10	10	10
good forecasting	0	0	0	0	2	1	0	0	1	0	1	0	1	0	1	1	0	0	0	0
best_network	-	-	-	-	5	5	-	-	7	-	5	-	8	-	4	7	-	-	-	-
best_for_mse	-	-	-	-	0,19143	0,19143	-	-	0,0886	-	0,16437	-	0,04297	-	0,07168	0,22023	-	-	-	-
best_for_mae	-	-	-	-	0,34504	0,34504	-	-	0,22619	-	0,27715	-	0,17027	-	0,21667	0,37681	-	-	-	-
# Neurons	41	42	43	44	45	46	47	48	49	50	Σύνολο	Ποσοστά								
bad fitting	0	2	1	4	8	9	9	10	10	10	120	24,00%								
good fitting	9	8	9	6	2	1	1	0	0	0	362	72,40%								
good forecasting	1	0	0	0	0	0	0	0	0	0	18	3,60%								
best_network	5	-	-	-	-	-	-	-	-	-	500									
best_for_mse	0,26941	-	-	-	-	-	-	-	-	-										
best_for_mae	0,43085	-	-	-	-	-	-	-	-	-										

Πίνακας 7.4: Αποτελέσματα δικτύου ενός κρυφού στρώματος για 40 σημεία εκπαίδευσης (Αλγόριθμος 2) (αδιάστατο)

• Αλγόριθμος 3

# Neurons	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
bad_fitting	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
good_fitting	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10
good_forecasting	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
best_network	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
best_for_mse	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
best_for_mae	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
# Neurons	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
bad_fitting	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
good_fitting	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10
good_forecasting	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
best_network	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
best_for_mse	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
best_for_mae	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
# Neurons	41	42	43	44	45	46	47	48	49	50	Σύνολο	Ποσοστά								
bad_fitting	0	0	0	0	0	0	0	0	0	0	0	0%								
good_fitting	10	10	10	10	10	10	10	10	10	10	500	100%								
good_forecasting	0	0	0	0	0	0	0	0	0	0	0	0%								
best_network	-	-	-	-	-	-	-	-	-	-	500	100%								
best_for_mse	-	-	-	-	-	-	-	-	-	-										
best_for_mae	-	-	-	-	-	-	-	-	-	-										

Πίνακας 7.5: Αποτελέσματα δικτύου ενός κρυφού στρώματος για 40 σημεία εκπαίδευσης (Αλγόριθμος 3) (αδιάστατο)

• Αλγόριθμος 4

# Neurons	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
bad_fitting	10	10	10	10	10	10	0	0	0	0	0	0	0	0	0	0	9	0	0	4
good_fitting	0	0	0	0	0	0	10	10	10	10	10	10	10	10	10	10	1	10	10	5
good_forecasting	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
best_network	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	6
best_for_mse	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,23974
best_for_mae	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,40341
# Neurons	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
bad_fitting	0	0	1	2	2	2	1	1	1	1	0	1	2	0	0	0	0	0	1	1
good_fitting	8	10	9	8	7	8	9	9	9	9	10	8	5	10	10	10	10	10	9	9
good_forecasting	2	0	0	0	1	0	0	0	0	0	0	1	3	0	0	0	0	0	0	0
best_network	9	-	-	-	4	-	-	-	-	-	-	8	5	-	-	-	-	-	-	-
best_for_mse	0,37407	-	-	-	0,05431	-	-	-	-	-	-	0,30009	0,06455	-	-	-	-	-	-	-
best_for_mae	0,43557	-	-	-	0,18361	-	-	-	-	-	-	0,36702	0,19374	-	-	-	-	-	-	-
# Neurons	41	42	43	44	45	46	47	48	49	50	Σύνολο	Ποσοστά								
bad_fitting	0	8	1	9	10	10	10	10	10	10	167	33,40%								
good_fitting	10	2	9	1	0	0	0	0	0	0	325	65,00%								
good_forecasting	0	0	0	0	0	0	0	0	0	0	8	1,60%								
best_network	-	-	-	-	-	-	-	-	-	-	500	100,00%								
best_for_mse	-	-	-	-	-	-	-	-	-	-										
best_for_mae	-	-	-	-	-	-	-	-	-	-										

Πίνακας 7.6: Αποτελέσματα δικτύου ενός κρυφού στρώματος για 40 σημεία εκπαίδευσης (Αλγόριθμος 4) (αδιάστατο)

• Αλγόριθμος 5

# Neurons	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
bad_fitting	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
good_fitting	10	3	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
good_forecasting	0	7	10	10	10	9	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10
best_network	-	2	7	7	6	4	3	10	2	7	6	9	2	2	9	2	2	9	10	9	
best_for_mse	-	6,9E-11	2,2E-11	2,9E-12	1,3E-12	1,3E-12	1E-12	5,1E-13	6,1E-13	1,2E-12	2,1E-13	8,8E-14	1,8E-13	2,3E-13	3,4E-13	4,6E-13	2,8E-13	1,2E-12	2,2E-12	4,2E-13	
best_for_mae	-	5,8E-06	3E-06	1,2E-06	7,9E-07	7,6E-07	6,5E-07	5,5E-07	5,5E-07	6,4E-07	2,8E-07	2,4E-07	2,7E-07	3,2E-07	4,1E-07	3,2E-07	3,2E-07	6,4E-07	9,4E-07	3,7E-07	
# Neurons	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	
bad_fitting	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
good_fitting	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
good_forecasting	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10
best_network	10	4	9	6	6	10	4	4	2	4	7	9	4	2	8	9	9	7	10	4	
best_for_mse	2,6E-13	1E-12	1,9E-13	1,2E-12	5E-13	7,7E-13	6,3E-13	8,4E-13	5,9E-13	4,6E-13	2,3E-13	2,9E-13	7E-13	2,2E-13	8,7E-14	3,6E-13	1,6E-13	3,4E-13	2,9E-13	0,00848	
best_for_mae	3,7E-07	5,4E-07	3,1E-07	7,7E-07	4,8E-07	6,1E-07	5,5E-07	4,6E-07	4,7E-07	4,6E-07	3,6E-07	3,7E-07	5,6E-07	3,6E-07	2,1E-07	3,8E-07	2,7E-07	3,9E-07	3E-07	0,06339	
# Neurons	41	42	43	44	45	46	47	48	49	50	Σύνολο	Ποσοστά									
bad_fitting	0	0	0	0	0	0	0	0	0	0	0	0,00%									
good_fitting	0	0	0	0	0	0	0	0	0	0	14	2,80%									
good_forecasting	10	10	10	10	10	10	10	10	10	10	486	97,20%									
best_network	2	9	5	5	3	1	8	9	8	4	500	100,00%									
best_for_mse	3,4E-13	5,1E-12	4,3E-12	3E-12	2,4E-12	1,8E-12	3,33E-12	1,8E-11	1,2E-11	1,4E-11											
best_for_mae	4,1E-07	1,5E-06	1E-06	8,9E-07	9,7E-07	8,1E-07	1,1E-06	2,2E-06	2E-06	2E-06											

Πίνακας 7.7: Αποτελέσματα δικτύου ενός κρυφού στρώματος για 40 σημεία εκπαίδευσης (Αλγόριθμος 5) (αδιάστατο)

Παραπάνω παρατέθηκαν οι πίνακες με τα αποτελέσματα που προέκυψαν κατά τις δοκιμές που έγιναν σε δίκτυα ενός κρυφού στρώματος. Από τα αποτελέσματα αυτά παρατηρούμε ότι, με εξαίρεση τον τελευταίο αλγόριθμο (για το οποίο υπάρχει εξήγηση) και τον τρίτο, ο αριθμός των δικτύων που πραγματοποιούν καλή πρόβλεψη μετά την εκπαίδευση είναι αρκετά χαμηλός. Το υψηλότερο ποσοστό τέτοιων δικτύων που παρατηρείται είναι στον αλγόριθμο 2 (3,6%-18 δίκτυα). Στις υπόλοιπες περιπτώσεις ο αριθμός αυτός είναι αρκετά χαμηλότερος (από 0.8% έως 1.60%). Επίσης για τον αλγόριθμο 1 παρατηρούμε ότι όσο αυξάνουμε τα σημεία εκπαίδευσης αυξάνεται και ο αριθμός των δικτύων καλής πρόβλεψης. Μάλιστα ο αριθμός των συγκεκριμένων δικτύων για 320 σημεία εκπαίδευσης (8 δίκτυα) είναι ο διπλάσιος από τον αντίστοιχο για 40 σημεία εκπαίδευσης (4 δίκτυα).

Σε όλους τους αλγορίθμους πλην του 5, το μεγαλύτερο ποσοστό δικτύων κατάφερε να προσομοιώσει το ημίτονο στα σημεία που του δόθηκαν για εκπαίδευση αλλά δεν πραγματοποίησε καλές προβλέψεις. Το χαμηλότερο ποσοστό τέτοιων δικτύων (εκτός του αλγορίθμου 3) παρατηρείται στον αλγόριθμο 4 (65%). Το υψηλότερο παρατηρείται στον αλγόριθμο 1 για 160 σημεία εκπαίδευσης (94%). Αυτό το ποσοστό είναι περίπου ίδιο για την περίπτωση των 320 σημείων (93.8%) ενώ μειώνεται στην περίπτωση των 40 σημείων (81.6 %) καθώς εκεί είναι αυξημένος ο αριθμός των δικτύων που δεν καταφέρνουν να προσομοιώσουν καθόλου το ημίτονο.

Μεταβολές στις προβλέψεις παρατηρούνται και βάση του αριθμού των κρυφών νευρώνων. Αυτό μπορεί να διαπιστωθεί αν παρατηρήσει κανείς στους πίνακες ότι τα δίκτυα τα οποία πραγματοποιούν καλές προβλέψεις εμφανίζονται στη συντριπτική τους πλειοψηφία για αριθμούς νευρώνων από 6 έως 35. Στους υπόλοιπους αριθμούς νευρώνων σπάνια εμφανίζεται κάποιο δίκτυο που να πραγματοποιεί καλή πρόβλεψη ανεξαρτήτως αλγορίθμου εκπαίδευσης. Παρατηρούμε επίσης ότι οι καλύτερες προβλέψεις για κάθε αλγόριθμο παρατηρούνται σε δίκτυα με αριθμό νευρώνων που κυμαίνονται στο παραπάνω διάστημα.

Όσον αφορά την ακρίβεια των προβλέψεων μπορούμε να παρατηρήσουμε ότι δεν παρουσιάζονται σημαντικές μεταβολές από αλγόριθμο σε αλγόριθμο στην τάξη μεγέθους της. Επίσης με μια πιο προσεκτική ματιά παρατηρεί κανείς ότι στους διαφόρους αλγορίθμους φαίνεται να μην επηρεάζεται ούτε από τη μεταβολή του αριθμού των κρυφών νευρώνων. Για παράδειγμα στον αλγόριθμο 2 η ακρίβεια πρόβλεψης είναι περίπου ίδια για δίκτυο 14 νευρώνων ($mse=0.24$ και $mae=0.35$), 26 νευρώνων ($mse=0.19$ και $mae=0.34$) και 36 νευρώνων ($mse=0.24$ και $mae=0.38$).

Όσον αφορά τους αλγορίθμους 3 και 5 φαίνεται ότι τα αποτελέσματα δεν είχαν τη μορφή των υπόλοιπων αποτελεσμάτων. Πιο συγκεκριμένα στον αλγόριθμο 3 όλα τα δίκτυα μετά την εκπαίδευσή παρουσίαζαν την ίδια απόκριση: Κατάφεραν να προσομοιώσουν σωστά το ημίτονο για τα σημεία στα οποία είχαν εκπαιδευτεί αλλά μετά από τα σημεία αυτά κανένα δίκτυο του συγκεκριμένου αλγορίθμου δεν πραγματοποίησε σωστή πρόβλεψη. Από την άλλη πλευρά τα αποτελέσματα του αλγορίθμου 5 είχαν αντίθετη μορφή καθώς όλα τα δίκτυα πλην δεκατεσσάρων πραγματοποίησαν ακριβείς προβλέψεις της μορφής του ημιτόνου ακόμα και μετά τα σημεία εκπαίδευσης. Η κύρια αιτία για αυτό είναι η μορφή του συγκεκριμένου αλγορίθμου ο οποίος είναι προσαρμοσμένος στα δεδομένα που του έχουν δοθεί περισσότερο από τους υπόλοιπους ενώ παράλληλα χρειάζεται και παρέμβαση από τον χειριστή για να του δώσει τον κατάλληλο αριθμό περιόδων στις οποίες εμφανίζεται εποχικότητα. Αυτοί είναι οι λόγοι για τους οποίους η ακρίβεια που παρουσιάζει αυτός ο αλγόριθμος είναι πολύ υψηλή σε σχέση με τους άλλους. Για αυτούς τους λόγους τα αποτελέσματα των αλγορίθμων 3 και 5 δεν θα ληφθούν υπόψη αλλά παρατίθενται για την πληρότητα της εργασίας.

Το δίκτυο εκείνο που πραγματοποίησε την καλύτερη πρόβλεψη θα επιλεγεί από τα δίκτυα που εκπαιδεύτηκαν σε σετ 40 σημείων δεδομένων. Στον αλγόριθμο 1 χρησιμοποιήθηκαν και σετ δεδομένων 160 και 320 σημείων για να πραγματοποιηθεί σύγκριση των αποτελεσμάτων με τα αντίστοιχα αποτελέσματα δικτύου 40 σημείων.

Οπότε από τα δίκτυα που πραγματοποίησαν προβλέψεις την πιο ακριβή πρόβλεψη πραγματοποίησε **το δίκτυο του αλγορίθμου 2 το οποίο είχε 19 νευρώνες** ($mse=0.0138$, $mae=0.0939$). Αυτό θα είναι το πρώτο δίκτυο που θα χρησιμοποιηθεί για την μελέτη της δεύτερης χρονοσειράς.

Στην περίπτωση δικτύου με δύο κρυφούς νευρώνες τα πράγματα δείχνουν να διαφέρουν σε σχέση με τα δίκτυα ενός κρυφού νευρώνα. Παρότι ο αριθμός των παραμέτρων προς βελτιστοποίηση σε κάθε δίκτυο δεν είναι πολύ διαφορετικός σε σχέση με εκείνων για δίκτυα ενός στρώματος τα αποτελέσματα δείχνουν ότι τέτοια δίκτυα δυσκολεύονται να πραγματοποιήσουν καλές προβλέψεις σε σχέση με τα δίκτυα ενός στρώματος.

Πιο συγκεκριμένα για τους αλγόριθμους με 40 σημεία εκπαίδευσης το υψηλότερο ποσοστό που παρατηρείται είναι το 0,8% (8 δίκτυα) στον Αλγόριθμο 2. Μικρότερο σε σχέση με το αντίστοιχο για τα δίκτυα ενός στρώματος είναι και το γενικά μεγαλύτερο ποσοστό τέτοιων δικτύων ανεξαρτήτως αριθμού εκπαίδευσης και το οποίο είναι 2,5% και παρουσιάζεται στον αλγόριθμο 1 για 160 σημεία εκπαίδευσης.

Αντίθετα με τα δίκτυα ενός κρυφού στρώματος δεν παρατηρείται αύξηση του αριθμού των δικτύων που πραγματοποιούν καλή πρόβλεψη στον αλγόριθμο 1 καθώς αυξάνεται ο αριθμός των σημείων του σετ δεδομένων. Όπως επισημάνθηκε και προηγουμένως περισσότερα δίκτυα πραγματοποίησαν καλές προβλέψεις για 160 σημεία εκπαίδευσης (25 δίκτυα) σε σχέση με τα αντίστοιχα για 320 σημεία (19). Ένα κοινό σημείο με τα αποτελέσματα δικτύων ενός κρυφού στρώματος για τον αλγόριθμο 1 είναι ότι όσο αυξάνεται ο αριθμός των σημείων στο σετ δεδομένων τόσο μειώνεται ο αριθμός των δικτύων που δεν καταφέρνουν να προσεγγίσουν καθόλου το ημίτονο (bad fitting) ενώ αυξάνεται ο αριθμός των δικτύων στις άλλες δύο κατηγορίες. Έτσι ενώ για 40 σημεία τα δίκτυα της πρώτης κατηγορίας είναι 601, όταν αυξήσουμε τα σημεία στα 160 και έπειτα στα 320 τα δίκτυα αυτά γίνονται 258 και 114 αντίστοιχα.

Η ακρίβεια με την οποία τα δίκτυα πραγματοποιούν προβλέψεις δεν φαίνεται να αλλάζει σημαντικά με την αλλαγή των αλγορίθμων και φαίνεται να είναι μειωμένη σε σχέση με την αντίστοιχη ακρίβεια πρόβλεψης των δικτύων ενός κρυφού στρώματος. Αυτό η πτώση στην ακρίβεια μπορεί να εξηγηθεί μετά από την οπτική παρατήρηση των δικτύων αυτών, στην οποία παρατηρείτε το φαινόμενο να επιτυγχάνεται καλή προσέγγιση στα αρχικά σημεία πρόβλεψης αλλά καθώς απομακρυνόμαστε από το σετ δεδομένων η ακρίβεια πρόβλεψης να μειώνεται αρκετά περισσότερο σε σχέση με τον τρόπο που μειωνόταν για τα δίκτυα ενός κρυφού στρώματος. Για αυτό εδώ παρατηρούμε ότι οι τιμές των συναρτήσεων απόδοσης προσεγγίζουν περισσότερο τη μονάδα.

Τώρα όσον αφορά την ακρίβεια για τις διάφορες περιπτώσεις του αλγορίθμου 1 εκεί παρατηρούμε ομοιότητες στην συμπεριφορά των δύο περιπτώσεων κρυφών στρωμάτων καθώς φαίνεται να μην εξαρτάται άμεσα από τον αριθμό των παρεχόμενων σημείων στο σετ δεδομένων. Όμοια συμπεριφορά παρατηρείται και στο γεγονός ότι στην καλύτερη πρόβλεψη για αυτές τις τρεις περιπτώσεις του αλγορίθμου 1 παρατηρείται βελτίωση της ακρίβειας με την αύξηση του αριθμού των σημείων. Έτσι ενώ στα 40 σημεία η ακρίβεια της καλύτερης πρόβλεψης είναι σχετικά χαμηλή ($mse=0.991$, $mae=0.866$) σταδιακά αυξάνεται παράλληλα με την αύξηση των

δοθέντων σημείων σε 160 ($mse=0.599$, $mae=0.681$) και 320 ($mse=0.489$, $mae=0.614$).

Εδώ επίσης παρατηρείται το φαινόμενο δύο από τους αλγορίθμους (3 και 4) να μην έχουν κάποιο δίκτυο το οποίο να καταφέρει να πραγματοποιήσει καλή πρόβλεψη. Ο αλγόριθμος 3 παρουσιάζει τα ίδια αποτελέσματα με την περίπτωση του ενός κρυφού στρώματος, δηλαδή όλα τα δίκτυα που εκπαιδεύτηκαν με τη συνδρομή αυτού του αλγορίθμου είχαν καλή προσέγγιση στα σημεία που του δόθηκαν για εκπαίδευση αλλά κανένα από αυτά δεν κατάφερε να πραγματοποιήσει καλή πρόβλεψη στα σημεία του test set. Από την άλλη πλευρά στον αλγόριθμο 4 παρατηρούμε μείωση τόσο του ποσοστού των δικτύων που πραγματοποίησαν καλή πρόβλεψη (0% εδώ, 1.6% για ένα κρυφό στρώμα) όσο και του ποσοστού των δικτύων που έκαναν καλή προσέγγιση στα σημεία προσομοίωσης του σετ εκπαίδευσης (36.9% εδώ, 65% για ένα κρυφό στρώμα).

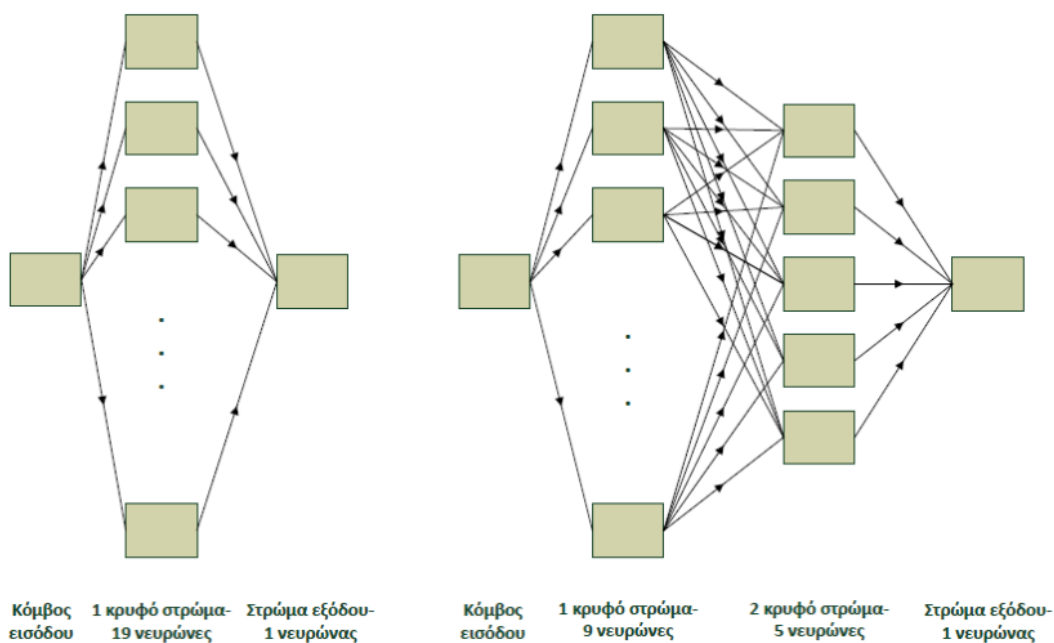
Ο αλγόριθμος 5 στην περίπτωση των δύο κρυφών στρωμάτων φαίνεται να παράγει πιο λογικά αποτελέσματα σε σχέση με την περίπτωση του ενός κρυφού στρώματος όπου σχεδόν όλα τα δίκτυα πέτυχαν πολύ καλή πρόβλεψη. Για την περίπτωση των δικτύων δύο κρυφών στρωμάτων η συντριπτική πλειοψηφία (99.5%) των δικτύων στον αλγόριθμο αυτό προσέγγισε καλά μόνο τα σημεία του training set ενώ υπήρχε και ένας μικρός αριθμός δικτύων (5 δίκτυα, 0.5% του συνόλου) τα οποία πραγματοποίησαν καλή πρόβλεψη. Παρόλα αυτά όπως και προηγουμένως τα αποτελέσματα του αλγορίθμου αυτού δεν θα ληφθούν υπόψη κατά την επιλογή του καλύτερου δικτύου για τους ίδιους λόγους που αναφέρθηκαν και στην περίπτωση του δικτύου ενός κρυφού στρώματος.

Να επισημάνουμε ότι εδώ τα περισσότερα δίκτυα με καλή πρόβλεψη παρουσιάστηκαν στον αλγόριθμο 2 όπως και το δίκτυο το οποίο είχε την πιο ακριβή πρόβλεψη σε σχέση με τα άλλα δίκτυα. Τα αποτελέσματα αυτά είναι κοινά με της περίπτωσης των δικτύων ενός κρυφού στρώματος καθώς και εκεί ο αλγόριθμος 2 ήταν εκείνος που έδωσε τα περισσότερα δίκτυα καλής πρόβλεψης και το πιο ακριβές δίκτυο.

Οπότε μετά από την εξέταση των αποτελεσμάτων παρατηρούμε ότι το δίκτυο το οποίο είχε την καλύτερη απόδοση κατά την πρόβλεψη είναι το δίκτυο το οποίο έχει στα **δύο κρυφά στρώματα του 9 και 5 νευρώνες διαδοχικά** και του οποίου οι συναρτήσεις απόδοσης είχαν τις τιμές $mse=0.4398$ και $mae=0.5537$.

7.3 Μελέτη περίπτωσης με πραγματικά δεδομένα

Μετά από την ανάλυση των αποτελεσμάτων που προέκυψαν από την διεξαγωγή των πειραμάτων καταλήξαμε στην επιλογή των δύο δικτύων (ένα για κάθε αριθμό κρυφών στρωμάτων) τα οποία πραγματοποίησαν την καλύτερη πρόβλεψη της συνάρτησης του ημιτόνου. Παρακάτω δίνονται σχηματικά τα δύο δίκτυα που επιλέχθηκαν.



Εικόνα7.1: Μορφή βέλτιστων δικτύων που επιλέχθηκαν

Αυτά τα δύο δίκτυα που επιλέχθηκαν προέκυψαν έπειτα από εκπαίδευση του δικτύου με τον Αλγόριθμο 2. Αυτά τα δύο δίκτυα θα εκπαιδευτούν στο σετ δεδομένων που περιλαμβάνει πραγματικές τιμές του ηλεκτρικού φορτίου του Οκτωβρίου του 2010. Πιο συγκεκριμένα τα δίκτυα θα πραγματοποιήσουν πρόβλεψη για τις τιμές του φορτίου το εικοσιτετράωρο της 20 Οκτωβρίου του 2010. Για το σκοπό αυτό καθένα από τα δύο δίκτυα θα προπονηθεί διαδοχικά με δεδομένα μίας ημέρας, δυο ημερών, μίας εβδομάδας, δύο εβδομάδων και όλων των προηγούμενων ημερών του Οκτωβρίου. Τα δεδομένα που θα χρησιμοποιηθούν φαίνονται στο Παράρτημα Α.

Οπότε όπως προηγουμένως στην περίπτωση του ημιτόνου χωρίζουμε τα δεδομένα σε δύο σετ. Στο σετ εκπαίδευσης είναι όλα τα δεδομένα των πρώτων 19 ημερών του Οκτωβρίου τα οποία και παρέχονται στο δίκτυο ανάλογα με τις μέρες που αφορά η περίπτωση που εξετάζουμε. Το δεύτερο σετ, το test set, περιλαμβάνει τα δεδομένα τις 20 Οκτωβρίου τα οποία και θα χρησιμοποιηθούν για να ελεγχθεί η ακρίβεια των προβλέψεων που πραγματοποιήθηκαν από τα δίκτυα.

Τα αποτελέσματα από την εκπαίδευση των δικτύων αυτών για τις διάφορες περιπτώσεις των δεδομένων στην είσοδό τους φαίνονται στον πίνακα παρακάτω. Να

σημειώσουμε ότι δίκτυο A είναι το δίκτυο ενός κρυφού στρώματος 19 νευρώνων και δίκτυο B το δίκτυο 2 κρυφών στρωμάτων.

MSE	1 ημέρα		2 ημέρες		1 εβδομάδα		2 εβδομάδες		19 ημέρες	
	Δίκτυο A	Δίκτυο B	Δίκτυο A	Δίκτυο B	Δίκτυο A	Δίκτυο B	Δίκτυο A	Δίκτυο B	Δίκτυο A	Δίκτυο B
1	7,58803E+11	750665,5934	1345355232	1914805,873	233728,1836	862845,2	917950,7167	847290,7	1032305	756217,7
2	27911796733	745367,4852	45226666,01	3218276,613	93138752,85	790291,3	5451300,143	811000,9	1911684	786365,2
3	1197943567	952854,1806	4882918,342	701196,9215	2374937245	757535,9	2077522,018	821195,3	1017989	756237,2
4	201801437,5	154294620,4	131088093,9	5335401,729	1,24485E+11	751385,3	937962,8564	768431,5	1232385	971643,3
5	137203177,3	964544,705	45867973,35	1346706,327	5413505936	751385,3	753188,8134	894486,9	1925156	754464,3
6	2679762,615	12573481,67	38737749,56	338806855,9	4017296890	1057294	1005242,797	760898,8	1166650	762444,7
7	26940876,08	2650245,826	29929708,03	22242827,74	176900504,5	782179,5	1639969,125	829904,2	714170,8	896160,4
8	17607481,63	646569,6586	5350197,93	642438281,5	285655148,9	921771,6	864409,0797	768431,5	1115393	903274,2
9	15672970,57	1239146,362	31159284,99	1306279,891	495735831,9	750504,8	1197581,027	829978,7	2035780	971540,2
10	5857641,285	789602,6159	42065653,88	7943281,694	53818695,55	840608,3	1334110,985	882030,6	1431841	751870,6
MAE	1 ημέρα		2 ημέρες		1 εβδομάδα		2 εβδομάδες		19 ημέρες	
Αριθμός Αρχικοποίησης	Δίκτυο A	Δίκτυο B	Δίκτυο A	Δίκτυο B	Δίκτυο A	Δίκτυο B	Δίκτυο A	Δίκτυο B	Δίκτυο A	Δίκτυο B
1	448489,2325	693,3895517	29094,65715	1197,162333	414,4818885	737,9743	835,6250723	741,1406	790,0159	724,5136
2	134250,0284	691,3575098	6157,269679	1549,367031	8726,683912	707,9537	2071,546342	727,5517	1288,972	719,8231
3	24773,31815	879,3375346	2030,200219	657,0176496	43236,17331	711,6273	1343,042209	725,0225	850,9633	724,5031
4	13571,17789	9045,80226	8728,771602	1888,433568	291859,0593	733,7061	771,6470417	719,7078	1019,216	777,6383
5	10281,85346	772,5773209	6499,274812	1012,010767	62123,46569	733,7061	727,9027672	748,3228	1302,19	725,5445
6	1401,729621	3003,492043	5802,05616	16245,45913	54895,62231	802,4634	899,3683305	722,3477	987,1378	721,7439
7	3221,135658	1168,429066	5179,092566	2693,095836	12371,61857	716,503	1195,516279	727,3142	702,5653	752,8131
8	3550,494525	643,5965422	2074,737214	20292,07162	14737,35853	754,4375	742,5990751	719,7078	956,767	754,3725
9	3671,859508	832,6845564	5107,816884	997,4199324	19262,39086	732,1855	1000,880598	729,2308	1334,575	773,4917
10	2149,094856	709,4831367	6189,18214	2367,120222	5659,295658	763,6767	930,2269413	745,0959	1104,995	724,1771

Πίνακας 7.22: Τιμές συναρτήσεων απόδοσης δικτύου για την περίπτωση της πραγματικής χρονοσειράς (αδιάστατα)

Όπως παρατηρείται στον παραπάνω πίνακα τα αποτελέσματα από την δοκιμή των δικτύων που επιλέχθηκαν στο προηγούμενο βήμα στη χρονοσειρά με τα πραγματικά δεδομένα του ηλεκτρικού φορτίου δεν ήταν επιτυχημένα. Αυτό γίνεται κατανοητό από τις υψηλές τιμές που λαμβάνουν οι συναρτήσεις απόδοσης στην πλειοψηφία των περιπτώσεων. Επίσης αυτό παρατηρήθηκε και από την εξέταση των γραφικών παραστάσεων που δημιουργήθηκαν, καθώς στις περισσότερες περιπτώσεις το δίκτυο κατάφερε επιτυχώς να προσεγγίσει τα σημεία που του δόθηκαν κατά τη διαδικασία εκπαίδευσης αλλά κανένα δίκτυο δεν πραγματοποίησε καλή πρόβλεψη των σημείων του test set.

Επίσης όπως παρατηρείται και από τα αποτελέσματα όσον αυξανόταν ο αριθμός των δεδομένων στο training set (δηλαδή όσες περισσότερες μέρες λαμβάναμε υπόψη) η απόδοση των δικτύων γινόταν όλο και καλύτερη καθώς προσέγγιζαν όλο και περισσότερο τα σημεία του training set και η απόκρισή τους για τα σημεία του test set δεν έπαιρνε ακραίες τιμές όπως γινόταν για μικρό αριθμό ημερών (1 και 2 ημέρες).

Στη μελέτη της συγκεκριμένης περίπτωσης παρατηρήθηκε και το εξής φαινόμενο: Τα δίκτυα που είχαν δύο κρυφά στρώματα νευρώνων είχαν καλύτερη απόκριση από τα αντίστοιχα του ενός κρυφού στρώματος για τις περισσότερες από τις περιπτώσεις. Αυτό το φαινόμενο έρχεται σε αντίθεση με τα όσα είχαμε παρατηρήσει κατά τη μελέτη των δικτύων με τη χρήση του ημιτόνου. Μάλιστα παρά την αλλαγή των στοιχείων που λαμβάνοντα υπόψη η απόκριση του δικτύου δύο κρυφών στρωμάτων παραμένει καλύτερη από την αντίστοιχη για ένα στρώμα. Αυτό μπορεί να οφείλεται και στον μεγαλύτερο αριθμό παραμέτρων που έχει το δίκτυο 2 κρυφών στρωμάτων (74 έναντι 58 του δικτύου 1 κρυφού στρώματος) που καθιστά το δίκτυο πιο ικανό να προσεγγίσει πιο περίπλοκες συναρτήσεις σε σχέση με το άλλο δίκτυο.

Παρακάτω δίνεται και ο πίνακας που περιλαμβάνει τις τιμές της συνάρτησης απόδοσης για το μέσο απόλυτο ποσοστιαίο σφάλμα (MAPE) για την καλύτερη κατανόηση των μεγεθών του σφάλματος κατά την πρόβλεψη.

MAPE (%)	1 ημέρα		2 ημέρες		1 εβδομάδα		2 εβδομάδες		19 ημέρες	
	Δίκτυο A	Δίκτυο B	Δίκτυο A	Δίκτυο B	Δίκτυο A	Δίκτυο B	Δίκτυο A	Δίκτυο B	Δίκτυο A	Δίκτυο B
1	7655,648038	14,75817224	500,8539674	20,89027848	7,983224677	15,44398	16,2590746	15,28624	16,74122	14,62472
2	2367,971259	14,33422358	111,1795504	26,71727098	154,4133652	14,6705	36,38587468	15,32629	23,47539	15,11281
3	480,0215264	16,1075856	38,83004743	13,89391946	764,27517	14,56517	24,44331528	15,1333	17,06102	14,62486
4	249,2829616	153,5952156	154,9129678	33,36349855	5054,673203	14,49829	16,02909208	14,69082	19,33343	16,69678
5	182,3237226	16,49253595	120,0118407	17,75160213	1080,757822	14,49829	14,62500268	15,66942	23,93362	14,61054
6	28,2382803	55,22267089	106,0101199	285,1552274	963,6702847	16,95057	17,20304728	14,65451	18,73955	14,66281
7	55,0618193	23,73932691	94,88975037	57,00226532	221,0288783	14,74919	21,94472134	15,15746	14,28318	16,07718
8	60,94000664	13,47184883	40,11761766	350,3637075	256,0887683	15,8154	15,42899516	14,69082	18,28778	16,12258
9	65,23689315	18,06674112	92,48550914	17,46432164	338,6986933	14,5192	19,02071254	15,30122	24,47403	16,58094
10	37,95936034	15,12466246	113,69695	40,50243914	96,666266	15,38453	19,84873657	15,69699	20,85608	14,58748

Πίνακας 7.23: Πίνακας με τις τιμές του MAPE για τα σημεία πρόβλεψης (%)

Εξετάζοντας τον παραπάνω μπορεί κανείς να παρατηρήσει ότι τα δίκτυα δύο κρυφών στρωμάτων έχουν πολύ χαμηλότερο ποσοστιαίο σφάλμα σε σχέση με εκείνα του ενός στρώματος. Επίσης ότι το σφάλμα στις περιπτώσεις του ενός στρώματος σταδιακά μειώνεται με την αύξηση των σημείων που δίνονται στο σετ εκπαίδευσης. Αντίθετα το σφάλμα του δικτύου δύο στρωμάτων δεν μεταβάλλεται πολύ. Είναι λογικό καθώς η απόκριση του δικτύου στα σημεία πρόβλεψης δεν αλλάζει με την αύξηση των σημείων στο training set.

ΚΕΦΑΛΑΙΟ 8

ΣΥΜΠΕΡΑΣΜΑΤΑ

8.1 Συμπεράσματα- Επίλογος διπλωματικής

Τα νευρωνικά δίκτυα αποτελούν ένα πολλά υποσχόμενο πεδίο στο χώρο των προβλέψεων. Τα μοναδικά χαρακτηριστικά που παρουσιάζουν όπως η προσαρμοστικότητα στα δεδομένα, η δυνατότητα προσέγγισης οποιασδήποτε συνάρτησης και η μη γραμμικότητα αποτελούν ισχυρά πλεονεκτήματα σε σχέση με τις κλασικές μεθόδους πρόβλεψης. Στον αντίποδα οι μεγάλες προσδοκίες που δημιουργούνται από αυτό το πεδίο συνοδεύονται και από μία μεγάλη δόση αβεβαιότητας για τη συμπεριφορά τους. Τα τεχνητά νευρωνικά δίκτυα αποτελούν ακόμα και σήμερα ένα είδος "μαύρου κουτιού" όσον αφορά τον τρόπο προσαρμογής τους στα δεδομένα πράγμα που δυσκολεύει την πραγματοποίηση συστηματικής έρευνας γύρω από αυτά και την εξαγωγή αποτελεσμάτων που θα αφορούν όλο το εύρος αυτού του πεδίου. Για αυτό παρά την πληθώρα ερευνών που έχουν πραγματοποιηθεί δεν υπάρχει βεβαιότητα ότι τα νευρωνικά δίκτυα είναι καλύτερα από τις κλασικές μεθόδους πρόβλεψης.

Στη συγκεκριμένη διπλωματική εργασία επιχειρήθηκαν δύο πράγματα. Το πρώτο από αυτά ήταν η διερεύνηση αλγορίθμων που έχουν σκοπό τη βελτιστοποίηση της απόδοσης των δικτύων κατά την πρόβλεψη. Το δεύτερο αφορούσε την ίδια τη φύση των νευρωνικών δικτύων και για αυτό διερευνήθηκε σε μικρότερο βαθμό σε σχέση με το πρώτο ερώτημα κατά πόσο η απόδοση των νευρωνικών δικτύων επηρεάζεται από τον αριθμό των νευρώνων, των κρυφών στρωμάτων καθώς και τον αριθμό των σημείων που παρέχονται στο σετ δεδομένων.

Για τη διερεύνηση των αλγορίθμων που αποσκοπούν στη βελτιστοποίηση της απόδοσης κατά την πρόβλεψη χρονοσειρών πραγματοποιήθηκε μία σειρά από υπολογιστικά πειράματα σε 5 αλγορίθμους. Κατά τα πειράματα αυτά δόθηκε στα δίκτυα ένα σετ δεδομένων 40 σημείων της συνάρτησης του ημιτόνου για τη διαδικασία της εκπαίδευσης. Στα δίκτυα αυτά μεταβάλλαμε τον αριθμό των κρυφών στρωμάτων καθώς και των νευρώνων σε κάθε κρυφό στρώμα με στόχο να ευρεθεί ο αλγόριθμος που παράγει τα βέλτιστα αποτελέσματα και παράλληλα το δίκτυο με την πιο ακριβή πρόβλεψη, μέσα από τη διαδικασία σφάλματος και διόρθωσης.

Έτσι κατά την εξέταση των αποτελεσμάτων παρατηρήθηκε ότι ο δεύτερος από τους εξεταζόμενους αλγορίθμους παράγει μεγαλύτερο πλήθος δικτύων που πραγματοποιούν καλές προβλέψεις ανεξαρτήτως αριθμού κρυφών στρωμάτων. Η απόδοση αυτού του αλγορίθμου είναι καλύτερη και από την αντίστοιχη του πρώτου αλγορίθμου, ο οποίος είναι ο κλασικός τρόπος εκπαίδευσης δικτύου, για το ίδιο σετ δεδομένων. Επίσης η ακρίβεια των παραγόμενων από τον αλγόριθμο 2 δικτύων είναι αρκετά καλύτερη από εκείνη των αντίστοιχων δικτύων για τους υπόλοιπους αλγορίθμους για 40 σημεία εκπαίδευσης. Μάλιστα τα δύο δίκτυα τα οποία πραγματοποίησαν τις πιο ακριβείς προβλέψεις για ένα και για δύο κρυφά στρώματα είχαν εκπαιδευτεί με την συμβολή του αλγορίθμου 2.

Από τους υπόλοιπους αλγορίθμους εκτός του πρώτου και του δεύτερου μόνο ο τέταρτος αλγόριθμος κατάφερε να παράγει δίκτυα τα οποία να πραγματοποιούν καλές

προβλέψεις. Παρόλα αυτά ο δεύτερος αλγόριθμος υπερτερεί των υπολοίπων και σε αριθμό αλλά και σε ακρίβεια πρόβλεψης.

Το δεύτερο σκέλος της επίδρασης ορισμένων παραγόντων στην απόδοση του νευρωνικού δικτύου διερευνήθηκε συμπληρωματικά με το πρώτο ερώτημα. Έτσι στον πρώτο αλγόριθμο, που αποτελεί και την κλασική μέθοδο εκπαίδευσης χωρίς την παρεμβολή κάποιου από τους αλγορίθμους βελτιστοποίησης, εξετάστηκαν επιπλέον και οι περιπτώσεις σετ δεδομένων εκπαίδευσης 160 και 320 σημείων εκτός από το σετ δεδομένων 40 σημείων που δόθηκε σε όλους τους αλγορίθμους. Παράλληλα μεταβάλλονταν και ο αριθμός των κρυφών στρωμάτων καθώς και ο αριθμός των νευρώνων.

Έτσι μετά την εκτέλεση των πειραμάτων και την εξέταση των αποτελεσμάτων εξήχθησαν ορισμένα συμπεράσματα. Στα δίκτυα του ενός κρυφού στρώματος ο αριθμός των δεδομένων δείχνει να διαδραματίζει σημαντικό ρόλο στην απόδοση του δικτύου. Αυτό το συμπέρασμα εξάγεται από το γεγονός ότι με την αύξηση του αριθμού των σημείων αυξήθηκαν τόσο ο αριθμός των δικτύων που πραγματοποιούν καλή πρόβλεψη όσο και η ακρίβεια των προβλέψεων. Έτσι τα δίκτυα που παρήγαγαν καλή πρόβλεψη είναι περισσότερα από τα αντίστοιχα για 160 και 40 σημεία ενώ η ακρίβεια των προβλέψεων δείχνει να βελτιώνεται όπως φαίνεται και από τις τιμές που λαμβάνουν οι συναρτήσεις απόδοσης.

Επιπλέον ο αριθμός των νευρώνων δείχνει να επηρεάζει την απόδοση του δικτύου καθώς η συντριπτική πλειοψηφία των δικτύων που πραγματοποιούν καλές προβλέψεις βρίσκονται μεταξύ για αριθμούς νευρώνων από 6 μέχρι 35. Αντίθετα η ακρίβεια των προβλέψεων δεν δείχνει να επηρεάζεται με την μεταβολή του αριθμού των νευρώνων για τα δίκτυα ενός κρυφού στρώματος.

Όσον αφορά τα δίκτυα δύο κρυφών στρωμάτων εκεί η επίδραση που έχει η μεταβολή του αριθμού των σημείων στην απόδοση του δικτύου δεν είναι τόσο ξεκάθαρη. Αυτό γιατί ο αριθμός των δικτύων που πραγματοποιούν καλή πρόβλεψη είναι μεγαλύτερος για τα δίκτυα 160 σημείων παρά για εκείνα των 320. Επίσης ο αριθμός κρυφών νευρώνων δείχνει να έχει την ίδια επίδραση που έχει και για τα δίκτυα του ενός στρώματος καθώς τα περισσότερα δίκτυα που κάνουν καλή πρόβλεψη εμφανίζονται για αριθμούς νευρώνων από 3 έως 8 για κάθε ένα από τα κρυφά στρώματα.

Τέλος όσον αφορά την επίδραση που έχει ο αριθμός κρυφών στρωμάτων στο δίκτυο φαίνεται ότι όσο αυξάνεται ο αριθμός των σημείων στο σετ δεδομένων τα δίκτυα δύο στρωμάτων που κάνουν καλή πρόβλεψη είναι περισσότερα από τα αντίστοιχα δίκτυα ενός κρυφού στρώματος. Παρόλα αυτά η ακρίβεια που επιτυγχάνουν τα δίκτυα ενός στρώματος είναι πολύ καλύτερη σε σχέση με εκείνης των δικτύων δύο στρωμάτων.

Τα δίκτυα που επιλέχθηκαν μετά από την εξέταση των αποτελεσμάτων χρησιμοποιήθηκαν για την πρόβλεψη των τιμών μιας χρονοσειράς η οποία περιελάμβανε τις ωριαίες τιμές του συνολικού φορτίου για κάθε μέρα του Οκτωβρίου του 2010. Η απόδοση των δικτύων στην πρόβλεψη της συγκεκριμένης χρονοσειράς

ήταν χαμηλή. Αυτό μπορεί να οφείλεται κατά κύριο λόγο στο γεγονός ότι η επιλογή των καλύτερων δικτύων έγινε με βάση την απόδοσή τους για κατά την πρόβλεψη της συνάρτησης του ημιτόνου, μίας κατά κάποιο τρόπο σύνθετης χρονοσειρά που διαφέρει από τις αντίστοιχες που έχουν πραγματικά δεδομένα.

Εν κατακλείδι τα συμπεράσματα που εξήχθησαν μέσα από την υπολογιστική διαδικασία επιβεβαιώνουν αυτά που αναφέρθηκαν στην πρώτη παράγραφο του κεφαλαίου σχετικά με την αδυναμία συστηματικής καταγραφής της συμπεριφοράς των νευρωνικών δικτύων. Παρόλο που τα νευρωνικά δίκτυα έχουν μία σειρά χαρακτηριστικών που τα καθιστούν ικανά να διαχειριστούν προβλήματα από μία μεγάλη ποικιλία πεδίων, ποτέ δεν θα αποτελέσουν πανάκεια. Επίσης καθίσταται σαφές ότι είναι πολύ δύσκολο να υπάρξει ένα συγκεκριμένο μοντέλο ή μία συγκεκριμένη αρχιτεκτονική δικτύου που θα είναι κατάλληλη και θα πραγματοποιεί καλές προβλέψεις για κάθε είδους πρόβλημα.

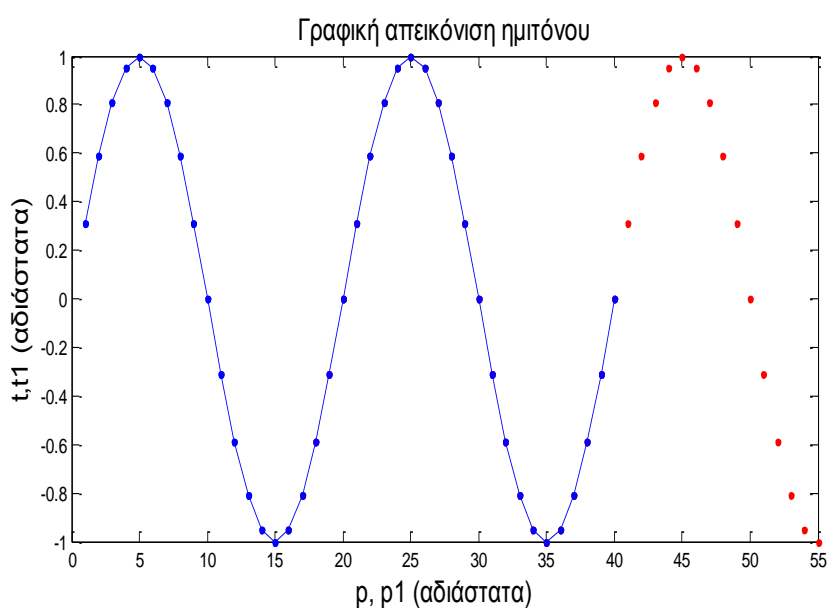
Παρά τα αρνητικά τους τα νευρωνικά δίκτυα αποτελούν ένα ισχυρό εργαλείο για την πρόβλεψη χρονοσειρών, λόγω κυρίως της δυνατότητας τους να εντοπίζουν μη γραμμικές σχέσεις μεταξύ των δεδομένων. Αν υπάρξει μία συστηματική καταγραφή των ιδιοτήτων τους και της συμπεριφοράς τους για τις διάφορες περιπτώσεις τους τότε μελλοντικά μπορούν να αποτελέσουν μια εξαιρετική περίπτωση μοντέλου για την πρόβλεψη χρονοσειρών και όχι μόνο.

ΠΑΡΑΡΤΗΜΑ

Παράρτημα Α: Σετ δεδομένων εκπαίδευσης.

- 1^ο σετ δεδομένων: Συνάρτηση ημιτόνου

Training set		Test set	
p	$t=\sin(0.1*p*\pi)$	$p1=p+15$	$t1=\sin(0.1*p1*\pi)$
1	0,3090	16	-0,9511
2	0,5878	17	-0,8090
3	0,8090	18	-0,5878
4	0,9511	19	-0,3090
5	1,0000	20	0,0000
6	0,9511	21	0,3090
7	0,8090	22	0,5878
8	0,5878	23	0,8090
9	0,3090	24	0,9511
10	0,0000	25	1,0000
11	-0,3090	26	0,9511
12	-0,5878	27	0,8090
13	-0,8090	28	0,5878
14	-0,9511	29	0,3090
15	-1,0000	30	0,0000
16	-0,9511	31	-0,3090
17	-0,8090	32	-0,5878
18	-0,5878	33	-0,8090
19	-0,3090	34	-0,9511
20	0,0000	35	-1,0000
21	0,3090	36	-0,9511
22	0,5878	37	-0,8090
23	0,8090	38	-0,5878
24	0,9511	39	-0,3090
25	1,0000	40	0,0000
26	0,9511	41	0,3090
27	0,8090	42	0,5878
28	0,5878	43	0,8090
29	0,3090	44	0,9511
30	0,0000	45	1,0000
31	-0,3090	46	0,9511
32	-0,5878	47	0,8090
33	-0,8090	48	0,5878
34	-0,9511	49	0,3090
35	-1,0000	50	0,0000
36	-0,9511	51	-0,3090
37	-0,8090	52	-0,5878
38	-0,5878	53	-0,8090
39	-0,3090	54	-0,9511
40	0,0000	55	-1,0000



Εικόνα Α.1: Γραφική απεικόνιση ημιτόνου

Πίνακας Α.1: Δεδομένα πρώτης υπολογιστικής διαδικασίας

- 2^ο σετ δεδομένων: Συνολικού ηλεκτρικό φορτίο (kWh) Οκτωβρίου 2010

Hours Date	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	Συνολική Κατανάλωση
1/10/2010	4746	4373	4195	4103	4071	4193	4599	5172	5583	6089	6260	6344	6387	6350	6032	5721	5563	5614	5837	6553	6856	6215	5609	5113	131580
2/10/2010	4806	4487	4284	4191	4146	4167	4271	4461	5015	5591	5908	6142	6300	6253	5880	5437	5230	5330	5563	6253	6518	6061	5516	5108	126916
3/10/2010	4681	4281	4093	3974	3888	3893	3977	4065	4357	4843	5311	5657	5713	5536	5090	4792	4689	4788	5122	5831	6152	5902	5437	5037	117108
4/10/2010	4560	4191	3962	3887	3905	3964	4461	5095	5528	6008	6217	6352	6438	6458	6214	5809	5611	5605	5838	6570	6861	6369	5713	5156	130770
5/10/2010	4688	4409	4256	4199	4152	4261	4684	5277	5726	6196	6394	6535	6573	6555	6210	5844	5664	5754	6043	6833	7093	6473	5780	5232	134829
6/10/2010	4743	4466	4284	4176	4191	4263	4731	5375	5783	6303	6528	6683	6755	6715	6403	5966	5803	5886	6063	6804	6986	6484	5780	5236	136408
7/10/2010	4791	4483	4301	4225	4202	4331	4735	5347	5819	6315	6514	6699	6813	6715	6289	5915	5695	5776	6094	6870	7108	6525	5866	5312	136738
8/10/2010	4847	4490	4306	4198	4154	4178	4609	5352	5786	6226	6380	6445	6469	6418	6138	5807	5633	5712	6070	6868	7074	6511	5865	5370	134907
9/10/2010	4976	4599	4328	4143	4084	4102	4231	4449	5063	5670	6006	6166	6271	6243	5872	5553	5425	5478	5778	6737	6873	6418	5838	5438	129739
10/10/2010	4935	4503	4233	4073	3975	3970	4005	4041	4413	4893	5362	5784	5981	5813	5305	4967	4924	5044	5460	6280	6545	6257	5716	5305	121784
11/10/2010	4744	4312	4040	3928	3909	4045	4546	5283	5756	6205	6368	6490	6606	6679	6485	6166	6008	6064	6368	7060	7205	6681	5921	5517	136384
12/10/2010	4961	4579	4350	4228	4200	4268	4706	5423	5891	6420	6715	6823	6932	6879	6530	6174	6002	6098	6454	7213	7350	6758	6041	5588	140581
13/10/2010	5003	4598	4335	4218	4179	4323	4800	5504	5967	6506	6690	6786	6845	6795	6536	6125	5898	5941	6259	7008	7094	6582	5912	5430	139335
14/10/2010	4881	4497	4227	4124	4063	4169	4610	5340	5782	6301	6494	6666	6749	6690	6303	5942	5846	5895	6056	6845	7014	6433	5769	5295	135992
15/10/2010	4761	4365	4140	3992	4000	4139	4563	5180	5633	6087	6240	6274	6293	6299	6001	5678	5512	5630	5945	6757	6775	6235	5607	5187	131291
16/10/2010	4825	4414	4214	4060	3994	4023	4183	4324	4857	5382	5777	5954	6118	6127	5758	5344	5156	5255	5564	6410	6429	6004	5428	5106	124706
17/10/2010	4691	4290	3996	3869	3807	3789	3897	3926	4249	4747	5196	5558	5701	5500	5031	4679	4594	4737	5174	5979	6052	5728	5315	5024	115529
18/10/2010	4494	4102	3826	3750	3744	3807	4272	5029	5496	6035	6254	6418	6458	6459	6152	5772	5689	5723	6002	6695	6805	6302	5576	5240	130098
19/10/2010	4660	4306	4118	4014	3984	4101	4583	5270	5601	6030	6165	6245	6309	6277	5940	5549	5366	5464	5856	6772	6842	6276	5618	5193	130535
20/10/2010	4660	4322	4083	3962	3948	4071	4538	5255	5632	6078	6202	6252	6278	6196	5957	5585	5370	5373	5701	6637	6726	6266	5620	5181	129892

Πίνακας Α.5: Πίνακας με τα στοιχεία συνολικού φορτίου σε kWh Οκτωβρίου 2010

Παράρτημα Β: Κώδικες

- Κώδικας υπολογισμού συναρτήσεων απόδοσης στα σημεία πρόβλεψης

```

load('C:\Users\Billy\Desktop\Dok\2_code\2_hidden_layers\Data
points(40)\p1','p1');
forecast_points_indices=find(p1>40);% Πίνακας με τις θέσεις των
σημείων πρόβλεψης στο p1
forecast_points_values=p1(forecast_points_indices);% Πίνακας με τις
τιμές που αντιστοιχούν στις θέσεις των σημείων πρόβλεψης του p1
max_Neurons_layer1=input('Dwse max gia layer 1: ');
max_Neurons_layer2=input('Dwse max gia layer 2: ');
arithmos_dokimhs=0;
for i=1:1:max_Neurons_layer1
    for k=1:1:max_Neurons_layer2
mkdir(strcat('C:\Users\Billy\Desktop\Dok\2_code\2_hidden_layers\Actua
l_forecast_error_measures\',strcat(num2str(i),'-',num2str(k))))
% Φορτώνεται το array με τα forecast_error
load(strcat('C:\Users\Billy\Desktop\Dok\2_code\2_hidden_layers\Data
points(40)\',strcat('L1=',num2str(i)),'\',strcat(num2str(i),'-
',num2str(k)),'\',strcat('forecast_error_array',strcat(num2str(i),'-
',num2str(k))),'forecast_error1')
        for j=1:1:10
            arithmos_dokimhs=arithmos_dokimhs+1;
            disp('Dokimh:')
            disp(arithmos_dokimhs)
            flag=forecast_error1{j};
%Εδώ καταχωρούνται στον πίνακα actual_forecast_error1 οι τιμές από
%τον πίνακα forecast_error που αφορούν τα σημεία πρόβλεψης.
            actual_forecast_error1{j}=flag(forecast_points_indices);
            actual_forecast_error=flag(forecast_points_indices);
            actual_forecast_mse(j)=mse(actual_forecast_error);
            actual_forecast_mae(j)=mae(actual_forecast_error);
        end
clear actual_forecast_error1 actual_forecast_mse actual_forecast_mae
best_forecast_mse best_forecast_mae place_for_mse place_for_mae flag
end
end
end

```

Στον κώδικα αυτό αφού βρεθούν οι θέσεις και οι τιμές στον πίνακα p1 που αντιστοιχούν στα σημεία πρόβλεψης ($p1 > 40$) ξεκινάει η επαναληπτική διαδικασία κατά την οποία για κάθε δίκτυο (δηλαδή για κάθε συνδυασμό νευρώνων πρώτου (i) και δεύτερο (k) κρυφού στρώματος) φορτώνεται η συστοιχία forecast_error1. Από εκεί βάση του εκάστοτε αριθμού αρχικοποίησης (εμφωλευμένος βρόγχος j) ανασύρεται από τη συστοιχία ο αντίστοιχος πίνακας που περιλαμβάνει τα forecast_error που υπολογίστηκαν για το συγκεκριμένο δίκτυο. Έπειτα από αυτόν τον πίνακα επιλέγονται οι τιμές του σφάλματος που αντιστοιχούν στα σημεία που πραγματοποιείται η πρόβλεψη και τα οποία έχουμε βρει προηγουμένως και αποθηκεύονται στη θέση που αντιστοιχεί στον αριθμό της αρχικοποίησης στη συστοιχία actual_forecast_error1. Τελικά βάση αυτών των τιμών υπολογίζονται οι τιμές των συναρτήσεων απόδοσης MAE και MSE. Αυτές οι τιμές αποθηκεύονται στους πίνακες actual_forecast_mae και actual_forecast_mse αντίστοιχα..

- **Κώδικας ταξινόμησης δικτύων**

```

for j=1:1:10
    if mae_fitting_table(j,k)>=0.60 && mae_forecasting_table(j,k)
>=0.50 && mse_forecasting_table(j,k)>=0.37 && mse_fitting_table
(j,k)>=0.42 && mae_fitting_table(j,k)<=0.67 &&
mae_forecasting_table(j,k)<=0.71 && mse_fitting_table(j,k)<=0.55
&& mse_forecasting_table(j,k)<=0.65

        bad_fitting_xl(1,k)=bad_fitting_xl(1,k)+1;
        bad_fitting(1,i)=bad_fitting(1,i)+1;

    elseif mae_fitting_table(j,k)>=0.70 &&
mae_forecasting_table(j,k)>=5 && mse_fitting_table(j,k)>=1.2 &&
mse_forecasting_table(j,k)>=10 && mae_fitting_table(j,k)<=1.45 &&
mae_forecasting_table(j,k)<=10 && mse_fitting_table(j,k)<=6.5 &&
mse_forecasting_table(j,k)<=70

        bad_fitting_xl(1,k)=bad_fitting_xl(1,k)+1;
        bad_fitting(1,i)=bad_fitting(1,i)+1;

    elseif mae_fitting_table(j,k)<0.60 &&
mse_fitting_table(j,k)<1.3
        if mae_forecasting_table(j,k)<1.2 &&
mse_forecasting_table(j,k)<1.5

            good_forecasting_xl(1,k)=good_forecasting_xl(1,k)+1;
            good_forecasting(1,i)=good_forecasting(1,i)+1;
        else
            good_fitting_xl(1,k)=good_fitting_xl(1,k)+1;
            good_fitting(1,i)=good_fitting(1,i)+1;

        end
    else
        others_xl(1,k)=others_xl(1,k)+1;
    end
end
end

```

Αρχικά να σημειώσουμε ότι από το συγκεκριμένο μέρος του κώδικα έχουν αφαιρεθεί οι εντολές οι οποίες αφορούν τη φόρτωση των απαραίτητων δεδομένων καθώς και της δημιουργίας των απαραίτητων πινάκων. Στο τμήμα του κώδικα αυτού που παρατίθεται υπάρχει η εμφωλευμένος βρόγχος του αριθμού αρχικοποίησης καθώς και δύο βρόγχοι υπόθεσης. Η λογική της κατάταξης είναι απλή: Κακό fitting και κακό forecasting (δηλαδή ψηλές τιμές των MAE και MSE) οδηγεί στην κατηγορία bad_fitting. Στην περίπτωση που έχουμε καλό fitting (χαμηλές τιμές του fitting_mae και mse) τότε:

- Αν έχουμε καλό forecasting (χαμηλές τιμές actual_forecast_mae/mse) τότε τοποθετείται στην κατηγορία good_forecasting
- Σε κάθε άλλη περίπτωση τοποθετείται στην κατηγορία good_fitting

Αν τα όρια των εξεταζόμενων μεγεθών δεν ανήκουν σε καμιά από τις παραπάνω κατηγορίες τότε τοποθετείται το δίκτυο στην κατηγορία other και εξετάζεται οπτικά

από το συγγραφέα για να ταξινομηθεί σε κάποια από τις παραπάνω κατηγορίες. Να σημειωθεί ότι οι πρώτες δύο περιπτώσεις του βρόγχου υποθέσεων αφορούν την ταξινόμηση στην κατηγορία `bad_fitting` καθώς πιο συνήθως σε αυτή την κατηγορία υπήρχαν δύο περιπτώσεις. Στην πρώτη η απόκριση του δικτύου δεν είχε κάποια συγκεκριμένη μορφή και απλώς απέκλινε από τη επιθυμητή. Στη δεύτερη η απόκριση ήταν συνήθως ευθεία γραμμή η οποία συνέπιπτε με τον άξονα των x ή ήταν παράλληλη με αυτόν και για τις οποίες τα μετρούμενα μεγέθη έπαιρναν τιμές που κινούνταν σε πολύ περιορισμένα διαστήματα για αυτό ήταν σχετικά εύκολο να ανιχνευθούν.

Βιβλιογραφία και διαδικτυακοί σύνδεσμοι

1. Grauppe D., Principles of artificial neural networks, 2nd Edition, World Scientific Publishing, 2007
2. <http://el.wikipedia.org/wiki/%CE%9D%CE%B5%CF%85%CF%81%CF%8E%CE%BD%CE%B1%CF%82>
3. Hornik K., Stinchcombe M., White Halbert , Multilayer Feedforward Networks are Universal Approximators, Neural Networks, Vol. 2 ,pp. 359-366 (1989)
4. Hornik K., Approximation Capabilities of Multilayer Feedforward Networks, Neural Networks, Vol.4, pp. 251-257 (1991)
5. Palmer A., Montano J. J., Sese A., Designing an artificial neural network for forecasting tourism time series, Tourism Management, Vol. 27, pp. 781-790 (2006).
6. Kuruvilla J., Gunavathi K., Lung cancer classification using neural networks for CT images, Computer Methods and Programs in Biomedicine, Vol. 113, pp. 202-209 (2014).
7. Neural Network Toolbox, Matlab 7.10.0 (R2010a)
8. Haykin S., Neural Networks : A Comprehensive Foundation, Second Edition, Prentice Hall International Inc., 1999.
9. Bishop C., Neural Networks for Pattern Recognition, Clarendon Press, Oxford,1995
10. M.R. Gallagher, Multi-layer Perceptron Error Surfaces: Visualization Structure and Modeling, PhD Thesis, Australia 1999
11. http://crsouza.blogspot.gr/2009/11/neural-network-learning-by-levenberg_18.html
12. D.J MacKay, Probable networks and plausible predictions- A review of practical Bayesian methods for supervised neural networks, Cavendish Laboratory, Cambridge
13. <ftp://ftp.sas.com/pub/neural/FAQ3.html>
14. W. Wang, P. Van Gelder, J.K. Vrijling, Some Issues About the Generalization of Neural Networks for Time Series Prediction, Lecture Notes in Computer Science, Vol. 3697, pp 559-564 (2005)
15. S. Amari, N. Murata, K.R Muller, M. Finke, H. H. Yank, Asymptotic statistical theory of overtraining and cross-validation, IEEE Transactions on Neural Networks, Vol.8, pp 985-996 (1997)
16. W. Sarle, Stopped Training and Other Remedies for Overfitting, Proceedings of the 27th Symposium on the Interface (1995)
17. S. Geman , E. Bienenstock, Neural Networks and the Bias/Variance Dilemma, Neural Computation, Vol. 4, pp 1-58
18. <http://scott.fortmann-roe.com/docs/BiasVariance.html>
19. G. Zhang, B.E. Patuwo , M. Y. Hu, Forecasting with artificial neural networks: The state of the art, International Journal of Forecasting, Vol. 14, pp 35-62 (1998)

20. H.S. Hippert, C.E. Pedreira, R.C. Souza, Neural Networks for short-term load forecasting: A review and evaluation, *IEEE Transactions on Power Systems*, Vol. 16, pp 44-55 (2001)
21. G Zhang, B.E. Patuwo, M.Y. Hu , A simulation study of artificial neural networks for nonlinear time-series forecasting, *Computer and Operations Research*, Vol. 28,pp 381-396 (2001)
22. Z. Tang, P.A. Fishwick, Feedforward neural nets as models for time series forecasting, *ORSA Journal on Computing*, Vol.5, pp 374-385
23. G. Zhang, Min Qi, Neural Network forecasting for seasonal and trend time series, *European Journal of Operational Research*, Vol. 160, pp 501-514 (2005)
24. M. Nelson, T. Hill, W. Remus, M. O'Connor, Time series forecasting using NN's: Should the data be deseasonalized first? , *Journal of Forecasting*, Vol. 18, pp 359-367 (1999).
25. C. Hamzacebi, Improving artificial neural network's performance in seasonal time series forecasting, *Information Sciences*, Vol. 178, pp 4550-4559 (2008)
26. S. Balkin and J. Ord, Automatic neural network modeling for univariate time series, *International Journal of Forecasting*, Vol. 16, pp 509-515 (2000)
27. W. Wong, M. Xia, W. Chu, Adaptive neural network model for time series forecasting, *European Journal of Operational Research*, Vol. 207, pp 807-816 (2010)
28. P. Zhang, A neural network ensemble method with jittered training data for time series forecasting, *Information Science*, Vol. 177,pp 5329-5346 (2007)
29. G. Darbellay and M. Slama, Forecasting the short-term demand for electricity: Do neural networks stand a better chance?, *International Journal of Forecasting*, Vol.16, pp 71-83 (2000)
30. L. Oikonomou, Greek long-term energy consumption prediction using artificial neural networks, *Energy*, Vol. 35, pp 512-517 (2010)
31. A.G. Bakirtzis, V. Petridis, S .J. Kiartzis, M.C. Alexiadis, A neural network short term load forecasting model for the Greek power system, *IEEE Transactions on Power System*, Vol.11,No. 2 (1996)
32. D. Singhal, K.S. Swarup, Electricity price forecasting using artificial neural networks, *Electrical Power and Energy Systems*, Vol.33, pp 550-555 (2011)
33. H. Pao, Forecasting electricity market pricing using artificial neural networks, *Energy Conversion and Management*, Vol. 48, pp 907-912 (2007)
34. H. Hippert, D. Bunn, R. Souza, Large neural networks for electricity load forecasting: Are they overfitted?, *International Journal of Forecasting*, Vol. 21, pp 425-434 (2005)
35. K. Ermis, A. Midilli, I. Dincer, M.A. Rosen, Artificial neural network analysis of world green energy use, *Energy Policy*, Vol. 35, pp 1731-1743 (2007)
36. W. Yeh, Y. Yeh, P. Chang, Y. Ke, V. Chung, Forecasting wind power in the Mai Liao Wind Farm based on the multi-layer perceptron artificial neural network

model with improved simplified swarm optimization, *Electrical Power and Energy Systems*, Vol. 55, pp 741-748 (2014)

37. P. Balestrassi, E. Popova, A. Paiva, J. Marangon Lima, Design of experiments on neural network's training for nonlinear time series forecasting, *Neurocomputing*, Vol. 72, pp 1160-1178 (2009)
38. D. Coleman, D. Montgomery, A Systematic Approach to Planning for a Designed Industrial Experiment, *Technometrics*, Vol. 35, pp 1-27 (1993)