

Πολυτεχνείο Κρήτης

Τμήμα Μηχανικών Παραγωγής και Διοίκησης



Πρόβλεψη της τιμής της Εθνικής Τράπεζας της Ελλάδος με χρήση νευρωνικών δικτύων

Διπλωματική Εργασία

Ανδρέας Ε. Σωμαράς

Εξεταστική επιτροπή

Ιωάννης Μαρινάκης (Επιβλέπων)

Κωνσταντίνος Ζοπουνίδης

Χρυσοβαλάντης Γαγάνης

XANIA 2009

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον λέκτορα κ. Ιωάννη Μαρινάκη για την επίβλεψη αυτής της διπλωματικής εργασίας. Επίσης τον καθηγητή κ. Κωνσταντίνο Ζοπουνίδη και τον λέκτορα κ. Χρυσοβαλάντη Γαγάνη για τη συμμετοχή τους στην επιτροπή. Ιδιαίτερα θα ήθελα να ευχαριστήσω την δρ. Μαγδαληνή Μαρινάκη για την πολύτιμη βοήθεια και καθοδήγηση της.

ΠΕΡΙΛΗΨΗ

Στην παρούσα διπλωματική εργασία παρουσιάζεται η ανάπτυξη και υλοποίηση ενός νευρωνικού δικτύου για την επίλυση του προβλήματος της πρόβλεψης της τιμής των μετοχών, και συγκεκριμένα της τιμής της Εθνικής Τράπεζας της Ελλάδος (Ε.Τ.Ε.) στην επόμενη συνεδρίαση του Χρηματιστηρίου Αξιών της Αθήνας (Χ.Α.Α.). Ως παράγοντες χρησιμοποιούνται διάφοροι χρηματοοικονομικοί δείκτες και μεταβλητές από τη διεθνή και την τοπική οικονομία. Ο αλγόριθμος που εκτελεί το συγκεκριμένο τεχνητό νευρωνικό δίκτυο (ΤΝΔ) υλοποιήθηκε στο πρόγραμμα προσομοίωσης MATLAB (R2007b) με χρήση του ενσωματωμένου toolbox για Νευρωνικά Δίκτυα, με τιμές που χρονικά πάρθηκαν από 2/1/2008 έως 2/5/2008, ένα σύνολο δηλαδή 462 παρατηρήσεων που αφορούν στους 6 παράγοντες που τελικά επιλέχθηκαν.

ΠΕΡΙΕΧΟΜΕΝΑ

Ευχαριστίες	2
ΠΕΡΙΛΗΨΗ.....	3
1. ΕΙΣΑΓΩΓΗ.....	7
1.1 Ορισμός του προβλήματος.....	7
1.2 Διάρθρωση της Διπλωματικής εργασίας.....	11
2. ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΕΠΙΣΚΟΠΗΣΗ	12
2.1 Εισαγωγή.....	12
2.2 Τρόποι πρόβλεψης της τιμής των μετοχών....	12
2.3 Μοντέλα νευρωνικών δικτύων.....	13
3. ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ.....	18
3.1 Εισαγωγή.....	18
3.2 Τεχνητά Νευρωνικά Δίκτυα.....	20
3.3 Τύποι Νευρωνικών Δικτύων.....	23
3.3.1 Δίκτυα με Απλή Τροφοδότηση.....	27
3.3.2 Κανόνας Δέλτα	32
3.3.3 Αλγόριθμος Εκμάθησης με βάση την Ανάστροφη Διάδοση του Σφάλματος (back propagation learning algorithm)	33
3.4 Στάδια Διαδικασίας Ανάπτυξης ενός ΤΝΔ.....	36

3.5	Μάθηση και Ανάκληση.....	37
3.5.1	Μάθηση υπό επίβλεψη (supervised learning).....	38
3.5.2	Μάθηση χωρίς επίβλεψη (unsupervised learning).....	38
3.5.3	Ανταγωνιστική μάθηση (competitive learning).....	39
3.5.4	Ενισχυτική μάθηση (reinforced learning).....	40
3.5.5	Βαθμολογημένη μάθηση (graded learning).....	41
3.5.6	Τυχαία Μάθηση (random learning)....	41
3.6	Βασικά Χαρακτηριστικά Τεχνητών Νευρωνικών Δικτύων.....	42
4.	ΧΡΗΣΗ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ ΣΤΗΝ ΠΑΡΟΥΣΑ ΕΡΓΑΣΙΑ.	44
4.1	Εισαγωγή.....	44
4.2	Επιλογή παραμέτρων του δικτύου.....	44
4.3	Προσομοίωση νευρωνικού δικτύου με χρήση διαφόρων συναρτήσεων εκμάθησης.....	47
4.3.1	Ο αλγόριθμος gradient descent	47
4.3.2	Προσαρμογή του ρυθμού εκμάθησης	49
4.3.3	Χρήση όρου Momentum	50

4.3.4	Αλγόριθμος Levenberg-Marquardt....	51
5.	ΑΠΟΤΕΛΕΣΜΑΤΑ.....	53
5.1	Εισαγωγή.....	53
5.2	Αλγόριθμος gradient descent	54
5.3	Με προσαρμογή του ρυθμού εκμάθησης.....	57
5.4	Με χρήση όρου Momentum.....	60
5.5	Αλγόριθμος Levenberg-Marquardt.....	63
6.	ΣΥΜΠΕΡΑΣΜΑΤΑ - ΕΠΕΚΤΑΣΕΙΣ	66
7.	ΒΙΒΛΙΟΓΡΑΦΙΑ.....	69

1. ΕΙΣΑΓΩΓΗ

1.1 Ορισμός του προβλήματος

Σε ένα συνεχώς αναπτυσσόμενο επενδυτικό περιβάλλον είναι ζωτικής σημασίας η ανάγκη πρόβλεψης του επενδυτικού μέλλοντος. Η ανάγκη που εμφανίζεται σε κάθε άνθρωπο να επενδύσει με σχετική σιγουριά είναι έμφυτη. Αυτό είναι και το κίνητρο για την πραγματοποίηση αυτής της διπλωματικής εργασίας. Τον τελευταίο αιώνα έχουν αναπτυχθεί πολλά και διαφορετικά επενδυτικά προϊόντα που αφορούν τις εθνικές και διεθνείς χρηματαγορές. Το πιο δημοφιλές προϊόν για επενδύσεις είναι οι μετοχές που είναι ελκυστικές τόσο για μικροεπενδυτές και μικροκαταθέτες όσο και για μεγάλα χαρτοφυλάκια και διεθνείς επενδυτικούς οίκους. Ο φόβος όμως της απώλειας μέρους του κεφαλαίου ή ακόμα και ολόκληρου του κεφαλαίου λόγω των συχνών και μεγάλων διακυμάνσεων των τιμών των μετοχών κάνει τους επενδυτές επιφυλακτικούς, σε αυτή την ανασφάλεια έρχεται να απαντήσει αυτή η διπλωματική εργασία. Η οποία ως σκοπό έχει την παραγωγή ενός συστήματος πρόβλεψης της τιμής των μετοχών δημιουργώντας έτσι το απόλυτο ανταγωνιστικό πλεονέκτημα στον επενδυτή που έχει πρόσβαση στην πληροφορία αυτή. Τώρα σε τι ποσοστό θα είναι ασφαλείς οι προβλέψεις θα προκύψει στην πορεία και θα είναι ανάλογη της επιτυχίας της μεθοδολογίας και των δεδομένων. Για τους λόγους που προαναφέρθηκαν έχουν υπάρξει στη διεθνή βιβλιογραφία πολλές προσπάθειες πρόβλεψης γενικότερα αλλά και συγκεκριμένα στην περίπτωση μας δηλαδή στις μετοχές. Πιο συγκεκριμένα όσον αφορά τις μετοχές ο λόγος που παράγονται συνεχώς καινούρια μοντέλα και μεθοδολογίες είναι ότι εξαιτίας της δυναμικότητας των αγορών αλλά και των διεθνών συνθηκών, ούτε οι συνθήκες είναι ίδιες, ούτε οι παράγοντες που επηρεάζουν τις τιμές είναι οι ίδιοι ανά διάστημα και

εποχή, πράγμα το οποίο θα προσπαθήσει να αποδείξει και η παρούσα εργασία. Επίσης στην παρούσα εργασία θα βάλουμε νέους παράγοντες όπως η τιμή του πετρελαίου και η ισοτιμία του κοινού ευρωπαϊκού νομίσματος (Ευρώ) με το αντίστοιχο των Ηνωμένων Πολιτειών Αμερικής (Δολάριο). Το παρόν πρόβλημα, η πρόβλεψη δηλαδή με την κατασκευή ενός μοντέλου στην παρούσα περίπτωση εμπεριέχει έναν αστάθμητο παράγοντα, τον άνθρωπο, καθώς οι τιμές των μετοχών και η αύξηση ή μείωση της τιμής τους είναι αποτέλεσμα αποφάσεων των επενδυτών, πράγμα που εισάγει έναν απρόβλεπτο παράγοντα που λέγεται ψυχολογία του ατόμου και της μάζας. Οι περισσότερες εργασίες έχουν πραγματοποιηθεί σε λεγόμενες ώριμες αγορές, η παρούσα όμως εργασία ασχολείται με το Χρηματιστήριο Αξιών Αθηνών (ΧΑΑ) το οποίο κατατάσσεται στις αναπτυσσόμενες αγορές και η νευρικότητα των επενδυτών καθώς και ο φόβος είναι μεγαλύτερος πράγμα που κάνει την παρούσα εργασία ακόμα πιο ενδιαφέρουσα.

Η παραγωγή ενός συστήματος πρόβλεψης της τιμής ή της τάσης που θα παρουσιάσει μια μετοχή ήταν όπως προείπαμε και παραμένει το αντικείμενο, μιας μεγάλης μερίδας ατόμων που ασχολούνται με τη επιστημονική έρευνα. Έχουν αναπτυχθεί διάφορες στρατηγικές, μοντέλα και μεθοδολογίες για την πρόβλεψη της τιμής ή της τάσης μιας μετοχής ή ενός χρηματιστηριακού δείκτη. Τις περισσότερες φορές η προβλέψεις αφορούν δείκτες και όχι μεμονωμένες μετοχές λόγω το ότι οι δείκτες έχουν μικρότερες διακυμάνσεις εξαιτίας ότι απαρτίζονται από πολλές επιμέρους μετοχές, επίσης απαιτούνται λιγότεροι παράγοντες για την πρόβλεψη της τιμής ή της τάσης ενός δείκτη παρά μιας μεμονωμένης μετοχής. Γνωρίζουμε και από την στατιστική ότι ο μέσος όρος έχει μικρότερη διασπορά σε σχέση με τα πρωτογενή προϊόντα. Πρέπει όμως να προσέξουμε ότι οι δείκτες είναι αποτέλεσμα μέσων όρων και δείχνουν την τάση της αγοράς πράγμα, που είναι και ο σκοπός της δημιουργίας τους, οπότε μπορεί και συχνά προκαλούν σύγχυση και δίνουν λανθασμένες εντυπώσεις για την απόδοση, μιας μεμονωμένης μετοχής (Ατσαλάκης 2006). Η αποτελεσματικότητα των δεικτών αυτών εξαρτάται από το ποσοστό της αγοράς που αντιπροσωπεύουν. Συμβαίνει πολύ συχνά, για παράδειγμα, ο γενικός δείκτης του χρηματιστηρίου να είναι ανοδικός και κάποιες

μετοχές να είναι καθοδικές και αντιστρόφως. Η παρούσα εργασία σκοπό έχει την ανάπτυξη ενός συστήματος πρόβλεψης της τάσης της τιμής μιας μετοχής κατά την επόμενη συνεδρίαση στο Χρηματιστήριο Αξιών Αθηνών.

Τα Τεχνητά Νευρωνικά Δίκτυα (ΤΝΔ) ή απλά Νευρωνικά Δίκτυα αποτελούν μια προσπάθεια προσέγγισης της λειτουργίας του ανθρώπινου εγκεφάλου από έναν ηλεκτρονικό υπολογιστή. Η αρχιτεκτονική τους βασίζεται στην αρχιτεκτονική των βιολογικών νευρωνικών δικτύων. Τα τελευταία χρόνια η χρήση των νευρωνικών δικτύων εκτείνεται από τον τομέα των ηλεκτρονικών καταναλωτικών αγαθών και τον έλεγχο των βιομηχανικών διαδικασιών μέχρι τα συστήματα υποστήριξης αποφάσεων και πρόβλεψης στις χρηματαγορές. Στην παρούσα εργασία θα γίνει πρόβλεψη της τιμής των μετοχών με χρήση νευρωνικών δικτύων και συγκεκριμένα γίνεται πρόβλεψη της τιμής της εθνικής τράπεζας της Ελλάδας (ΕΤΕ), όπου προβλέπουμε την τιμή που θα έχει η μετοχή την επόμενη μέρα.

Λόγο του μεγάλου αριθμού μετοχών που διαπραγματεύονται στο Χρηματιστήριο Αξιών Αθηνών (ΧΑΑ) είναι από απίθανο ως ακατόρθωτο, στο πλαίσιο μίας διπλωματικής εργασίας, να επιτευχτεί πρόβλεψη για το σύνολο των μετοχών και αφού όπως προείπαμε, έχουμε να κάνουμε με μία νευρική από τη φύση της αναπτυσσόμενη αγορά έπρεπε να επικεντρωθούμε σε μία μετοχή. Σκοπός της παρούσας εργασίας είναι να προβλέψει με ικανοποιητική ακρίβεια την τάση της τιμής της μετοχής της ΕΤΕ κατά την επόμενη συνεδρίαση του ΧΑΑ με τη χρήση νευρωνικών δικτύων. Ο λόγος που επιλέχθηκε η συγκεκριμένη μετοχή κάθε άλλο από τυχαίος είναι. Είναι μια μετοχή που ανήκει στη κύρια αγορά του ΧΑΑ και έχει μεγάλη συμμετοχή στο γενικό δείκτη του ΧΑΑ, αποτελεί δηλαδή το 13,6% του γενικού δείκτη. Επίσης συμμετέχει και στους δείκτες (FTSE), (ΣΑΓΔ), (FTSEA), (ΔΤΡ), (FTSEI), (FTSEB) με ποσοστά 19,69%, 13,60%, 16,87%, 39,02% , 17,43% , 48,10% αντίστοιχα. Παράλληλα σκοπό έχουμε την ελαχιστοποίηση του χρόνου εκπαίδευσης και πρόβλεψης του συστήματος, ώστε η πρόβλεψη μας να καταστεί ένας ευέλικτος, γρήγορος και αποτελεσματικός τρόπος πρόβλεψης ώστε να είναι ένα ελκυστικό προϊόν για τους μελλοντικούς του χρήστες.

Λαμβάνοντας ως βάση τη θεωρία ότι στις τιμές των μετοχών αντανακλάται η επίδραση κάθε γεγονότος που μπορεί να επηρεάσει την αξία μιας μετοχής επιλέχθηκαν ως είσοδοι του συστήματος οι παρακάτω:

1. τιμή της μετοχής (ETE)
2. ισοτιμία Ευρώ – Δολαρίου (€-\$)
3. τιμή του πετρελαίου (BRENT)
4. τιμή κλεισίματος του γενικού δείκτη (ΓΔ) του Χρηματιστηρίου Αξιών Αθηνών (ΧΑΑ)
5. τιμή κλεισίματος του δείκτη FTSE 100 του Χρηματιστηρίου Αξιών του Λονδίνου
6. τιμή κλεισίματος του δείκτη DOW JONES του Χρηματιστηρίου Αξιών της Νέας Υόρκης (NYSE)

Τα δεδομένα που χρησιμοποιούνται για την ανάπτυξη του συστήματος αφορούν τις ημερήσιες τιμές κλεισίματος της μετοχής, των δεικτών, της ισοτιμίας Ευρώ – Δολαρίου, του πετρελαίου από 2/1/2008 έως 2/5/2008, συνολικά 462 παρατηρήσεις σε ένα διάστημα 4 μηνών. Είναι σημαντικό σε αυτό το σημείο να πούμε ότι την περίοδο που πραγματοποιείτε η εργασία η τιμή του πετρελαίου ήταν το χρηματιστηριακό προϊόν το οποίο όχι μόνο καθόριζε τις οικονομικές εξελίξεις αλλά και τις αναπτυξιακές πολιτικές και γενικότερες πολιτικές, όλων ανεξαιρέτως των κρατών του κόσμου. Το νόμισμα που γίνονται παγκοσμίως η συναλλαγές του λεγόμενου μαύρου χρυσού (πετρέλαιο) είναι το δολάριο των Ηνωμένων Πολιτειών Αμερικής (Η.Π.Α.) αυτός είναι και ο κυριότερος λόγος που συμπεριελήφθη σαν παράγοντας στην διαμόρφωση του μοντέλου. Η διαμόρφωση της ισοτιμίας μεταξύ των νομισμάτων των χωρών είναι σημαντικός παράγοντας που επηρεάζει τις εξελίξεις και τις τιμές των μετοχών και έρχεται και σε συνέχεια με το δεδομένο ότι οι τιμές του πετρελαίου είναι σε δολάρια ενώ στην περίπτωση μας ελέγχουμε την τιμή της μετοχής της ETE σε Ευρώ.

1.2 Διάρθρωση της Διπλωματικής εργασίας

Η δομή της διπλωματικής αυτής εργασίας είναι η ακόλουθη. Αποτελείται από 6 κεφάλαια ως εξής :

Στο 1^ο κεφάλαιο υπάρχουν τα κίνητρα, ο ορισμός, η μεθοδολογία επίλυσης του προβλήματος καθώς και ο σκοπός και η διάθρωση της παρούσας διπλωματικής εργασίας

Στο 2^ο κεφάλαιο παρουσιάζεται μια βιβλιογραφική επισκόπηση από τις σημαντικότερες εργασίες που έχουν πραγματοποιηθεί στον χώρο.

Στο 3^ο κεφάλαιο γίνεται περιγραφή της αρχών λειτουργίας των Τεχνητών Νευρωνικών Δικτύων. Επίσης, αναλύονται οι περισσότεροι γνωστές μέθοδοι Τεχνητών Νευρωνικών Δικτύων και οι εφαρμογές τους.

Στο 4^ο κεφάλαιο παρουσιάζεται ο τρόπος προσέγγισης του προβλήματος της παρούσας εργασίας με χρήση τεχνητών νευρωνικών δικτύων.

Στο 5^ο κεφάλαιο είναι συγκεντρωμένα τα αποτελέσματα των νευρωνικών δικτύων που αναπτύχθηκαν στο κεφάλαιο 4.

Στο 6^ο κεφάλαιο παρουσιάζονται τα συμπεράσματα που προκύπτουν από την παρούσα εργασία και προτείνονται μελλοντικές προεκτάσεις της.

Τέλος, στο 7^ο κεφάλαιο υπάρχει η βιβλιογραφία που χρησιμοποιήθηκε για την παρούσα εργασία.

2. ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΕΠΙΣΚΟΠΗΣΗ

2.1 Εισαγωγή

Η ενασχόληση με χρηματοοικονομικούς τομείς έχει απασχολήσει τελευταία έντονα την επιστημονική κοινότητα με θέματα όπως η πρόβλεψη χρηματιστηριακών τιμών, η πρόβλεψη τιμών συναλλάγματος, η πρόβλεψη τιμών πετρελαίου, η επιλογή χαρτοφυλακίου και πολλοί άλλοι τομείς. Στο κεφάλαιο που ακολουθεί πραγματοποιείται μια επισκόπηση μεθόδων πρόβλεψης γενικότερα και ειδικότερα μεθόδων πρόβλεψης με χρήση νευρωνικών δικτύων.

2.2 Τρόποι πρόβλεψης της τιμής των μετοχών

Από την αναζήτηση που έγινε προκύπτει ότι οι τρόποι πρόβλεψης ομαδοποιούνται σε έξι (6) κατηγορίες που είναι οι ακόλουθες :

1. Συμβατικά μοντέλα
2. Μη συμβατικές μέθοδοι πρόβλεψης
3. Ασαφή μοντέλα
4. Νευρο-ασαφή
5. Συστήματα που χρησιμοποιούν γενετικούς αλγορίθμους
6. Συστήματα που χρησιμοποιούν νευρωνικά δίκτυα

Αφού όμως στη παρούσα εργασία πραγματεύεται πρόβλεψη της τιμής των μετοχών με χρήση νευρωνικών δικτύων εκεί έχει δοθεί και περισσότερο βάρος.

2.3 Μοντέλα νευρωνικών δικτύων

Πολλοί ερευνητές έχουν ασχοληθεί με την πρόβλεψη με χρήση νευρωνικών δικτύων, μετά από σχετική έρευνα για χρήση τους σε χρηματιστήρια και χρηματιστηριακές μετοχές και δείκτες αναφέρονται κάποιες από αυτές τις εργασίες (Ατσαλάκης 2006).

Ο Wikowska D. (1995) έκανε πρόβλεψη για τις τιμές μετοχών σε τρεις γνωστές εταιρείες (Wedel Works, Okocim, Zywiec), εισηγμένες στο χρηματιστήριο της Βαρσοβίας, ανέπτυξε ένα τεχνητό νευρωνικό δίκτυο, με σκοπό να βρει τιμές για τις δέκα επόμενες συνεδριάσεις. Ο συγγραφέας ανέπτυξε 6 νευρωνικά δίκτυα τα οποία αποτελούνται από ένα κρυμμένο επίπεδο. Στα οποία χρησιμοποιεί μεταβλητές εισόδου όπως τιμές 4 μετοχών και την ισοτιμία δολαρίου και ζλότυ. Τελικά περιγράφει τις σχέσεις μεταξύ των τιμών των μετοχών και της τιμής του δολαρίου. Τα νευρωνικά δίκτυα έδωσαν πολύ καλά αποτελέσματα δηλαδή προβλέψεις αρκετά κοντά στις πραγματικές τιμές.

Ο Walezak S. (1999) έφτιαξε ένα νευρωνικό δίκτυο για να αξιολογήσει αναπτυσσόμενες χρηματιστηριακές αγορές, όπως της Σιγκαπούρης. Χρησιμοποίησε εννέα διαφορετικά σεντ μεταβλητών εισόδου στο μοντέλο πρόβλεψης, για να διερευνηθεί η αποτελεσματικότητα της χρηματιστηριακής αγοράς της Σιγκαπούρης. Πέτυχε ετήσιο ποσοστό απόδοσης 62% που μας δείχνει πόσο αποτελεσματικά είναι τα νευρωνικά δίκτυα.

Οι Chenoweth T. et al. (1996) έφτιαξαν ένα υβριδικό σύστημα για την πρόβλεψη του δείκτη S&P500 με τη χρήση δύο νευρωνικών δικτύων, από τα οποία το ένα εκπαιδεύεται με δεδομένα από ανοδική αγορά και το άλλο εκπαιδεύεται με δεδομένα από καθοδική αγορά. Συγκρίνοντας τα αποτελέσματα του υβριδικού συστήματος με ένα απλό σύστημα, το υβριδικό αποκτά υψηλότερες αποδόσεις με λιγότερες συναλλαγές.

Οι Kimoto T. et al. (1990) είχαν σαν στόχο να προβλέψουν την καλύτερη χρονική στιγμή αγοράς και πώλησης. Έτσι έφτιαξαν ένα νευρωνικό δίκτυο με τρία επίπεδα. Το σύστημα πρόβλεψης έδωσε ικανοποιητικά αποτελέσματα τα οποία όμως φάνηκαν με προσομοίωση.

Οι Podding T. et al. (1996) έφτιαξαν ένα παγκόσμιο μοντέλο με τεχνητά νευρωνικά δίκτυα αποτελούμενο από μετοχές, ομόλογα και συνάλλαγμα των Η.Π.Α., της Ιαπωνίας και της Γερμανίας, με σκοπό να καθοριστεί η συνολική τιμή ισορροπίας των περιουσιακών στοιχείων (asset). Συνέκριναν την απόδοση των μεμονωμένων μοντέλων των αγορών, με την απόδοση των ολοκληρωμένων μοντέλων και βρήκαν ότι το παγκόσμιο μοντέλο είχε καλύτερες αποδόσεις από τα μεμονωμένα.

Ο Kim S.S. (1998) έφτιαξε ένα νευρωνικό δίκτυο με χρονική υστέρηση ως μοντέλο πρόβλεψης για το χρηματιστήριο της Κορέας και το συνέκρινε με κάποια άλλα. Τα αποτελέσματα του το καταστούν πιο ακριβές από τα άλλα συγκρινόμενα μοντέλα.

Ο Min Q. (1999) εξέτασε τον δείκτη S&P500 χρησιμοποιώντας γραμμική παλινδρόμηση και μη γραμμικά νευρωνικά δίκτυα χρησιμοποιώντας εννέα μεταβλητές και έδειξε ότι το νευρωνικό μοντέλο υπερτερεί του γραμμικού μοντέλου καθώς δίνει μικρότερο σφάλμα.

Οι Baba N. et al. (1992) εφάρμοσαν ένα νευρωνικό δίκτυο που είχε δύο κρυφά επίπεδα, 15 εισόδους και μία έξοδο προβλέποντας απλά αν θα αυξηθεί ή θα μειωθεί η τιμή μιας μετοχής με ικανοποιητικά αποτελέσματα.

Οι Kim S.H. et al. (1998) μελέτησαν με τη χρήση ενός arrayed probabilistic network APN, το Χρηματιστήριο της Σιγκαπούρης. Το APN δίκτυο που χρησιμοποίησαν ήταν πολύ καλύτερο συγκρινόμενο με τα δίκτυα οπίσθιας τροφοδότησης και το recurrent νευρωνικό δίκτυο.

Οι Baba N. et al. (2002) χρησιμοποίησαν ένα νευρωνικό δίκτυο για το χρηματιστήριο του Τόκιο. Επέλεξαν 16 μετοχές, χρησιμοποίησαν δεκατέσσερις μεταβλητές και το σύστημα παρήγαγε ικανοποιητικά κέρδη στο χρηματιστήριο του Τόκιο ακόμα και την περίοδο που το χρηματιστήριο έπεφτε.

Οι Wittkemper H.G. et al. (1996) συνέκριναν επτά παραδοσιακά μοντέλα, με δύο νευρωνικά δίκτυα, με αποτέλεσμα, το γενικό παλίνδρομο νευρωνικό δίκτυο να έχει καλύτερη απόδοση από το δίκτυο το οποίο χρησιμοποιούσε θεμελιώδεις μεταβλητές.

Οι Harvey C.R. et al. (2000) συνέκριναν νευρωνικά δίκτυα με γραμμικά μοντέλα πρόβλεψης για τις αναπτυσσόμενες αγορές χρησιμοποίησαν τρία μέτρα απόδοσης : τις συνολικές αποδόσεις, την κατεύθυνση της αγοράς και τη μεταβολή της κατεύθυνσης της αγοράς. Το νευρωνικό δίκτυο σε σχέση και με τις άλλες στρατηγικές, έχει πολύ καλύτερα αποτελέσματα.

Οι Grudnitski G. et al. (1993) έκανα χρήση νευρωνικών δικτύων για το χρηματιστήριο της Αμερικής και το δείκτη S&P500. Το μοντέλο έχει σαν εισόδους : τους μηνιαίους ρυθμούς ανάπτυξης της συνολικής προσφοράς χρήματος, τις μεταβολές της διακύμανσης του S&P και των χρυσών συμβολαίων μελλοντικής εκπλήρωσης και τα ποσοστά προμηθειών στο τέλος του μήνα των μεγάλων κερδοσκοπών, των μεγάλων αντισταθμιστών και των μικρών συναλλασσομένων. Έξοδος του μοντέλου είναι η μέση μηνιαία αλλαγή της τιμής και τα αποτελέσματα δείχνουν, μια σημαντική δυνατότητα πρόβλεψης της κατεύθυνσης της μηνιαίας τιμής των συμβολαίων μελλοντικής εκπλήρωσης του δείκτη S&P.

Οι Quah T.S. et al. (1999) χρησιμοποίησαν ένα νευρωνικό δίκτυο για το χρηματιστήριο της Σιγκαπούρης. Χρησιμοποιήθηκαν ως είσοδοι χρηματοοικονομικές μεταβλητές και ως έξοδος χρησιμοποιήθηκε η απόδοση της μετοχής, σαν σκοπό είχαν να επιλέξουν μετοχές που θα υπερβαίνουν σε απόδοση τη συνολική αγορά. Η έξοδος ήταν η διαφορά μεταξύ της απόδοσης των μετοχών και της απόδοσης της συνολικής αγοράς. Τα αποτελέσματα ήταν πολύ καλά καθώς ξεπέρασαν κατά ένα μεγάλο ποσοστό τη συνολική αγορά.

Οι Motiwalla L. et al. (2000) έκαναν χρήση ενός νευρωνικού δικτύου και ενός μοντέλου παλινδρόμησης. Σαν αποτέλεσμα είχαν ότι το νευρωνικό μοντέλο είναι περισσότερο αποδοτικό σε σχέση με το μοντέλο παλινδρόμησης.

Οι Yiwen Y. et al. (2000) χρησιμοποίησαν ένα νευρωνικό δίκτυο οπίσθιας τροφοδότησης με ρυθμό μάθησης 0,1 με moment παράμετρο 0,09 και με δομή επιπέδων 6-10-1, λαμβάνοντας τα επαναδημιουργηθέντα διανύσματα ως πολλαπλές εισόδους για το χρηματιστήριο της Σαγκάης και το δείκτη SSE. Τα αποτελέσματα τους ήταν ικανοποιητικά.

Ο Casas C.A. (2001) χρησιμοποίησε ένα νευρωνικό δίκτυο, για να υποστηρίξει μια τακτική τοποθέτησης κεφαλαίων σε μετοχές, σε ομόλογα και στην αγορά χρήματος. Έξοδος ήταν το ποσοστό της απόδοσης της κάθε τοποθέτησης. Με χρήση προσομοίωσης υπήρξε καλό αποτέλεσμα.

Οι Atiya A. et al. (1997) προσπάθησαν πρόβλεψη των τιμών των μετοχών με νευρωνικά δίκτυα. Ως εισόδους είχαν : τα κέρδη ανά μετοχή, τον πολλαπλασιαστής κερδών, τα μερίσματα, τις πωλήσεις και τα περιθώρια κέρδους. Τα αποτελέσματα του νευρωνικού δικτύου είναι πάρα πολύ καλά συγκρινόμενα με άλλες μεθόδους.

Ο Andreou A. S. (2000) χρησιμοποίησε νευρωνικά δίκτυα, για να προβλέψει τις τιμές μετοχών στο χρηματιστήριο της Κύπρου. Προσπάθησε να δείξει την επίδραση οικονομικών και πολιτικών παραγόντων από το εθνικό και διεθνές περιβάλλον στις τιμές

των μετοχών. Χρησιμοποίησε το γενικό δείκτη CSE και δεδομένα μετοχών από πέντε κλάδους της αγοράς. Τα αποτελέσματα του ήταν ικανοποιητικά.

Οι Fernandez-Rodriguez F. et al. (2000) έδειξαν πόση είναι η κερδοφορία που μπορεί να επιτύχουν τα νευρωνικά δίκτυα στις αγορές μετοχών και συγκεκριμένα στο χρηματιστήριο της Μαδρίτης. Σαν είσοδοι χρησιμοποιήθηκαν παρελθοντικές τιμές, ενώ η έξοδος είναι μια τιμή στο διάστημα $(-1, 1)$. Όταν έχουμε θετικό σημαίνει να αγοράσουμε, ενώ όταν είναι αρνητικό σημαίνει πώληση.

Οι Chen A.S. et al. (2003) με χρήση νευρωνικού δικτύου προβλέπουν την κατεύθυνση του δείκτη του χρηματιστηρίου της Taiwan.

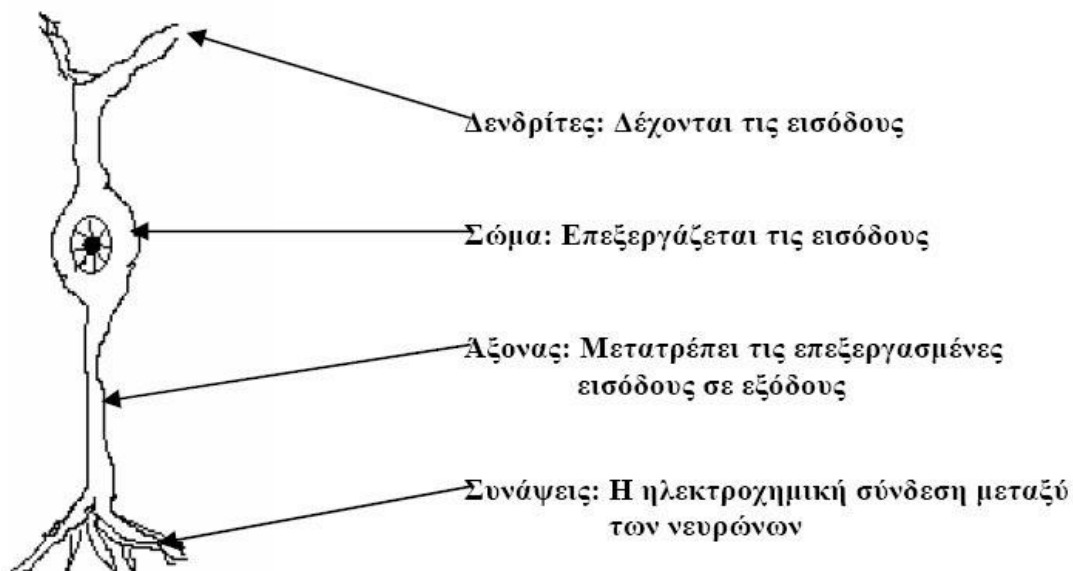
Οι Thawornwong S. et al. δημιούργησαν ένα νευρωνικό δίκτυο που έχει ένα κρυφό επίπεδο και χρησιμοποιεί ως συνάρτηση μεταφοράς μια σιγμοειδή υπερβολική εφαπτόμενη συνάρτηση και το συγκρίνει με ένα μοντέλο γραμμικής παλινδρόμησης.

Οι Schumann M. et al. (1993) εφάρμοσαν ένα νευρωνικό για το γερμανικό δείκτη DAX και τη μετοχή Hochtief. Άλλα τα αποτελέσματα τους δεν ήταν ξεκάθαρα.

3. ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

3.1 Εισαγωγή

Τα τεχνητά νευρωνικά δίκτυα ανήκουν στον τομέα της Μη Συμβολικής Τεχνητής Νοημοσύνης (non symbolic AI), η οποία προσομοιώνει βιολογικές διαδικασίες, όπως τη λειτουργία του εγκεφάλου ή τη διαδικασία εξέλιξης των ειδών, και η οποία διαφέρει από τη Συμβολική Τεχνητή Νοημοσύνη (symbolic AI), που προσομοιώνει τον τρόπο σκέψης του ανθρώπου, χρησιμοποιώντας ως δομικές μονάδες τα σύμβολα (ένα σύμβολο μπορεί να αναπαριστά μια έννοια ή μια σχέση ανάμεσα σε έννοιες). Η ικανότητα του ανθρώπου να σκέφτεται, να θυμάται και να επιλύει προβλήματα εντοπίζεται στον εγκέφαλό του. Όπως είναι γνωστό από τη Βιολογία, η δομική μονάδα του εγκεφάλου είναι ο νευρώνας. Ένας τυπικός βιολογικός νευρώνας αποτελείται από το σώμα, που συνιστά τον πυρήνα του, τους δενδρίτες, μέσω των οποίων λαμβάνει σήματα από γειτονικούς νευρώνες (σημεία εισόδου), και τον άξονα, που αποτελεί την έξοδο του νευρώνα και το μέσο σύνδεσής του με τους άλλους νευρώνες (Σχήμα 3.1). Σε κάθε δενδρίτη υπάρχει ένα απειροελάχιστο κενό που ονομάζεται συνάψη. Οι συνάψεις μέσω χημικών διαδικασιών επιταχύνουν ή επιβραδύνουν τη ροή ηλεκτρικών φορτίων προς το σώμα του νευρώνα. Η ικανότητα μάθησης και μνήμης που εμφανίζει ο εγκέφαλος οφείλεται στην ικανότητα των συνάψεων να μεταβάλλουν την αγωγιμότητά τους. Τα ηλεκτρικά σήματα που εισέρχονται στο σώμα μέσω των δενδριτών συνδυάζονται και, εφόσον το αποτέλεσμα ξεπερνά κάποια τιμή κατωφλίου, το σήμα διαδίδεται με τη βοήθεια του άξονα προς άλλους νευρώνες (Wasserman 1989).



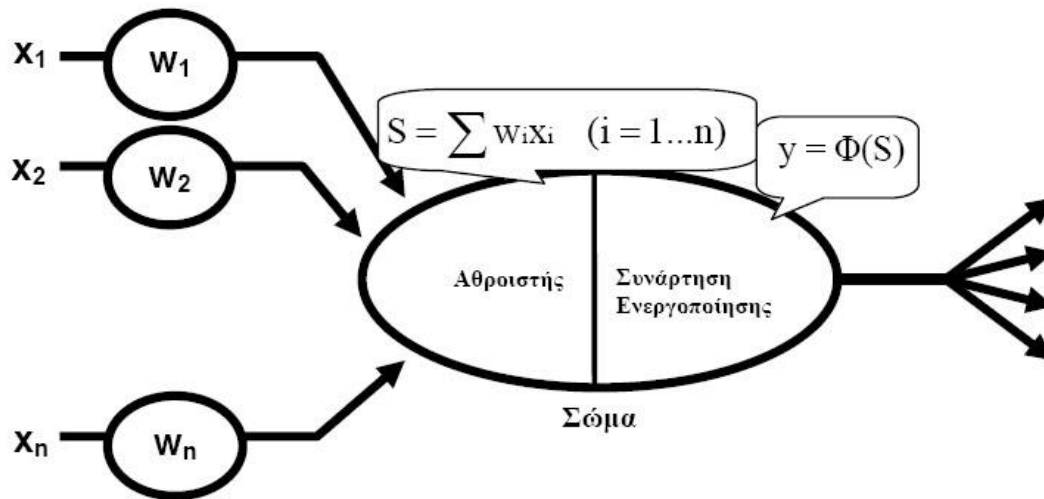
Σχήμα 3.1: Απεικόνιση απλού νευρώνα

Ο εγκέφαλος είναι σε θέση να λαμβάνει πολύπλοκες αποφάσεις εκπληκτικά γρήγορα παρά το ότι ο χρόνος απόκρισης των βιολογικών νευρώνων είναι της τάξης των χιλιοστών του δευτερολέπτου. Κατά μια άποψη, αυτό οφείλεται στο ότι η υπολογιστική ικανότητα του εγκεφάλου και η πληροφορία που περιέχει είναι διαμερισμένα σε όλο του τον όγκο. Πρόκειται, δηλαδή, για ένα παράλληλο και κατανεμημένο υπολογιστικό σύστημα. Αυτά τα χαρακτηριστικά διαμορφώνουν το κυριότερο κίνητρο πίσω από την επιθυμία να μοντελοποιηθεί ο ανθρώπινος εγκέφαλος με τα τεχνητά νευρωνικά δίκτυα.

3.2 Τεχνητά Νευρωνικά Δίκτυα

Ο τεχνητός νευρώνας (artificial neuron) αποτελεί το πρωταρχικό δομικό στοιχείο των τεχνητών νευρωνικών δικτύων και είναι ένα υπολογιστικό μοντέλο τα μέρη του οποίου αντιστοιχίζονται άμεσα με αυτά του βιολογικού νευρώνα. Όπως απεικονίζεται στο σχήμα 3.2, ένας τεχνητός νευρώνας δέχεται κάποια σήματα εισόδου $x_0, x_1, \dots, x_n$, τα οποία, σε αντίθεση με τους ηλεκτρικούς παλμούς του εγκεφάλου, αντιστοιχούν σε συνεχείς μεταβλητές. Κάθε τέτοιο σήμα εισόδου σταθμίζεται με κάποιο βάρος w_i (weight), ο ρόλος του οποίου είναι αντίστοιχος με εκείνον της σύναψης του βιολογικού εγκεφάλου. Η τιμή βάρους μπορεί να είναι θετική ή αρνητική σε αντιστοιχία με την επιταχυντική ή επιβραδυντική λειτουργία της σύναψης. Το σώμα του τεχνητού νευρώνα χωρίζεται σε δύο μέρη:

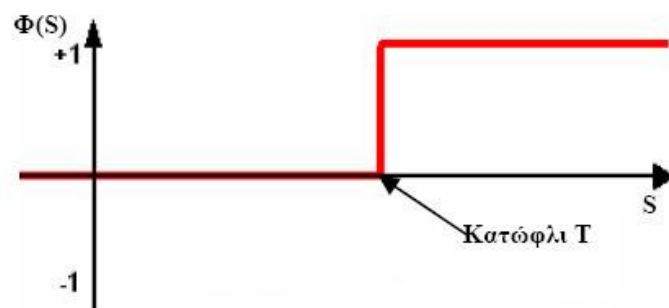
1. Τον αθροιστή (sum), ο οποίος προσθέτει τα επηρεασμένα από τα βάρη σήματα εισόδου και παράγει την ποσότητα $S = \sum w_i x_i, i = 1 \dots n$
2. Τη συνάρτηση ενεργοποίησης ή κατωφλίου (activation ή threshold ή transformation function), που αποτελεί ένα μη γραμμικό φίλτρο το οποίο διαμορφώνει την τελική τιμή του σήματος εξόδου y , σε συνάρτηση με την ποσότητα S .



Σχήμα 3.2: Μοντέλο Τεχνητού Νευρώνα

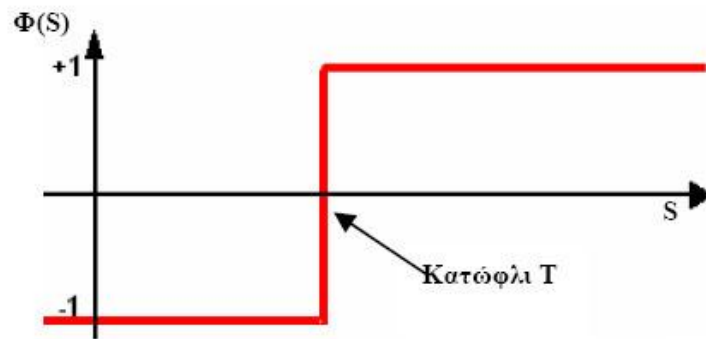
Στα Σχήματα 3.3 - 3.5 παρουσιάζονται τρεις τυπικές μορφές που μπορεί να λάβει η συνάρτηση ενεργοποίησης:

1. Η βηματική (step) συνάρτηση η οποία δίνει στην έξοδο αποτέλεσμα (συνήθως 1), εφόσον η τιμή που υπολογίζει ο αθροιστής είναι μικρότερη (ή μεγαλύτερη) από μια τιμή κατωφλίου T .



Σχήμα 3.3: Βηματική Συνάρτηση

2. Η συνάρτηση πρόσημου (sign), η οποία δίνει στην έξοδο αρνητική (ή θετική) πληροφορία, εφόσον η τιμή που υπολογίζει ο αθροιστής είναι μικρότερη (ή μεγαλύτερη) από μια τιμή κατωφλίου T.

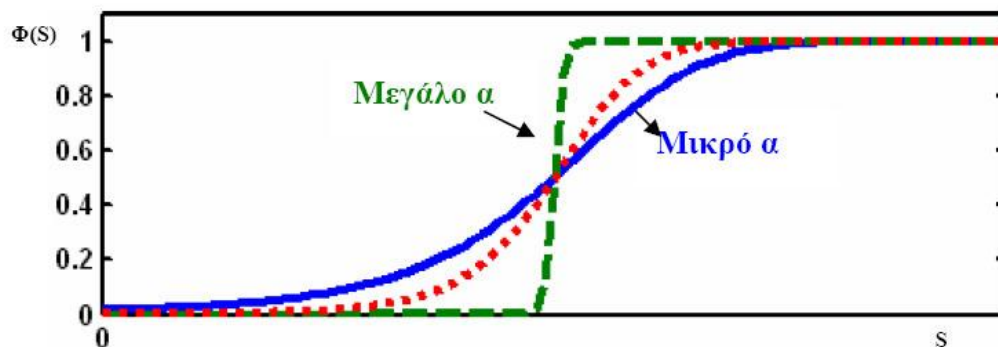


Σχήμα 3.4: Συνάρτηση Πρόσημου

3. Η σιγμοειδής (sigmoid) συνάρτηση, η οποία εκφράζεται από τη γενική σχέση:

$$\Phi = \frac{1}{1 + e^{-a s}} \quad (3.1)$$

όπου a είναι ένας συντελεστής ρύθμισης της ταχύτητας μετάβασης μεταξύ των δύο ασύμπτωτων τιμών. Η σιγμοειδής συνάρτηση είναι σημαντική γιατί παρέχει μη γραμμικότητα στο νευρώνα, η οποία είναι απαραίτητη για τη μοντελοποίηση μη γραμμικών φαινομένων.



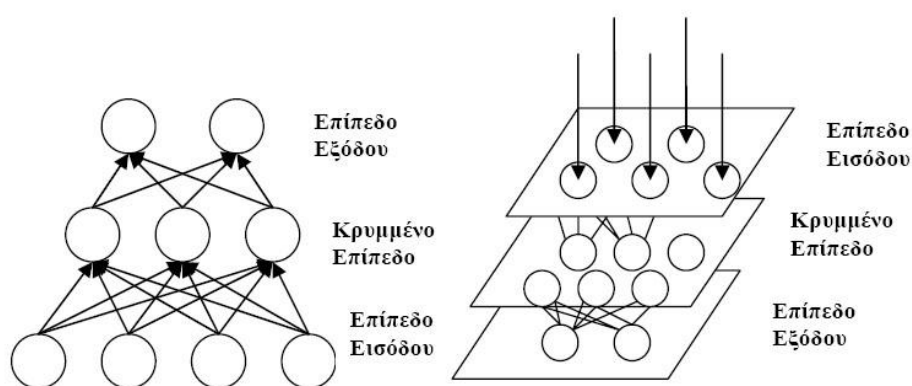
Σχήμα 3.5: Σιγμοειδής Συνάρτηση

3.3 Τύποι Νευρωνικών Δικτύων

Τα Τεχνητά Νευρωνικά Δίκτυα (Artificial Neural Networks) ή ΤΝΔ, χαρακτηρίζονται ως συστήματα επεξεργασίας δεδομένων που αποτελούνται από ένα πλήθος τεχνητών νευρώνων οργανωμένων σε δομές παρόμοιες με αυτές του ανθρώπινου εγκεφάλου. Συνήθως οι τεχνητοί νευρώνες είναι οργανωμένοι σε μία σειρά από στρώματα ή επίπεδα (layers). Το πρώτο από αυτά τα επίπεδα ονομάζεται επίπεδο εισόδου (input layer) και χρησιμοποιείται για την εισαγωγή των δεδομένων. Τα στοιχεία, που συνιστούν το επίπεδο εισόδου, δεν είναι ουσιαστικά νευρώνες, καθώς δεν εκτελούν κάποιον υπολογισμό (δεν έχουν ούτε βάρη εισόδου, ούτε συναρτήσεις ενεργοποίησης). Στη συνέχεια μπορούν να υπάρχουν, προαιρετικά, ένα ή περισσότερα ενδιάμεσα ή κρυφά επίπεδα (hidden layers). Τέλος ακολουθεί ένα επίπεδο εξόδου (output layer).

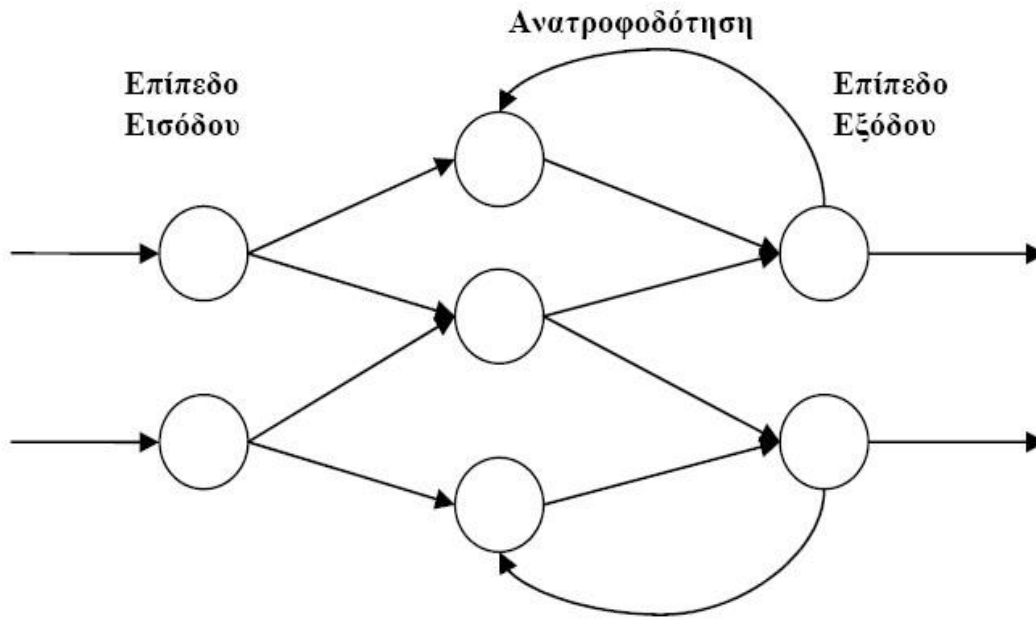
Οι νευρώνες των διάφορων στρωμάτων μπορεί να είναι πλήρως ή μερικώς συνδεδεμένοι. Πλήρως συνδεδεμένοι (fully connected) είναι εκείνοι οι οποίοι συνδέονται με όλους τους νευρώνες του επόμενου επιπέδου. Σε κάθε άλλη περίπτωση οι νευρώνες είναι μερικώς συνδεδεμένοι (partially connected). Όταν δεν υπάρχουν συνδέσεις μεταξύ

νευρώνων ενός επιπέδου και νευρώνων προηγούμενου επιπέδου (όταν δηλαδή η ροή πληροφορίας είναι μιας κατεύθυνσης) τα ΤΝΔ χαρακτηρίζονται ως δίκτυα με απλή (ή πρόσθια) τροφοδότηση (feedforward). Στην αντίθετη περίπτωση, καθώς και στην περίπτωση συνδέσεων μεταξύ νευρώνων του ίδιου επιπέδου, τα ΤΝΔ χαρακτηρίζονται ως δίκτυα με ανατροφοδότηση (feedback ή recurrent). Ο τύπος του δικτύου με ανατροφοδότηση διαφέρει από τον τύπο της απλούς τροφοδότησης στο ότι περιλαμβάνει ένα βρόχο ανάδρασης, όπου κάθε νευρώνας τροφοδοτεί το σήμα της εξόδου του στις εισόδους όλων των άλλων νευρώνων (Αργυράκης 2001).



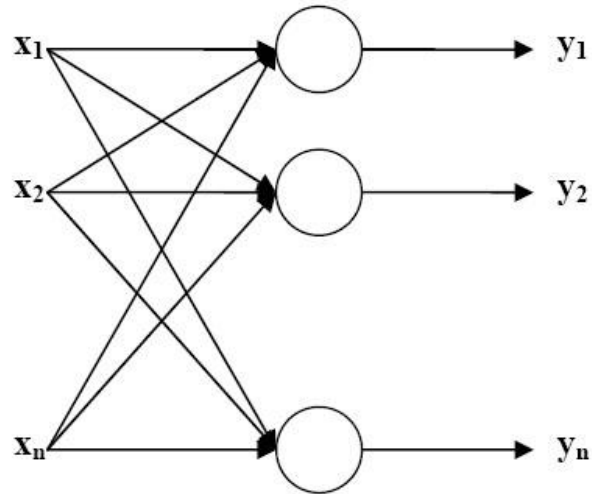
Σχήμα 3.6: Παραδείγματα πλήρως διασυνδεδεμένων ΤΝΔ απλής τροφοδότησης

Τα ΤΝΔ του Σχήματος 3.6 αποτελούν παραδείγματα πλήρως διασυνδεδεμένων ΤΝΔ απλής τροφοδότησης, ενώ στο Σχήμα 3.7 φαίνεται ένα παράδειγμα δικτύου με ανατροφοδότηση.

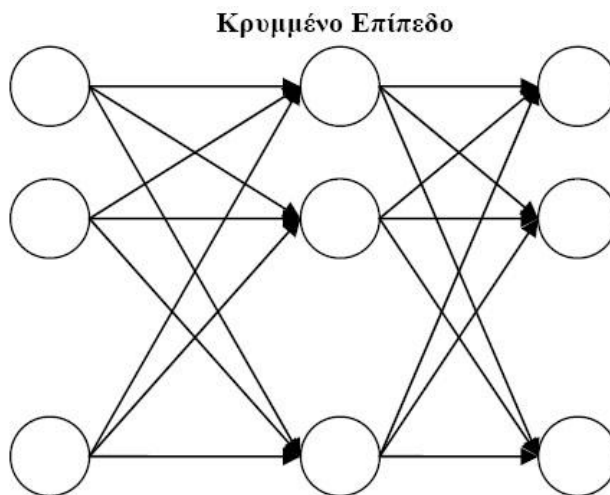


Σχήμα 3.7: Παραδείγματα ΤΝΔ με ανατροφοδότηση

Έχουν αναπτυχθεί συγκεκριμένα ΤΝΔ, πάνω στα πρότυπα ορισμένων βιολογικών λειτουργιών που είναι γνωστά ως paradigms στη διεθνή βιβλιογραφία, και αποτελούν ουσιαστικά τους διαφορετικούς τύπους ΤΝΔ. Τα ΤΝΔ στις περισσότερες περιπτώσεις είναι δομημένα από στρώματα νευρώνων. Κάποια από αυτά έχουν ένα επίπεδο (SLNN=Single-Layer Neural Networks) (Σχήμα 3.8) και κάποια έχουν περισσότερα (MLNN=Multi-Layer Neural Networks) (Σχήμα 3.9).



Σχήμα 3.8: ΤΝΔ με ένα επίπεδο



Σχήμα 3.9: ΤΝΔ με πολλαπλά επίπεδα

Τα πρώτα χρησιμοποιούν ως κανόνα μάθησης κυρίως τον κανόνα Hebb (Hebb rule) και τον κανόνα Δέλτα (Delta rule). Όσον αφορά στα MLNN οι Rumelhart, Hinton,

και Williams (1986) εισήγαγαν τον αλγόριθμο εκμάθησης με ανάστροφη διάδοση του σφάλματος (back-propagation training algorithm), που θα αναλυθεί παρακάτω. Επιπλέον τα MLNN χρησιμοποιούν σε ορισμένες περιπτώσεις την ανταγωνιστική μάθηση (competitive learning). Παρακάτω παρουσιάζονται οι πιο γνωστοί τύποι ΤΝΔ, και κάποιοι από τους αλγόριθμους εκπαίδευσής τους.

3.3.1 Δίκτυα με Απλή Τροφοδότηση

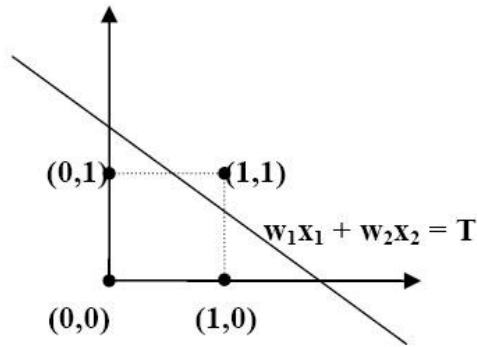
Τα ΤΝΔ με απλή τροφοδότηση (feedforward) συνιστούν την πιο απλή μορφή νευρωνικών δικτύων και αποτελούνται από ένα επίπεδο εισόδου, ένα επίπεδο εξόδου και προαιρετικά, ένα ή περισσότερα ενδιάμεσα, κρυφά επίπεδα. Δύο είναι τα θέματα τα οποία ανακύπτουν στην υλοποίηση των νευρωνικών δικτύων του τύπου αυτού. Το πρώτο αφορά στη μάθηση, δηλαδή στον τρόπο που εκπαιδεύεται το δίκτυο για να έχει την επιθυμητή συμπεριφορά. Στα δίκτυα απλής τροφοδότησης χρησιμοποιούνται μέθοδοι μάθησης με επίβλεψη, οι σημαντικότερες από τις οποίες εξετάζονται στη συνέχεια. Το δεύτερο θέμα αφορά στην τοπολογία του δικτύου, δηλαδή στο πόσα κρυφά επίπεδα θα έχει το δίκτυο και από πόσους νευρώνες θα αποτελείται το κάθε επίπεδο. Θα πρέπει να αναφερθεί ότι έχουν προταθεί αλγόριθμοι οι οποίοι μπορούν να δημιουργήσουν μια τοπολογία ΤΝΔ βάσει των εισόδων και των επιθυμητών εξόδων. Παρακάτω παρουσιάζεται η πιο απλή μορφή δικτύου απλής τροφοδότησης που ονομάζεται perceptron, καθώς και ο ομώνυμος αλγόριθμος εκπαίδευσης Perceptron και Multilayer Perceptron (MLP).

Το perceptron συνιστά την πιο απλουστευμένη τοπολογία δικτύου με απλή τροφοδότηση και αποτελεί ιστορικά μια πρώτη προσέγγιση τεχνητών νευρωνικών δικτύων που την εισήγαγε ο Rosenbalt (1958, 1962). Πρόκειται ουσιαστικά για ένα και

μοναδικό τεχνητό νευρώνα, ο οποίος χρησιμοποιεί ως συνάρτηση κατωφλίου τη βηματική συνάρτηση. Η μάθηση στο perceptron συνίσταται στην επιλογή κατάλληλων τιμών βαρών, έτσι ώστε, δεδομένου ενός διανύσματος εισόδου, να παραχθεί η επιθυμητή έξοδος. Πρόκειται δηλαδή για μια απλή μορφή μάθησης υπό επίβλεψη. Ο αλγόριθμος μεταβολής των βαρών έχει ως εξής:

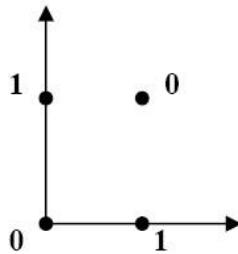
1. Εάν για τη συγκεκριμένη είσοδο παράγεται το επιθυμητό αποτέλεσμα, τότε δεν γίνεται καμία μεταβολή.
2. Εάν το αποτέλεσμα είναι 1, ενώ θα έπρεπε να είναι 0, τότε μειώνονται τα βάρη των ενεργών γραμμών (εκείνων που στο συγκεκριμένο πρότυπο έχουν είσοδο 1) κατά μια τιμή d , η οποία ονομάζεται ρυθμός μάθησης (learning rate).
3. Εάν το αποτέλεσμα είναι 0, ενώ θα έπρεπε να είναι 1, τότε αυξάνονται τα βάρη των ενεργών γραμμών κατά d .

Ένας perceptron με n γραμμές εισόδου μπορεί να θεωρηθεί ότι αναπαριστά ένα υπερεπίπεδο $n-1$ διαστάσεων, που διαχωρίζει τα διανύσματα εισόδου σε δύο ομάδες, τοποθετώντας από τη μια πλευρά όσα παράγουν έξοδο 1, και από την άλλη όσα παράγουν έξοδο 0. Προβλήματα των οποίων οι τιμές εισόδου-εξόδου υπόκεινται σε αυτόν τον κανόνα ονομάζονται γραμμικώς διαχωρίσιμα (linearly separable) και έχει αποδειχτεί ότι μπορούν να μοντελοποιηθούν με τη χρήση του perceptron. Επομένως μετά από πεπερασμένο χρόνο εκπαίδευσης και εφόσον η συνάρτηση που συνδέει την είσοδο με την έξοδο είναι γραμμική, επέρχεται η σύγκλιση του perceptron. Στο Σχήμα 3.10 απεικονίζεται γραφικά η έννοια του γραμμικώς διαχωρίσιμου για τη συνάρτηση AND η οποία μπορεί να μοντελοποιηθεί με perceptron. Η ευθεία T χωρίζει τα ζευγάρια εισόδου –εξόδου σε δύο περιοχές. Πάνω και δεξιά της T βρίσκονται αυτά που έχουν έξοδο 1 και κάτω αριστερά αυτά που έχουν έξοδο 0. Εκτός από το AND, το perceptron είναι σε θέση να μοντελοποιήσει και τους στοιχειώδεις λογικούς συνδέσμους OR και NOT.



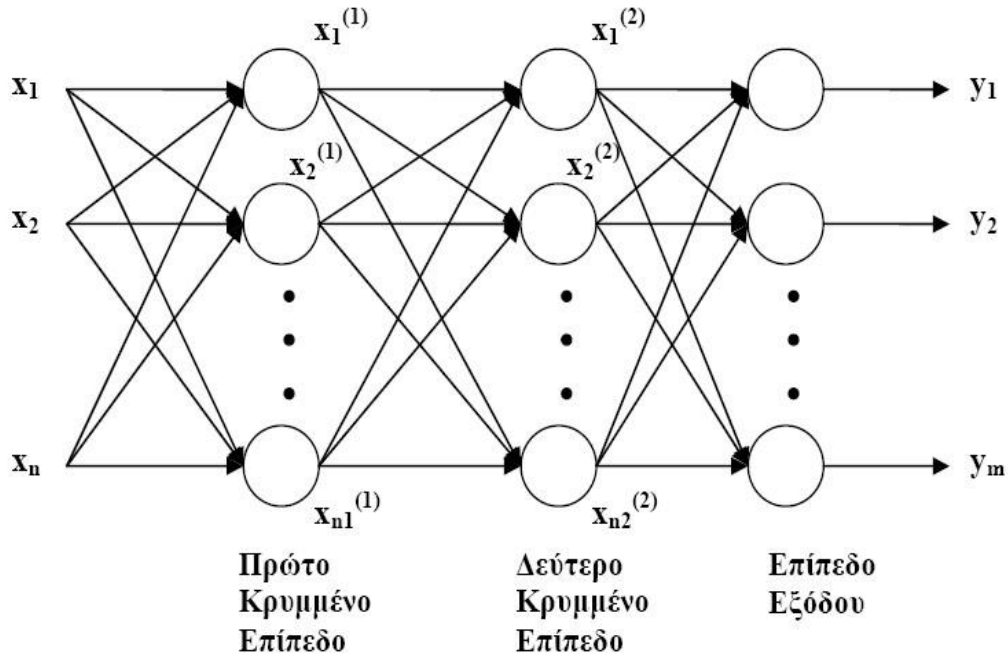
Σχήμα 3.10: Συνάρτηση AND γραμμικώς διαχωρίσιμη

Προφανώς υπάρχουν προβλήματα στα οποία τα ζευγάρια εισόδου-εξόδου δεν είναι γραμμικώς διαχωρίσιμα. Τέτοια προβλήματα δε μπορούν να μοντελοποιηθούν με perceptron και απαιτούν τη χρήση ΤΝΔ με ενδιάμεσα κρυφά επίπεδα. Κλασικό παράδειγμα τέτοιας συνάρτησης είναι η XOR, της οποίας ο μη γραμμικώς διαχωρίσιμος χαρακτήρας απεικονίζεται στο Σχήμα 3.11 (Αργυράκης 2001).



Σχήμα 3.11: Συνάρτηση XOR μη-γραμμικώς διαχωρίσιμη

Στην κατηγορία των νευρωνικών δικτύων πρόσθιας τροφοδότησης υπάγεται ο Multilayer Perceptron (MLP), ο οποίος δεν χρησιμοποιεί παρελθούσες τιμές των εξόδων του ή άλλου είδους εσωτερικές μεταβλητές για τον υπολογισμό της τρέχουσας εξόδου. Ο MLP αποτελείται από πολλαπλά επίπεδα απλοποιημένων σιγμοειδών κόμβων (επεξεργαστικών στοιχείων) ή νευρώνων, που αλληλεπιδρούν χρησιμοποιώντας σταθμισμένες συνδέσεις. Η βασική δομή του MLP παρουσιάζεται στο σχήμα 3.12.



Σχήμα 3.12: Πολυεπίπεδος Perceptron

Οι κύκλοι αναπαριστούν τους νευρώνες (βάρη, κατώφλι και συνάρτηση ενεργοποίησης) και οι γραμμές αναπαριστούν τις συνδέσεις μεταξύ των εισόδων και των νευρώνων, καθώς και μεταξύ των νευρώνων σε ένα επίπεδο και εκείνων στο επόμενο επίπεδο. Το συγκεκριμένο δίκτυο αποτελεί έναν τριών-επιπέδων perceptron, καθώς αποτελείται από τρία στάδια νευρωνικής επεξεργασίας μεταξύ των εισόδων και των εξόδων. Ο MLP έχει εισόδους τα x_i , $i = 1, 2, \dots, n$ και εξόδους τα y_j , $j = 1, 2, \dots, m$. Ο αριθμός των νευρώνων στο πρώτο κρυμμένο επίπεδο είναι $n1$. Στο δεύτερο κρυμμένο επίπεδο υπάρχουν $n2$ νευρώνες, και στο επίπεδο εξόδου υπάρχουν m νευρώνες. Οι νευρώνες στο πρώτο επίπεδο του MLP εκτελούν υπολογισμούς, των οποίων τα αποτελέσματα δίνονται από τον παρακάτω τύπο:

$$x_j^{(1)} = f_j^{(1)} \left(\left(\sum_{i=1}^n w_{ij}^{(1)} x_i \right) - \theta_j^{(1)} \right) \quad (3.2)$$

όπου $j = 1, 2, \dots, n_1$. Οι νευρώνες στο δεύτερο επίπεδο του MLP εκτελούν υπολογισμούς, το αποτέλεσμα των οποίων δίνεται από:

$$x_j^{(2)} = f_j^{(2)} \left(\left(\sum_{i=1}^{n_1} w_{ij}^{(2)} x_i^{(2)} \right) - \theta_j^{(2)} \right) \quad (3.3)$$

όπου $j = 1, 2, \dots, n_2$. Οι νευρώνες στο επίπεδο εξόδου του MLP εκτελούν υπολογισμούς, το αποτέλεσμα των οποίων δίνεται από:

$$y_j = f_j \left(\left(\sum_{i=1}^{n_2} w_{ij} x_i^{(2)} \right) - \theta_j \right) \quad (3.4)$$

όπου $j = 1, 2, \dots, m$.

Οι παράμετροι $w_{ij}^{(1)}$ ονομάζονται βάρη του πρώτου κρυμμένου επιπέδου. Τα $w_{ij}^{(2)}$ ονομάζονται βάρη του δεύτερου κρυμμένου επιπέδου. Τα w_{ij} ονομάζονται βάρη του επιπέδου εξόδου. Οι παράμετροι $\theta_j^{(1)}$ και $\theta_j^{(2)}$ ονομάζονται κατώφλια του πρώτου και δεύτερου κρυμμένου επιπέδου αντίστοιχα, ενώ τα θ_j ονομάζονται τα κατώφλια του επιπέδου εξόδου. Οι συναρτήσεις f_j (για το επίπεδο εξόδου), $f_j^{(2)}$ (για το δεύτερο κρυμμένο επίπεδο), και $f_j^{(1)}$ (για το πρώτο κρυμμένο επίπεδο) αναπαριστούν τις συναρτήσεις ενεργοποίησης (activation functions). Οι συναρτήσεις ενεργοποίησης μπορούν να διαφοροποιούνται μεταξύ των νευρώνων του MLP.

3.3.2 Κανόνας Δέλτα

Ο κανόνας Δέλτα, (Delta rule) αποτελεί γενίκευση του αλγορίθμου εκπαίδευσης του perceptron, με την έννοια ότι μπορεί να εφαρμοστεί και σε άλλα ΤΝΔ, τα οποία όμως επίσης δεν έχουν κρυφά επίπεδα. Παρακάτω περιγράφεται η λειτουργία του αλγορίθμου. Έστω ότι σε ένα ΤΝΔ η συνάρτηση ενεργοποίησης παίρνει πραγματικές τιμές, όπως είναι για παράδειγμα η σιγμοειδής συνάρτηση:

$$g(z) = \frac{1}{1 + e^z} \quad (3.5)$$

Τότε η έξοδος ενός νευρώνα i μπορεί να υπολογιστεί βάσει της σχέσης:

$$a_i = g \cdot \sum_{j=0}^n w_j a_j \quad (3.6)$$

όπου a_i είναι η έξοδος του νευρώνα i , w_j το βάρος της σύνδεσης j και a_j η είσοδος j από το νευρώνα του προηγούμενου επιπέδου. Η μεταβολή του βάρους w_j υπολογίζεται από τον τύπο:

$$w_j = w_{j-\text{old}} - d \cdot (a_i - a) a_j \quad (3.7)$$

όπου a_i είναι η έξοδος του νευρώνα, a είναι η επιθυμητή έξοδος, w_j το βάρος της σύνδεσης j , a_j η είσοδος j και d ο ρυθμός μάθησης. Το συνολικό σφάλμα, που αποτελεί ένα μέτρο της απόστασης από την επιθυμητή κατάσταση, υπολογίζεται από τη σχέση:

$$\text{Error} = \frac{1}{2} \sum (a_i - a)^2 \quad (3.8)$$

Αν και ο αλγόριθμος αυτός αποτελεί βελτίωση εκείνου που εφαρμόζεται στα perceptrons, δε μπορεί να εφαρμοστεί σε δίκτυα τα οποία έχουν κρυφά επίπεδα, καθώς για κάθε νευρώνα πρέπει να είναι γνωστή με ακρίβεια η έξοδός του, κάτι που δεν είναι δυνατό όταν υπάρχουν κρυμμένα επίπεδα.

3.3.3 Αλγόριθμος Εκμάθησης με βάση την Ανάστροφη Διάδοση του Σφάλματος (back propagation learning algorithm)

Η ανάστροφη διάδοση του λάθους (back propagation) αποτελεί την πιο γνωστή μέθοδο εκπαίδευσης νευρωνικών δικτύων πολλών επιπέδων. Την εισήγαγαν οι Werbos (1974), Rumelhart et al. (1986) και Parker (1987). Η βασική ιδέα είναι να καθοριστεί το «ποσοστό» του συνολικού σφάλματος που αντιστοιχεί στα βάρη του κάθε νευρώνα. Με αυτόν τον τρόπο γίνεται δυνατό να υπολογιστούν οι διορθώσεις στα βάρη του κάθε νευρώνα ξεχωριστά, το οποίο είναι αρκετά πολύπλοκο για τα κρυφά επίπεδα, καθώς η έξοδός τους επηρεάζει πολλούς νευρώνες ταυτόχρονα. Στην ανάστροφη διάδοση του λάθους υπολογίζεται αρχικά το σφάλμα για τους νευρώνες του επιπέδου εξόδου κατά τρόπο παρόμοιο με εκείνον του κανόνα Δέλτα. Το σφάλμα που υπολογίζεται χρησιμοποιείται για να υπολογιστούν τα σφάλματα στο τελευταίο κρυφό επίπεδο. Στη συνέχεια η διαδικασία επαναλαμβάνεται αναδρομικά προς το πρώτο επίπεδο. Με βάση δηλαδή τη διάδοση του σφάλματος προς τα πίσω (back propagation), γίνεται ένας υπολογισμός της συνεισφοράς κάθε βάρους των νευρώνων στο ολικό σφάλμα. Κατόπιν τα σφάλματα που υπολογίστηκαν για τους νευρώνες κάθε επιπέδου χρησιμοποιούνται για να μεταβάλουν τα βάρη του κάθε νευρώνα κατά τρόπο αντίστοιχο με εκείνον του κανόνα Δέλτα. Η διαδικασία επαναλαμβάνεται έως ότου το σφάλμα να πάρει τιμή στα

όρια ανοχής που έχει θέσει ο χρήστης. Έστω ότι έχουμε s παραδείγματα, κάθε ένα από τα οποία περιγράφεται από ένα διάνυσμα εισόδου $X_i=(x_{i1},x_{i2},\dots,x_{in})$ και ένα διάνυσμα εξόδου $D_i=(d_{i1},d_{i2},\dots,d_{im}), 1\leq i\leq s$.

Ο αλγόριθμος ανάστροφης διάδοσης του σφάλματος αποτελείται από δύο φάσεις:

1. Στη φάση πρόσθιας διάδοσης (forward propagation), το X_i τροφοδοτείται στο επίπεδο εισόδου (input layer) και η έξοδος $Y_i=(y_{i1},y_{i2},\dots,y_{im})$ δημιουργείται μέσα στα πλαίσια του χώρου W , μέσα στον οποίο ο αλγόριθμος εκπαίδευσης θα ψάξει, προκειμένου να καθορίσει τα βάρη που αρμόζουν καλύτερα στα δεδομένα παραδείγματα. Έπειτα, υπολογίζοντας το τετραγωνικό σφάλμα $(y_{ij} - d_{ij})^2, 1 \leq i \leq m$, για κάθε μονάδα εξόδου, η τιμή Y_i συγκρίνεται με την πραγματική (ή επιθυμητή) έξοδο D_i . Οι διαφορές που προκύπτουν αθροίζονται. Με αυτόν τον τρόπο προκύπτει η συνάρτηση σφάλματος E , που ορίζεται ως:

$$E = \sum_{i=1}^s \sum_{j=1}^m \frac{(y_{ij} - d_{ij})^2}{2} \quad (3.9)$$

Σκοπός είναι να ελαχιστοποιηθεί το E μεταβάλλοντας το W έτσι ώστε όλα τα διανύσματα εισόδου να αντιστοιχηθούν σωστά με τα αντίστοιχα διανύσματα εξόδου τους. Επομένως η διαδικασία μάθησης μπορεί να θεωρηθεί ως ένα πρόβλημα ελαχιστοποίησης με αντικειμενική συνάρτηση την E που ορίζεται στο διάστημα W .

2. Στη δεύτερη φάση (backward propagation) επιδιώκεται να εντοπισθεί η βέλτιστη λύση στο χώρο που ορίζονται τα βάρη, μέσω της μεθόδου φθίνουσας κλίσης (gradient descend method). Η κατεύθυνση και το μέγεθος αλλαγής Δw_{ij} του κάθε w_{ij} υπολογίζεται ως εξής:

$$\Delta w_{ij} = -\frac{\partial E}{\partial w_{ij}} \epsilon \quad (3.10)$$

όπου $0 < \varepsilon < 1$, είναι μια παράμετρος που ελέγχει το ρυθμό σύγκλισης του αλγορίθμου. Το συνολικό τετραγωνικό σφάλμα που υπολογίστηκε στην πρώτη φάση «διαδίδεται» προς τα πίσω, από επίπεδο σε επίπεδο, από τις μονάδες εξόδου προς τις μονάδες εισόδου, κατά τη δεύτερη φάση. Οι προσαρμογές των βαρών προσδιορίζονται από τον τρόπο που η διάδοση (propagation) εξελίσσεται σε κάθε επίπεδο. Αν I_i (input function), O_i (output function) και E , είναι συνεχείς και παραγωγίσιμες συναρτήσεις, τότε η τιμή της παραπάνω εξίσωσης σε κάθε επίπεδο μπορεί να υπολογιστεί με βάση τον κανόνα της αλυσίδας (chain rule):

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial Q_i} \frac{\partial Q_i}{\partial I_i} \frac{\partial I_i}{\partial w_{ij}} \quad (3.11)$$

Η όλη διαδικασία μπορεί να θεωρηθεί ως μια αναζήτηση του ολικού ελάχιστου της συνάρτησης σφάλματος, η οποία έχει ως παραμέτρους τις τιμές των βαρών. Η διόρθωση που μπορεί να γίνει κάθε φορά προσπαθεί να ελαχιστοποιήσει το σφάλμα επιλέγοντας να κάνει εκείνες τις αλλαγές που τείνουν να το μειώνουν τοπικά. Πρόκειται δηλαδή για μια αναζήτηση τύπου αναρρίχησης «λόφου». Υπάρχουν ωστόσο και περιπτώσεις, όπου ένα δίκτυο που εκπαιδεύεται με αυτόν τον τρόπο, δεν αποδίδει τα αναμενόμενα. Πρόκειται για τις περιπτώσεις κατά τις οποίες το δίκτυο πέφτει σε τοπικά ελάχιστα (local minima) και/ή παραλύει τελείως (network paralysis). Στην πρώτη περίπτωση πρόκειται για εγγενή αδυναμία της αναζήτησης αναρρίχησης λόφου να βρει το ολικό ελάχιστο, δηλαδή το διάνυσμα των βαρών για το οποίο ελαχιστοποιείται το σφάλμα. Στη δεύτερη περίπτωση, το νευρωνικό δίκτυο πέφτει σε στάσιμη κατάσταση γιατί ένα ή περισσότερα βάρη έχουν σταθερά υψηλές απόλυτες τιμές και δεν τροποποιούνται σημαντικά σε κάθε διόρθωση. Για την αποφυγή αυτών των ανεπιθύμητων καταστάσεων έχουν προταθεί διάφορες παραλλαγές του βασικού αλγόριθμου (Dorsey et al., 1994, Craven, 1997).

3.4 Στάδια Διαδικασίας Ανάπτυξης ενός ΤΝΔ

Η διαδικασία ανάπτυξης σχηματισμού ενός ΤΝΔ αποτελείται από 9 στάδια. Στο πρώτο στάδιο συλλέγονται τα δεδομένα που θα χρησιμοποιηθούν για την εκπαίδευση και τον έλεγχο του δικτύου. Στο δεύτερο στάδιο το δείγμα εκμάθησης πρέπει να αναγνωριστεί και ένα πλάνο πρέπει να δημιουργηθεί αναφορικά με τον έλεγχο της αποτελεσματικότητας του δικτύου. Επομένως κατά τα δύο πρώτα στάδια τα δεδομένα χωρίζονται σε δείγματα εκμάθησης (training test), που σκοπό έχουν την προσαρμογή των βαρών των συνδέσεων του ΤΝΔ, και σε δείγματα ελέγχου (testing set), που σκοπό έχουν την εξακρίβωση της εγκυρότητας του δικτύου. Γενικά όσο περισσότερα είναι τα δεδομένα, τόσο το καλύτερο, καθώς τα μεγάλα σύνολα δεδομένων, μπορεί από τη μια πλευρά να αυξάνουν τον χρόνο επεξεργασίας του δικτύου, αλλά από την άλλη βελτιώνουν την ακρίβεια της εκμάθησης και συχνά οδηγούν σε ταχύτερη σύγκλιση σε ένα καλό σύνολο σταθμίσεων των βαρών. Στα στάδια 3 και 4, επιλέγεται η αρχιτεκτονική του δικτύου και ο αλγόριθμος εκπαίδευσης του δικτύου. Ο ακριβής αριθμός νευρώνων και επιπέδων (layers) θεωρείται σημαντικός παράμετρος αυτών των σταδίων. Μέρος της διαδικασίας του πέμπτου σταδίου αποτελεί να τεθούν τα αρχικά βάρη συνδέσεων των νευρώνων του δικτύου και να καθοριστούν οι παράμετροι για το επιθυμητό επίπεδο αποτελεσματικότητας της διαδικασίας μάθησης. Το έκτο στάδιο μετατρέπει τα δεδομένα της εφαρμογής στην απαιτούμενη από το νευρωνικό δίκτυο κατάλληλη μορφή. Αυτό μπορεί να σημαίνει ότι απαιτείται λογισμικό για την προεπεξεργασία των δεδομένων. Σε αυτό το στάδιο σχεδιάζονται τεχνικές και διαδικασίες αποθήκευσης και χειρισμού δεδομένων για την αποτελεσματική και ευχερή επανεκπαίδευση του δικτύου, εφόσον βέβαια αυτή απαιτείται. Στα στάδια 7 και 8 διεξάγονται διαδοχικά η εκπαίδευση και ο έλεγχος του δικτύου με την εισαγωγή των δεδομένων εισόδου και των επιθυμητών ή γνωστών δεδομένων εξόδου σε αυτό. Το δίκτυο υπολογίζει τα αποτελέσματα και προσαρμόζει τα βάρη μέχρι το σφάλμα απόκλισης των αποτελεσμάτων, που προκύπτει ως προς τα δεδομένα εξόδου των δεδομένων περιπτώσεων που χρησιμοποιήθηκαν, να

πάρει τιμή στα όρια ανοχής που έχει θέσει ο χρήστης. Οι επιθυμητές εξόδοι του ΤΝΔ και οι σχέσεις τους ως προς τις εισόδους του εξάγονται από ιστορικά στοιχεία (που αποτελούν μέρος των δεδομένων που συλλέξαμε στο πρώτο στάδιο). Το δίκτυο είναι πλέον σε θέση να αναπαράγει τις επιθυμητές εξόδους με βάση εισόδους παρόμοιες με εκείνες του δείγματος εκμάθησης. Το δίκτυο τότε θεωρείται έτοιμο για χρήση.

3.5 Μάθηση και Ανάκληση

Τα ΤΝΔ χρησιμοποιούν δύο βασικές λειτουργίες: τη μάθηση και την ανάκληση. Ως μάθηση (learning) αποκαλείται η διαδικασία της τροποποίησης της τιμής των βαρών του δικτύου, ώστε, δοθέντος ενός συγκεκριμένου διανύσματος εισόδου, να παραχθεί συγκεκριμένο διάνυσμα εξόδου. Η διαδικασία αυτή ονομάζεται επίσης και εκπαίδευση (training) του ΤΝΔ. Ως ανάκληση (recall) αποκαλείται η διαδικασία του υπολογισμού ενός διανύσματος εξόδου για συγκεκριμένο διάνυσμα εισόδου και τιμές βαρών. Ο γενικός τρόπος με τον οποίο γίνεται η τροποποίηση των βαρών ενός ΤΝΔ κατά την εκπαίδευσή του, επιτρέπει τη διάκριση διαφορετικών ειδών μάθησης στα ΤΝΔ, που είναι τα εξής όπως αναλύονται στα κεφάλαια που ακολουθούν : α) Μάθηση υπό επίβλεψη (supervised learning), β) Μάθηση χωρίς επίβλεψη (unsupervised learning), γ) Ανταγωνιστική μάθηση (competitive learning), δ) Ενισχυτική μάθηση (reinforced learning), ε) Βαθμολογημένη μάθηση (graded learning), στ) Τυχαία μάθηση (random learning)

3.5.1 Μάθηση υπό επίβλεψη (supervised learning)

Στη μάθηση υπό επίβλεψη (supervised learning) το δίκτυο τροφοδοτείται με ζεύγη διανυσμάτων εισόδου – επιθυμητής εξόδου. Το ΤΝΔ, παράγει μία έξοδο με βάση την τρέχουσα κατάσταση βαρών, η οποία αρχικά διαφέρει από την επιθυμητή έξοδο. Αυτή η διαφορά ονομάζεται σφάλμα (error) και βάσει αυτής, καθώς και ενός αλγορίθμου εκπαίδευσης, γίνεται συνήθως η αναπροσαρμογή των βαρών. Στην πράξη στις περισσότερες εφαρμογές ΤΝΔ χρησιμοποιείται μάθηση υπό επίβλεψη, για την οποία υπάρχουν αρκετοί αλγόριθμοι. Στον αλγόριθμο που βασίζεται στον κανόνα Δέλτα (Delta rule learning), η διαφορά μεταξύ πραγματικής και επιθυμητής εξόδου ελαχιστοποιείται μέσω μιας διαδικασίας ελαχίστων τετραγώνων. Στον αλγόριθμο ανάστροφης διάδοσης του σφάλματος (back propagation) η μεταβολή των βαρών βασίζεται στον υπολογισμό της συνεισφοράς του κάθε βάρους στο συνολικό σφάλμα. Σε αυτήν την περίπτωση το σφάλμα μεταξύ επιθυμητής και πραγματικής εξόδου ανατροφοδοτείται στα βάρη των προηγούμενων επιπέδων του ΤΝΔ, χρησιμοποιώντας τη διαφορική της τιμή ή τις προσεγγίσεις αυτής.

3.5.2 Μάθηση χωρίς επίβλεψη (unsupervised learning)

Στη μάθηση χωρίς επίβλεψη, (unsupervised learning) η απόκριση του δικτύου βασίζεται στην ικανότητά του να αυτό-οργανώνεται με βάση τα διανύσματα εισόδου (π.χ. δίκτυα με βάση τον κανόνα του Kohonen, υπό τον οποίο αναπτύσσονται διανύσματα εξόδου με βάση εκείνα της εισόδου, χωρίς να απαιτείται η ύπαρξη προτύπων εξόδου). Αυτή η εσωτερική οργάνωση γίνεται έτσι ώστε σε συγκεκριμένο

σύνολο εισόδων να αντιδρά ισχυρά ένας συγκεκριμένος νευρώνας. Τέτοια σύνολα εισόδων αντιστοιχούν σε έννοιες και χαρακτηριστικά του πραγματικού κόσμου, τα οποία το ΤΝΔ καλείται να μάθει. Καθώς το δίκτυο δεν γνωρίζει τις πραγματικές τιμές εξόδου, εξετάζει τα πρότυπα εισόδου με βάση μετρούμενες ποσότητες ομοιότητας ή βαθμούς ποιότητας των δεδομένων εισόδου. Με αυτόν τον τρόπο διαμερίζεται το σύνολο εισόδων σε έναν αριθμό από ομάδες οι οποίες αυτό-προσαρμόζονται. Η συνάρτηση σφάλματος των νευρωνικών δικτύων που εκπαιδεύονται με μάθηση χωρίς επίβλεψη βασίζεται στις μετρήσεις ομοιότητας ή ποιότητας για όλες τις εισόδους. Το αντικείμενο της εκπαίδευσης είναι η ελαχιστοποίηση της συνάρτησης αυτής, ώστε να δημιουργηθούν ομάδες προτύπων με παρόμοιες ιδιότητες. Οι ομάδες αυτές μπορούν να θεωρηθούν ως κατηγορίες εξόδων προερχόμενες από μια ταξινόμηση προτύπων. Για το σχηματισμό των ομάδων, πάντως, απαιτούνται ορισμένες κατευθυντήριες οδηγίες σχετικά με το ποια χαρακτηριστικά θα χρησιμοποιηθούν στην ομαδοποίηση αυτή, διαφορετικά η ομαδοποίηση μπορεί να μην είναι επιτυχής.

3.5.3 Ανταγωνιστική μάθηση (competitive learning)

Στην ανταγωνιστική μάθηση (competitive learning), που αποτελεί ιδιαίτερη μορφή μάθησης υπό επίβλεψη, οι τεχνητοί νευρώνες συναγωνίζονται, κατά κάποιο τρόπο, μεταξύ τους και μόνο αυτός με τη μεγαλύτερη απόκριση σε δοθείσα είσοδο τροποποιεί τα βάρη του. Με αυτόν τον τρόπο επιτελείται ένα είδος ανταγωνισμού στους νευρώνες εξόδου ενός δικτύου, που έχει ως αποτέλεσμα σε κάθε χρονική στιγμή να είναι ενεργός μόνο ένας νευρώνας. Ένας τέτοιος νευρώνας εξόδου που κερδίζει τον ανταγωνισμό ονομάζεται «ο νικητής τα παίρνει όλα» (winner-takes-all). Μια ειδικής μορφής κατηγορία νευρωνικών δικτύων προβολής δεδομένων, που βασίζεται στην ανταγωνιστική μάθηση, είναι οι αυτό-οργανωμένοι χάρτες χαρακτηριστικών (self-organizing feature maps). Η

κύρια εργασία των δικτύων αυτών είναι η μεταφορά των προτύπων από τη διάσταση που ορίζουν τα χαρακτηριστικά τους σε μια χαμηλότερη διάσταση (συνήθως επίπεδο) μέσω μίας τοπολογικής μορφής απεικόνισης. Για αυτόν το σκοπό χρησιμοποιείται ένας τοπογραφικός χάρτης των προτύπων εισόδου, στον οποίο οι συντεταγμένες των νευρώνων πάνω στο πλέγμα που σχηματίζεται αντιστοιχούν σε χαρακτηριστικά των προτύπων. Στην γενική του μορφή, ο αλγόριθμος μάθησης που χρησιμοποιούν οι αυτό-οργανωμένοι χάρτες χαρακτηριστικών βασίζεται στους αλγόριθμους εκπαίδευσης του πλησιέστερου γείτονα (k-nearest neighbor) και στη γειτνίαση των κόμβων του πλέγματος που ορίζει ο χάρτης. Έτσι κατά την τροφοδότηση ενός προτύπου x στο δίκτυο, υπολογίζεται ο κοντινότερός του κόμβος j . Θεωρώντας ότι κάθε κόμβος j στο πλέγμα χαρακτηρίζεται από ένα διάνυσμα βάρους w_j , η παρακάτω ενημέρωση του δικτύου λαμβάνει χώρα.

$$w_i \leftarrow w_i + h_{ij}(t) \cdot x - w_i \quad (3.12)$$

για όλους της γείτονες i του κόμβου j . Η συνάρτηση h_{ij} μπορεί είτε να είναι σταθερή (ρυθμός μάθησης) ή να εξαρτάται από τον βαθμό γειτνίασης του κόμβου i ως προς το j .

3.5.4 Ενισχυτική μάθηση (reinforced learning)

Η ενισχυτική μάθηση (reinforced learning) αποτελεί μια διαφορετική προσέγγιση στη διεργασία μάθησης σε σχέση με τη μάθηση υπό επίβλεψη και τη μάθηση χωρίς επίβλεψη. Το κυριότερο χαρακτηριστικό της ενισχυτικής μάθησης βρίσκεται στην ύπαρξη ενός ενισχυτικού σήματος (reinforcement signal) με βάση το οποίο αποτιμάται η αποτελεσματικότητα του συστήματος. Το ενισχυτικό σήμα μπορεί να είναι συνεχές ή

διακριτό στην περιοχή τιμών $[0,1]$ και απεικονίζει την απόκριση του περιβάλλοντος στην εκτέλεση μιας ενέργειας, με την μορφή ενός βαθμού επιτυχίας (reward) ή αποτυχίας (penalty). Κατά τη διάρκεια της ενισχυτικής μάθησης σε ένα νευρωνικό δίκτυο, οι παράμετροι του δικτύου (βάρη συνάψεων) προσαρμόζονται με τέτοιο τρόπο, ώστε να επιτρέπουν την κατάλληλη επιλογή ενέργειας για κάθε σήμα εισόδου, που εκτιμάται από το ενισχυτικό σήμα. Ένα σήμα επιτυχίας έχει με αυτόν τον τρόπο ως αποτέλεσμα την ακόμα μεγαλύτερη ενίσχυση της ενέργειας που επιλέχτηκε, ενώ ένα σήμα αποτυχίας θα αποτρέψει την επανάληψη της επιλογής ενέργειας κατά την επόμενη εμφάνιση του αντίστοιχου σήματος εισόδου.

3.5.5 Βαθμολογημένη μάθηση (graded learning)

Στη βαθμολογημένη μάθηση (graded learning) η έξοδος χαρακτηρίζεται ως «καλή» ή «κακή» με βάση μια αριθμητική κλίμακα και τα βάρη αναπροσαρμόζονται με βάση αυτόν τον χαρακτηρισμό.

3.5.6 Τυχαία Μάθηση (random learning)

Στην τυχαία μάθηση (random learning), οι μεταβολές στα βάρη εισάγονται τυχαία και ανάλογα με το αν η έξοδος βελτιώνεται ή όχι με βάση κάποια προκαθορισμένα από το χρήστη κριτήρια, οι μεταβολές αυτές υιοθετούνται ή απορρίπτονται.

3.6 Βασικά Χαρακτηριστικά Τεχνητών Νευρωνικών Δικτύων

Υπάρχουν τέσσερα χαρακτηριστικά που είναι άρρηκτα συνδεδεμένα με τα ΤΝΔ. Αυτά είναι:

1. Η ικανότητά τους να μαθαίνουν μέσω παραδειγμάτων (learn by example)
2. Η δυνατότητα θεώρησής τους ως κατανεμημένη μνήμη (distributed memory) και ως μνήμη συσχέτισης (associative memory)
3. Η μεγάλη τους ανοχή σε σφάλματα (fault tolerant)
4. Η ικανότητά τους για αναγνώριση προτύπων (pattern recognition)

Αν και τα ΤΝΔ δεν είναι τα μόνα συστήματα με ικανότητα μάθησης μέσω παραδειγμάτων, εν τούτοις διακρίνονται για την ικανότητά τους να οργανώνουν την πληροφορία των δεδομένων εισόδου σε χρήσιμες μορφές. Αυτές οι μορφές αποτελούν στην ουσία ένα μοντέλο που αναπαριστά τη σχέση που ισχύει μεταξύ των δεδομένων εισόδου και εξόδου. Το γεγονός αυτό τα κατέστησε χρήσιμα στην εξόρυξη σε δεδομένα (data mining) μέσα στα πλαίσια εύρεσης γνώσης σε βάσεις δεδομένων (knowledge discovery in databases). Ο χαρακτηρισμός των ΤΝΔ ως κατανεμημένες μνήμες, πηγάζει από το ότι η πληροφορία που κωδικοποιούν είναι κατανεμημένη σε όλα τα βάρη της συνδεσμολογίας τους. Για τον ίδιο λόγο τα ΤΝΔ χαρακτηρίζονται και ως μνήμες συσχέτισης. Μια μνήμη συσχέτισης αποθηκεύει πληροφορία συσχετίζοντας αποθηκευμένα δεδομένα μεταξύ τους. Η ανάκληση της πληροφορίας γίνεται με βάση το περιεχόμενο και όχι τη διεύθυνση, ότι δηλαδή συμβαίνει με τον ανθρώπινο εγκέφαλο. Στην κατανεμημένη αναπαράσταση της γνώσης ανάμεσα στους συνδέσμους οφείλεται η ικανότητα αναγνώρισης προτύπων των ΤΝΔ. Τα ΤΝΔ έχουν μεγάλη ανοχή σε δομικά σφάλματα. Αυτό σημαίνει ότι η κακή λειτουργία ή καταστροφή ενός νευρώνα ή κάποιων συνδέσεων δεν είναι ικανή να διαταράξει σημαντικά τη λειτουργία τους καθώς, όπως αναφέρθηκε, η πληροφορία που εσωκλείουν δεν είναι εντοπισμένη σε συγκεκριμένο

σημείο, αλλά διάχυτη σε όλο το δίκτυο. Γενικά το μέγεθος του σφάλματος λόγω «δομικών αστοχιών» είναι ανάλογο του ποσοστού των κατεστραμμένων συνδέσεων. Το χαρακτηριστικό αυτό καθιστά τα ΤΝΔ ιδανικά για χρήση σε αυτοματισμούς που θα λειτουργήσουν σε αντίξοες συνθήκες (π.χ. διάστημα, χώρους με ραδιενέργεια κλπ). Τα ΤΝΔ διακρίνονται κυρίως για τη δυνατότητα παράλληλης επεξεργασίας και για την ικανότητά τους να αναπαριστούν μη γραμμικές συμπεριφορές. Από την άλλη πλευρά όμως, τα αποτελέσματα που προκύπτουν δεν αιτιολογούνται, με αποτέλεσμα η διαδικασία που ακολουθούν να έχει χαρακτηριστεί ως ένα «μαύρο κουτί» (black box), όπου το εξαγόμενο αποτέλεσμα δεν γίνεται γνωστό πώς προέκυψε. Ένα άλλο μειονέκτημα των ΤΝΔ είναι ότι κατά τη φάση της εκμάθησής τους παρατηρείται αυξημένος υπολογιστικός φόρτος.

4. ΧΡΗΣΗ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ ΣΤΗΝ ΠΑΡΟΥΣΑ ΕΡΓΑΣΙΑ

4.1 Εισαγωγή

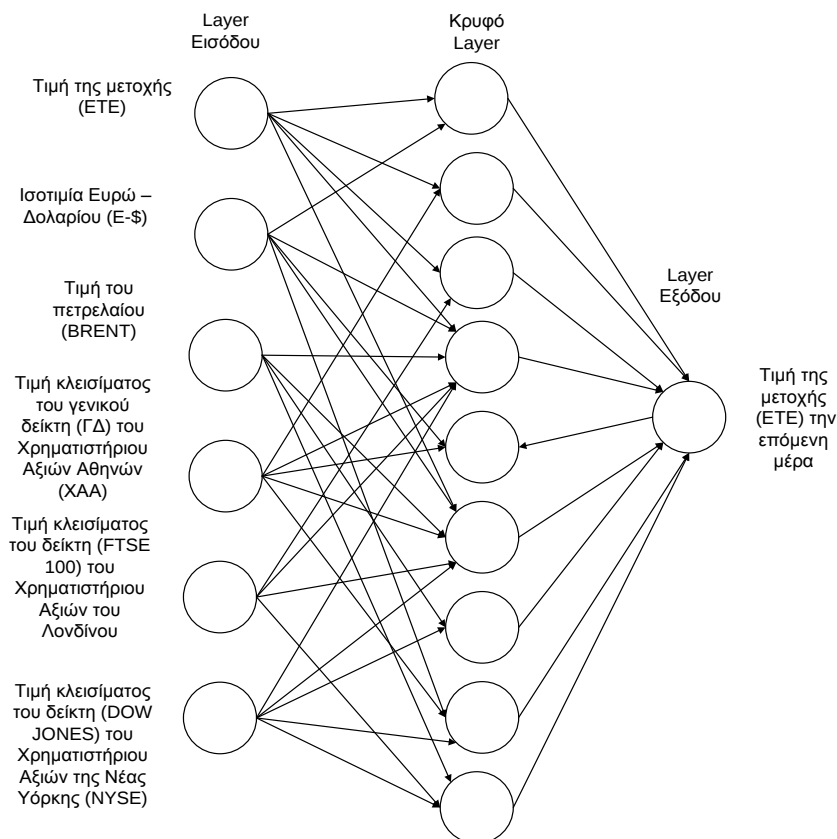
Στην παρούσα εργασία γίνεται η χρήση Τεχνητών Νευρωνικών Δικτύων για ένα πρόβλημα σχετικό με το Χρηματιστήριο άξιων Αθηνών και συγκεκριμένα για την πρόβλεψη της τιμής της μετοχής της Εθνικής Τράπεζας. Στις επόμενες παραγράφους περιγράφονται η αρχιτεκτονική του ΤΝΔ που χρησιμοποιείται στην παρούσα εργασία καθώς και οι διάφορες τεχνικές για τη συνάρτηση εκμάθησης, ενώ τα αντίστοιχα αποτελέσματα παρουσιάζονται στο 5^ο κεφάλαιο.

4.2 Επιλογή παραμέτρων του δικτύου

Για τον υπολογισμό της τιμής της μετοχής της Εθνικής Τράπεζας, μετά από δοκιμές και με βάση όσα αναφέρθηκαν προηγουμένως, απαιτούνται έξι παράμετροι. Συγκεκριμένα, απαιτείται η τιμή της μετοχής (ΕΤΕ), η ισοτιμία Ευρώ – Δολαρίου (€-\$), η τιμή του πετρελαίου (BRENT), η τιμή κλεισίματος του γενικού δείκτη (ΓΔ) του Χρηματιστηρίου Αξιών Αθηνών (ΧΑΑ), η τιμή κλεισίματος του δείκτη (FTSE 100) του Χρηματιστηρίου Αξιών του Λονδίνου, καθώς και η τιμή κλεισίματος του δείκτη (DOW JONES) του Χρηματιστηρίου Αξιών της Υόρκης (NYSE), που αποτελούν και τις εισόδους του ΤΝΔ.

Η έξοδος του ΤΝΔ θα αποτελείται μόνο από μία παράμετρο, την τιμή της Εθνικής Τράπεζας την επόμενη συνεδρίαση του ΧΑΑ.

Για την εκμάθηση του δικτύου επιλέχθηκε ο αλγόριθμος Ανάστροφης Διάδοσης του Σφάλματος (Back Propagation), ο οποίος περιγράφεται στην παράγραφο 3.3.3 καθώς αποτελεί μια πολύ επιτυχημένη μέθοδο εκπαίδευσης ΤΝΔ με πληθώρα εφαρμογών σε διαφορετικού είδους προβλήματα και πολύ καλά αποτελέσματα στην πλειοψηφία των περιπτώσεων. Όπως περιγράφηκε παραπάνω στο επίπεδο εισόδου χρησιμοποιούνται 6 νευρώνες για τις 6 γνωστές παραμέτρους εισόδου, ενώ στο επίπεδο εξόδου χρησιμοποιείται ένας νευρώνας για τη ζητούμενη έξοδο. Μετά από δοκιμές επιλέχθηκε η χρήση ενός κρυφού επιπέδου με 9 νευρώνες που δίνει και το μικρότερο σφάλμα σε σχέση με χρήση λιγότερων ή περισσότερων νευρώνων στο κρυφό επίπεδο ή και τη χρήση παραπάνω κρυφών επιπέδων. Το γενικό διάγραμμα του ΤΝΔ της παρούσας εργασίας δίνεται στο διάγραμμα που ακολουθεί όπου κάθε είσοδος αντιστοιχίζεται σε όλους τους νευρώνες του κρυφού επιπέδου και κάθε νευρώνας καταλήγει στην έξοδο.



Σχήμα 4.1 Γενικό Διάγραμμα του ΤΝΔ που χρησιμοποιείται

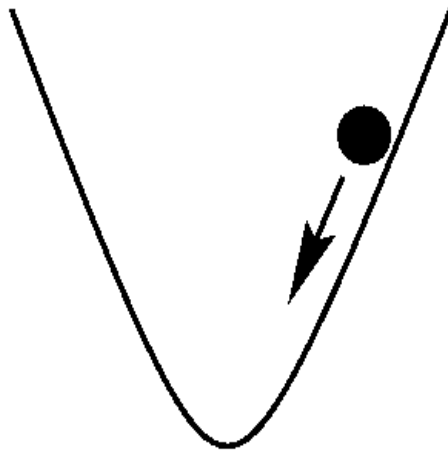
Ο αλγόριθμος που εκτελεί το συγκεκριμένο ΤΝΔ υλοποιήθηκε στο πρόγραμμα προσομοίωσης MATLAB R2007b με χρήση του ενσωματωμένου toolbox για Νευρωνικά Δίκτυα και μέσω των συναρτήσεων newff (για τη δημιουργία του ΤΝΔ) και train (για την εκμάθησή του).

4.3 Προσομοίωση νευρωνικού δικτύου με χρήση διαφόρων συναρτήσεων εκμάθησης

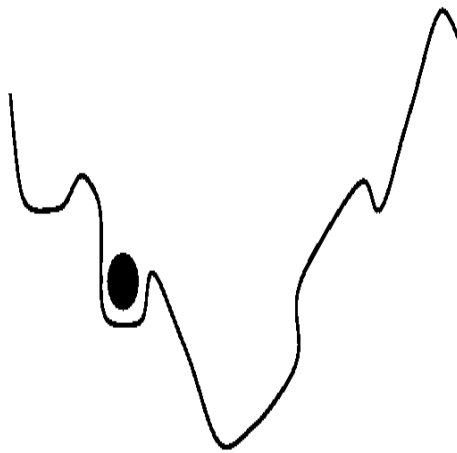
Η προσομοίωση του ΤΝΔ έγινε για την αρχιτεκτονική που περιγράφηκε στην παράγραφο 4.2 και για διάφορους αλγορίθμους εκμάθησης. Παρακάτω περιγράφονται οι αλγόριθμοι που χρησιμοποιήθηκαν, ενώ παρατίθενται τα αποτελέσματα των αντιστοίχων προσομοιώσεων τους σε σχηματική μορφή στο κεφάλαιο 5.

4.3.1 Ο αλγόριθμος gradient descent

Ο αλγόριθμος gradient descent ξεκινάει από ένα σημείο της συνάρτησης σφάλματος των βαρών και προσπαθεί να εντοπίσει το ολικό ελάχιστο της συνάρτησης. Στην απλοποιημένη περίπτωση του Σχήματος 4.2 η διαδικασία αυτή φαίνεται αρκετά απλή. Ωστόσο, για συνήθη προβλήματα η διακύμανση του σφάλματος είναι αρκετά πιο σύνθετη και μπορεί να ομοιάζει με την κατάσταση που απεικονίζεται στο Σχήμα 4.3. Σε τέτοιες περιπτώσεις μπορεί να υπάρχουν πολλαπλά τοπικά ελάχιστα. Για να υπάρξει πρόοδος σε ένα τέτοιο σύστημα πρέπει να γίνει άνοδος σε υψηλότερο σημείο πριν την κάθοδο στο ολικό ελάχιστο.



Σχήμα 4.2: Απλή διακύμανση σφάλματος χωρίς τοπικά ελάχιστα



Σχήμα 4.3: Σύνθετη διακύμανση σφάλματος με τοπικά ελάχιστα

Ο αλγόριθμος gradient descent υλοποιήθηκε με χρήση της συνάρτησης `traingd` του Matlab. Ο αλγόριθμος gradient descent μεταβάλλει τα βάρη του ΤΝΔ σύμφωνα με την ακόλουθη σχέση:

$$dw = \mu \cdot dE / dw \quad (4.3)$$

Όπου:

μ είναι ο ρυθμός εκμάθησης και

dE η μεταβολή της συνάρτησης σφάλματος.

Η εκμάθηση τερματίζεται όταν ο αριθμός των επαναλήψεων (epochs) συμπληρωθεί ή όταν το σφάλμα έχει μειωθεί στην επιθυμητή τιμή.

4.3.2 Προσαρμογή του ρυθμού εκμάθησης

Μια γνωστή μέθοδος για την προσαρμογή του ρυθμού εκμάθησης μ είναι ο αλγόριθμος bold driver. Η λειτουργία του είναι αρκετά απλή. Μετά από κάθε εποχή (epoch) γίνεται σύγκριση του σφάλματος με την προηγούμενη εποχή και αν το σφάλμα έχει μειωθεί, τότε το μ αυξάνεται κατά ένα μικρό ποσοστό (περίπου 1%-5%). Αν το σφάλμα έχει αυξηθεί περισσότερο από ένα μικρό κατώφλι (π.χ. 10^{-10}) τότε αναιρείται η τελευταία αλλαγή βαρών και το μ αλλάζει αρκετά (π.χ. 50%).

Μια άλλη μέθοδος προσέγγισης του ελαχίστου είναι η σταδιακή μείωση του ρυθμού εκμάθησης. Ένας απλός τρόπος για να γίνει αυτό είναι η μέθοδος search-then-converge όπου:

$$\mu_t = \mu_0 / 1 + t/T \quad (4.4)$$

Η ονομασία αυτής της μεθόδου προέρχεται από το γεγονός ότι η τιμή του μ παραμένει σχεδόν σταθερή για τα πρώτα T δείγματα (training patterns) επιτρέποντας στο δίκτυο να βρει μια γενικότερη θέση του ελαχίστου πριν σταθεροποιήσει την αναζήτηση με πολύ μικρό βήμα, το οποίο εγγυάται σύγκλιση στο ελάχιστο. Η

παράμετρος χρόνου T είναι μια παράμετρος που καθορίζεται μέσω της δοκιμής και του σφάλματος. Στην παρούσα προσομοίωση χρησιμοποιείται ο πρώτος αλγόριθμος προσαρμογής του ρυθμού εκμάθησης μέσω της συνάρτησης `traingda`. Η εκμάθηση τερματίζεται όταν ο αριθμός των `epochs` συμπληρωθεί ή όταν το σφάλμα έχει μειωθεί στην επιθυμητή τιμή.

4.3.3 Χρήση όρου Momentum

Μια τεχνική βελτίωσης της ικανότητας του ΤΝΔ να μην παγιδεύεται σε τοπικά ελάχιστα είναι η χρήση ενός όρου «momentum». Αυτή είναι πιθανώς η πιο χρησιμοποιούμενη προέκταση του αλγορίθμου `backpropagation` καθώς σπανίως δε χρησιμοποιείται. Με ένα momentum m , η μεταβολή των βαρών σε ένα χρόνο t ορίζεται ως:

$$\Delta w_{ij}(t) = \mu_i \delta_i y_i + m \Delta w_{ij}(t-1) \quad (4.5)$$

όπου $0 < m < 1$ είναι μια νέα παράμετρος που ορίζεται μέσω της δοκιμής και του σφάλματος. Το momentum απλά αθροίζει ένα τμήμα της προηγούμενης μεταβολής στην παρούσα. Αν ο gradient επιμένει προς μια κατεύθυνση αυτό μπορεί να αυξήσει το μέγεθος του βήματος προς το ελάχιστο. Είναι συχνά απαραίτητη η μείωση του ολικού ρυθμού εκμάθησης όταν χρησιμοποιείται μεγάλη τιμή του m (κοντά στη μονάδα). Επίσης, αν ο gradient επιμένει σε αλλαγή κατευθύνσεων το momentum μπορεί να μειώσει τη διαφοροποίηση. Για την προσομοίωση του ΤΝΔ της παρούσας εργασίας με τον αλγόριθμο εκμάθησης `gradient-descent` με προσαρμογή του ρυθμού εκμάθησης και χρήση momentum χρησιμοποιήθηκε η συνάρτηση `traingdx`.

4.3.4 Αλγόριθμος Levenberg-Marquardt

Στον αλγόριθμο Levenberg-Marquardt (LM) η συνάρτηση απόδοσης (performance index) $F(w)$ που πρέπει να ελαχιστοποιηθεί ορίζεται ως το άθροισμα των μέσων τετραγωνικών σφαλμάτων μεταξύ των γνωστών εξόδων (target outputs) και των εξόδων του ΤΝΔ:

$$F(w) = e^T e \quad (4.6)$$

όπου το $w = [w_1, w_2, \dots, w_N]$ αποτελείται από όλα τα βάρη του ΤΝΔ και e είναι το διάνυσμα που περιέχει το σφάλμα για όλα τα δείγματα εκμάθησης. Όταν γίνεται εκμάθηση με τη μέθοδο LM η μεταβολή των βαρών Δw υπολογίζεται από τη σχέση:

$$\Delta w = [J^T J + \mu I]^{-1} J^T e \quad (4.7)$$

όπου J είναι ο πίνακας Jacobian, μ είναι ο ρυθμός εκμάθησης (learning rate) ο οποίος προσαρμόζεται με τη χρήση του β ανάλογα με το αποτέλεσμα. Έτσι ο μ πολλαπλασιάζεται με το ρυθμό εξασθένησης (decay rate) β ($0 < \beta < 1$) όταν το $F(w)$ μειώνεται ενώ ο μ διαιρείται με το β όταν το $F(w)$ αυξάνεται στο επόμενο βήμα. Η διαδικασία εκμάθησης LM μπορεί να περιγραφεί με τα παρακάτω βήματα:

1. Αρχικοποίηση των βαρών και της παραμέτρου μ ($\mu = 0.01$ είναι μια αποδεκτή τιμή).
2. Υπολογισμός του $F(w)$ για όλες τις εισόδους.
3. Υπολογισμός των μεταβολών Δw
4. Υπολογισμός ξανά του αθροίσματος των μέσων σφαλμάτων $F(w)$ χρησιμοποιώντας το άθροισμα $w + \Delta w$ ως «δοκιμαστικό» w , και εκτέλεση του παρακάτω ελέγχου:

Αν το νέο $F(w) < F(w)$ του βήματος 2 τότε $w = w + \Delta w$ και $\mu = \mu * \beta$ ($\beta=0.1$) και επαναφορά στο βήμα 2, αλλιώς $\mu = \mu / \beta$ και επαναφορά στο βήμα 4.

Η εκπαίδευση του ΤΝΔ με τον αλγόριθμο εκμάθησης LM έγινε τη συνάρτηση `trainlm`.

5. ΑΠΟΤΕΛΕΣΜΑΤΑ

5.1 Εισαγωγή

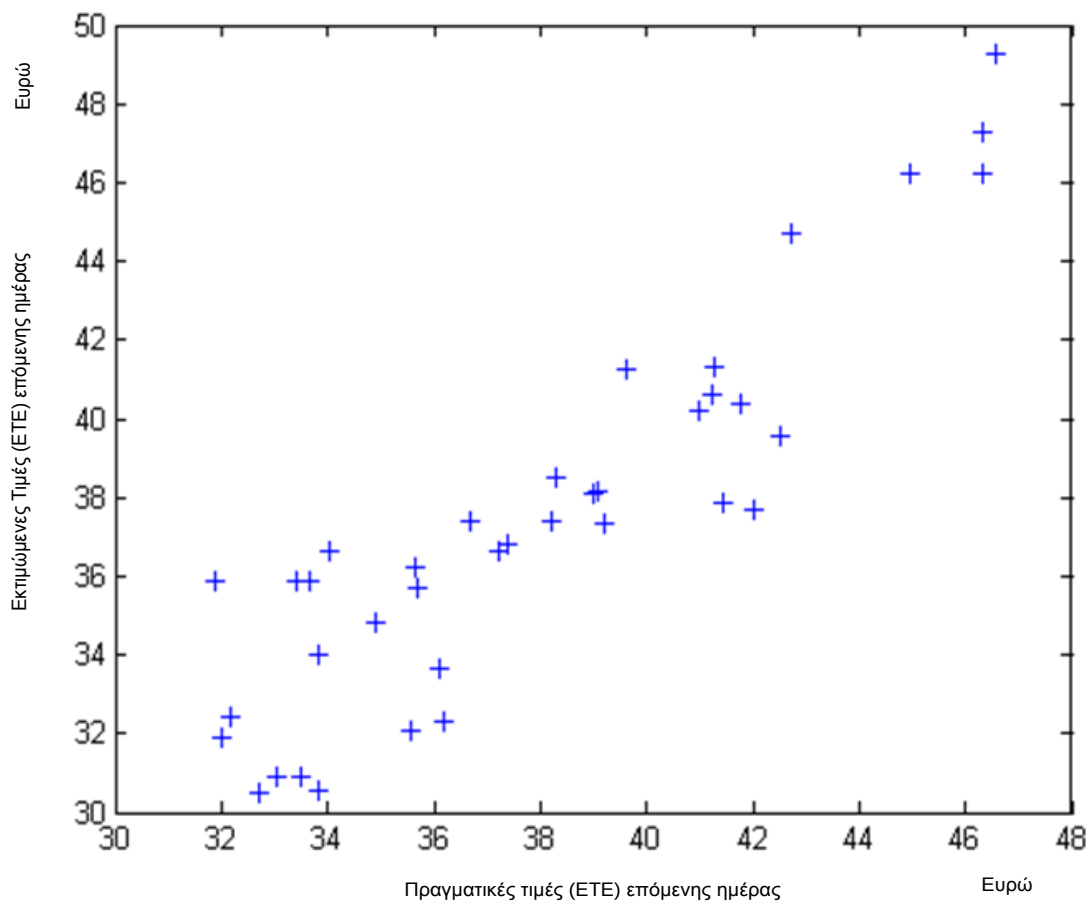
Για την τελειοποίηση του συγκεκριμένου νευρωνικού έγιναν πολλές δοκιμές τόσο σε επίπεδο διαφορετικών μετοχών και δεικτών όσο και σε παράγοντες που θεωρούνται όχι αμιγώς χρηματιστηριακοί. Οι δοκιμές με άλλους παράγοντες επανελήφθησαν και με την τελική μορφή του νευρωνικού, δηλαδή με έξι εισόδους, ένα κρυφό επίπεδο με εννιά νευρώνες και μία έξοδο. Ενδεικτικά αναφέρονται άλλοι παράγοντες που χρησιμοποιήθηκαν (τιμή χρυσού, οι δείκτες των χρηματιστηρίων, Παρισιου (Cac 40), Νέας Υόρκης (Nasdaq composite), (Sp 500), Χονγκ Κονγκ (Hang Seng), Τόκυο (Nikkei 225), επίσης δοκιμάστηκαν και άλλες ισοτιμίες του ευρώ όπως οι ισοτιμίες με την λίρα Αγγλίας και το Γιεν Ιαπωνίας. Σε όλες τις προαναφερθείσες περιπτώσεις τα αποτελέσματα είτε εμπεριείχαν μεγαλύτερο σφάλμα, είτε τα βάρη των συγκεκριμένων παραγόντων ήταν αρνητικά στο μεγαλύτερο κομμάτι του train set, πράγμα που σημαίνει ότι επηρεάζουν αντίστροφα ή καθόλου την τιμή τις εξόδου, αυτός ήταν και ο λόγος που δεν επιλέχθηκαν.

Για την εκπαίδευση και τη δοκιμή του νευρωνικού χρησιμοποιήθηκαν όπως αναφέρεται και στο πρώτο κεφάλαιο 462 δεδομένα εισόδου που αντιπροσωπεύουν 77 τιμές για κάθε ένα από τους παράγοντες εισόδου. Οι τιμές αυτές ομαδοποιήθηκαν σε δύο σετ για την εκπαίδευση του νευρωνικού και την δοκιμή του, πάρθηκαν δηλαδή για την δημιουργία του train set τα τελικά δεδομένα με βήμα 3 και για την δημιουργία του test set με βήμα 2 από το σύνολο των δεδομένων, αυτό έγινε για να υπάρχει πιο αντιπροσωπευτικό δείγμα από όλο το διάστημα των τιμών.

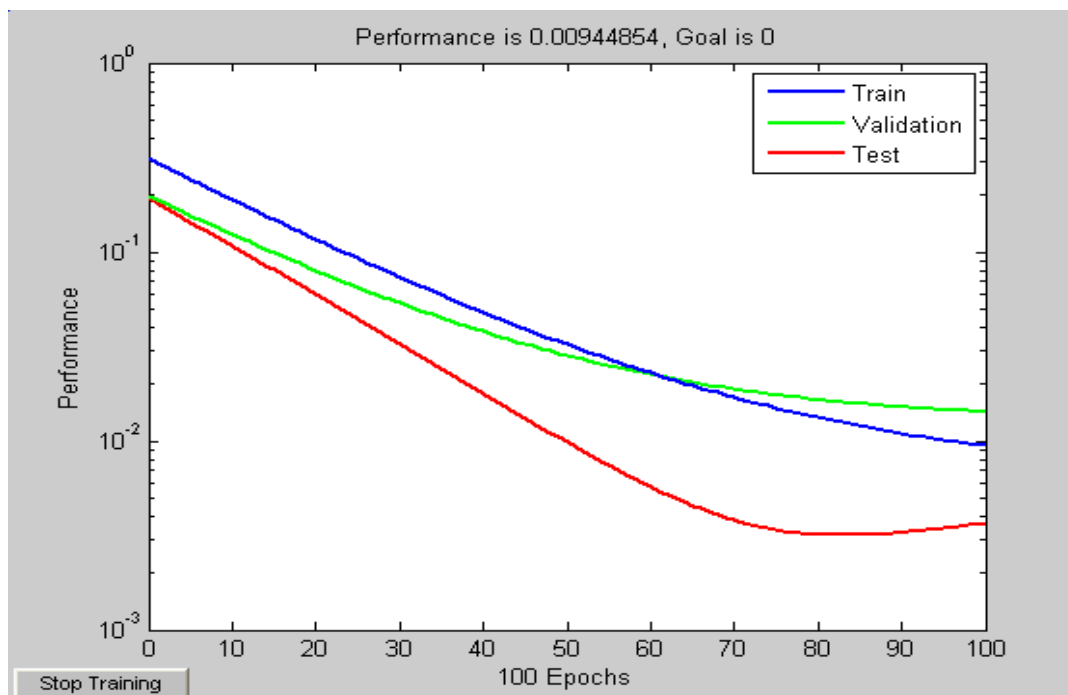
Αξίζει επίσης σε αυτό το σημείο να πούμε ότι για την τελειοποίηση του νευρωνικού έγιναν πολλά τρεξίματα και τα αποτελέσματα που παρουσιάζονται στη συνέχεια είναι μόνο τα τελικά και αποτελούν ένα μικρό μόνο ποσοστό των αποτελεσμάτων που ελέχθησαν.

5.2 Αλγόριθμος gradient descent

Για τη συνάρτηση εκμάθησης με αλγόριθμο gradient descent έχουμε τα αποτελέσματα των Σχημάτων 5.1 και 5.2, που προέκυψαν από την υλοποίηση του αλγορίθμου όπως προκύπτει από την παράγραφο 4.4.1. Συγκεκριμένα, για κάθε προσομοίωση δίνεται ένα διάγραμμα με τις εξόδους του ΤΝΔ για τα δείγματα του Test σε σχέση με τις επιθυμητές εξόδους (target outputs). Όλες οι προσομοιώσεις έγιναν για 100 epochs ώστε να υπάρχει δυνατότητα εύκολης σύγκρισης των διαφόρων αλγορίθμων. Παρατηρούμε ότι για τον αλγόριθμο gradient descent έγιναν και οι 100 epochs πράγμα που σημαίνει ότι δεν φτάσαμε σε κριτήριο τερματισμού.



Σχήμα 5.1 Έξοδος ΤΝΔ σε σχέση με την επιθυμητή έξοδο για διάφορα δείγματα με χρήση αλγορίθμου gradient descent.

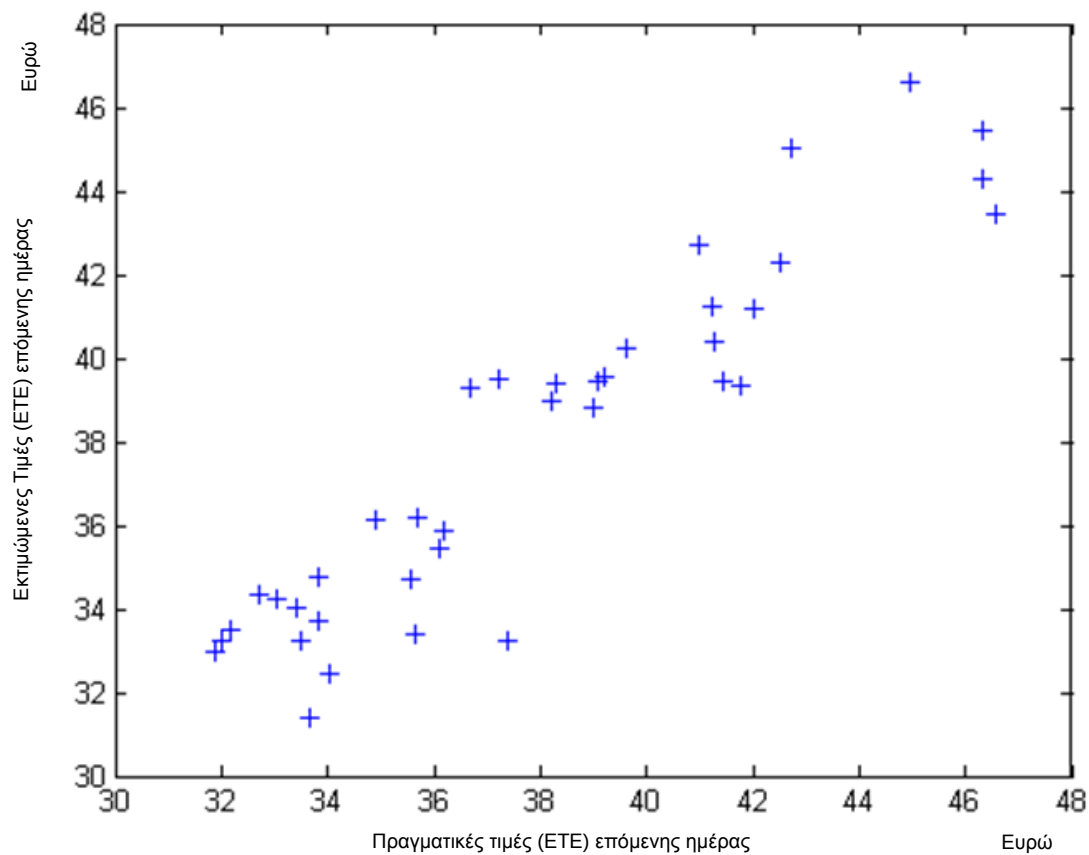


Σχήμα 5.2 Διακύμανση συνάρτησης σφάλματος (Performance) σε σχέση με τον αριθμό των epochs με χρήση αλγορίθμου gradient descent.

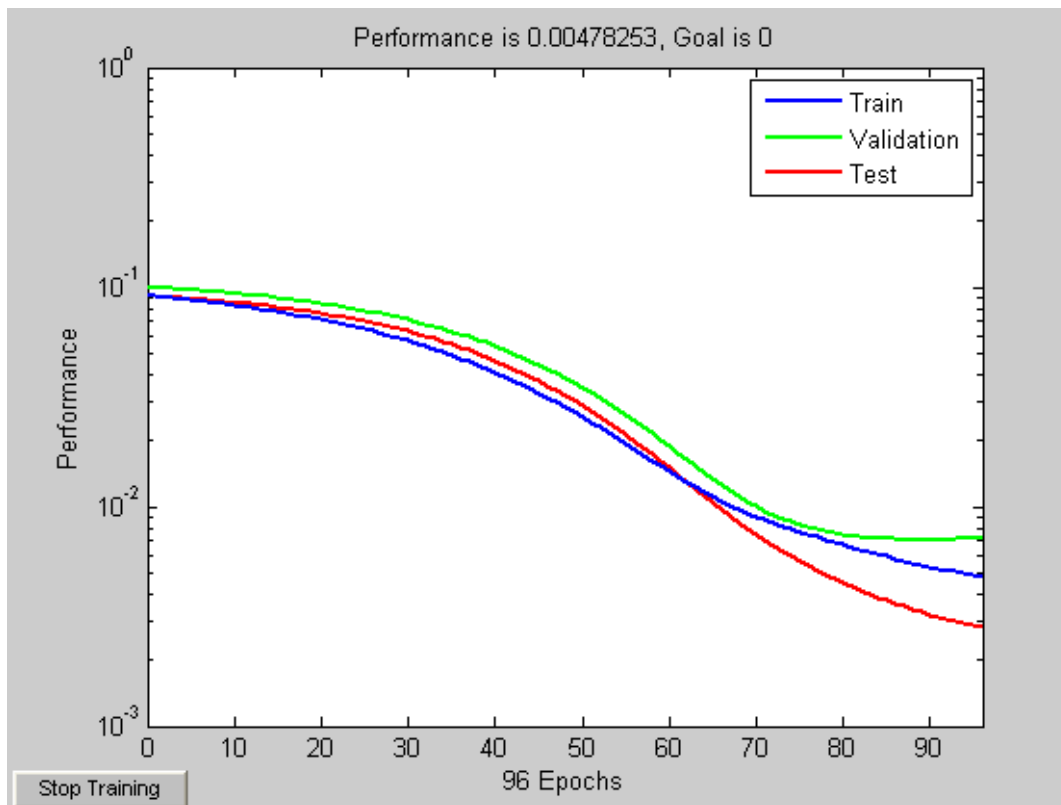
Στο Σχήμα 5.1 παρατηρείται ότι η απόδοση του αλγορίθμου gradient-descent δεν είναι ικανοποιητική καθώς σε μεγάλο αριθμό δειγμάτων παρατηρούνται μεγάλες αποκλίσεις από την έξοδο του ΤΝΔ σε σχέση με την επιθυμητή τιμή της εξόδου. Σε αυτό το σχήμα θα έπρεπε ιδανικά όλα τα σημεία να βρίσκονται πάνω στην ευθεία $y=x$. Επίσης στο Σχήμα 5.2 παρατηρείται ότι το σφάλμα είναι σχετικό μεγάλο (της τάξης του 10^{-2}).

5.3 Με προσαρμογή του ρυθμού εκμάθησης

Στα Σχήματα 5.3 και 5.4 παρουσιάζονται τα αποτελέσματα για εφαρμογή του αλγορίθμου gradient-descent με προσαρμογή του ρυθμού εκμάθησης, που προέκυψαν από την υλοποίηση του αλγορίθμου όπως περιγράφεται στην παράγραφο 4.4.2. Συγκεκριμένα, για κάθε προσομοίωση δίνεται ένα διάγραμμα με τις εξόδους του ΤΝΔ για τα δείγματα του Test σε σχέση με τις επιθυμητές εξόδους (target outputs). Εδώ ο αλγόριθμος τερμάτισε το την εκπαίδευση σου στις 96 epochs επειδή έφτασε σε κριτήριο τερματισμού.



Σχήμα 5.3 Έξοδος ΤΝΔ σε σχέση με την επιθυμητή έξοδο για διάφορα δείγματα με χρήση αλγορίθμου gradient descent με προσαρμογή του ρυθμού εκμάθησης

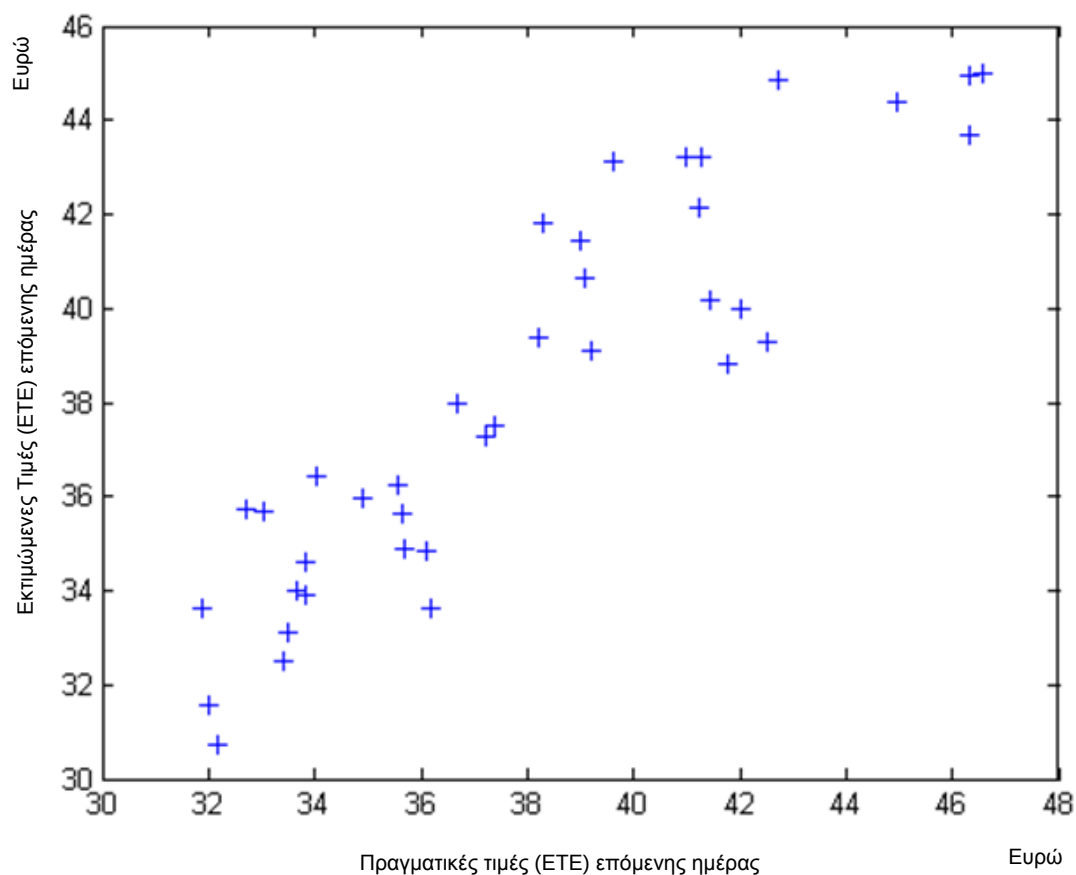


Σχήμα 5.4 Διακύμανση συνάρτησης σφάλματος (Performance) σε σχέση με τον αριθμό των epochs με χρήση αλγορίθμου gradient descent. με προσαρμογή του ρυθμού εκμάθησης

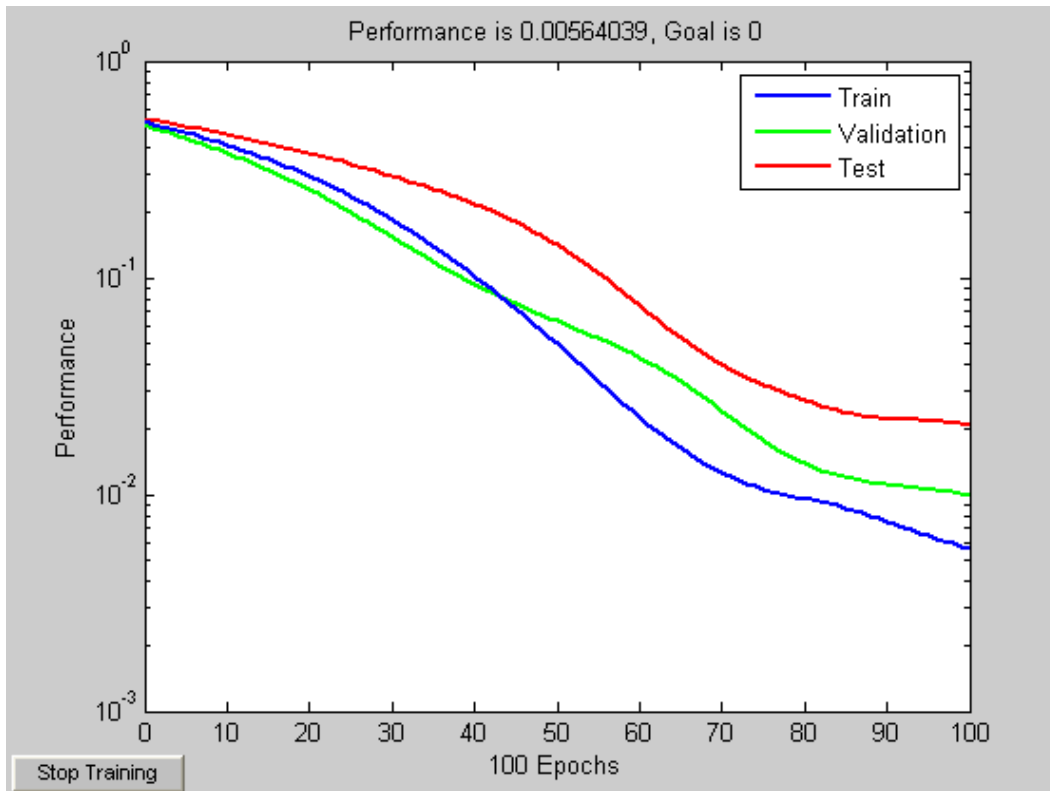
Στα Σχήματα 5.3 και 5.4 παρατηρούμε ότι υπάρχει σαφής βελτίωση του αλγορίθμου με τη χρήση αλγορίθμου προσαρμογής του ρυθμού εκμάθησης. Αυτό είναι εμφανές από το ότι τα σημεία στο Σχήμα 5.3 είναι πιο κοντά στην ευθεία $y=x$, ενώ στο Σχήμα 5.4 παρατηρείται ότι υπάρχει μείωση του σφάλματος.

5.4 Με χρήση όρου Momentum

Στα Σχήματα 5.5 και 5.6 παρουσιάζονται τα αποτελέσματα που προέκυψαν από την υλοποίηση του αλγορίθμου όπως περιγράφεται στην παράγραφο 4.4.3 με χρήση του όρου Momentum. Συγκεκριμένα, δίνεται ένα διάγραμμα με τις εξόδους του ΤΝΔ για τα δείγματα του Test σε σχέση με τις επιθυμητές εξόδους (target outputs). Στην παρούσα περίπτωση έγιναν 100 epochs.



Σχήμα 5.5 Έξοδος TNA σε σχέση με την επιθυμητή έξοδο για διάφορα δείγματα με χρήση αλγορίθμου gradient descent με προσαρμογή του ρυθμού εκμάθησης και χρήση momentum

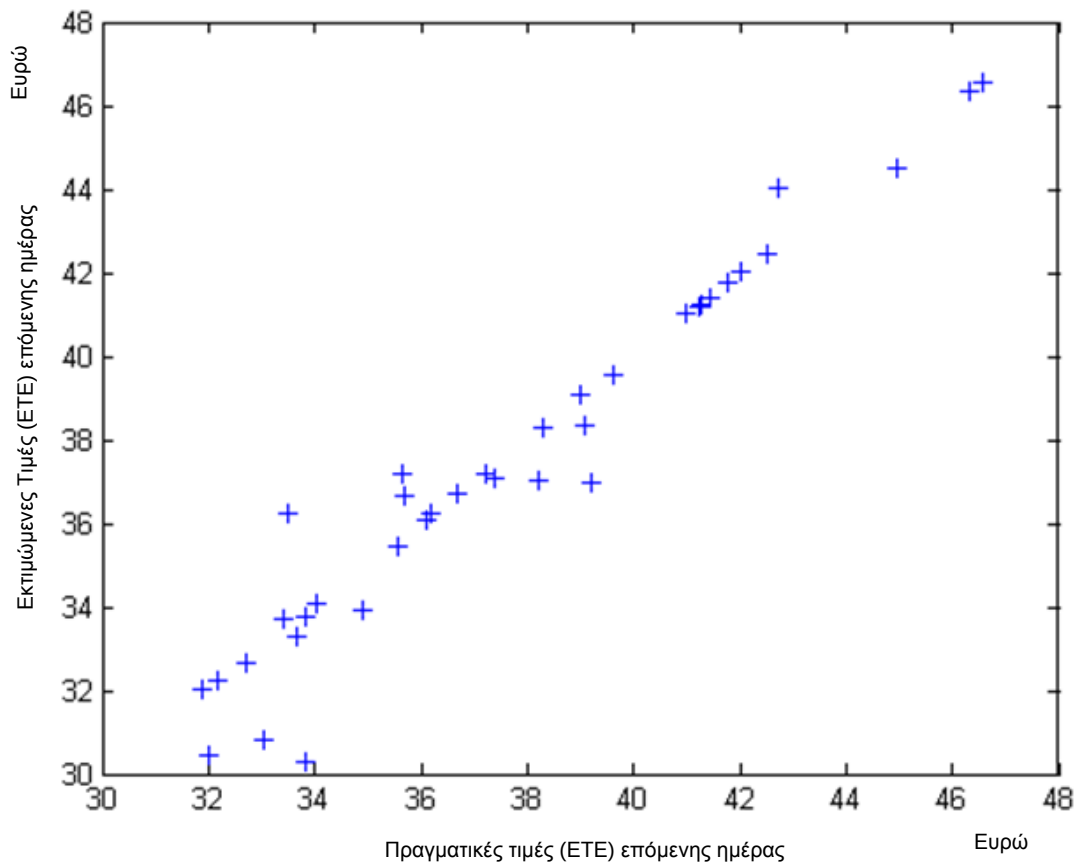


Σχήμα 5.6 Διακύμανση συνάρτησης σφάλματος (Performance) σε σχέση με τον αριθμό των epochs με χρήση αλγορίθμου gradient descent. με προσαρμογή του ρυθμού εκμάθησης και χρήση momentum

Στα Σχήματα 5.5 και 5.6 παρατηρείται ότι με χρήση momentum στο συγκεκριμένο πρόβλημα τα αποτελέσματα είναι ελαφρώς χειρότερα από ότι χωρίς τη χρήση του, τόσο σε επίπεδο σωστής σύγκλισης (στο Σχήμα 5.5 κάποια δείγματα αποκλίνουν αρκετά) όσο και σε επίπεδο συνολικού σφάλματος (είναι ελαφρώς αυξημένο).

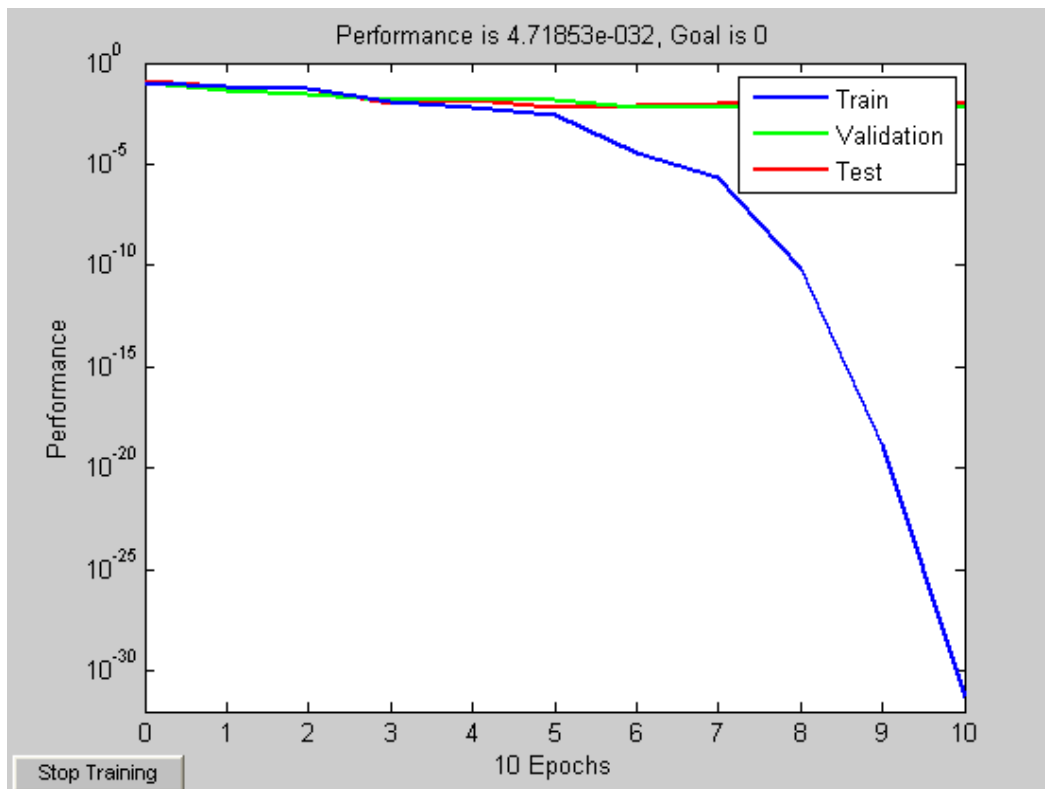
5.5 Αλγόριθμος Levenberg-Marquardt

Τα αποτελέσματα της εφαρμογής του αλγορίθμου εκπαίδευσης Levenberg-Marquardt παρουσιάζονται στα Σχήματα 5.7 και 5.8, όπως περιγράφεται στην παράγραφο 4.4.4. Στη συγκεκριμένη περίπτωση, έχουμε το διάγραμμα με τις εξόδους του ΤΝΔ για τα δείγματα του Test σε σχέση με τις επιθυμητές εξόδους (target outputs) και παρατηρούμε ότι ο αλγόριθμος Levenberg-Marquardt χρειάστηκε μόνο 10 epochs.



Σχήμα 5.7 Έξοδος TNA σε σχέση με την επιθυμητή έξοδο για διάφορα δείγματα με χρήση αλγορίθμου Levenberg-Marquardt

Στο Σχήμα 5.7 παρατηρείται ότι με χρήση του εν λόγω αλγορίθμου τα αποτελέσματα του TNA συγκλίνουν σε μεγάλο βαθμό με τις επιθυμητές εξόδους των δειγμάτων. Έτσι, όπως φαίνεται, τα δείγματα βρίσκονται πολύ κοντά στην ιδανική ευθεία $y=x$.



Σχήμα 5.8 Διακύμανση συνάρτησης σφάλματος (Performance) σε σχέση με τον αριθμό των epochs με χρήση αλγορίθμου Levenberg-Marquardt.

Στο Σχήμα 5.8 παρατηρείται ότι με χρήση του αλγορίθμου Levenberg-Marquardt το σφάλμα φτάνει σε πολύ μικρές τιμές (της τάξης του 10^{-32}), ενώ μειώνεται σε μεγάλο βαθμό για μικρό αριθμό epochs. Είναι, λοιπόν, σαφές ότι ο συγκεκριμένος αλγόριθμος υπερτερεί έναντι των τριών άλλων που χρησιμοποιήθηκαν και προαναφέρθηκαν.

6. ΣΥΜΠΕΡΑΣΜΑΤΑ - ΕΠΕΚΤΑΣΕΙΣ

Στο αβέβαιο οικονομικό περιβάλλον που ζούμε είναι επιτακτική ανάγκη να ανακαλύψουμε τρόπους πρόβλεψης των οικονομικών μεγεθών όπως είναι οι μετοχές. Σε αυτή την ανάγκη έρχεται ή τουλάχιστον προσπαθεί να δώσει απαντήσεις η παρούσα διπλωματική εργασία.

Σαν πρώτο και πρώιμο συμπέρασμα μπορούμε να πούμε ότι η πρόβλεψη της τιμής των μετοχών είναι εφικτή, αρκεί να επιλεχθούν τα σωστά δεδομένα και να βρεθούν οι παράγοντες που επηρεάζουν την τιμή της κάθε μετοχής ξεχωριστά.

Με χρήση τεχνικών νευρωνικών δικτύων αποδεικνύεται εδώ ότι, όχι μόνο μπορούμε να κάνουμε μια πολύ καλή πρόβλεψη της τιμής μιας μετοχής, της εθνικής τράπεζας της Ελλάδος στην προκειμένη περίπτωση, αλλά και να πούμε ποιοι είναι οι παράγοντες που επηρεάζουν την άνοδο ή την πτώση της, στην παρούσα εργασία είναι οι εξής έξι:

1. τιμή της μετοχής (ETE),
2. ισοτιμία Ευρώ – Δολαρίου (€-\$),
3. τιμή του πετρελαίου (BRENT),
4. τιμή κλεισίματος του γενικού δείκτη (ΓΔ) του Χρηματιστηρίου Αξιών Αθηνών (ΧΑΑ),
5. τιμή κλεισίματος του δείκτη FTSE 100 του Χρηματιστηρίου Αξιών του Λονδίνου και
6. τιμή κλεισίματος του δείκτη DOW JONES του Χρηματιστηρίου Αξιών της Νέας Υόρκης (NYSE).

Μετά την υλοποίηση του νευρωνικού δικτύου και την εκπαίδευση του με τα επιλεγμένα δεδομένα εισόδου αυτό μας απέδωσε μια έξοδο διαφορετική για κάθε αλγόριθμο που χρησιμοποιήθηκε για την εκπαίδευση του νευρωνικού δικτύου.

Ανατρέχοντας στο κεφάλαιο 5 στο οποίο και παραθέτονται τα αποτελέσματα των διαφόρων αλγορίθμων γίνεται σαφές ότι ακριβέστερα αποτελέσματα πρόβλεψης προέκυψαν με τη χρήση του αλγορίθμου Levenberg-Marquardt για του οποίου τη χρήση, παρατηρείται ότι το σφάλμα φτάνει σε πολύ μικρές τιμές (της τάξης του 10^{-32}), ενώ με τους άλλους αλγόριθμους που αναπτύχθηκαν τα αποτελέσματα ήταν πολύ χειρότερα.

Άρα καταλήγουμε στην πρόταση του συγκεκριμένου αλγορίθμου δηλαδή του αλγορίθμου Levenberg-Marquardt για την εκπαίδευση του νευρωνικού δικτύου και προτείνουμε την χρήση των παραγόντων που προαναφέρθηκαν.

Όπως σε κάθε εργασία έτσι και σε αυτήν εδώ, πάντα υπάρχει κάτι ακόμα για να γίνει και καινούργια πράγματα να ερευνηθούν.

Η αμέσως επόμενη μελέτη που θα είχε πρακτική και ουσιαστική αξία, θα ήταν να ελεγχθεί η συμπεριφορά άλλων μετοχών από την υψηλή, και όχι μόνο, κεφαλαιοποίηση για το ίδιο διάστημα και με τους ίδιους παράγοντες, ώστε να αποδειχθεί ότι η λεγόμενη πετρελαϊκή κρίση που έλαβε μέρος την περίοδο που μελετήθηκε επηρέασε εξίσου όλες τις μετοχές του ΧΑΑ, αξίζει σε αυτό το σημείο να υπενθυμίσουμε ότι στο διάστημα που μελετήθηκε η τιμή του Brent πετρελαίου βρέθηκε από 88,10€ στα 115,34€, μια διακύμανση της τάξης του 31%, πράγμα που έκανε τις οικονομίες όλου του κόσμου να αντιδρούν με μεγάλη νευρική δραστηριότητα.

Ένας άλλος δρόμος για τη δημιουργία σωστότερων αποτελεσμάτων και επιβεβαίωσης της ορθότητας της μεθόδου, είναι να επιλεγούν διαστήματα, διατηρώντας τους υπόλοιπους παράγοντες σταθερούς, στα οποία το ΧΑΑ και συγκεκριμένα η μετοχή της ΕΤΕ να είναι ανοδική καθώς την περίοδο που επιλέχθηκε η τιμή της ΕΤΕ από τα 46,67€ που ξεκίνησε έφτασε τα 32,00€.

Επίσης θα μπορούσε να μελετηθεί με τη χρήση άλλων παραγόντων, όπως της τιμής άλλων προϊόντων ή μετάλλων, τις τιμές κλεισίματος άλλων χρηματιστηρίων από

όλο τον κόσμο, επίσης να γίνει η πρόβλεψη της τιμής της ΕΤΕ άλλα και άλλων μετοχών με τα νέα δεδομένα, ακόμα όμως μεγαλύτερο ενδιαφέρον θα είχε η σύγκριση πολλών αποτελεσμάτων για διάφορες μετοχές από όλα τα χρηματιστήρια τόσο των λεγόμενων ώριμων όσο και των αναπτυσσόμενων αγορών.

7. ΒΙΒΛΙΟΓΡΑΦΙΑ

1. Γ. Ατσαλάκης, Σύστημα πρόβλεψης της βραχυχρόνιας τάσης της τιμής των μετοχών με χρήση νευρο-ασαφών μεθόδων, Χανιά 2006
2. Wasserman, P.D. (1989). Neural Computing: Theory and Practice. Van Nostrand, Reinhold, New York.
3. Wikowska, D., Neural Networks As A Forecasting Instrument For The Polish Stock Exchange, International Advances In Economic Research, Vol.1, Issue 3, 232-242, ISSN:10830898 (1995).
4. Walezak, S., Gaining Competitive Advantage For Trading In Emerging Capital Markets With Neural Networks, Journal Of Management Information Systems, Vo. 16, Issue 2, 178-194 (1999).
5. Chenoweth, T. and Z. Obradovic, A Multi-Component Nonlinear Prediction System For The S & P 500 Index, Neurocomputing Vol. 10, 275-290 (1996).
6. Kimoto, T., K. Asakawa , M. Yoda and M. Takeoka, Stock Market Prediction System With Modular Neural Networks, Proceedings of the International Joint Conference on Neural Networks (1990).
7. Poddig T., and H. Rehkugler, A “world” model of integrated financial markets using artificial neural networks. Neurocomputing, 10, pg. 251-273, (1996).
8. Kim, S.-S., Time-Delay Recurrent Neural Network For Temporal Correlations And Prediction, Neurocomputing, Vol. 20, 253-263 (1998).
9. Min Q., Nonlinear Predictability of Stock Returns Using Financial and Economic Variables. Journal of Business & Economic Statistics, vol.17, pg. 419, (1999).
10. Baba, N. and M. Kozaki, An Intelligent Forecasting System Of Stock Price Using Neural Networks, Proceedings Of The IEEE International Joint Conference On Neural Networks, 371-377 (1992).
11. Kim, S. H., Y. H. Kim, K. B. Sim, and H. T. Jeon, On developing an adaptive neural-fuzzy control system. Proc. IEEE/RSJ Conference on intelligent robots and systems, Yokohama, Japan, pp 950-957, (1993).

12. Baba N., and H. Suto, Utilization of artificial neural networks and the TD-learning method for constructing intelligent decision support systems. *European Journal of Operational Research*, 122, pg. 501-508, (2002).
13. Wittkemper, H.-G and M. Steiner, Using Neural Networks To Forecast The Systematic Risk Of Stocks, *European Journal Of Operational Research*, Vol. 90, 577-588 (1996).
14. Harvey, C. R., K. E. Travens and M. J. Costa, Forecasting Emerging Market Returns Using Neural Networks, *Emerging Markets Quarterly*, Vol. 4, Issue 2, 43-55 (2000).
15. Grudnitski, G. and L. Osburn, Forecasting S & P And Gold Futures Prices : An Application Of Neural Networks, *The Journal Of Future Markets*, Vol. 13, 631- 643 (1993).
16. Quah, T.-S. and B. Srinivasan, Improving Returns On Stock Investment Through Neural Network Selection, *Expert Systems With Applications*, vol. 17, 295-301 (1999).
17. Motiwalla, L. and M. Wahab, Predictable Variation And Profitable Trading Of US Equities: A Trading Simulation Using Neural Networks, *Computer And Operations Research*, Vol 27, 1111-1129 (2000).
18. Yiwen,Y., L. Guizhong and Z. Zongping, Stock Market Trend Prediction Based On Neural Networks, *Multiresolution Analysis And Dynamical Reconstruction Proceedings of the IEEE/IAFE/INFORMS 2000 Conference on Computational Intelligence for Financial Engineering*, 155-156, ISBN: 0-7803-6429-5 (2000).
19. Casas, C.A. Tactical Asset allocation: An Artificial Neural Network Based Model, *Proceedings of the International Joint Conference on Neural Networks*, 1811-1816, ISBN: 6678288 (2001).
20. Atiya, A., Noha Talaat and Samir Shaheen, An Efficient Stock Market Forecasting Model Using Neural Networks, *Proceedings Of The IEEE International Conference On Neural Networks*, ISSN: 2112-2115 (1997).
21. Andreou, A.S., C.C. Neocleous, C.N. Schizas and C. Toumpouris, Testing The Predictability Of The Cyprus Stock Exchange : The Case Of An Emerging Market,

Proceedings Of The International Joint Conference On Neural Networks, 360-365. ISBN: 5350978 (2000).

22. Fernandez-Rodriguez, F., C. Gonzalez-Martel and S. Sosvilla-Rivebo, On The Profitability Of Technical Trading Rules Based On Artificial Neural Networks: Evidence From The Madrid Stock Market, Economics Letters, Vol. 69, 89-94 (2000).

23. Chen, A.S., M.T. Leung and H. Daouk, Application Of Neural Networks To An Emerging Financial Market : Forecasting And Trading The Taiwan Stock Index, Computers And Operations Research, Vol. 30, 901-923 (2003).

24. Thawornwong,S. and D. Enke, The Adaptive Selection Of Financial And Economic Variables For Use With Artificial Neural Networks, Neurocomputing, Vol. 56, 205-232 (2004).

25. Schumann, M. and T. Lohrbach, Comparing Artificial Neural Networks With Statistical Methods Within The Field Of Stock Market Prediction, Proceedings of the Twenty-Sixth International Conference on System Sciences, Vol. 4, 597-606, ISBN: 0-8186-3230-5 (1993).

26. Π. Αργυράκη, Νευρωνικά Δίκτυα και Εφαρμογές, Ελληνικό Ανοικτό Πανεπιστήμιο, Σχολή Θετικών Επιστημών & Τεχνολογίας, Πάτρα 2001.

27. Rumelhart, D.E., Hinton, G.E., Williams, R.J. ,1986. Learning internal representation by back-propagating errors. In: Rumelhart, D.E., McClelland, J.L., the PDP Research Group (Eds.), Parallel Distributed Processing:Explorations in the Microstructure of cognition. MIT Press, MA

28. Harry Wechsler, Neural networks for perception, Boston Academic Press, 1992.

29. Parker, D. B. (1987). Optimal Algorithms For Adaptive Networks: Second Order Back 31 Propagation, Second Order Direct Propagation, And Second Order Hebbian Learning. Proceedings of IEEE International Conference on Neural Networks, 593-600.

30. J. A. Anderson, An Introduction to Neural Networks, MIT Press, Cambridge (1995)

31. J. Hertz, A. Krogh an R. G. Palmer, Introduction to the Theory of Neural Computation, Addison-Wesley, Reading (Mass), (1999)

32. <http://tonto.eia.doe.gov/dnav/pet/hist/rbrted.htm> (τιμές πετρελαίου)

33. <http://www.in.gr> (Γενικός δείκτης, Τιμές μετοχών)
34. <http://www.ase.gr> (Σύσταση δεικτών)
35. <http://finance.yahoo.com> (Δείκτες Ξένων Χρηματιστηρίων)
36. <http://www.oanda.com/convert/fxhistory> (Ισοτιμία Ευρώ – Δολαρίου)