



ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ

Τμήμα Μηχανικών Παραγωγής και Διοίκησης

Εργαστήριο Δυναμικών Συστημάτων και Προσομοίωσης

**ΜΑΚΡΟΟΙΚΟΝΟΜΙΚΗ ΠΡΟΒΛΕΨΗ ΜΕ ΧΡΗΣΗ
ΓΡΑΜΜΙΚΩΝ ΜΟΝΤΕΛΩΝ ΚΑΙ ΤΕΧΝΗΤΩΝ
ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΑΡΒΑΝΙΤΑΚΗ ΑΦΡΟΔΙΤΗ - ΔΗΜΗΤΡΑ

ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ

Δρ. ΚΟΣΜΑΤΟΠΟΥΛΟΣ ΗΛΙΑΣ

ΧΑΝΙΑ 2005

ΠΕΡΙΕΧΟΜΕΝΑ

ΚΕΦΑΛΑΙΟ 1: ΜΑΚΡΟΟΙΚΟΝΟΜΙΚΗ ΠΡΟΒΛΕΨΗ

1.1	Προσέγγιση της πρόβλεψης	1
1.2	Μέθοδοι μακροοικονομικής πρόβλεψης	2
1.3	Προβλήματα στη μακροοικονομική πρόβλεψη	3
1.4	Ταξινόμηση των προβλέψεων	5
1.5	Μέθοδοι Χρονοσειρών	6
1.5.1	Εποχική Αποσύνθεση	7
1.5.2	Εκθετική Εξομάλυνση	8
1.5.3	Γραμμική Παλινδρόμηση	8
1.5.4	Αυτοπαλινδρομικά Μοντέλα	9
1.5.5	Τεχνητά Νευρωνικά Δίκτυα	9
1.6	Τα κίνητρα μακροοικονομικής πρόβλεψης της οικονομίας μιας χώρας	10
1.7	Η πρόσφατη οικονομική κατάσταση στις ΗΠΑ	12
1.8	Σκοπός της εργασίας	15
1.9	Περιεχόμενα της εργασίας	16

ΚΕΦΑΛΑΙΟ 2: ΠΡΟΣΕΓΓΙΣΗ ΤΗΣ ΜΑΚΡΟΟΙΚΟΝΟΜΙΚΗΣ ΠΡΟΒΛΕΨΗΣ ΜΕ ΓΡΑΜΜΙΚΑ ΜΟΝΤΕΛΑ

2.1	Η μέθοδος της πολλαπλής γραμμικής παλινδρόμησης	17
2.2	Υποθέσεις και Περιορισμοί	19
2.3	Μαθηματικό Υπόβαθρο	21
2.4	Διερεύνηση σε μορφή πινάκων	22
2.5	Πρόβλεψη της Αμερικανικής Οικονομίας με πολλαπλή γραμμική παλινδρόμηση	25
2.6	Αξιολόγηση του μοντέλου	27

ΚΕΦΑΛΑΙΟ 3: ΤΕΧΝΗΤΑ ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

3.1	Πρόβλεψη με νευρωνικά δίκτυα	30
3.2	Ο ανθρώπινος εγκέφαλος	33
3.3	Εισαγωγή στα τεχνητά νευρωνικά δίκτυα	35
3.4	Η απλούστερη μορφή τεχνητού νευρωνικού δικτύου	36
3.4.1	Νευρωνικά δίκτυα με μη γραμμικές συναρτήσεις ενεργοποίησης	39
3.4.2	Νευρωνικά δίκτυα με κρυμμένα επίπεδα	41
3.4.3	Αυξητικά νευρωνικά δίκτυα	44
3.5	Νευρωνικά Δίκτυα Πολλαπλών Επιπέδων	45

3.6	Νευρωνικά δίκτυα υψηλής τάξης	48
3.6.1	Δίκτυα sigma-pi	52
3.6.2	Δίκτυα pi-sigma	55
3.6.3	Η κατάλληλη αρχιτεκτονική για τα δίκτυα υψηλής τάξης	57
3.6.4	Σύγκριση δικτύων υψηλής τάξης με τα δίκτυα πολλαπλών επιπέδων	58
3.7	Μεθοδολογία κατασκευής ΤΝΔ για πρόβλεψη	58
3.8	Πλεονεκτήματα και μειονεκτήματα της χρήσης ΤΝΔ για πρόβλεψη σε σχέση με στατιστικές μεθόδους	73

ΚΕΦΑΛΑΙΟ 4: ΠΡΟΣΕΓΓΙΣΗ ΤΗΣ ΜΑΚΡΟΟΙΚΟΝΟΜΙΚΗΣ ΠΡΟΒΛΕΨΗΣ ΜΕ ΤΕΧΝΗΤΑ ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

4.1	Περιγραφή του προβλήματος	78
4.2	Εισαγωγή – Προετοιμασία των δεδομένων στη matlab	79
4.3	Σχολιασμός του κώδικα/ διαδικασίας του προβλήματος	79
4.4	Σχολιασμός των Αποτελεσμάτων	90

ΚΕΦΑΛΑΙΟ 5: ΣΥΜΠΕΡΑΣΜΑΤΑ

105

ΠΑΡΑΡΤΗΜΑ Ι ΔΕΔΟΜΕΝΑ ΧΡΟΝΟΣΕΙΡΩΝ

107

ΠΑΡΑΡΤΗΜΑ ΙΙ ΚΩΔΙΚΑΣ

110

ΒΙΒΛΙΟΓΡΑΦΙΑ

114

ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα να ευχαριστήσω εγκάρδια τον επιβλέποντα καθηγητή της παρούσας εργασίας, δόκτορα Κοσματοπούλο Ηλία για την ευκαιρία που μου έδωσε να συνεργαστούμε καθώς και για την εμπιστοσύνη που μου έδειξε όλη αυτή την περίοδο. Αποτέλεσε πηγή έμπνευσης για εμένα, ενώ η καθοδήγηση και οι πολύτιμες παρατηρήσεις του κατά τη διεξαγωγή της εργασίας, αποδείχθηκαν αποφασιστικής σημασίας για την περάτωσή της. Η ευφυΐα του πνεύματος και η πραότητα του χαρακτήρα του, συνετέλεσαν καθοριστικά στο αποτέλεσμα, το οποίο ελπίζω ότι δικαιώνει την προσπάθειά μας.

ΠΕΡΙΛΗΨΗ

Οι σύγχρονες εξελίξεις και εφαρμογές στον τομέα της τεχνητής νοημοσύνης, καθιστούν τα τεχνητά νευρωνικά δίκτυα ισχυρό εργαλείο στον τομέα της μακροοικονομικής πρόβλεψης, εξαιτίας των δυνατοτήτων που έχουν να ανιχνεύουν και να αναπαράγουν γραμμικές και μη γραμμικές σχέσεις στα σύνολα των μεταβλητών. Η παρούσα εργασία εξετάζει κατά πόσο είναι δυνατή η μακροοικονομική πρόβλεψη με τη χρήση μιας συγκεκριμένης δομής νευρωνικών δικτύων, τα νευρωνικά δίκτυα υψηλής τάξης και συγκεκριμένα τα δίκτυα sigma-pi, ενώ ταυτόχρονα παραλληλίζει και συγκρίνει την εφαρμογή των εν λόγω δικτύων με παραδοσιακές οικονομετρικές τεχνικές. Τα γραμμικά και μη γραμμικά μοντέλα που εξετάζονται στοχεύουν στην πρόβλεψη της μεταβολής του πραγματικού προϊόντος, των καθαρών εξαγωγών και του πληθωρισμού με δεδομένα χρονοσειρών για την εθνική οικονομία των ΗΠΑ.

Η χρήση τεχνητών νευρωνικών δικτύων ως εργαλείο μακροοικονομικής πρόβλεψης αποτελεί σχεδόν καινούριο πεδίο εφαρμογής. Τα αποτελέσματα των ερευνών που έχουν διεκπαιραιωθεί είναι ενθαρυντικά, ωστόσο συνιστάται επιφυλακτικότητα, εξαιτίας του μικρού αριθμού αξιόπιστων μελετών. Φυσικά, όπως και με οποιοδήποτε μαθηματικό μοντέλο, οι προβλέψεις των νευρωνικών δικτύων εξαρτώνται από τα δεδομένα που εισάγονται στο μοντέλο. Τα νευρωνικά δίκτυα, καθώς και τα παραδοσιακά οικονομετρικά μοντέλα πρόβλεψης, δεν μπορούν να προβλέψουν αναπάντεχα γεγονότα, όπως οικονομική ύφεση εξαιτίας αναπροσαρμογής των τιμών του πετρελαίου ή πολιτικές κρίσεις. Συνεπώς, τόσο τα μοντέλα γραμμικής παλινδρόμησης, όσο και τα τεχνητά νευρωνικά δίκτυα θα πρέπει να αντιμετωπίζονται ως εργαλεία πρόβλεψης σε συνδιασμό με επιπλέον εργαλεία και μεθόδους.

ΚΕΦΑΛΑΙΟ 1^ο

Μακροοικονομική πρόβλεψη

“Δεν υπάρχει ποτέ μια κρυστάλλινη σφαίρα όταν τη χρειάζεσαι”

Washington Times, τίτλος πρωτοσέλιδου σχετικά με την πιθανότητα οικονομικής ύφεσης στις ΗΠΑ, Ιούνιος 2001.

1.1 Προσέγγιση της πρόβλεψης

Πρόβλεψη είναι οποιαδήποτε δήλωση για το μέλλον. Τέτοιες δηλώσεις μπορεί να είναι ισχυρά θεμελιωμένες ή να υστερούν σταθερών βάσεων, ακριβείς ή ανακριβείς, βασισμένες σε μοντέλα ή ανεπίσημες. Οι προβλέψεις παράγονται με ποικίλες μεθόδους από διάφορα εμπειριστατωμένα συστήματα με εκατοντάδες εξισώσεις, έως μοντέλα που δεν παρουσιάζουν κάποια αξιοσημείωτη βάση. Συνεπώς, η πρόβλεψη αποτελεί ευρύ αντικείμενο έρευνας.

Εφόσον αποτελεί δήλωση για το μέλλον, οτιδήποτε μπορεί να προβλεφθεί, χωρίς ωστόσο να ισχυριζόμαστε ότι τα αποτελέσματα θα είναι ουσιώδη και χρήσιμα. Η εγκυρότητα των προβλέψεων εξαρτάται από το πόσο καλά δομημένο είναι το μοντέλο πρόβλεψης. Απλοϊκές υποθέσεις δεν εμπνέουν ιδιαίτερη εμπιστοσύνη, ενώ προβλέψεις από εκτενώς δοκιμασμένες προσεγγίσεις αντιμετωπίζονται πιο αισιόδοξα. Δυστυχώς ακόμη και ο δεύτερος τύπος προβλέψεων ορισμένες φορές αποδεικνύεται ανεπαρκής, καθώς το μέλλον είναι αβέβαιο.

“Εξαιτίας των γεγονότων που αγνοούμε ότι αγνοούμε το μέλλον είναι ευρέως απρόβλεπτο. Ωστόσο μπορούμε να προβλέψουμε ή έστω να φανταστούμε ορισμένες εξελίξεις βάσει της υπάρχουσας γνώσης”. (Maxine Singer, “Thoughts of a nonmillenarian”).

Ελάχιστη πρόοδος μπορεί να πραγματοποιηθεί σχετικά με την αβεβαιότητα που πηγάζει από τα γεγονότα που “αγνοούμε ότι αγνοούμε”. Ωστόσο, η προφανής τυχαιότητα των αποτελεσμάτων μέσα στα πλαίσια που κατανοούμε, η αποκαλούμενη “μετρήσιμη αβεβαιότητα”, δύναται να επεξεργασθεί από τον ερευνητή που διεξάγει την πρόβλεψη. Αυτό πραγματοποιείται συνήθως με τη μορφή του διαστήματος

(interval) πρόβλεψης γύρω από ένα “σημείο” πρόβλεψης, το οποίο θεωρείται ως η κεντρική τάση ή “το πιθανότερο” αποτέλεσμα [1].

1.2 Μέθοδοι μακροοικονομικής πρόβλεψης

Υπάρχουν πολυάριθμοι τρόποι πρόβλεψης. Αυτοί περιλαμβάνουν: επίσημες στατιστικές αναλύσεις βασισμένες σε καλώς δομημένα μοντέλα, στατιστικές αναλύσεις που δε βασίζονται σε παραμετρικά μοντέλα, ανεπίσημους εμπειρικούς υπολογισμούς, απλά συμπεράσματα, αναλύσεις αγοράς, προσωπικές κρίσεις, ρύψη ενός νομίσματος, υποθέσεις και προαισθήματα.

Στην οικονομική επιστήμη οι μέθοδοι προβλέψεις περιλαμβάνουν:

1. Εικασίες, εμπειρικοί κανόνες ή ανεπίσημα μοντέλα
2. Κρίσεις εμπειρογνομόνων (expert judgment)
3. Εξαγωγή συμπερασμάτων (extrapolation)
4. Κυρίαρχοι δείκτες (leading indicators)
5. Μελέτες έρευνας αγοράς (survey)
6. Μοντέλα χρονοσειρών
7. Οικονομετρικά συστήματα

Οι εικασίες και οι σχετικές μέθοδοι βασίζονται εξολοκλήρου στην τύχη. Αν και αυτό αποτελεί την ελάχιστη υπόθεση συγκριτικά με τις υπόλοιπες μεθόδους, οι εικασίες δεν αποτελούν χρήσιμη μέθοδο: επιτυχείς εικασίες αναφέρονται συχνά, ενώ ανεπιτυχείς αγνοούνται σιωπηλά. Επίσης η αβεβαιότητα που σχετίζεται με την κάθε εικασία είναι αδύνατο να υπολογισθεί εκ των προτέρων. Εάν πλήθος οικονομολόγων αρχίσουν τις υποθέσεις οι πιθανότητες είναι ότι κάποιος από αυτούς θα αποδειχθεί σωστός. Αλλά αυτό δε θα μπορούσε ποτέ να αποτελέσει εμπειριστατωμένη επιστημονική προσέγγιση.

Οι κρίσεις εμπειρογνομόνων είναι συνήθως μέρος της προσέγγισης πρόβλεψης, ωστόσο στερείται εγκυρότητας όταν είναι το μοναδικό στοιχείο της διαδικασίας. Η εξαγωγή συμπερασμάτων έχει ισχύ μόνο όσο παραμένουν οι τάσεις, κάτι το οποίο είναι καθεαυτό αμφίβολο. Η πρόβλεψη που βασίζεται σε κυρίαρχους δείκτες προϋποθέτει σταθερή σχέση μεταξύ των μεταβλητών που “καθοδηγούν” (“lead”) και των μεταβλητών που “καθοδηγούνται” (“led”). Όταν τα αίτια της

κυριαρχίας είναι σαφή, όπως συμβαίνει με τις παραγγελίες που προηγούνται της παραγωγής, τότε οι δείκτες ίσως είναι χρήσιμοι, αλλιώς είναι πιθανό να δώσουν παραπλανητικές πληροφορίες.

Οι μελέτες έρευνας αγοράς μπορεί να παρέχουν πληροφορίες σχετικά με μελλοντικά γεγονότα, αλλά εξαρτώνται από σχέδια που πρέπει να πραγματοποιηθούν. Τα μοντέλα χρονοσειρών που περιγράφουν τα ιστορικά πρότυπα των δεδομένων αποτελούν δημοφιλείς μεθόδους προβλέψεις και συχνά έχει παρατηρηθεί ότι είναι ανταγωνιστικά συγκρίσιμα με τα οικονομετρικά συστήματα εξισώσεων πολλαπλών μεταβλητών. Αλλά όπως και όλες οι μέθοδοι εστιάζουν αποκλειστικά στη “μετρήσιμη αβεβαιότητα”.

Τα οικονομετρικά συστήματα εξισώσεων είναι το κύριο εργαλείο της οικονομικής πρόβλεψης. Οι εξισώσεις που περιλαμβάνουν μοντελοποιούν τη συμπεριφορά διακριτών ομάδων (καταναλωτές, παραγωγοί, εργάτες, επενδυτές) υποθέτοντας κάποιο βαθμό ορθολογισμού και ιστορικής εμπειρίας. Τα πλεονεκτήματα της χρήσης επίσημων οικονομετρικών συστημάτων εθνικών οικονομιών είναι ότι λαμβάνουν υπόψη την υπάρχουσα εμπειρική και θεωρητική γνώση του τρόπου λειτουργίας των οικονομιών, παρέχουν το πλαίσιο εργασίας για προοδευτική έρευνα στρατηγικής και τέλος εξασφαλίζουν προβλέψεις και συμβουλές πολιτικής. Τα οικονομετρικά μοντέλα και τα μοντέλα χρονοσειρών αποτελούν τις κύριες μεθόδους οικονομικής πρόβλεψης, ωστόσο οι κρίσεις, οι δείκτες και ακόμη και οι εικασίες ενδέχεται να τροποποιήσουν τα αποτελέσματά τους.

1.3 Προβλήματα στην μακροοικονομική πρόβλεψη

Ένα από τα κυριότερα προβλήματα στην οικονομική πρόβλεψη είναι το γεγονός ότι οι οικονομίες εξελίσσονται στον χρόνο και είναι υποκείμενες σε ασυνεχείς απρόσμενες διαταραχές. Η οικονομική εξέλιξη έχει τις βάσεις της σε επιστημονικές ανακαλύψεις που οδηγούν στην τεχνολογική πρόοδο η οποία ενσωματώνεται στο κεφάλαιο και οδηγεί σε αύξηση της πραγματικής παραγόμενης εξόδου. Επιπλέον δομικές μεταβολές (structural breaks) μπορεί να επιταχυνθούν από αιφνίδιες μεταβολές στη νομοθεσία, την οικονομική πολιτική ή από πολιτικές αναταραχές. Επομένως, προηγούμενες ευσταθείς σχέσεις μεταξύ οικονομικών μεταβλητών είναι υποκείμενες σε μεταβολές και εάν χρησιμοποιήθηκαν για την

πραγματοποίηση προβλέψεων σε τέτοιες περιόδους μπορεί να έχουν ως αποτέλεσμα μεγάλα και εξακολουθητικά σφάλματα.

Επιπλέον, τα εμπειρικά οικονομετρικά μοντέλα που χρησιμοποιούνται για την κατανόηση και την πρόβλεψη διαδικασιών πολύπλοκων όσο οι εθνικές οικονομίες απέχουν αρκετά από την τέλεια αναπαράσταση συμπεριφοράς. Οι επιστήμονες που πραγματοποιούν την πρόβλεψη γνωρίζουν αμυδρά τις αλλαγές που αναπτύσσονται και ακόμη και όταν είναι σε θέση να τις οραματιστούν είναι δύσκολο να ποσοτικοποιήσουν τις πιθανές επιπτώσεις. Αυτές οι δυσκολίες συνεπάγονται ότι η οικονομική πρόβλεψη είναι γεμάτη προβλήματα και στην πράξη η αποτυχία της πρόβλεψης – σημαντική απόκλιση της απόδοσης σχετικά με το αναμενόμενο αποτέλεσμα – είναι σύνηθες φαινόμενο.

Τα οικονομετρικά μοντέλα πρόβλεψης είναι συστήματα που προϋποθέτουν σχέση μεταξύ μεταβλητών όπως GDP, πληθωρισμός, τιμή συναλλάγματος κτλ. Οι εξισώσεις τους εκτιμώνται από τα διαθέσιμα δεδομένα κυρίως συναθροιστικές χρονοσειρές. Τέτοια μοντέλα θεωρείται ότι έχουν τρεις συνιστώσες: αιτιοκρατικούς όρους για τη σύλληψη των μέσων και σταθερών αυξήσεων, με σταθερή μελλοντική αξία, στοχαστικές μεταβλητές με άγνωστη μελλοντική τιμή και απαρατήρητα σφάλματα των οποίων όλες οι τιμές (παρελθοντικές, παροντικές και μελλοντικές είναι άγνωστες). Η σχέση μεταξύ των οποιωνδήποτε εκ των τριών συνιστώσες πιθανόν να διατυπωθεί ακατάλληλα, να εκτιμηθεί ανακριβώς, ή να μεταβληθεί απρόσμενα. Τις περισσότερες φορές η αποτυχία της πρόβλεψης οφείλεται στη συμπεριφορά των αιτιοκρατικών όρων και συγκεκριμένα στις απροσδόκητες μεταβολές των τιμών τους.

Τα διαστήματα (forecast intervals) πρόβλεψης στοχεύουν στη μέτρηση της αβεβαιότητας της πρόβλεψης αλλά μπορούν μόνο να απορρίψουν τη “γνωστή αβεβαιότητα” – που απορρέει από τη εκτίμηση του μοντέλου, την υπόθεση ότι οι μελλοντικές διαταραχές μοιάζουν με τις παρελθοντικές – ενώ οι απρόσμενες αιτιοκρατικές μεταβολές παρουσιάζονται περιοδικά στην οικονομία. Εφόσον “δεν γνωρίζουμε τι δεν γνωρίζουμε” είναι δύσκολη η αντιμετώπιση της “άγνωστης αβεβαιότητας”.

Ωστόσο, υπάρχουν τρόποι αποφυγής της συστηματικής αποτυχίας οικονομικής πρόβλεψης που είναι υποκείμενη σε μεγάλες και απρόσμενες μεταβολές. Όταν οι μεταβολές είναι απροσδόκητες, μόνο ένας μάγος θα ήταν σε θέση να αποφύγει τα σφάλματα σε προβλέψεις που πραγματοποιήθηκαν πριν συμβούν οι

μεταβολές. Δεδομένης της αδυναμίας πρόβλεψης των αναπάντεχων μεταβολών, απαιτείται προσαρμογή του μοντέλου μόλις παρουσιαστεί η μεταβολή ώστε να αποφευχθεί η ακολουθία των ανεπαρκών προβλέψεων. Ορισμένα μοντέλα προσαρμόζονται ταχύτερα από άλλα. Ακόμη και όταν ένα μοντέλο δεν προσαρμόζεται γρήγορα, υπάρχουν μηχανισμοί που επιταχύνουν τη διαδικασία. Επίσης προσπάθειες γίνονται στην μελέτη σημάτων “πρώιμης προειδοποίησης” (early warning signals). Τέτοιες μέθοδοι προβλέπουν τις μεταβολές παρατηρώντας τις να συμβαίνουν αλλού, σε “κυρίαρχες περιοχές” ή σε υψηλότερη συχνότητα δεδομένων (πχ. εβδομαδιαία, όταν χρησιμοποιούνται μηνιαία δεδομένα) .

Τα οικονομετρικά μοντέλα για να πραγματοποιούν επιτυχείς προβλέψεις πρέπει να μιμούνται την προσαρμοστικότητα των καλύτερων μηχανισμών πρόβλεψης, διατηρώντας παράλληλα τα θεμέλιά τους στην οικονομική ανάλυση. Παρά τα πολλά προβλήματα που αντιμετωπίζει η οικονομική πρόβλεψη, τα εν λόγω μοντέλα συντελούν ως μέσο κατανόησης της οικονομικής συμπεριφοράς καθώς και επιπλέον μάθησης από την αποτυχία.

Συνοψίζοντας θα λέγαμε ότι στην πράξη η οικονομική πρόβλεψη αποτελείται από μεικτά στοιχεία επιστήμης – βασισμένα σε οικονομετρικά συστήματα που έχουν εκτιμηθεί προσεκτικά ενσωματώνοντας την οικονομική γνώση – και τέχνης, δηλαδή υποκειμενικές αποφάσεις και κρίσεις σχετικά με διαταραχές που προήλθαν από πρόσφατα αναπάντεχα γεγονότα [2].

1.4 Ταξινόμηση των προβλέψεων

Οι τεχνικές προβλέψεων χωρίζονται σε δύο μεγάλες κατηγορίες: στις ποσοτικές (quantitative) και στις κριτικές (judgmental).

Οι ποσοτικές μέθοδοι πρόβλεψης αποσκοπούν στην μαθηματική μοντελοποίηση της διαδικασίας πρόβλεψης. Εφαρμόζονται όταν ισχύουν οι παρακάτω προϋποθέσεις:

- Υπάρχει διαθέσιμη πληροφορία για το παρελθόν
- Υπάρχει η δυνατότητα ποσοτικοποίησης της πληροφορίας αυτής με την μορφή αριθμητικών δεδομένων
- Θεωρείται ότι κάποια στοιχεία του προτύπου συμπεριφοράς θα διατηρηθούν και στο μέλλον

Οι ποσοτικές μέθοδοι χωρίζονται στις ακόλουθες δύο μεγάλες κατηγορίες:

- *Μέθοδοι Χρονοσειρών*
- *Αιτιοκρατικές Μέθοδοι*

1.5 Μέθοδοι Χρονοσειρών

Πρόκειται για το πιο διαδεδομένο είδος ποσοτικού μοντέλου πρόβλεψης. Η εφαρμογή του είναι δυνατή όταν υπάρχουν στοιχεία για την τιμή του υπό πρόβλεψη μεγέθους (ιστορικά δεδομένα) σε προηγούμενες και σταθερές χρονικές περιόδους (π.χ. ημέρες, μήνες, έτη κ.λ.π.). Βασίζεται στην υπόθεση ότι η μεταβολή της τιμής του μεγέθους ακολουθεί ένα συγκεκριμένο πρότυπο (λανθάνον πρότυπο) που επαναλαμβάνεται στον χρόνο και παραμένει σταθερό. Οι προβλέψεις πραγματοποιούνται με την αναγνώριση του ακολουθούμενου προτύπου και την προέκτασή του στο μέλλον.

Με τον όρο χρονοσειρά ή χρονολογική σειρά ονομάζουμε μία ακολουθία από παρατηρήσεις των δεδομένων του υπό εξέταση μεγέθους στη διάρκεια του χρόνου. Για παράδειγμα, μηνιαίες πωλήσεις αυτοκινήτων, ημερήσιες τιμές μετοχών, ετήσια κέρδη, ετήσια παραγωγή, ημερήσιες μέγιστες θερμοκρασίες, κ.α. Στην επιστήμη των προβλέψεων, ζητούμενο είναι να εκτιμηθεί πως αυτή η ακολουθία παρατηρήσεων θα συνεχιστεί στο μέλλον.

Τα βασικά ποιοτικά χαρακτηριστικά μιας χρονοσειράς είναι τα ακόλουθα:

- Στασιμότητα (Stationary)
- Τάση (Trend)
- Εποχικότητα (Seasonal)
- Κυκλικότητα (Cyclical)
- Τυχαιότητα (Randomness)

Οι συνηθέστερες ανωμαλίες που μπορούν να εμφανιστούν στα ιστορικά δεδομένα μιας χρονοσειράς είναι οι εξής:

- Αλλαγή Επιπέδου (Level Shift)
- Αλλαγή στην Βασική τάση (Change in Basic Trend)
- Ασυνήθιστες τιμές (Outlier)

Ειδική περίπτωση ασυνήθιστης τιμής είναι όταν η ασυνήθιστη τιμή είναι η τελευταία (ULO: Unusual Last Observation) αφού επηρεάζει σημαντικά τις παραγόμενες προβλέψεις – προεκτάσεις της χρονοσειράς.

Ελλιπής χαρακτηρίζεται μια χρονοσειρά που για κάποιο άγνωστο λόγο λείπουν κάποια δεδομένα της σε μη περιοδικά διαστήματα. Η έλλειψη των δεδομένων οφείλεται συνήθως σε μη καταγραφή και όχι σε μηδενική κίνηση του υπό εξέταση προϊόντος. Στη δεύτερη περίπτωση θα έπρεπε να είχε καταγραφεί τιμή ίση με μηδέν.

Σαν εποχικές χαρακτηρίζονται οι χρονοσειρές που εμφανίζουν έντονο το χαρακτηριστικό της εποχικότητας. Για την ανάλυση αυτών των χρονοσειρών απαιτείται να γίνει διαδικασία εποχικής αποσύνθεσης.

Είναι συνηθισμένο στην ανάλυση χρονοσειρών να γίνονται ομαδοποιήσεις – κατηγοριοποιήσεις τυποποιήσεις – ώστε να γίνεται ομαδική ανάλυση αυτών.

Η τυποποίηση γίνεται συνήθως βάσει των ακόλουθων:

- Χαρακτηριστικών
- Περιγραφής
- Τα βασικά εργαλεία ανάλυσης χρονοσειρών είναι:
- Γραφήματα
- Στατιστικοί Δείκτες (η μέση τιμή, η τυπική απόκλιση S, η MAD)
- Δείκτες σφάλματος (ME, MAE, MSE, MAPE, SMAPE, Average Ranking, Percentage Better)
- Μετασχηματισμοί

1.5.1 Εποχική Αποσύνθεση

Η εποχιακή αποσύνθεση είναι μια μαθηματική διαδικασία η οποία αποσκοπεί στην απομόνωση των εξής συνιστωσών μιας χρονοσειράς: τάση (Trend component), κύκλος (Cyclical component), εποχικότητα (Seasonal component) και τυχαιότητα (Irregular or random component).

Η μοντελοποίηση και πρόβλεψη της εποχικότητας - εποχική αποσύνθεση είναι μια μέθοδος χρονοσειρών η οποία όμως δεν αποσκοπεί στην προέκταση της χρονοσειράς στο μέλλον αλλά στην κατασκευή μιας φιλτραρισμένης χρονοσειράς, απαλλαγμένης από τις εποχιακές επιδράσεις.

Βασικά συμπεράσματα για την διαδικασία αυτή είναι τα ακόλουθα:

- Η αποεποχικοποίηση - μέρος της διαδικασίας αποσύνθεσης, αποσκοπεί στην απομόνωση της συνιστώσας εποχικότητας της χρονοσειράς με σκοπό την παραγωγή της αποεποχικοποιημένης χρονοσειράς, μιας φιλτραρισμένης από εποχιακές επιδράσεις χρονοσειρά.
- Η διαδικασία αποεποχικοποίησης είναι ιδιαίτερα σημαντική καθότι οι μέθοδοι προβλέψεων πρέπει να εφαρμόζονται σε μη εποχικές χρονοσειρές, να παράγονται οι σημειακές προβλέψεις και στην συνέχεια αυτές να επαναεποχικοποιούνται. Η πρακτική αυτή αυξάνει την ακρίβεια των παραγόμενων προβλέψεων.

Έχουν παρουσιαστεί διάφορες μέθοδοι αποσύνθεσης, οι επικρατέστερες παραμένουν όμως η μέθοδος Census και η κλασσική μέθοδος αποσύνθεσης.

1.5.2 Εκθετική εξομάλυνση

Αναλύονται τα ακόλουθα μοντέλα:

- *Εκθετική Εξομάλυνση Σταθερού Επιπέδου*
- *Εκθετική Εξομάλυνση Γραμμικής Τάσης*
- *Εκθετική Εξομάλυνση με σταθερή Γραμμική Τάση*
- *Εκθετική Εξομάλυνση μη Γραμμικής Τάσης*

Τα βασικά χαρακτηριστικά των μοντέλων αυτών είναι τα ακόλουθα:

- Δίνουν μεγάλο βάρος στις τελευταίες παρατηρήσεις, τονίζοντας έτσι την βραχυπρόθεσμη συμπεριφορά.
- Έχουν δώσει πολύ καλά αποτελέσματα στους διαγωνισμούς πρόβλεψης (M1, M2, M3)
- Είναι αρκετά απλά στην υλοποίηση και την εκτέλεση τους
- Αναγνωρίζουν επίπεδο, τάση και εποχιακούς δείκτες

1.5.3 Γραμμική Παλινδρόμηση

Η χρησιμοποίηση της γραμμικής παλινδρόμησης είναι συνήθης πρακτική στο λογισμικό προβλέψεων. Τα βασικά χαρακτηριστικά της διαδικασίας αυτής είναι τα ακόλουθα:

- Τονίζει την μακροπρόθεσμη συμπεριφορά.
- Δεν έχει δώσει πολύ καλά αποτελέσματα στους διαγωνισμούς πρόβλεψης (M1, M2, M3)
- Είναι αρκετά απλή στην υλοποίηση και την εκτέλεση της
- Χρησιμοποιείται κυρίως για πρόβλεψη σε μακρινούς ορίζοντες

Στην παρούσα εργασία επιλέχθηκε να χρησιμοποιηθεί η πολλαπλή γραμμική παλινδρόμηση ως χαρακτηριστική εφαρμογή προσέγγισης της μακροοικονομικής πρόβλεψης με γραμμικά μοντέλα.

1.5.4 Αυτοπαλινδρομικά Μοντέλα

Τα ολοκληρωμένα Αυτοπαλινδρομικά μοντέλα κινητού μέσου όρου (AutoRegressive-Integrated-Moving Average) είναι στοχαστικά μαθηματικά μοντέλα με τα οποία προσπαθούμε να περιγράψουμε τη διαχρονική εξέλιξη κάποιου φυσικού μεγέθους.

Η χρησιμοποίηση αυτοπαλινδρομικών μοντέλων είναι συνήθης πρακτική στο λογισμικό προβλέψεων. Τα βασικά χαρακτηριστικά των μοντέλων αυτών είναι τα ακόλουθα:

- Εφαρμόζονται μόνο σε στάσιμες χρονοσειρές.
- Δεν έχουν δώσει καλά αποτελέσματα στους διαγωνισμούς πρόβλεψης (M1, M2, M3)
- Είναι αρκετά δύσκολα στην υλοποίηση και την εκτέλεση τους
- Όταν προσδιοριστεί το κατάλληλο μοντέλο από την οικογένεια ARIMA για μια χρονοσειρά (κάτι ιδιαίτερα δύσκολο) τότε οι παραγόμενες προβλέψεις είναι πολύ αξιόπιστες

1.5.5 Τεχνητά Νευρωνικά Δίκτυα

Τα Τεχνητά Νευρωνικά Δίκτυα ανήκουν στο γνωστικό πεδίο της Υπολογιστικής Νοημοσύνης και η χρήση τους έγκειται σε επίλυση προβλημάτων Τεχνητής Νοημοσύνης, χρησιμοποιώντας αριθμητικά μοντέλα.

Τεχνητό Νευρωνικό Δίκτυο είναι μια αρχιτεκτονική δομή αποτελούμενη από έναν αριθμό τεχνητών νευρώνων. Κάθε νευρώνας χαρακτηρίζεται από εισόδους και εξόδους και υλοποιεί έναν απλό υπολογισμό. Κάθε σύνδεση μεταξύ δύο μονάδων

χαρακτηρίζεται από μία τιμή βάρους. Οι τιμές των βαρών των συνδέσεων αποτελούν την αποθηκευμένη γνώση στο δίκτυο και καθορίζουν πλήρως την λειτουργικότητά του. Η έξοδος κάθε νευρώνα καθορίζεται από τον τύπο του νευρώνα, τη διασύνδεση με τους υπόλοιπους και πιθανώς κάποιες εξωτερικές εισόδους.

Τα βασικά χαρακτηριστικά των μοντέλων αυτών είναι τα ακόλουθα:

- Δεν έχουν δώσει καλά αποτελέσματα στους διαγωνισμούς πρόβλεψης (M3)
- Είναι πολύ δύσκολα στην υλοποίηση και την εκτέλεση τους
- Μπορούν να προσαρμοστούν σε οποιασδήποτε μορφής ιστορικά δεδομένα (γραμμικά, μη γραμμικά, πολυδιάστατα)

Στην παρούσα εργασία μελετήθηκε μια συγκεκριμένη εφαρμογή τεχνητών νευρωνικών δικτύων, τα δίκτυα υψηλής τάξης, για την προσέγγιση της μακροοικονομικής πρόβλεψης με τη χρήση μη γραμμικών μοντέλων.

1.6 Τα κίνητρα μακροοικονομικής πρόβλεψης της οικονομίας μιας χώρας

Τα βασικά αίτια για την πρόβλεψη των τάσεων της οικονομίας οποιασδήποτε χώρας είναι η διαβεβαίωση ότι η κυβέρνηση (και όσο είναι δυνατό η αστική κοινωνία) ακολουθεί πορεία που συντελεί στη διατήρηση της μακροπρόθεσμης ανάπτυξης της οικονομίας και τη μείωση της φτώχειας, χωρίς να είναι υποκείμενη σε τρέχουσες πιέσεις/ ανάγκες. Έχουν καταγραφεί τρεις τύποι τέτοιων πιέσεων:

- Εγχώριοι πολιτικοί παράγοντες: Αντιπροσωπευτικό παράδειγμα πολιτικών παραγόντων που απαντάται εξίσου σε αναπτυγμένες και αναπτυσσόμενες χώρες είναι η προεκλογική πίεση για την αύξηση των κυβερνητικών δαπανών και την περικοπή των φόρων ή των επιτοκίων. Άλλα τέτοια παραδείγματα είναι οι φορολογικές μεταρρυθμίσεις, μεταβολές των δασμών εισαγόμενων προϊόντων και αναπροσαρμογή των μισθών.
- Εξωτερικοί πολιτικοί παράγοντες: Για πολλές χώρες οι κυριότεροι εξωτερικοί πολιτικοί παράγοντες είναι η συμμόρφωση στα πρότυπα και τις απαιτήσεις των διεθνών οικονομικών ιδρυμάτων/ θεσμών/ δοτών. Ανεξαιρέτως όλες οι κυβερνήσεις είναι υποκείμενες και υπόλογες στις αρχές των διεθνών οικονομικών συμφωνιών.

- Εγχώριες ή εξωτερικές διαταραχές (“shocks”) της οικονομίας: Με τον όρο διαταραχή ονομάζουμε κάθε αναπάντεχο γεγονός που αποπροσανατολίζει την υπάρχουσα πολιτική τάξη και πιθανότατα διαστρέφει μόνιμα τη βέλτιστη πορεία της οικονομίας. Τέτοια εγχώρια παραδείγματα αποτελούν ακραία περιβαλλοντικά φαινόμενα, ενώ εξωτερικά παραδείγματα απαντώνται σε περιπτώσεις απρόσμενης μεταβολής των τιμών εισαγόμενων εμπορευμάτων/αγαθών.

Αναμφίβολα, οι προαναφερθέντες παράγοντες πυροδοτούν αλληλοσχετιζόμενες διαταραχές που επιβάλλουν ζήτηση για ταχεία αντίδραση από εγχώριους και εξωτερικούς φορείς. Οι εν λόγω αντιδράσεις – ανταποκρίσεις πιθανόν να είναι μακροχρόνιες, ανεπαρκείς ή εσφαλμένα σχεδιασμένες, εάν η κυβέρνηση αποτύχει να προβλέψει τον αντίκτυπο των εν λόγω διαταραχών μέσω της μακροοικονομικής πρόβλεψης. Προφανώς, σκοπός της κυβέρνησης δεν είναι η εξάλειψη των παραπάνω παραγόντων (πράγμα ανέφικτο για οποιαδήποτε χώρα), αλλά η πρόβλεψή τους ώστε στην ουσία να μειωθούν αν όχι εξαλειφθούν οι αρνητικές επιδράσεις.

Εξίσου κρίσιμης σημασίας είναι η κατανόηση ότι η μακροοικονομική πρόβλεψη των θετικών προεκτάσεων δύναται να επηρεάσει ευμενώς την οικονομία και εν τέλει τα πολιτικά δρώμενα (ειδικά σε χώρες χαμηλού εισοδήματος που διανύουν περιόδους φιλελευθεροποίησης και προσαρμογής). Συγκεκριμένα μπορεί να συμβεί:

- Ενίσχυση των προοπτικών για οικονομική σταθεροποίηση θεμελιώνοντας ρεαλιστικούς στόχους σχετικά με τις ανταποκρίσεις αποθέματος, την επαγρύπνηση για προβλέψιμες ελλείψεις πόρων και τη συστηματική προστασία έναντι αρνητικών διαταραχών. Συνεπώς προκύπτει:
- Αύξηση και ενδυνάμωση της κυβερνητικής δέσμευσης στη μακροπρόθεσμη προσαρμογή της χώρας σε πορεία ανάπτυξης της οικονομίας και μείωσης της φτώχειας, αποφεύγοντας παράλληλα ελλείψεις πόρων που οδηγούν σε επίφοβη “κρίση διαχείρισης”.
- Ενθάρρυνση εξωτερικών και εγχώριων δοτών/ πιστωτών/ φορολογουμένων/ επενδυτών στη χρηματοδότηση εγκαταστάσεων και έργων εξαιτίας της αξιόπιστης οικονομικής πολιτικής.

- Εξασφάλιση ότι η κυβέρνηση στοχεύει σε περαιτέρω ανάπτυξη όπως μείωση της φτώχειας και της ανεργίας, προστασία του περιβάλλοντος, αύξηση των επενδύσεων και προληπτικά μέτρα αντιστάθμισης μελλοντικών διαταραχών.

Λαμβάνοντας όλα τα παραπάνω υπόψη καθίσταται φανερό ότι η πρόβλεψη της οικονομίας μιας χώρας είναι αποφασιστικής σημασίας για την ανάπτυξη και την ευημερία σε όλους τους τομείς [4].

1.7 Η πρόσφατη οικονομική κατάσταση στις ΗΠΑ

Στη σύγχρονη ιστορία των ΗΠΑ παρουσιάστηκε το φαινόμενο της οικονομικής ύφεσης κατά τις περιόδους 1969-70, 1973-75 και 1981-82. Το 1990 η οικονομία των ΗΠΑ και πάλι διήνυσε περίοδο οικονομικής ύφεσης. Οι περίοδοι που μεσολάβησαν ήταν καιροί οικονομικής ανάπτυξης και ευημερίας. Η οικονομική ύφεση, καθώς και περίοδοι οικονομικής επέκτασης εμφανίζονται από την αρχή της οικονομικής ιστορίας που έχει καταγραφεί αν και τα αίτια, το μέγεθος και η διάρκεια τους διαφοροποιούνται σημαντικά.

Η οικονομική ύφεση του 1981-1982 ήταν το επακόλουθο του υψηλού πληθωρισμού κατά τέλη της δεκαετίας του '70 που δημιουργήθηκε κυρίως λόγω των αυξήσεων στις τιμές του πετρελαίου που επέβαλε ο Οργανισμός Χωρών που Εξάγουν Πετρέλαιο κατά τη διετία 1978-1980 (από 15\$ σε 32\$ ανά βαρέλι). Αφού το ποσοστό του πληθωρισμού αποτελεί συνιστώσα των επιτοκίων (interest rate), το ελάχιστο επιτόκιο δανεισμού (prime rate ο τόκος με τον οποίο δανείζουν οι τράπεζες βραχυπρόθεσμα στους καλύτερους πελάτες τους) αυξήθηκε από 6% το 1977 σε 21,5% το 1980. Ενώ πολλές επιχειρήσεις ήταν σε θέση να περάσουν στους πελάτες τους τα υψηλότερα κόστη, ορισμένοι μικροί εμπορικοί οίκοι απέτυχαν. Επενδυτές με πόρους σε ασφάλειες πάγιου εισοδήματος (όπως προθεσμιακά ομόλογα >10 ετών) έχασαν μεγάλα κεφάλαια, ιδρύματα χρεώθηκαν, οι καταθέσεις και τα δάνεια ήταν απαγορευτικά εξαιτίας των υψηλών επιτοκίων και πόροι χάθηκαν. Αυτές οι συνθήκες οδήγησαν την Ομοσπονδία Κρατικού Αποθέματος (Federal Reserve) να λάβει αυστηρά μέτρα έναντι του πληθωρισμού και να συντονίσει το απόθεμα χρηματικού πλούτου ώστε νέα “χρήματα” αντικατοπτρίζοντας την πραγματική οικονομική ανάπτυξη και όχι την πολιτική εκμετάλλευση της οικονομίας, να εισαχθούν στην αγορά.

Το αποτέλεσμα ήταν η μείωση του πληθωρισμού από 12,7% το 1980 σε 2,9% το 1983. Η εν λόγω μείωση επηρέασε αρνητικά επιχειρήσεις με μακροπρόθεσμα συμβόλαια εξόδων (όπως εργασίας) που προϋπέθεταν υψηλά επίπεδα μελλοντικού πληθωρισμού και βραχυπρόθεσμα συμβόλαια για έσοδα που προσαρμόζονταν άμεσα στις μεταβολές των τιμών. Πολλές μεγάλες επιχειρήσεις ευδοκίμησαν την περίοδο του υψηλού πληθωρισμού, αλλά οι λειτουργίες τους ήταν προβληματικές αφού το κέρδος ήταν πιο σημαντικό από την παραγωγικότητα. Τα υψηλά επιτόκια επίσης εμπόδιζαν τις επενδύσεις σε νέους εξοπλισμούς και εγκαταστάσεις, έτσι πολλές εταιρίες είχαν απαρχαιωμένο εξοπλισμό ειδικά σε σύγκριση με τους ξένους ανταγωνιστές. Καινούρια έργα και κατασκευές καθυστερούσαν υπερβολικά στις αρχές της δεκαετίας του '80.

Η κυβέρνηση των ΗΠΑ το 1981 επισημείωσε το Νόμο Οικονομικής Ανάκαμψης ώστε να παρακινήσει τις επενδύσεις σε εξοπλισμό και εγκαταστάσεις επιτρέποντας ταχεία απόσβεση των μακροπρόθεσμων κεφαλαίων παραγωγής (συμπεριλαμβανομένου κτιρίων) και φορολογικές πιστώσεις σε ορισμένες επενδύσεις. Αυτές οι μεταβολές αναζωογόνησαν την οικονομία των ΗΠΑ που διήνυσε μια μακρά περίοδο άνθισης ώστε η υπόλοιπη δεκαετία του 1980 να αποτελέσει περίοδο συνεχούς επέκτασης. Σημειώθηκε ραγδαία αύξηση στην κατασκευή οικοδομών εξαιτίας της ευνοϊκής φορολογίας, της πραγματικής οικονομικής ανάπτυξης, των χαμηλών τιμών ενέργειας και του επιθετικού δανεισμού από οικονομικά ιδρύματα (που ήταν σε θέση έπειτα να επενδύσουν περισσότερους πόρους σε εμπορική ακίνητη περιουσία). Ωστόσο κατά τη δεκαετία του 1980, αρκετές περιοχές αντιμετώπισαν ουσιαστική εξασθένηση, για παράδειγμα περιοχές που η οικονομία τους βασιζόταν στο πετρέλαιο (Ντάλλας, Χιούστον), καθώς η τιμή του πετρελαίου μειώθηκε περαιτέρω. Αυτές οι αγορές αναγνωρίζονται από μια αιφνίδια αύξηση της ανεργίας καθώς οι εργοδότες προσαρμόζονταν στις νέες οικονομικές συνθήκες. Οι συνέπειες αυτών των αλλαγών στους επαγγελματίες στην περιοχή εξαρτώταν από τη φύση της υποκείμενης ζήτησης για αγαθά/υπηρεσίες. Για παράδειγμα, για τους αρχιτέκτονες μειώθηκε η ζήτηση σε όλες τις κατηγορίες ακινήτων λόγω του κορεσμού που προήλθε εξαιτίας της συνεχούς οικοδόμησης την περίοδο της άνθισης.

Το 1986 το Κογκρέσο των ΗΠΑ μεταβάλλει για άλλη μια φορά τη φορολογία – υπό το πρόσχημα της αποκαλούμενης φορολογικής μεταρρυθμίσεις – απομακρύνοντας την προνομιακή μεταχείριση της ακίνητης περιουσίας καταργώντας

τους μειωμένους φόρους για τα κεφαλαιοποιήσιμα κέρδη και παρατείνοντας την απόσβεση των ακινήτων. Για παράδειγμα για την απόσβεση ενός εμπορικού ακινήτου το 1981 χρειάζονταν 15 έτη, ενώ το 1986 31,5 έτη. Το δεύτερο μισό της δεκαετίας του 1980 ήταν φανερό ότι πολλές από τις εμπορικές τράπεζες και ασφαλιστικές εταιρίες ήταν σε οικονομικό κίνδυνο με τα χαρτοφυλάκια δανείων ασφαλισμένα σε ακίνητα με μειωμένες τιμές κοινής αποδοχής (συνέπεια της υπερ-οικοδόμησης, μειωμένης ζήτησης και της Φορολογικής Μεταρρύθμισης του 1986). Το δυνατό Αμερικανικό δολάριο παρήγαγε ευρύ εμπορικό έλλειμμα και η σχετικά σταθερή οικονομία, ο χαμηλός πληθωρισμός και τα υψηλά πραγματικά επιτόκια συντέλεσαν σε μαζική εισαγωγή κεφαλαίου από άλλες χώρες. Εν μέρει, αυτά τα υψηλά πραγματικά επιτόκια προκλήθηκαν από την ανικανότητα του Κογκρέσου να ισοσταθμίσει τον προϋπολογισμό, έτσι οι οφειλές χρησιμοποιούνταν κυρίως για τη χρηματοδότηση του Ομοσπονδιακού ελλείμματος. Η οικονομική ύφεση και τα υψηλά πραγματικά επιτόκια κατά το τέλος της δεκαετίας του 1980 και τις αρχές του 1990 ήταν οι κύριοι παράγοντες αναδόμησης της οικονομίας των ΗΠΑ. Οι μικρές εταιρίες προσαρμόστηκαν στο μεγαλύτερο ανταγωνισμό από τις αμερικάνικες και τις ξένες επιχειρήσεις αυξάνοντας την αποδοτικότητά τους, κάνοντας περικοπές, κυρίως μειώνοντας τον αριθμό των υπαλλήλων.

Στο νέο αυτό περιβάλλον, τα δάνεια για την ανάπτυξη και επένδυση σε ακίνητα περικόπηκαν από τις τράπεζες και τα κατασκευαστικά έργα μειώθηκαν ραγδαία. Οι συνέπειες στις τεχνικές εταιρίες διαφοροποιούνταν στις ΗΠΑ, ανάλογα με τη φύση και τις συνθήκες των τοπικών οικονομιών και την έκταση της υπερ-προμήθειας δομικών υλικών που είχαν συσσωρευτεί τα προηγούμενα έτη. Μέχρι το 1991 οι βορειοανατολικές πολιτείες (κυρίως η Νέα Αγγλία) επλήγησαν περισσότερο, κυρίως λόγω των μειώσεων στις κατασκευές και την επιβράδυνση των πωλήσεων υψηλής τεχνολογίας και αμυντικού εξοπλισμού. Η Δυτική Ακτή, επίσης σημαντικός κατασκευαστής αγαθών/υπηρεσιών υψηλής τεχνολογίας και άμυνας επηρεάστηκε ομοίως.

Οι κατηφείς οικονομικές συνθήκες διαδραμάτισαν πρωταγωνιστικό ρόλο στην εκλογή Δημοκρατικού Προέδρου και Κογκρέσου. Το 1993, το Κογκρέσο επισημείωσε προϋπολογισμό που αύξανε τους φόρους στους πολίτες και τις επιχειρήσεις με εισοδήματα της ανώτερης βαθμίδας, μειώνοντας περαιτέρω την προνομιακή φορολογία του ακίνητου πλούτου. Μέχρι το τέλος του 1993, τα συνεχή χαμηλά επιτόκια και η αναβολή της ζήτησης προκάλεσαν ανάκαμψη απόδειξη της οποίας

ήταν τα χαμηλότερα επίπεδα ανεργίας σε ορισμένους τομείς και ορισμένες περιοχές της χώρας [5].

1.8 Σκοπός της εργασίας

Έπειτα από την εισαγωγή στην έννοια της πρόβλεψης, την περιγραφή και ταξινόμηση των μεθόδων μακροοικονομικής πρόβλεψης, τη σύντομη αναφορά στις κυριότερες μεθόδους χρονοσειρών εκ των οποίων θα χρησιμοποιήσουμε τη γραμμική παλινδρόμηση και τα τεχνητά νευρωνικά δίκτυα, κρίνεται αναγκαίο να εξηγήσουμε το σκοπό της παρούσας εργασίας.

Όπως προαναφέραμε το αντικείμενο της μακροοικονομικής πρόβλεψης είναι ευρύ, ενώ η κάθε μέθοδος χαρακτηρίζεται από συγκεκριμένα πλεονεκτήματα και μειονεκτήματα. Τα οφέλη από την επιτυχή πρόβλεψη της εθνικής οικονομίας είναι ποικίλα για την πολιτική και κοινωνική ευημερία κάθε χώρας. Στόχος της εργασίας είναι να εξετάσουμε κατά πόσο είναι εφικτή η μακροοικονομική πρόβλεψη με συγκεκριμένα δεδομένα χρονοσειρών, εφαρμόζοντας παροδοσιακές τεχνικές, όπως είναι τα γραμμικά μοντέλα, αλλά και καινοτόμες μεθόδους, όπως είναι τα νευρωνικά δίκτυα υψηλής τάξης. Για τη μελέτη μας χρησιμοποιήσαμε δεδομένα χρονοσειρών 64 ετών της Αμερικανικής οικονομίας για τη μεταβολή της πρόβλεψης των μακροοικονομικών μεταβλητών: εθνικό προϊόν (GDP), ακαθάριστες εξαγωγές και πληθωρισμός.

Τα νευρωνικά δίκτυα υψηλής τάξης και συγκεκριμένα τα δίκτυα sigma-pi που κατασκευάσαμε έχει αποδειχθεί ότι διαθέτουν ισχυρές ικανότητες πρόβλεψης και γενίκευσης. Αναπτύσσουμε κώδικα σε περιβάλλον MATLAB και μελετάμε κατά πόσο τα δεδομένα τα οποία είχαμε στη διάθεσή μας, είναι κατάλληλα για επιτυχή πρόβλεψη. Αξιοσημείωτο είναι ότι το γραμμικό μοντέλο ταυτίζεται απόλυτα με το νευρωνικό δίκτυο πρώτης τάξης. Τα αποτελέσματα τόσο του γραμμικού μοντέλου που μελετάτε, όσο και του υψηλού νευρωνικού δικτύου που αναπτύσσεται εξαρτώνται από την ποιότητα των δεδομένων. Η μακροοικονομική πρόβλεψη αποτελεί πρόκληση στον τομέα των προβλέψεων εξαιτίας της μη γραμμικότητας που χαρακτηρίζει τα δεδομένα, αλλά και των αναπάντεχων διαταραχών που οφείλονται σε εξωγενείς παράγοντες.

1.9 Περιεχόμενα της εργασίας

Τα παρόν κεφάλαιο αναλύει την έννοια της μακροοικονομικής πρόβλεψης, τις μεθόδους και τα κίνητρά της. Παρουσιάζεται επίσης μια σύντομη αναδρομή στην ιστορία της Αμερικανικής οικονομίας, με σκοπό να σημειωθούν οι χρονολογίες στις οποίες εμφανίστηκαν απρόσμενες διακυμάνσεις (“shocks”), οι οποίες είναι αναπόφευκτο να επηρεάσουν τα αποτελέσματα της πρόβλεψης.

Ακολουθεί η παρουσίαση μιας παραδοσιακής τεχνικής μακροοικονομικής πρόβλεψης, της πολλαπλής γραμμικής παλινδρόμησης. Αναπτύσσεται το μοντέλο και υπολογίζονται οι συντελεστές του με τη μέθοδο των ελαχίστων τετραγώνων. Τέλος, μέσω διαγραμμάτων αξιολογείται η απόδοση της μεθόδου.

Το τρίτο κεφάλαιο αποτελεί την εισαγωγή στην έννοια των τεχνητών νευρωνικών δικτύων. Παρουσιάζονται τα κυριότερα είδη, ανάμεσά τους και τα υψηλής τάξης νευρωνικά δίκτυα. Ακολουθεί μεθοδολογία κατασκευής δικτύων, όπου αναλύονται οι σημαντικότερες συνιστώσες των δικτύων, καθώς και ζητήματα εμπειρικής φύσης. Τέλος, συγκρίνονται τα εν λόγω μοντέλα με τα μοντέλα γραμμικής παλινδρόμησης.

Στο τέταρτο κεφάλαιο αναλύεται και παρουσιάζεται το μοντέλο νευρωνικών δικτύων υψηλής τάξης που κατασκευάστηκε, σχολιάζεται ο κώδικας, αλλά και η λογική του μοντέλου. Η αξιολόγηση πραγματοποιείται μέσω γραφικών παραστάσεων, ενώ ακολουθούν τα συμπεράσματα της παρούσας μελέτης.

ΚΕΦΑΛΑΙΟ 2^ο

Προσέγγιση της μακροοικονομικής πρόβλεψης με γραμμικά μοντέλα

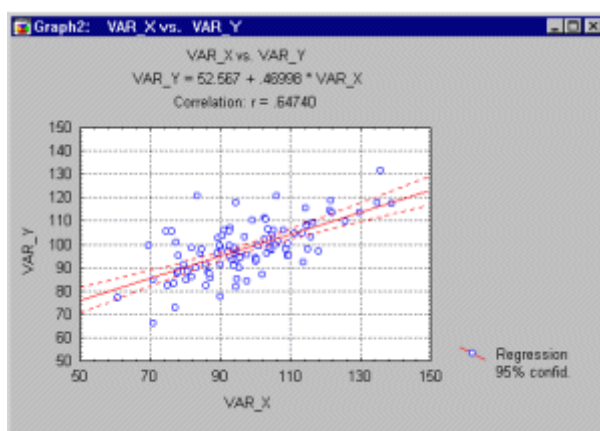
“Ορισμένοι απεχθάνονται ακόμη και την ίδια την ονομασία της στατιστικής επιστήμης, αλλά εγώ τη βρίσκω γεμάτη ομορφιά και ενδιαφέρον. Όταν μεταχειριζόμαστε τις ανώτερες μεθόδους της με σύνεση και τις ερμηνεύουμε με περίσκεψη, η δύναμή τους να διαχειρίζονται πολύπλοκα φαινόμενα είναι εκπληκτική”

Sir Francis Galton, ο επιστήμονας που ανακάλυψε την έννοια της συσχέτισης.

2.1 Η Μέθοδος της πολλαπλής γραμμικής παλινδρόμησης

Ο κύριος σκοπός της πολλαπλής παλινδρόμησης (ο όρος χρησιμοποιήθηκε για πρώτη φορά από τον Pearson το 1908) είναι η μάθηση σχετικά με τη σχέση μεταξύ πολλών ανεξάρτητων μεταβλητών ή μεταβλητών πρόβλεψης (predictors) και μιας εξαρτημένης μεταβλητής ή μεταβλητής κριτηρίου.

Στις κοινωνικές και φυσικές επιστήμες οι διαδικασίες πολλαπλής παλινδρόμησης χρησιμοποιούνται ευρέως για ερευνητικούς σκοπούς. Γενικά, η πολλαπλή παλινδρόμηση επιτρέπει στον ερευνητή να διατυπώσει (και πιθανότατα να απαντήσει) στη γενική ερώτηση “ποιος είναι ο καλύτερος τρόπος πρόβλεψης”. Το γενικό υπολογιστικό πρόβλημα που επιλύεται στην ανάλυση πολλαπλής παλινδρόμησης είναι η προσέγγιση μιας ευθείας γραμμής που περνάει από ένα πλήθος σημείων.



Σχήμα 2.1: Γράφημα γραμμικής παλινδρόμησης.

Στην απλούστερη περίπτωση που έχουμε μία εξαρτημένη και μία ανεξάρτητη μεταβλητή μπορούμε να κατασκευάσουμε το παραπάνω γράφημα. Στο γράφημα έχουμε την ανεξάρτητη μεταβλητή X και την εξαρτημένη μεταβλητή Y . Ο στόχος της διαδικασίας γραμμικής παλινδρόμησης είναι η προσέγγιση μιας ευθείας μέσω των σημείων. Συγκεκριμένα το πρόγραμμα θα υπολογίσει μια γραμμή ώστε να ελαχιστοποιούνται οι τετραγωνικές αποκλίσεις των σημείων των παρατηρήσεων από αυτή τη γραμμή. Συνεπώς η γενική διαδικασία συνήθως αναφέρεται ως εκτίμηση των ελαχίστων τετραγώνων.

Η εξίσωση της ευθείας γραμμής στο δισδιάστατο χώρο είναι η: $Y=a+b*X$ που δηλώνει ότι η μεταβλητή Y μπορεί να εκφρασθεί σε σχέση με μια σταθερά a και την κλίση b επί τη μεταβλητή X . Η σταθερά επίσης αναφέρεται ως *intercept* (συντελεστής ανάσχεσης) και η κλίση ως *συντελεστής παλινδρόμησης* ή *συντελεστής B* .

Στην περίπτωση που έχουμε πολλές μεταβλητές, όπου υπάρχουν περισσότερες από μία ανεξάρτητες μεταβλητές, η γραμμή της παλινδρόμησης δεν μπορεί να απεικονισθεί στο δισδιάστατο χώρο, αλλά μπορεί να υπολογισθεί με ευκολία. Η εξίσωση της ευθείας θα είναι τότε:

$$Y = a + b_1*X_1 + b_2*X_2 + \dots + b_p*X_p$$

Αξιοσημείωτο είναι ότι στη παραπάνω εξίσωση οι συντελεστές παλινδρόμησης (ή συντελεστές B) εκφράζουν την ανεξάρτητη συμβολή της κάθε ανεξάρτητης μεταβλητής, για την πρόβλεψη της εξαρτημένης μεταβλητής. Με άλλα λόγια μπορούμε να πούμε ότι για παράδειγμα η μεταβλητή X_1 είναι συσχετισμένη με τη μεταβλητή Y , αφού ελέγξαμε όλες τις άλλες ανεξάρτητες μεταβλητές. Αυτός ο τύπος συσχέτισης αναφέρεται επίσης ως *μερική συσχέτιση* (ο όρος χρησιμοποιήθηκε για πρώτη φορά από τον Yule το 1907).

Η γραμμή παλινδρόμησης εκφράζει την καλύτερη πρόβλεψη της εξαρτημένης μεταβλητής (Y), δεδομένης της εξαρτημένης μεταβλητής (X). Ωστόσο, τα πραγματικά προβλήματα δεν είναι ποτέ τέλεια προβλέψιμα και υπάρχει συνήθως ουσιαστική απόκλιση των σημείων από τη γραμμή παλινδρόμησης (όπως φαίνεται και στο γράφημα). Η απόκλιση ενός συγκεκριμένου σημείου από τη γραμμή παλινδρόμησης (την προβλεπόμενη τιμή του) ονομάζεται τιμή *κατάλοιπου* (*residual value*).

Όσο μικρότερη είναι η διακύμανση των τιμών καταλοίπων γύρω από τη γραμμή παλινδρόμησης σε σχέση με την συνολική διακύμανση, τόσο καλύτερη είναι η πρόβλεψη. Παραδείγματος χάριν, εάν δεν υπάρχει σχέση μεταξύ των μεταβλητών X και Y τότε η αναλογία της διακύμανσης καταλοίπων της μεταβλητής Y προς την

αρχική διακύμανση είναι 1,0. Εάν οι X και Y είναι τέλεια συσχετισμένες τότε δεν υπάρχει διακύμανση καταλοίπων και η αναλογία της διακύμανσης θα είναι 0,0. Στις περισσότερες περιπτώσεις η αναλογία κυμαίνεται μεταξύ αυτών των ακραίων τιμών, δηλαδή μεταξύ 0,0 και 1,0. Ο αριθμός 1,0 μείον αυτή την αναλογία ονομάζεται *R-square* ή *συντελεστής προσδιορισμού (coefficient of determination)*. Αυτή η τιμή μεταφράζεται ως εξής: Εάν έχουμε ένα *R-square* που ισούται με 0,4 τότε γνωρίζουμε ότι η διακύμανση της τιμής του Y γύρω από τη γραμμή παλινδρόμησης είναι $1 - 0,4$ φορές την αρχική διακύμανση. Με άλλα λόγια έχουμε εξηγήσει το 40% της αρχικής διακύμανσης και απομένει το 60% της διακύμανσης καταλοίπων. Στην ιδανική περίπτωση θα θέλαμε να εξηγήσουμε το μεγαλύτερο μέρος αν όχι όλη την αρχική διακύμανση. Η τιμή *R-square* είναι μια ένδειξη του κατά πόσο καλά το μοντέλο προσεγγίζει τα δεδομένα (π.χ., η τιμή *R-square* κοντά στο 1,0 δηλώνει ότι έχουμε εκτιμήσει σχεδόν όλη τη διακύμανση των μεταβλητών που ορίζονται στο μοντέλο).

Συνήθως, ο βαθμός στον οποίο δύο ή περισσότερες μεταβλητές πρόβλεψης (predictors – ανεξάρτητες ή μεταβλητές X) σχετίζονται με την εξαρτημένη μεταβλητή (Y) εκφράζεται με το συντελεστή συσχέτισης R , που είναι η τετραγωνική ρίζα του *R-square*. Στην πολλαπλή παλινδρόμηση το R κυμαίνεται μεταξύ 0 και 1. Για να ερμηνεύσουμε την κατεύθυνση της σχέσης μεταξύ των μεταβλητών προσέχουμε τα σύμβολα (συν ή μείον) των συντελεστών παλινδρόμησης ή συντελεστών B . Εάν ο συντελεστής B είναι θετικός, τότε η σχέση αυτής της μεταβλητής με την εξαρτημένη μεταβλητή είναι θετική. Εάν ο συντελεστής B είναι αρνητικός τότε η σχέση είναι αρνητική. Φυσικά, εάν ο συντελεστής B είναι μηδενικός τότε δεν υφίσταται σχέση μεταξύ των μεταβλητών.

2.2 Υποθέσεις και περιορισμοί της μεθόδου

Κατ' αρχήν, όπως είναι φανερό από την ονομασία πολλαπλή γραμμική παλινδρόμηση, υποτίθεται ότι η σχέση μεταξύ των μεταβλητών είναι γραμμική. Στην πράξη, αυτή η υπόθεση δεν επαληθεύεται σχεδόν ποτέ. Ευτυχώς οι διαδικασίες πολλαπλής παλινδρόμησης δεν επηρεάζονται σημαντικά από μικρές αποκλίσεις από την υπόθεση γραμμικότητας. Ωστόσο, ο κανόνας είναι να κοιτάμε πάντα σε διμετάβλητο γράφημα τις μεταβλητές που μας ενδιαφέρουν. Εάν απεικονίζεται καμπύλωση στις σχέσεις, συνιστάται είτε μετασχηματισμός των μεταβλητών, είτε σαφής παραδοχή μη γραμμικών συνιστωσών.

Στη μέθοδο της πολλαπλής παλινδρόμησης πραγματοποιείται η υπόθεση ότι τα κατάλοιπα (προβλεπόμενη τιμή μείον την τιμή στις παρατηρήσεις) κατανέμονται κανονικά. Και πάλι, αν και οι περισσότερες δοκιμές (ειδικά τα F -test) είναι αρκετά εύρωστες σε σχέση με παρεκκλίσεις από την υπόθεση, ενδείκνυται πάντα, πριν τη διεξαγωγή τελικού συμπεράσματος, να λάβουμε υπόψη τις κατανομές των κυριότερων μεταβλητών που μας ενδιαφέρουν. Για αυτό το σκοπό χρησιμοποιούνται συνήθως ιστογράμματα και γραφήματα κανονικής κατανομής.

Ο κύριος περιορισμός όλων των τεχνικών παλινδρόμησης είναι ότι μπορούμε με ευκολία να εξακριβώσουμε τις σχέσεις, αλλά ποτέ δεν μπορούμε να είμαστε σίγουροι σχετικά με τον υποκείμενο αιτιολογικό μηχανισμό. Στην πράξη, κατά τη διερεύνηση της συσχέτισης των μεταβλητών, εναλλακτικές αιτιολογικές εξηγήσεις, συχνά δε λαμβάνονται υπόψη, με αποτέλεσμα μια εξωτερική μεταβλητή που παραλείψαμε να μελετήσουμε, να ευθύνεται για το αποτέλεσμα.

Σχετικά με το πλήθος των μεταβλητών οι περισσότεροι ερευνητές προτείνουν ότι πρέπει να έχουμε τουλάχιστον 10 ή 20 φορές περισσότερες σε αριθμό παρατηρήσεις συγκριτικά με το πλήθος των υπό μελέτη μεταβλητών. Διαφορετικά η εκτίμηση της γραμμής παλινδρόμησης θα είναι ιδιαίτερος ασταθής και θα είναι εξαιρετικά απίθανο να παραχθεί η ίδια ακριβώς γραμμή εάν η διαδικασία επαναληφθεί.

Το συνηθέστερο πρόβλημα που αντιμετωπίζεται κατά την ανάλυση της συσχέτισης των μεταβλητών είναι όταν οι μεταβλητές πρόβλεψης είναι πλεοναστικές και καθίσταται αναγκαίο για τη διαδικασία της πολλαπλής παλινδρόμησης να καθορισθεί ποια από τις ανεξάρτητες μεταβλητές αποτελεί την καταλληλότερη μεταβλητή πρόβλεψης. Όταν εμπλέκονται πολλές μεταβλητές συνήθως δεν είναι εμφανές ότι υπάρχει πρόβλημα και μπορεί να παρουσιαστεί μόνο αφού πολλές μεταβλητές έχουν ήδη εισαχθεί στην εξίσωση παλινδρόμησης. Ωστόσο, όταν το παρουσιαστεί πρόβλημα, σημαίνει ότι τουλάχιστον μία από τις μεταβλητές πρόβλεψης είναι πλεοναστική. Υπάρχουν πολλοί στατιστικοί δείκτες του εν λόγω τύπου πλεονασμού (ανοχή, semi - partial R κτλ.) όπως και τρόποι διόρθωσης (πχ. *ridge regression*).

Η προσαρμογή πολυωνύμων υψηλής τάξης μιας ανεξάρτητης μεταβλητής με έναν αριθμητικό μέσο διάφορο του μηδενός μπορεί να προκαλέσει σημαντικά προβλήματα πολυσυγγραμμικότητας. Συγκεκριμένα, τα πολυώνυμα θα είναι υψηλά συσχετισμένα εξαιτίας του μέσο της προηγούμενης ανεξάρτητης μεταβλητής. Το

πρόβλημα μπορεί να αποδειχθεί ιδιαίτερα σημαντικό όταν εμπλέκονται μεγάλα νούμερα και αν δε ληφθούν μέτρα, μπορεί να προκαλέσει εσφαλμένα αποτελέσματα. Η λύση είναι η “επικέντρωση” της ανεξάρτητης μεταβλητής (ορισμένες φορές η διαδικασία αναφέρεται ως “κεντρικά πολυνύμα”), συγκεκριμένα η αφαίρεση του μέσου και έπειτα ακολουθεί ο υπολογισμός των πολυνύμων.

Μολονότι οι περισσότερες υποθέσεις της πολλαπλής παλινδρόμησης δε δύνανται να δοκιμασθούν διεξοδικά, καταφανείς παραβιάσεις των περιορισμών μπορούν να αναγνωρισθούν και πρέπει να διαχειρίζονται κατάλληλα. Συγκεκριμένα *outliers* (δηλαδή ακραίες καταστάσεις) ενδέχεται να προκαλέσουν σημαντική απόκλιση των αποτελεσμάτων έλκοντας ή απωθώντας τη γραμμή παλινδρόμησης προς μια συγκεκριμένη κατεύθυνση, οδηγώντας σε συντελεστές παλινδρόμησης με διαφορετικές τιμές. Συχνά, η αφαίρεση μίας μόνο ακραίας περίπτωσης αποφέρει τελείως διαφορετικά αποτελέσματα [6].

2.3 Μαθηματικό Υπόβαθρο της μεθόδου

Η πολλαπλή παλινδρόμηση είναι το στατιστικό εργαλείο για την κατανόηση της σχέσης μεταξύ δύο ή περισσότερων μεταβλητών. Περιλαμβάνει τη μεταβλητή που θα εξηγηθεί – την εξαρτημένη μεταβλητή – και επιπλέον εξηγηματικές μεταβλητές που θεωρείται ότι παράγουν ή σχετίζονται με μεταβολές της εξαρτημένης μεταβλητής.

Θεωρούμε τις δύο ανεξάρτητες μεταβλητές x και y και την εξαρτημένη μεταβλητή z στη γραμμική σχέση:

$$z = a + bx + cy$$

Για το σύνολο δεδομένων $(x_1, y_1, z_1), (x_2, y_2, z_2), \dots, (x_n, y_n, z_n)$, όπου $n \geq 3$, η βέλτιστη καμπύλη $f(x)$ έχει το εξής σφάλμα ελάχιστων τετραγώνων:

$$\Pi = \sum_{i=1}^n [z_i - f(x_i, y_i)]^2 = \sum_{i=1}^n [z_i - (a + bx_i + cy_i)]^2 = \min.$$

Σημειώνουμε ότι τα a, b, c είναι άγνωστοι συντελεστές, ενώ τα x_1, y_1, z_1 δίδονται. Για εξασφαλίσουμε το σφάλμα ελαχίστων τετραγώνων οι άγνωστοι πρέπει οι πρώτες παράγωγοι των συντελεστών a, b και c να είναι μηδενικές.

$$\begin{cases} \frac{\partial \Pi}{\partial a} = 2 \sum_{i=1}^n [z_i - (a + bx_i + cy_i)] = 0 \\ \frac{\partial \Pi}{\partial b} = 2 \sum_{i=1}^n x_i [z_i - (a + bx_i + cy_i)] = 0 \\ \frac{\partial \Pi}{\partial c} = 2 \sum_{i=1}^n y_i [z_i - (a + bx_i + cy_i)] = 0 \end{cases}$$

Επεκτείνοντας τις παραπάνω εξισώσεις έχουμε:

$$\begin{cases} \sum_{i=1}^n z_i = a \sum_{i=1}^n 1 + b \sum_{i=1}^n x_i + c \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i z_i = a \sum_{i=1}^n x_i + b \sum_{i=1}^n x_i^2 + c \sum_{i=1}^n x_i y_i \\ \sum_{i=1}^n y_i z_i = a \sum_{i=1}^n y_i + b \sum_{i=1}^n x_i y_i + c \sum_{i=1}^n y_i^2 \end{cases}$$

Συνεπώς οι άγνωστοι συντελεστές a , b και c υπολογίζονται από την επίλυση των παραπάνω γραμμικών εξισώσεων.

2.4 Διερεύνηση σε μορφή πινάκων

Όταν η μεταβλητή απόκρισης ή εξαρτημένη Y επηρεάζεται από περισσότερες από μία ερμηνευτικές ή ανεξάρτητες μεταβλητές X_1, X_2, \dots, X_{p-1} έστω, το γραμμικό μοντέλο που περιγράφει την εξάρτηση της Y από τις X_1, X_2, \dots, X_{p-1} θα έχει τη μορφή:

$$Y = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_{p-1} X_{p-1} + \varepsilon$$

για κάποιες παραμέτρους b_0, b_1, \dots, b_{p-1} .

Για να διερευνηθεί η σχέση μεταξύ της Y και των X_1, X_2, \dots, X_{p-1} λαμβάνεται δείγμα μεγέθους n και για κάθε παρατήρηση του δείγματος

καταγράφονται οι τιμές των συγκεκριμένων μεταβλητών. Π.χ. για την i -παρατήρηση του δείγματος καταγράφονται οι τιμές $(Y_i, X_{i1}, \dots, X_{i,p-1})$, $i = 1, 2, \dots, n$.

Εφαρμόζουμε λοιπόν το μοντέλο:

$$Y_i = b_0 + b_1 X_{i1} + b_2 X_{i2} + \dots + b_{p-1} X_{i,p-1} + \varepsilon_i, \quad i = 1, 2, \dots, n$$

όπου τα σφάλματα $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ θεωρούνται ανεξάρτητες τ.μ. από την $N(0,1)$, ενώ οι ερμηνευτικές μεταβλητές X_1, X_2, \dots, X_{p-1} δε θεωρούνται τυχαίες.

Το παραπάνω μοντέλο γράφεται με τη μορφή πινάκων στην απλούστερη μορφή:

$$\mathbf{Y} = \mathbf{X}\mathbf{b} + \boldsymbol{\varepsilon}$$

Όπου:

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & X_{11} & \dots & X_{1,p-1} \\ 1 & X_{21} & & X_{2,p-1} \\ \vdots & & & \\ 1 & X_{n1} & \dots & X_{n,p-1} \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_{p-1} \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}.$$

Ο πίνακας \mathbf{X} καλείται και πίνακας σχεδιασμού. Στο πολλαπλό μοντέλο θεωρούμε ότι τα σημεία δε βρίσκονται κοντά σε μία ευθεία αλλά κοντά σε ένα (υπερ)επίπεδο στις p διαστάσεις.

Το τυχαίο διάνυσμα $\boldsymbol{\varepsilon}$ αποτελείται από n ανεξάρτητες $N(0, \sigma^2)$ τυχαίες μεταβλητές και επομένως θα έχει από κοινού σ.π.π. $N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$, δηλαδή ακολουθεί μία πολυδιάστατη κανονική κατανομή όπου \mathbf{I}_n είναι ο μοναδιαίος πίνακας διάστασης n . Ως συνέπεια το τυχαίο διάνυσμα $\mathbf{Y} = [Y_1, Y_2, \dots, Y_n]^T$ θα ακολουθεί και αυτό πολυδιάστατη κανονική $N(\mathbf{X}\mathbf{b}, \sigma^2 \mathbf{I}_n)$, με συνάρτηση πιθανοφάνειας :

$$L(\mathbf{b}, \sigma^2) = f(y_1, y_2, \dots, y_n; \mathbf{b}, \sigma^2) = \frac{1}{(2\pi)^{n/2} (\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\mathbf{b})^T (\mathbf{y} - \mathbf{X}\mathbf{b})}$$

Η οποία μεγιστοποιείται ως προς \mathbf{b} (ώστε να πάρουμε την ε.μ.π. του \mathbf{b}) όταν ελαχιστοποιείται το:

$$(\mathbf{Y} - \mathbf{X}\mathbf{b})^T (\mathbf{Y} - \mathbf{X}\mathbf{b}) = \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} = \sum_{i=1}^n \varepsilon_i^2.$$

Το παραπάνω γράφεται:

$$(\mathbf{Y}^T - \mathbf{b}^T \mathbf{X}^T)(\mathbf{Y} - \mathbf{Xb}) = \mathbf{Y}^T \mathbf{Y} - \mathbf{Y}^T \mathbf{Xb} - \mathbf{b}^T \mathbf{X}^T \mathbf{Y} + \mathbf{b}^T \mathbf{X}^T \mathbf{Xb}$$

Παραγωγίζοντας ως προς $\mathbf{b} = (\frac{df}{db} = (\frac{\partial f}{\partial b_0}, \dots, \frac{\partial f}{\partial b_{p-1}}))$ είναι ίσο (από ιδιότητες

πινάκων) με:

$$\frac{d}{db} (\mathbf{Y} - \mathbf{Xb})^T (\mathbf{Y} - \mathbf{Xb}) = -2\mathbf{X}^T \mathbf{Y} + 2\mathbf{X}^T \mathbf{Xb}$$

Η παραπάνω παράγωγος (δηλ. το διάνυσμα των μερικών παραγώγων) είναι ίση με $\mathbf{0}$ όταν:

$$\mathbf{X}^T \mathbf{Xb} = \mathbf{X}^T \mathbf{Y}.$$

Το παραπάνω σύστημα των p εξισώσεων με p αγνώστους (το διάνυσμα \mathbf{b}) έχει μοναδική λύση όταν υπάρχει ο αντίστροφος του $\mathbf{X}^T \mathbf{X}$ και σε αυτή την περίπτωση οι εκτιμήτριες μέγιστης πιθανοφάνειας για το $\mathbf{b} = [b_0, b_1, \dots, b_{p-1}]^T$ θα είναι:

$$\hat{\mathbf{b}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$$

(αποδ. ότι πράγματι αποτελεί ελάχιστο της $(\mathbf{Y} - \mathbf{Xb})^T (\mathbf{Y} - \mathbf{Xb})$). Επίσης, ε.μ.π. του σ^2 είναι η

$$\hat{\sigma}^2 = \frac{1}{n} (\mathbf{Y} - \mathbf{X}\hat{\mathbf{b}})^T (\mathbf{Y} - \mathbf{X}\hat{\mathbf{b}}) = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{b}_0 - \hat{b}_1 X_{i1} - \dots - \hat{b}_{p-1} X_{i,p-1})^2.$$

Καλούμε *προβλέψεις* των Y_i (\mathbf{Y} predicted) ή *προσαρμοσμένες* (πάνω στο εκτιμημένο επίπεδο παλινδρόμησης) *τιμές* των Y_i τις εκτιμήσεις των $E(Y_i) = b_0 + b_1 X_{i1} + \dots + b_{p-1} X_{i,p-1}$:

$$\hat{Y}_i = \hat{b}_0 + \hat{b}_1 X_{i1} + \dots + \hat{b}_{p-1} X_{i,p-1}, i = 1, 2, \dots, n$$

ή συνοπτικά,

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\mathbf{b}} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} = \mathbf{PY} \quad (\mathbf{P} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T).$$

Οι διαφορές των προσαρμοσμένων \hat{Y}_i καλούνται *κατάλοιπα* (residuals) ή *εκτιμημένα σφάλματα* και συμβολίζονται με

$$\hat{\varepsilon}_i = Y_i - \hat{Y}_i, \quad i=1, 2, \dots, n$$

ή συνοπτικά,

$$\hat{\varepsilon} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{X}\hat{\mathbf{b}} = \mathbf{Y} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} = (\mathbf{I}_n - \mathbf{P})\mathbf{Y}$$

όπου ο \mathbf{P} είναι γνωστός ως πίνακας ορθής προβολής (projection ή hat matrix) [7].

2.5 Πρόβλεψη της Αμερικανικής Οικονομίας με πολλαπλή γραμμική παλινδρόμηση

Για την πρόβλεψη της Αμερικανικής οικονομίας με πολλαπλή γραμμική παλινδρόμηση θα χρησιμοποιηθεί το γραμμικό μοντέλο $\Delta\mathbf{x} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} + \mathbf{C}\mathbf{z} + \varepsilon$ καθώς και δεδομένα από την αμερικανική οικονομία. Οι χρονοσειρές των δεδομένων προήλθαν από το <http://www.economicswebinstitute.org> και είναι από το έτος 1932 έως 1995. Οι μεταβλητές κατάστασης που θα μπουν στο μοντέλο θα είναι:

πραγματικό προϊόν (real gdp), καθαρές εξαγωγές (real net exports), και πληθωρισμός (inflation rate, cpi). Οι μεταβλητές ελέγχου είναι κυβερνητικές δαπάνες (real government exp.), exchange rate, short interest rate (I_Short). Η εξωγενής: velocity of M1.

Στο γραμμικό μοντέλο που μελετάμε \mathbf{x} είναι το διάνυσμα των μεταβλητών κατάστασης $-(3 \times 1)$ - που περιέχει τις μεταβλητές: πραγματικό προϊόν, καθαρές εξαγωγές και πληθωρισμός. $\Delta\mathbf{x}$ είναι το διάνυσμα $-(3 \times 1)$ - με τις μεταβολές στις μεταβλητές κατάστασης $\mathbf{x}_{t+1} - \mathbf{x}_t$. \mathbf{u} είναι το διάνυσμα με τις μεταβλητές ελέγχου: δαπάνες (real government exp.), συνάλλαγμα (exchange rate), επιτόκια (short interest rate ή I_Short) και \mathbf{z} είναι το διάνυσμα με τις εξωγενείς μεταβλητές (M1). Τα $\mathbf{A}, \mathbf{B}, \mathbf{C}$ είναι πίνακες $3 \times 3, 3 \times 3, 3 \times 1$ με τους **σταθερούς στο χρόνο** συντελεστές αυτών των μεταβλητών στο μοντέλο. Το ε είναι το 3×1 διάνυσμα των θορύβων. Οι θόρυβοι είναι λευκοί (κατανέμονται κανονικά με μέσο 0 και τυπική απόκλιση 1 και είναι ανεξάρτητοι μεταξύ τους).

Το αντίστοιχο οικονομετρικό μοντέλο που πρέπει να εκτιμηθεί είναι :

$$\Delta x_1 = a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + b_{11}u_1 + b_{12}u_2 + b_{13}u_3 + c_1z_1 + \varepsilon_1$$

$$\Delta x_2 = a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + b_{21}u_1 + b_{22}u_2 + b_{23}u_3 + c_2z_1 + \varepsilon_2$$

$$\Delta x_3 = a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + b_{31}u_1 + b_{32}u_2 + b_{33}u_3 + c_3z_1 + \varepsilon_3$$

Για την εκτίμηση των παραπάνω εξισώσεων θα χρησιμοποιηθεί η μέθοδος των ελαχίστων τετραγώνων. Η εξίσωση που υπολογίζει τους συντελεστές όπως διερευνήθηκε και στην παραπάνω παράγραφο δίδεται από τη σχέση:

$$\hat{b} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$$

Όλοι οι υπολογισμοί πραγματοποιήθηκαν σε περιβάλλον matlab και προέκυψαν οι ακόλουθοι πίνακες:

$$d_1 = \begin{bmatrix} -0.0066 \\ -1.1294 \\ 0.8426 \\ 0.0195 \\ 0.3262 \\ -3.6601 \\ 6.1239 \end{bmatrix} \quad d_2 = \begin{bmatrix} 0.0038 \\ 0.3195 \\ 0.0676 \\ 0.0008 \\ -0.1440 \\ -0.4935 \\ 2.5837 \end{bmatrix} \quad d_3 = \begin{bmatrix} 0.0008 \\ -0.0169 \\ 0.2185 \\ -0.0078 \\ 0.0043 \\ -0.3663 \\ 0.3672 \end{bmatrix}$$

Οι παραπάνω πίνακες d_1 , d_2 , d_3 , που αντιστοιχούν στην πρώτη (Δx_1), στη δεύτερη (Δx_2) και στην τρίτη (Δx_3) παλινδρόμηση αντίστοιχα περιέχουν τα στοιχεία των ζητούμενων πινάκων A,B,C. Συγκεκριμένα έχουμε:

$$d_1 = \begin{bmatrix} a_{11} \\ a_{12} \\ a_{13} \\ b_{11} \\ b_{12} \\ b_{13} \\ c_1 \end{bmatrix} \quad d_2 = \begin{bmatrix} a_{21} \\ a_{22} \\ a_{23} \\ b_{21} \\ b_{22} \\ b_{23} \\ c_2 \end{bmatrix} \quad d_3 = \begin{bmatrix} a_{31} \\ a_{32} \\ a_{33} \\ b_{31} \\ b_{32} \\ b_{33} \\ c_3 \end{bmatrix}$$

Άρα οι ζητούμενοι πίνακες που δίδουν τους συντελεστές του γραμμικού μοντέλου που μελετάμε είναι:

$$A = \begin{bmatrix} -0.0066 & -1.1294 & 0.8426 \\ 0.0038 & 0.3195 & 0.0676 \\ 0.0008 & -0.0169 & 0.2185 \end{bmatrix}$$

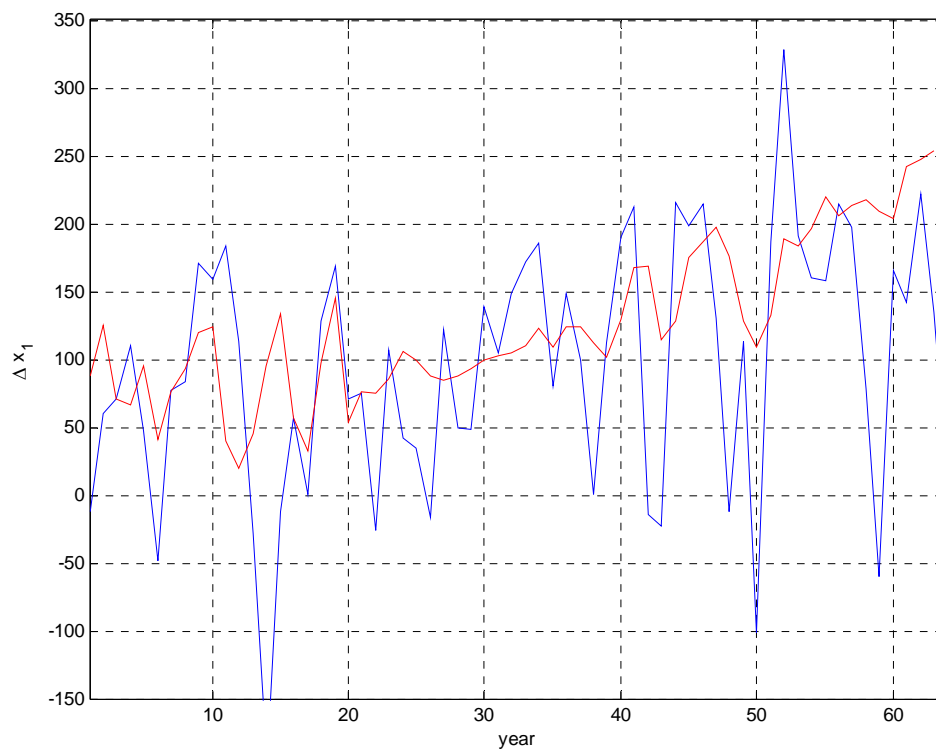
$$B = \begin{bmatrix} 0.0195 & 0.3262 & -3.6601 \\ 0.0008 & -0.1440 & -0.4935 \\ -0.0078 & 0.0043 & -0.3663 \end{bmatrix}$$

$$C = \begin{bmatrix} 6.1239 \\ 2.5837 \\ 0.3672 \end{bmatrix}$$

2.6. Αξιολόγηση του μοντέλου

Το μοντέλο πολλαπλής γραμμικής παλινδρόμησης έπειτα από τον υπολογισμό των συντελεστών κατέδειξε τα εξής αποτελέσματα όπως αυτά παρουσιάζονται στα παρακάτω διαγράμματα.

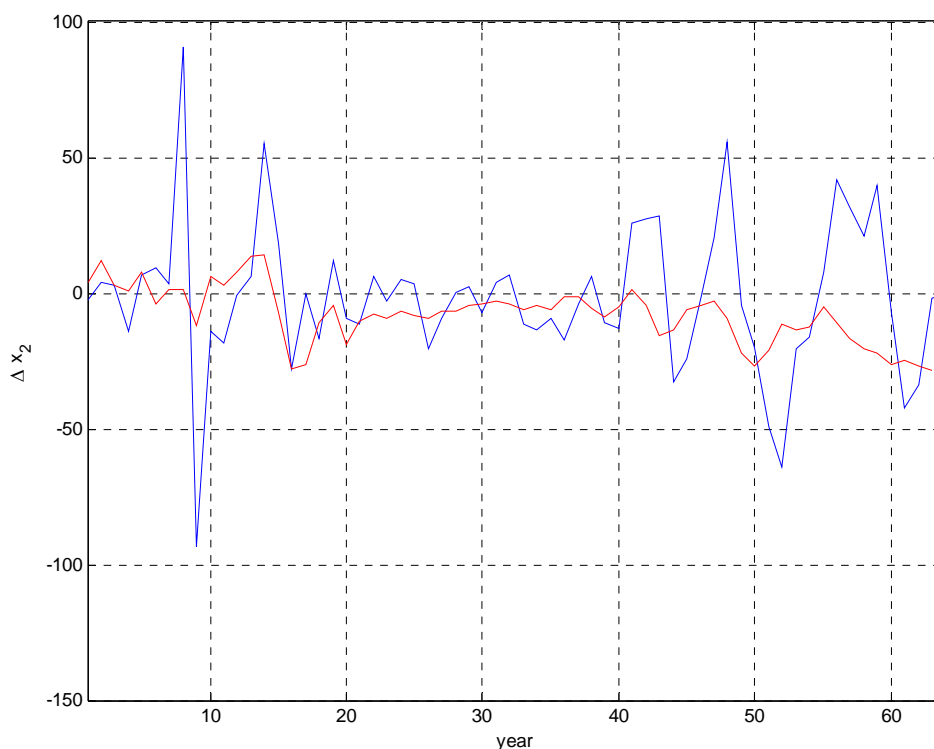
Πιο συγκεκριμένα για την πρόβλεψη του πραγματικού προϊόντος έχουμε:



Διάγραμμα 2.1: Πρόβλεψη του GDP με το μοντέλο παλινδρόμησης.

Στο παραπάνω διάγραμμα, με μπλε χρώμα απεικονίζεται η πραγματική μεταβολή του πραγματικού προϊόντος από το έτος 1932 έως 1995 και με κόκκινο η τιμή της μεταβολής που πρόβλεψε το μοντέλο πολλαπλής γραμμικής παλινδρόμησης. Είναι φανερό πως η πρόβλεψη δεν είναι ικανοποιητική.

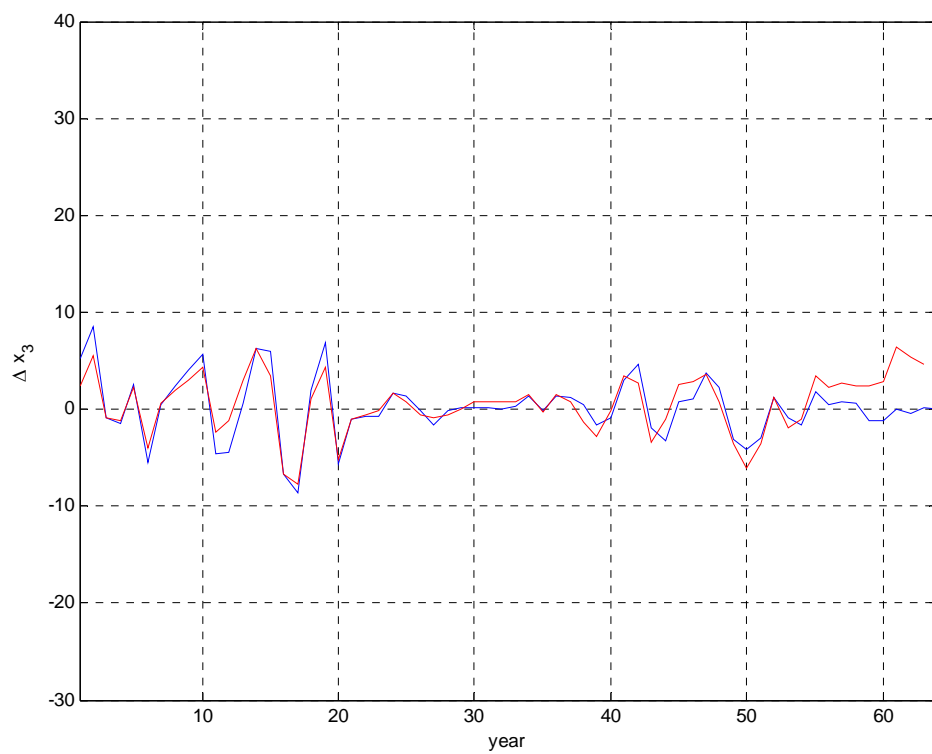
Για την πρόβλεψη των καθαρών εξαγωγών έχουμε:



Διάγραμμα 2.2: Πρόβλεψη των real net exports με το μοντέλο παλινδρόμησης.

Στο παραπάνω διάγραμμα γίνεται προφανές πως η πραγματική μεταβολή των καθαρών εξαγωγών (μπλε χρώμα) και η προβλεπόμενη μεταβολή από το μοντέλο (κόκκινο χρώμα), συμπίπτουν σε ελάχιστα σημεία και συνεπώς η πρόβλεψη είναι ανεπαρκής.

Τέλος το ακόλουθο διάγραμμα που απεικονίζει την πραγματική μεταβολή του πληθωρισμού (μπλε χρώμα) και την προβλεπόμενη τιμή της (κόκκινο χρώμα) αποτελεί την καλύτερη πρόβλεψη που πραγματοποίησε το μοντέλο πολλαπλής γραμμικής παλινδρόμησης για τις τρεις εξαρτημένες μεταβλητές. Ωστόσο στα τελευταία έτη η απόκλιση είναι σημαντική.



Διάγραμμα 2.3: Πρόβλεψη του πληθωρισμού με το μοντέλο παλινδρόμησης

ΚΕΦΑΛΑΙΟ 3^ο**Τεχνητά Νευρωνικά Δίκτυα**

“Pluralitas non est ponenda sine necessitate”, William of Occam

(Είναι ματαιοδοξία να προσπαθείς να πετύχεις με τα πολλά, αυτό που μπορείς με τα λίγα)

Αρχή γνωστή ως *Occam's razor*, στο πνεύμα της οποίας στηρίζεται η μάθηση των νευρωνικών δικτύων

3.1 Πρόβλεψη με Νευρωνικά Δίκτυα

Στον τομέα της χρηματοοικονομικής πρόβλεψης έχουν παρατηρηθεί σύγχρονες εξελίξεις σχετικά με τα μοντέλα πρόβλεψης. Παραδοσιακά, δημοφιλείς τεχνικές πρόβλεψης περιλαμβάνουν ανάλυση μέσω παλινδρόμησης (regression analysis), ανάλυση χρονοσειρών, μέθοδοι εξομάλυνσης (smoothing methods) καθώς και μέθοδοι του κινούμενου μέσου. Ωστόσο, όλες έχουν το ίδιο μειονέκτημα καθώς απαιτούν υποθέσεις σχετικά με τη μορφή της κατανομής του δείγματος. Τα μοντέλα παλινδρόμησης, παραδείγματος χάριν υποθέτουν ότι το υποκείμενο δείγμα ακολουθεί την κανονική κατανομή.

Τα Τεχνητά Νευρωνικά Δίκτυα ανήκουν στην ίδια οικογένεια στατιστικών τεχνικών με τα ευέλικτα μη γραμμικά μοντέλα παλινδρόμησης, τα διακριτά μοντέλα (discriminant models), τα μοντέλα απαλοιφής δεδομένων (data reduction models) και τα μη γραμμικά δυναμικά συστήματα. Αποτελούν εκπαιδεύσιμα αναλυτικά εργαλεία που επιχειρούν να μιμηθούν τα πρότυπα επεξεργασίας πληροφοριών του εγκεφάλου. Καθώς δεν απαιτούν απαραίτητα υποθέσεις για την κατανομή του δείγματος, οικονομολόγοι και μαθηματικοί χρησιμοποιούν τα Τεχνητά Νευρωνικά Δίκτυα για την ανάλυση δεδομένων ολοένα και περισσότερο. Όχι μόνο δεν απαιτούν υποθέσεις για την κατανομή του υποκείμενου δείγματος, αλλά επιπλέον συνιστούν ισχυρά εργαλεία πρόβλεψης συγκεντρώνοντας τις πλέον πρόσφατες εξελίξεις στην έρευνα τεχνητής νοημοσύνης.

Σύγχρονες ερευνητικές δραστηριότητες σχετικά με τα τεχνητά νευρωνικά δίκτυα (ΤΝΔ) κατέδειξαν ότι τα ΤΝΔ έχουν ισχυρές ιδιότητες ταξινόμησης και

αναγνώρισης προτύπων. Εμπνευσμένα από τα βιολογικά συστήματα, ιδιαίτερα από την έρευνα του ανθρωπίνου εγκεφάλου, τα ΤΝΔ είναι σε θέση να μαθαίνουν και να γενικεύουν από την εμπειρία. Επί του παρόντος τα ΤΝΔ χρησιμοποιούνται σε ένα ευρύ φάσμα εφαρμογών σε διάφορους τομείς όπως οι επιχειρήσεις, η βιομηχανία και η επιστήμη.

Από τις κυριότερες περιοχές εφαρμογών των ΤΝΔ αποτελεί ο τομέας της πρόβλεψης, καθώς προσφέρουν ένα ισχυρό εναλλακτικό εργαλείο για τους μελετητές και τους επαγγελματίες των προβλέψεων. Αρκετά μοναδικά χαρακτηριστικά των ΤΝΔ τα καθιστούν πολύτιμα και ελκυστικά για προγνωστικές εργασίες. Πιο συγκεκριμένα τα ΤΝΔ έχουν τα ακόλουθα χαρακτηριστικά:

- Πρώτον, αντίθετα από τις παραδοσιακές μεθόδους, τα ΤΝΔ συνιστούν αυτοπροσαρμόσιμες αντλούμενες από τα δεδομένα μεθόδους κατά το ότι υπάρχουν ελάχιστες εκ των προτέρων υποθέσεις σχετικά με τα υπό μελέτη μοντέλα. Μαθαίνουν από τα παραδείγματα και συλλαμβάνουν περίπλοκες λειτουργικές σχέσεις μεταξύ των δεδομένων, ακόμη και εάν οι υποκείμενες σχέσεις είναι άγνωστες ή δύσκολες στην περιγραφή. Συνεπώς τα ΤΝΔ είναι κατάλληλα για προβλήματα των οποίων οι λύσεις απαιτούν γνώσεις δύσκολα καθορισμένες για τις οποίες τα δεδομένα ή οι παρατηρήσεις είναι ανεπαρκή. Υπό αυτή την έννοια συγκαταλέγονται στις πλέον μη γραμμικές, μη παραμετρικές με πολλαπλές μεταβλητές στατιστικές μεθόδους. Η προσέγγιση μοντελοποίησης που παρέχουν με την ικανότητα μάθησης από την εμπειρία αποδεικνύεται ιδιαίτερα χρήσιμη για πολλά πρακτικά προβλήματα, καθώς συχνά είναι ευκολότερο να έχουμε δεδομένα αντί να έχουμε καλές θεωρητικές υποθέσεις για τους υποκείμενους νόμους στους οποίους το σύστημα υπακούει και από τους οποίους δημιουργούνται τα δεδομένα. Το πρόβλημα με την προσέγγιση μοντελοποίησης οδηγούμενη από τα δεδομένα είναι ότι οι υποκείμενοι κανόνες δεν είναι πάντα ευδιάκριτοι και οι παρατηρήσεις είναι συχνά λανθάνουσες από θόρυβο. Παρόλα αυτά παρέχουν έναν πρακτικό και σε ορισμένες περιπτώσεις το μόνο εφικτό τρόπο επίλυσης πραγματικών προβλημάτων.
- Δεύτερον τα ΤΝΔ έχουν τη δυνατότητα γενίκευσης. Έπειτα από τη διαδικασία μάθησης των δεδομένων, τα ΤΝΔ έχουν συχνά τη δυνατότητα εξαγωγής τεκμηριωμένων συμπερασμάτων των μη ορατών μερών του δείγματος ακόμη

και εάν το δείγμα περιέχει πληροφορίες με θόρυβο. Καθώς η πρόβλεψη πραγματοποιείται μέσω προγνωστικών της μελλοντικής συμπεριφοράς (το μη ορατό μέρος) από παραδείγματα παρελθοντικής συμπεριφοράς, αποτελεί ιδανικό τομέα εφαρμογής των νευρωνικών δικτύων.

- Τρίτον, τα ΤΝΔ είναι προσεγγιστές γενικών συναρτήσεων. Έχει παρατηρηθεί ότι ένα δίκτυο είναι δυνατόν να προσεγγίσει κάθε συνεχή συνάρτηση με οποιαδήποτε επιθυμητή ακρίβεια. Τα ΤΝΔ μπορούν να χειρισθούν πιο γενικές και ευέλικτες μορφές συναρτήσεων από τις παραδοσιακές στατιστικές μεθόδους. Κάθε μοντέλο πρόβλεψης υποθέτει ότι υπάρχει μια υποκείμενη (γνωστή ή άγνωστη) σχέση μεταξύ των εισόδων (οι παρελθοντικές τιμές των χρονοσειρών και/ ή των υπόλοιπων μεταβλητών) και των εξόδων (οι μελλοντικές τιμές). Συχνά, τα παραδοσιακά στατιστικά μοντέλα πρόβλεψης έχουν περιορισμούς στην εκτίμηση της υποκείμενης συνάρτησης εξαιτίας της πολυπλοκότητας των πραγματικών συστημάτων. Τα ΤΝΔ αποτελούν την καλύτερη εναλλακτική μέθοδο όσον αφορά την αναγνώριση τέτοιων συναρτήσεων.
- Τέλος τα ΤΝΔ είναι μη γραμμικά. Η πρόβλεψη αποτελούσε ανέκαθεν περιοχή εφαρμογής της γραμμικής στατιστικής. Οι παραδοσιακές προσεγγίσεις στην πρόγνωση χρονοσειρών, όπως οι μέθοδοι Box-Jenkins και ARIMA, προϋποθέτουν ότι οι υπό μελέτη χρονοσειρές προέρχονταν από γραμμικές διαδικασίες. Τα γραμμικά μοντέλα έχουν το πλεονέκτημα της κατανόησης και ανάλυσής τους με κάθε λεπτομέρεια καθώς και τη δυνατότητα εύκολης εξήγησης και υλοποίησης. Ωστόσο, μπορεί να αποδειχθούν εντελώς ακατάλληλα εάν ο υποκείμενος μηχανισμός είναι μη γραμμικός. Είναι παράλογη η υπόθεση *a priori* (εκ των προτέρων) ότι οι συγκεκριμένες χρονοσειρές παράγονται από γραμμικές διαδικασίες. Στην πραγματικότητα, τα σύγχρονα συστήματα και ιδιαίτερα τα μακροοικονομικά είναι τις περισσότερες φορές μη γραμμικά. Ωστόσο αυτά τα μη γραμμικά μοντέλα αντιμετωπίζουν κάποιους περιορισμούς υπό την έννοια ότι είναι αναγκαίο να υποτεθεί μια ρητή σχέση για τα υπό μελέτη δεδομένα, όταν ελάχιστη γνώση του υποκείμενου νόμου είναι διαθέσιμη. Στην πραγματικότητα, η διαμόρφωση ενός μη γραμμικού μοντέλου για ένα συγκεκριμένο σύνολο δεδομένων αποτελεί ένα πολύ δύσκολο εγχείρημα, εφόσον υπάρχουν πολλά πιθανά μη

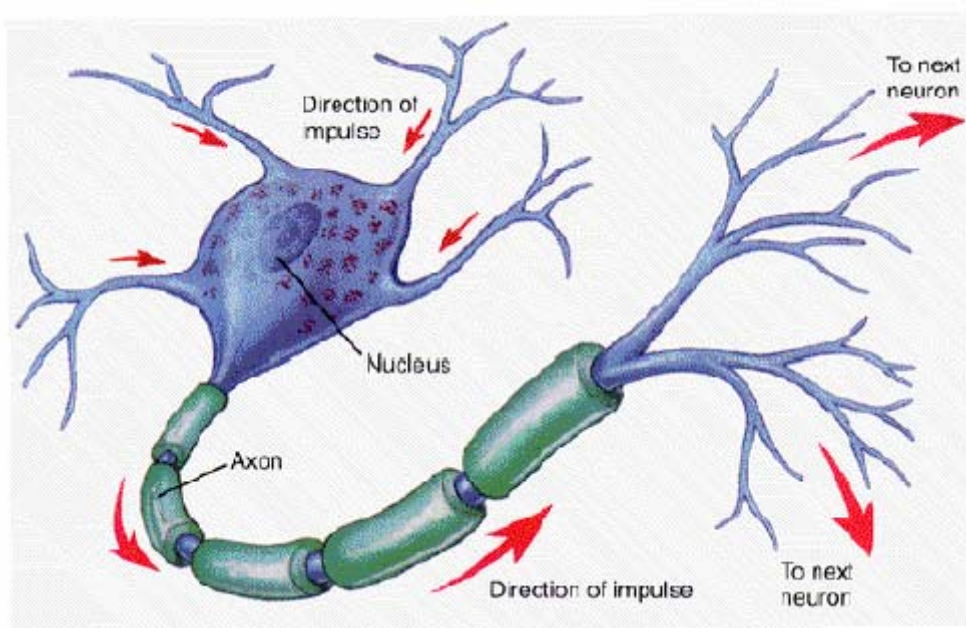
γραμμικά πρότυπα και ένα προκαθορισμένο μη γραμμικό μοντέλο μπορεί να αποδειχθεί όχι αρκετά γενικό για την σύλληψη όλων των σημαντικών χαρακτηριστικών. Τα τεχνητά νευρωνικά δίκτυα, που αποτελούν μη γραμμικές οδηγούμενες από τα δεδομένα προσεγγίσεις σε αντίθεση με τα παραπάνω μη γραμμικά μοντέλα, είναι ικανά για την εκτέλεση μη γραμμικής μοντελοποίησης χωρίς να απαιτούν a priori γνώση σχετικά με τη σχέση των μεταβλητών εισόδου και εξόδου. Συνεπώς συνιστούν ένα πιο γενικό και ευέλικτο εργαλείο μοντελοποίησης για την πρόβλεψη.

Η ιδέα της χρήσης ΤΝΔ στην πρόβλεψη δεν είναι καινούρια. Η πρώτη εφαρμογή χρονολογείται το 1964 όταν ο Hu χρησιμοποίησε το Widrow προσαρμόσιμο γραμμικό δίκτυο για την πρόβλεψη του καιρού. Λόγω της τότε απουσίας αλγόριθμου εκπαίδευσης για τα γενικά δίκτυα πολλαπλών επιπέδων, η έρευνα ήταν ιδιαιτέρως περιορισμένη. Μόλις το 1986 όταν ο αλγόριθμος backpropagation πρωτοεμφανίστηκε, παρουσιάστηκε αξιοσημείωτη ανάπτυξη στη χρήση ΤΝΔ για πρόβλεψη. Ο Werbos (1974), (1988) μορφοποιεί για πρώτη φορά τη μέθοδο backpropagation και διαπιστώνει ότι τα ΤΝΔ που εκπαιδεύονται με αυτή τη μέθοδο υπερτερούν των παραδοσιακών στατιστικών μεθόδων όπως η παλινδρόμηση και η Box-Jenkins. Οι Lapedes και Farber (1987) διεξάγουν μια μελέτη προσομοίωσης και καταλήγουν στο συμπέρασμα ότι τα ANNs μπορούν να χρησιμοποιηθούν για τη μοντελοποίηση και πρόβλεψη χρονοσειρών. Το σύνολο των ερευνητικών μελετών σχετικά με τα νευρωνικά δίκτυα είναι πολυπληθές και δυναμικά αυξανόμενο [8],[9].

3.2 Ο ανθρώπινος εγκέφαλος

Οι γνωστικοί επιστήμονες, στην προσπάθειά τους να κατανοήσουν το πλέον πολύπλοκο υπολογιστικό σύστημα που η ανθρωπότητα μέχρι σήμερα γνωρίζει, τον ανθρώπινο εγκέφαλο, τις ιδιότητες και τις λειτουργίες του, προσδιόρισαν ορισμένα θεμελιώδη χαρακτηριστικά με εξέχουσα σημασία για την επιτυχή λειτουργία του εγκεφάλου. Τα εν λόγω χαρακτηριστικά αποτέλεσαν έπειτα τη βάση για την κατασκευή τεχνητών νευρωνικών δικτύων. Συνεπώς για την καλύτερη κατανόηση αυτών των δικτύων, θα είναι χρήσιμο να εξετάσουμε τα βασικά χαρακτηριστικά του ανθρώπινου εγκεφάλου.

Ο εγκέφαλος αποτελείται από δισεκατομμύρια απλές μονάδες που καλούνται νευρώνες και αποτελούν στο σύνολό τους ένα ευρύ δίκτυο. Η βιολογική έρευνα υποδηλώνει ότι οι νευρώνες εκτελούν τη σχετικά απλή εργασία της επιλεκτικής μετάδοσης ηλεκτρικών σημάτων σε γειτονικούς νευρώνες. Όταν ένας νευρώνας λάβει σήματα από τους γειτονικούς νευρώνες, η αντίδρασή του θα ποικίλει ανάλογα με την ένταση των σημάτων και τη δική του ευαισθησία σχετικά με τους νευρώνες που έστειλαν τα σήματα. Ορισμένοι νευρώνες δε θα αντιδράσουν καθόλου σε συγκεκριμένα σήματα. Όταν ένας νευρώνας αντιδράσει (ή ενεργοποιηθεί) θα στείλει με τη σειρά του σήματα σε άλλους νευρώνες. Η ένταση των σημάτων που εκπέμφθισαν θα είναι ανάλογη της έντασης των σημάτων που ελήφθησαν. Καθώς τα σήματα μεταδίδονται μεταξύ των νευρώνων, τελικά ένα “νέφος” από νευρώνες ενεργοποιείται ταυτόχρονα, δημιουργώντας έτσι τις σκέψεις και τα συναισθήματα.



Σχήμα 3.1: Πρότυπο βιολογικού νευρώνα.

Η δύναμη του εγκεφάλου προέρχεται από αυτές τις πολύπλοκες δικτυακές συνδέσεις μεταξύ των νευρώνων και από τον τρόπο με τον οποίο η δραστηριότητα εκατομμυρίων νευρώνων συγχρονίζεται και συνδυάζεται σε κλάσματα δευτερολέπτου [10].

3.3 Εισαγωγή στα Τεχνητά Νευρωνικά Δίκτυα

Τα τεχνητά νευρωνικά δίκτυα, τα οποία αρχικά αναπτύχθηκαν για τη μίμηση βασικών βιολογικών νευρικών συστημάτων – πιο συγκεκριμένα του ανθρώπινου εγκεφάλου, αποτελούνται από αλληλοσυνδεδεμένα απλά λειτουργικά στοιχεία που καλούνται νευρώνες ή κόμβοι. Κάθε κόμβος λαμβάνει ένα σήμα εισόδου, που αποτελεί τη συνολική «πληροφορία» από τους άλλους κόμβους ή το εξωτερικό ερέθισμα, το οποίο επεξεργάζεται τοπικά μέσω της συνάρτησης ενεργοποίησης ή μεταφοράς και παράγει ένα μετασηματοποιημένο σήμα εξόδου προς τους άλλους κόμβους ή τις εξωτερικές εξόδους. Μολονότι κάθε μεμονωμένος νευρώνας εκτελεί τη λειτουργία του αρκετά αργά και ατελώς, συλλογικά ένα δίκτυο μπορεί να πραγματοποιήσει ικανοποιητικά πλήθος εργασιών. Αυτό το χαρακτηριστικό επεξεργασίας πληροφοριών καθιστά τα ΤΝΔ ισχυρούς υπολογιστικούς μηχανισμούς οι οποίοι είναι σε θέση να μάθουν από παραδείγματα και έπειτα να γενικεύσουν σε παραδείγματα που δεν έχουν ξανασυναντήσει.

Οι νευρώνες ατομικά αποτελούν απλές δομές και η χρησιμότητα τους έγκειται στην ικανότητα τους να συνδέονται μεταξύ τους ως μέρος ενός μεγαλύτερου δικτύου. Τα τεχνητά νευρωνικά δίκτυα αποτελούνται από το σύνολο δεδομένων x_i , τα βάρη w_i , το κατώφλι u , τη συνάρτηση μεταφοράς ή ενεργοποίησης f και το δεδομένο εξόδου y , όπου i είναι ο βαθμός (ο αριθμός των εισόδων) του νευρώνα. Τα βάρη αντιπροσωπεύουν την ενίσχυση ή την από-ενίσχυση της διαδικασίας.

Τυπικώς, οι τιμές των δεδομένων εισόδου, x_i , είναι εξωτερικές και προκαλούνται από το περιβάλλον. Εναλλακτικά είναι δυνατό να δημιουργούνται από τα δεδομένα εξόδου των άλλων τεχνητών νευρώνων. Μπορεί να είναι διακριτές τιμές ενός συνόλου, όπως το $[0,1]$ ή πραγματικοί αριθμοί.

Τα βάρη, w_i , είναι πραγματικοί αριθμοί που καθορίζουν τη συνεισφορά κάθε δεδομένου εισόδου στο συνολικό άθροισμα βαρών του νευρώνα και τελικά στο δεδομένο εξόδου του. Ο στόχος των αλγορίθμων εκπαίδευσης των νευρωνικών δικτύων είναι ο προσδιορισμός του καλύτερου δυνατού συνόλου βαρών για το κάθε πρόβλημα. Η εύρεση του βέλτιστου συνόλου είναι τις περισσότερες φορές συνδυασμός υπολογιστικού χρόνου και ελαχιστοποίησης του σφάλματος του δικτύου.

Το κατώφλι, u , είναι ένα πραγματικός αριθμός που αφαιρείται από το συνολικό άθροισμα των βαρών των δεδομένων εισόδου. Ορισμένες φορές το κατώφλι αναφέρεται ως τιμή κλήσης. Σ' αυτή την περίπτωση, ο πραγματικός αριθμός

προστίθεται στο άθροισμα των βαρών. Χάριν απλότητας, το κατώφλι μπορεί να θεωρηθεί ως ένα επιπλέον ζεύγος δεδομένων εισόδου/ βάρους, όπου $w_0 = u$ και $x_0 = -1$.

Η συνάρτηση ενεργοποίησης, f , μπορεί να είναι ιδιαίτερα απλή, παραδείγματος χάριν μπορεί να είναι η βηματική (unit step) συνάρτηση. Ωστόσο, το μοντέλο του τεχνητού νευρώνα έχει επεκταθεί ώστε να περιλαμβάνει και άλλες συναρτήσεις όπως τη σιγμοειδή, την τμηματικά γραμμική (piecewise linear), και τη συνάρτηση Gauss.

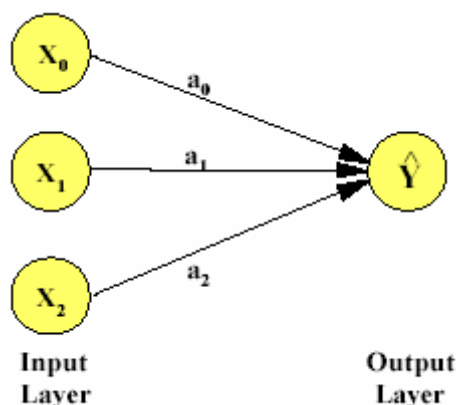
Τα δεδομένα εξόδου του τεχνητού νευρώνα, y , υπολογίζουν τα αποτελέσματα σύμφωνα με τις εξισώσεις των παραπάνω συναρτήσεων. Αυτή είναι η έξοδος της συνάρτησης ενεργοποίησης για το τρέχων άθροισμα βαρών, μείον το κατώφλι. Αυτή η τιμή μπορεί να είναι διακριτή ή πραγματική, ανάλογα με τη συνάρτηση ενεργοποίησης που χρησιμοποιήθηκε. Μόλις υπολογιστεί η έξοδος μεταβιβάζεται σε έναν άλλο νευρώνα (ή σύνολο νευρώνων) ή χρησιμοποιείται ως παράδειγμα στο εξωτερικό περιβάλλον. Η ερμηνεία της εξόδου του νευρώνα εξαρτάται από το εξεταζόμενο πρόβλημα.

Πολλά διαφορετικά μοντέλα ΤΝΔ έχουν προταθεί από το 1980. Ίσως τα πλέον σημαντικά μοντέλα είναι τα perceptrons πολλαπλών επιπέδων MLP, τα δίκτυα Hopfield και τα δίκτυα αυτοοργάνωσης Kohonen. Ο Hopfield (1982) προτείνει ένα περιοδικό νευρωνικό δίκτυο το οποίο δουλεύει με συνδετική μνήμη. Τα δίκτυα Hopfield δεν έχουν επίπεδα και μεταξύ των κόμβων παρουσιάζεται πλήρης αλληλοσύνδεση. Οι έξοδοι του δικτύου δεν αποτελούν απαραίτητα της συναρτήσεις των εισόδων, αλλά συνήθως σταθερές καταστάσεις μιας επαναληπτικής διαδικασίας. Οι χάρτες χαρακτηριστικών του Kohonen παρακινούνται από την αυτο-οργανωτική συμπεριφορά του ανθρώπινου εγκεφάλου [8, 9].

3.4 Η απλούστερη μορφή Τεχνητού Νευρωνικού Δικτύου

Όπως ο εγκέφαλος, ένα νευρωνικό δίκτυο είναι ουσιαστικά ένα σύνολο από αλληλοσυνδεδεμένους νευρώνες, ομαδοποιημένους σε *επίπεδα* (layers), που ανταλλάσσουν πληροφορίες. Η απλούστερη μορφή δικτύου έχει μόνο δύο επίπεδα: το *επίπεδο εισόδου* και το *επίπεδο εξόδου*. Το δίκτυο λειτουργεί ως σύστημα εισόδου-εξόδου, χρησιμοποιώντας τις τιμές των νευρώνων εισόδου για τον υπολογισμό μιας τιμής για το νευρώνα εξόδου. Το σχήμα 3.2 απεικονίζει την τυπική γραφική

αναπαράσταση ενός νευρωνικού δικτύου. Κάθε νευρώνας απεικονίζεται με έναν κύκλο, ενώ οι συνδέσεις μεταξύ των νευρώνων περιγράφονται με τόξα. Η έξοδος Y και οι εισόδους X_0 , X_1 και X_2 είναι $n \times 1$ διανύσματα, όπου n είναι ο αριθμός των παρατηρήσεων. Στο παρών παράδειγμα η πληροφορία κατευθύνεται αποκλειστικά από τις εισόδους προς τις εξόδους και εξαιτίας αυτού του γεγονότος προέρχεται ο όρος δίκτυο *feedforward*.



Σχήμα 3.2: Το βασικό feedforward νευρωνικό δίκτυο.

Κάθε σύνδεση μεταξύ εισόδου- εξόδου χαρακτηρίζεται από τα *βάρη* a_i που εκφράζουν τη σχετική σημασία μιας συγκεκριμένης εισόδου στον υπολογισμό της εξόδου. Για τον υπολογισμό της τιμής εξόδου για την παρατήρηση t , ο νευρώνας εξόδου συλλέγει τις τιμές από κάθε νευρώνα εισόδου για την παρατήρηση t και έπειτα πολλαπλασιάζει κάθε μία από αυτές με το βάρος που αντιστοιχεί στη σχετική σύνδεση. Αυτά τα γινόμενα έπειτα αθροίζονται δίνοντας την ακόλουθη τιμή:

$$a_0X_{0t} + a_1X_{1t} + a_2X_{2t} \quad (1)$$

Στη συνέχεια ο νευρώνας εξόδου επεξεργάζεται αυτή την τιμή χρησιμοποιώντας μια *συνάρτηση ενεργοποίησης*, $f(x)$. Στην απλούστερη μορφή feedforward νευρωνικού δικτύου, η συνάρτηση ενεργοποίησης είναι η ταυτότητα, δηλαδή η $f(x) = x$. Στην περίπτωση αυτή η τιμή της σχέσης (1) αποτελεί την τελική έξοδο του δικτύου για την παρατήρηση t :

$$Y_t = a_0X_{0t} + a_1X_{1t} + a_2X_{2t} \quad (2)$$

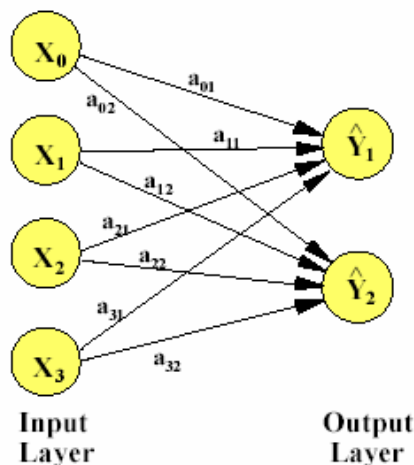
Συνήθως, μία από τις εισόδους, που ονομάζεται *σταθερό βάρος* (bias), ισούται με 1 για όλες τις παρατηρήσεις. Υποθέτοντας ότι το X_0 είναι το σταθερό βάρος, η έξοδος του δικτύου δίδεται από τη σχέση:

$$Y_t = a_0 + a_1X_{1t} + a_2X_{2t} \quad (3)$$

Εν γένει, οι ερευνητές τροφοδοτούν το δίκτυο με την *επιθυμητή τιμή εξόδου* (target output value) την οποία το δίκτυο πρέπει να προσπαθήσει να αναπαράγει μέσω των υπολογισμών του, δεδομένων των τιμών εισόδου. Ακολουθώς υπολογίζεται το σφάλμα πρόβλεψης για κάθε παρατήρηση ως η διαφορά μεταξύ της πραγματικής και της επιθυμητής τιμής. Χρησιμοποιώντας ποικίλους επαναληπτικούς αλγορίθμους τα βάρη του δικτύου τροποποιούνται ωστότου το σφάλμα πρόβλεψης για όλο το δείγμα να ελαχιστοποιηθεί, όπως εκτιμάται από το άθροισμα των τετραγωνικών σφαλμάτων ή του μέσου απόλυτου σφάλματος. Καθώς τα βάρη μεταβάλλονται σε κάθε επανάληψη, το δίκτυο *μαθαίνει*.

Από τα παραπάνω, είναι φανερό ότι το feedforward νευρωνικό δίκτυο δύο επιπέδων είναι ταυτόσημο με το μοντέλο γραμμικής παλινδρόμησης. Οι νευρώνες εισόδου ισοδυναμούν με τις ανεξάρτητες μεταβλητές, ενώ ο νευρώνας εξόδου είναι η εξαρτημένη μεταβλητή. Τα διάφορα βάρη του δικτύου αντιστοιχούν στους εκτιμημένους συντελεστές του μοντέλου παλινδρόμησης και το σταθερό βάρος είναι απλώς ο συντελεστής ανάσχεσης (intercept). Αξιοσημείωτο είναι ότι στις εξισώσεις (2) και (3) ο όρος σφάλματος *ei* παραλείπεται καθώς εκφράζεται μόνο η μαθηματική έκφραση της υπολογισμένης τιμής εξόδου.

Ορισμένα μοντέλα έχουν περισσότερες από μία εξόδους εάν ο ερευνητής ενδιαφέρεται για περισσότερες από μία εξαρτημένες μεταβλητές (Σχήμα 3.3).



Σχήμα 3.3: Απλό feedforward νευρωνικό δίκτυο με δύο εξόδους,

Στο σχήμα 3.3 τα a_{ij} εκφράζουν τα βάρη που συνδέουν την είσοδο i με την έξοδο j . Υποθέτοντας και πάλι ότι το X_0 είναι ο όρος σταθερού βάρους, οι έξοδοι του δικτύου δίδονται:

$$\begin{aligned} Y_1 &= a_{01} + a_{11}X_1 + a_{21}X_2 + a_{31}X_3 \\ Y_2 &= a_{02} + a_{12}X_1 + a_{22}X_2 + a_{32}X_3 \end{aligned} \quad (4)$$

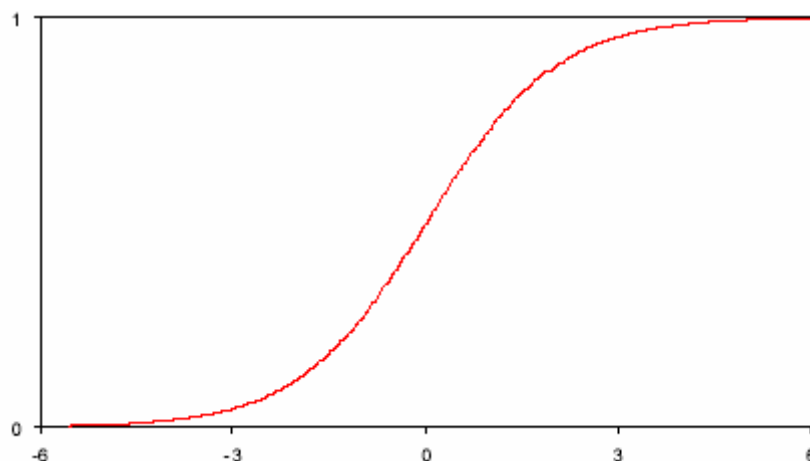
Παίρνουμε ένα σύστημα γραμμικών εξισώσεων παρόμοιο με το σύστημα φαινομενικά ασυσχέτιστων εξισώσεων παλινδρόμησης (seemingly unrelated regression equations) (*à la* Zellner). Όταν υπάρχουν δεδομένα χρονοσειρών, μπορούμε να κατασκευάσουμε νευρωνικά δίκτυα ισοδύναμα με τα μοντέλα διανύσματος αυτοπαλινδρόμησης (vector autoregressive models) προσθέτοντας απλά στο σύνολο των εισόδων, τιμές χρονικών στιγμών πίσω (lagged values) των εξαρτημένων και των ανεξάρτητων μεταβλητών. Εισάγοντας μια σύνδεση μεταξύ των Y_1 και Y_2 αποκτάμε νευρωνικά δίκτυα που αντιστοιχούν στα συστήματα ταυτόχρονων εξισώσεων (systems of simultaneous equations).

3.4.1 Νευρωνικά δίκτυα με μη γραμμικές συναρτήσεις ενεργοποίησης

Στα παραπάνω παραδείγματα έγινε ή υπόθεση ότι η συνάρτηση ενεργοποίησης στους νευρώνες εξόδου είναι η ταυτότητα. Για να διερευνήσουμε πραγματικά τις δυνατότητες των νευρωνικών δικτύων πρέπει να χρησιμοποιήσουμε μια μη γραμμική συνάρτηση ενεργοποίησης. Σχεδόν όλα τα νευρωνικά δίκτυα χρησιμοποιούν μη γραμμικές συναρτήσεις ενεργοποίησης σε κάποιο σημείο του δικτύου. Έτσι το δίκτυο μπορεί να αναπαράγει μη γραμμικά υποδείγματα σε πολύπλοκα σύνολα δεδομένων. Στην ιδανική περίπτωση, η συνάρτηση ενεργοποίησης πρέπει να είναι συνεχής, διαφορήσιμη και μονότονη, καθώς έτσι θα διευκολύνει το έργο του αλγόριθμου βελτιστοποίησης στην εύρεση των κατάλληλων βαρών. Η συνηθέστερη συνάρτηση ενεργοποίησης είναι η σιγμοειδής συνάρτηση (logistic):

$$f(x) = \frac{1}{1 + e^{-x}} \quad (5)$$

Η συνάρτηση logistic κυμαίνεται μεταξύ 0 και 1, όπως φαίνεται στο σχήμα 3.4.



Σχήμα 3.4: Η συνάρτηση logistic.

Χρησιμοποιώντας μία πεπερασμένη συνάρτηση, οι ερευνητές προσπάθησαν να αναπαράγουν την κατάσταση ενεργοποίησης του πραγματικού νευρώνα. Όταν η συνάρτηση είναι κοντά στο 1, τα σήματα που λαμβάνονται από το νευρώνα οδηγούν σε υψηλό επίπεδο ενεργοποίησης. Όταν η συνάρτηση είναι κοντά στο 0, ο νευρώνας μόλις που αντιδρά στο σήμα που λαμβάνει.

Εάν κάνουμε πρόβλεψη για μια μεταβλητή που μπορεί να παίρνει αρνητικές τιμές είναι προτιμότερο να χρησιμοποιούμε την συνάρτηση της υπερβολικής εφαπτομένης:

$$f(x) = \frac{(e^x - e^{-x})}{(e^x + e^{-x})} \quad (6)$$

Η συνάρτηση υπερβολικής εφαπτομένης έχει την ίδια μορφή με τη συνάρτηση logistics, αλλά κυμαίνεται μεταξύ -1 και 1 .

Επιστρέφοντας στο απλό feedforward δίκτυο του σχήματος 3.2, η συνάρτηση ενεργοποίησης logistic στον νευρώνα εξόδου θα είχε ως αποτέλεσμα την ακόλουθη έξοδο για την παρατήρηση t :

$$Y_t = f(a_0 + a_1 X_{1t} + a_2 X_{2t}) = \frac{1}{1 + e^{-(a_0 + a_1 X_{1t} + a_2 X_{2t})}} \quad (7)$$

Το δίκτυο που προκύπτει είναι όμοιο με το δυαδικό λογιστικό μοντέλο πιθανότητας (binary logit probability model). Εάν η συνάρτηση ενεργοποίησης ήταν η συνάρτηση κανονικής αθροιστικής κατανομής, θα παίρναμε ένα δυαδικό μοντέλο πιθανότητας (binary probit model). Η χρήση άλλων πεπερασμένων συναρτήσεων θα

οδηγούσε σε πολλά άλλα είδη δικτύων ικανά να διαχειριστούν μη γραμμικά προβλήματα όπου η εξαρτημένη μεταβλητή είναι πεπερασμένη.

Κατά τη διαχείριση εξαρτημένων μεταβλητών που δεν είναι πεπερασμένες, είναι δυνατή η επιλογή μιας μη πεπερασμένης μη γραμμικής συνάρτησης ενεργοποίησης όπως η $f(x) = x^3$. Ωστόσο, οι ερευνητές προτιμούν να διατηρούν πεπερασμένες συναρτήσεις ενεργοποίησης και να επιτρέπουν μη πεπερασμένες εξαρτημένες μεταβλητές προσθέτοντας *κρυμμένα επίπεδα* στη δομή του δικτύου.

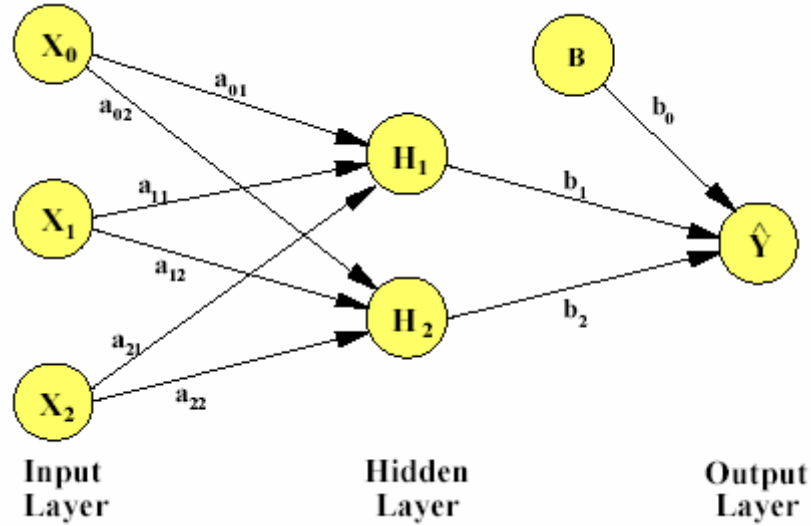
3.4.2 Νευρωνικά δίκτυα με κρυμμένα επίπεδα

Τα δίκτυα που περιγράφηκαν μέχρι στιγμής είχαν μια πολύ απλή δομή δύο επιπέδων που συνέδεαν τις εισόδους με τις εξόδους. Στις πραγματικές εφαρμογές, η αρχιτεκτονική του δικτύου είναι πιο πολύπλοκη. Οι ερευνητές σχεδόν πάντα σχεδιάζουν δομές που περιλαμβάνουν ένα ή περισσότερα *κρυμμένα επίπεδα*, όπως στο σχήμα 3.5. Σ' αυτό το σχήμα τα a_{ij} δηλώνουν τα βάρη που συνδέουν την είσοδο i με την κρυμμένη μονάδα j . Υποθέτουμε ότι X_0 είναι ο όρος σταθερού βάρους (δηλαδή ο συντελεστής ανάσχεσης) για τις κρυμμένες μονάδες, ενώ το B είναι ο όρος σταθερού βάρους για τη μονάδα εξόδου.

Εν αντιθέσει με τις μονάδες εισόδου και εξόδου, οι κρυμμένες μονάδες δεν αναπαριστούν κάποια πραγματική έννοια. Δεν έχουν ούτε ερμηνεία ούτε νόημα. Είναι απλά το ενδιάμεσο αποτέλεσμα στη διαδικασία υπολογισμού της τιμής εξόδου. Συνεπώς δεν έχουν αντίστοιχη έννοια στην οικονομετρία. Οι κρυμμένες μονάδες συμπεριφέρονται όπως οι μονάδες εξόδου, δηλαδή υπολογίζουν το σταθμικό άθροισμα των μεταβλητών εισόδου και έπειτα επεξεργάζονται το αποτέλεσμα χρησιμοποιώντας τη συνάρτηση ενεργοποίησης, που είναι σχεδόν πάντα η συνάρτηση logistic. Στο δίκτυο που απεικονίζεται στο σχήμα 3.5, το αποτέλεσμα που παράγουν οι κρυμμένες μονάδες θα είναι:

$$H_1 = f(a_{01} + a_{11}X_1 + a_{21}X_2) = \frac{1}{1 + e^{-(a_{01} + a_{11}X_1 + a_{21}X_2)}} \quad (8)$$

$$H_2 = f(a_{02} + a_{21}X_1 + a_{22}X_2) = \frac{1}{1 + e^{-(a_{02} + a_{21}X_1 + a_{22}X_2)}} \quad (9)$$



Σχήμα 3.5: Το feedforward νευρωνικό δίκτυο με ένα κρυμμένο επίπεδο

Τοποθετώντας τη συνάρτηση ενεργοποίησης logistic στις κρυμμένες μονάδες και όχι στη μονάδα εξόδου, το δίκτυο δεν περιορίζεται πλέον στην παραγωγή εκτιμήσεων πεπερασμένων μεταβλητών. Εάν η εξαρτημένη μεταβλητή είναι μη πεπερασμένη, η μονάδα εξόδου συνήθως θα χρησιμοποιήσει ως συνάρτηση ενεργοποίησης την ταυτότητα, δηλαδή η έξοδος θα ισούται με το σταθμικό άθροισμα των τιμών των κρυμμένων μονάδων, σταθμισμένο από τους συντελεστές b_j . Αυτό συνεπάγεται συνεχή, μη γραμμική, μη πεπερασμένη έξοδο όπως εκφράζεται στην εξίσωση (10):

$$Y = b_0 + b_1 H_1 + b_2 H_2$$

$$Y = b_0 + \frac{b_1}{1 + e^{-(a_{01} + a_{11}X_1 + a_{21}X_2)}} + \frac{b_2}{1 + e^{-(a_{02} + a_{12}X_1 + a_{22}X_2)}} \quad (10)$$

Εάν η εξαρτημένη μεταβλητή είναι πεπερασμένη, η έξοδος χρησιμοποιεί συνήθως τη συνάρτηση ενεργοποίησης logistic, παράγοντας έτσι μια πεπερασμένη έξοδο, όπως στη σχέση (11):

$$Y = f(b_0 + b_1 H_1 + b_2 H_2)$$

$$Y = \frac{1}{1 + e^{-(b_0 + b_1 H_1 + b_2 H_2)}}$$

$$Y = \frac{1}{1 + e^{-\left(b_0 + \frac{b_1}{1 + e^{-(a_{01} + a_{11}X_1 + a_{21}X_2)}} + \frac{b_2}{1 + e^{-(a_{02} + a_{12}X_1 + a_{22}X_2)}}\right)}} \quad (11)$$

Συνεπώς το δίκτυο έχει τη δυνατότητα παραγωγής πεπερασμένης ή μη εξόδου, διατηρώντας ταυτόχρονα τα μη γραμμικά χαρακτηριστικά του [11].

Ο συνυπολογισμός κρυμμένων μονάδων στο δίκτυο προσφέρει ένα ακόμη σημαντικό πλεονέκτημα. Πολλοί ερευνητές απέδειξαν πέραν κάθε αμφιβολίας ότι ένα νευρωνικό δίκτυο τριών επιπέδων με τη συνάρτηση ενεργοποίησης *logistic* στις κρυμμένες μονάδες, όπως στη σχέση (10), είναι μια γενική εκτιμήτρια (*universal approximator*). Αυτό σημαίνει ότι εάν συμπεριληφθεί ένας σημαντικός αριθμός κρυμμένων μονάδων, το δίκτυο είναι σε θέση να προσδιορίσει σχεδόν οποιαδήποτε γραμμική ή μη συνάρτηση σε ένα επιθυμητό επίπεδο ακρίβειας. Επομένως τα νευρωνικά δίκτυα δύναται να χρησιμοποιηθούν ως ισχυρό εργαλείο για την αναγνώριση και την αναπαραγωγή πολύπλοκων μη γραμμικών δεδομένων σε χρονοσειρές. Είτε μελετάμε το πραγματικό προϊόν (GDP), τον πληθωρισμό, την ανεργία ή την τιμή συναλλάγματος, θεωρητικά το νευρωνικό δίκτυο πρέπει να είναι σε θέση να ανιχνεύσει και να αντιγράψει οποιοδήποτε πολύπλοκο μη γραμμικό υπόδειγμα από τα δεδομένα. Επιπλέον, δεν απαιτείται *a priori* γνώση της διαδικασίας δημιουργίας των δεδομένων, όπως συμβαίνει στα μοντέλα μη γραμμικής παλινδρόμησης. Αρκεί μόνο η χρήση μιας γενικής συναρτησιακής μορφής, όπως στη σχέση (10), αλλά με μεγαλύτερο αριθμό κρυμμένων μονάδων [12, 13].

Δεν υπάρχει θεωρητική βάση για τον προσδιορισμό του αριθμού των κρυμμένων μονάδων ή επιπέδων στο δίκτυο. Βάσει της ιδιότητας γενικής εκτιμήτριας που περιγράφηκε παραπάνω, είναι λογική η χρήση μεγάλου αριθμού κρυμμένων μονάδων. Εντούτοις, εάν προστεθούν πολλά κρυμμένα επίπεδα, το δίκτυο γίνεται επιρρεπές στο πρόβλημα της υπερεκπαίδευσης (*overfitting*) των δεδομένων. Αυτό σημαίνει ότι το δίκτυο μπορεί να πετύχει μια υψηλότερη ακρίβεια πρόβλεψης για την περίοδο που εκτιμάται, αλλά θα παράγει φτωχές προβλέψεις εκτός δείγματος. Επιπλέον, το πλήθος των βαρών του δικτύου αυξάνεται ραγδαία, καθώς προστίθενται περισσότερες κρυμμένες μονάδες, αυξάνοντας έτσι το χρόνο εκτίμησης του μοντέλου.

Στην πράξη, η σχεδίαση της αρχιτεκτονικής του δικτύου είναι μια μονότονη διαδικασία δοκιμής και σφάλματος. Οι ερευνητές συνήθως εκτιμούν πολλά διαφορετικά δίκτυα και επιλέγουν εκείνο που οδηγεί στο μικρότερο σφάλμα πρόβλεψης [14].

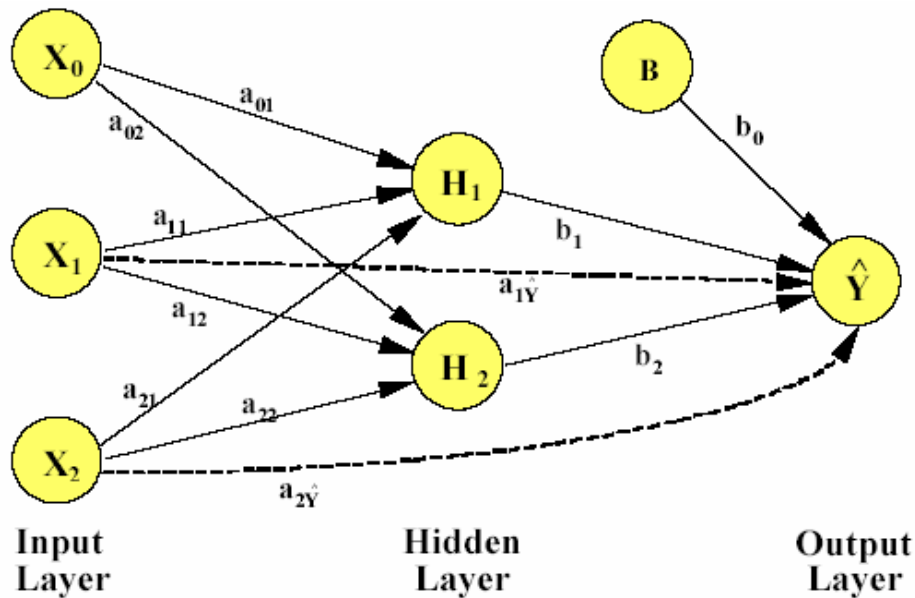
3.4.3 Αυξητικά Νευρωνικά Δίκτυα (Augmented Neural Networks)

Ένα ακόμη είδος νευρωνικών δικτύων μπορεί να κατασκευασθεί με την πρόσθεση άμεσων συνδέσεων από τις εισόδους προς την έξοδο (Σχήμα 3.6). Στην εν λόγω δομή, που καλείται *αυξητικό* (*augmented*) νευρωνικό δίκτυο, οι εισοδοί συνδέονται άμεσα με την έξοδο μέσω των βαρών a_{1Y} και a_{2Y} . Δεν είναι απαραίτητο ο όρος σταθερού βάρους X_0 να συνδέεται με την έξοδο, καθώς η έξοδος έχει ήδη έναν όρο σταθερού βάρους B .

Υποθέτοντας ως συνάρτηση ενεργοποίησης την ταυτότητα για το νευρώνα εξόδου, το αυξητικό νευρωνικό δίκτυο έχει μεγάλο ενδιαφέρον, καθώς περικλείει το μοντέλο γραμμικής παλινδρόμησης. Συνεπώς αποτελεί την συνηθέστερη αρχιτεκτονική στα νευρωνικά δίκτυα που πραγματοποιούν μακροοικονομική πρόβλεψη. Χρησιμοποιώντας τη σχέση (10), μπορούμε να εξάγουμε άμεσα την έξοδο για το αυξητικό δίκτυο:

$$Y = b_0 + a_{1Y}X_1 + a_{2Y}X_2 + \frac{b_1}{1 + e^{-(a_{01} + a_{11}X_1 + a_{21}X_2)}} + \frac{b_2}{1 + e^{-(a_{02} + a_{12}X_1 + a_{22}X_2)}} \quad (12)$$

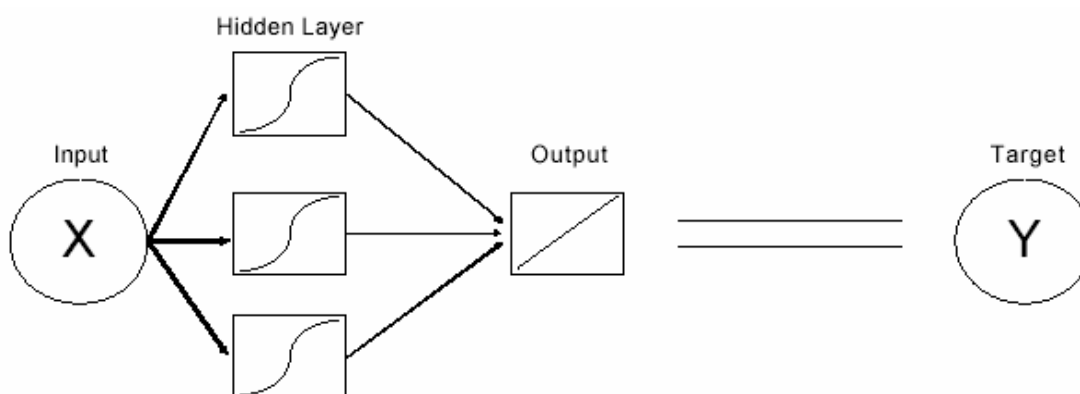
Επομένως το αυξητικό δίκτυο μπορεί να θεωρηθεί ως ένα πρότυπο μοντέλο γραμμικής παλινδρόμησης εμπλουτισμένο με μη γραμμικούς όρους. Εάν η εξαρτημένη μεταβλητή που μελετάμε δεν παρουσιάζει μη γραμμικές ιδιότητες, οι συντελεστές b_1 και b_2 ισοδυναμούν με 0, συνεπώς παίρνουμε ένα πρότυπο γραμμικό μοντέλο [11, 15].



Σχήμα 3.6: Το αυξητικό νευρωνικό δίκτυο

3.5 Νευρωνικά Δίκτυα Πολλαπλών Επιπέδων (MLP)

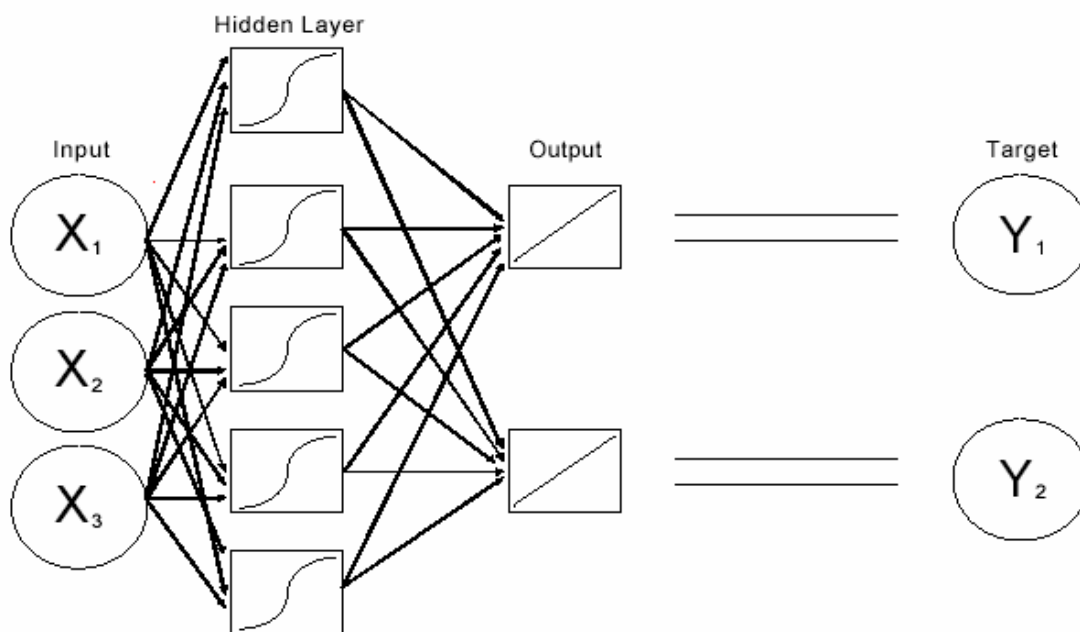
Μία από τις χρησιμότερες και πλέον επιτυχημένες εφαρμογές των νευρωνικών δικτύων στην ανάλυση δεδομένων και στην πρόβλεψη αποτελεί το μοντέλο πολλαπλών επιπέδων perceptron (multilayer perceptron model) MLP. Τα μοντέλα πολλαπλών επιπέδων perceptron είναι μη γραμμικά μοντέλα νευρωνικών δικτύων που μπορούν να χρησιμοποιηθούν σχεδόν με όλες τις συναρτήσεις με υψηλό βαθμό ακρίβειας. Τα MLP περιέχουν ένα ή περισσότερα κρυμμένα επίπεδα νευρώνων που χρησιμοποιούν μη γραμμικές συναρτήσεις ενεργοποίησης, όπως συναρτήσεις logistic. Στο σχήμα 3.7 απεικονίζεται ένα MLP με ένα κρυμμένο επίπεδο και μία μόνο είσοδο και έξοδο. Το MLP στο σχήμα 3.7 αναπαριστά μια απλή μη γραμμική παλινδρόμηση.



Σχήμα 3.7: Perceptron Πολλαπλών επιπέδων με μία μόνο είσοδο και έξοδο

Ο αριθμός των εισόδων και εξόδων στο MLP, μπορεί να διαχειριστεί κατάλληλα ώστε να αναλύει ποικίλους τύπους δεδομένων. Στο σχήμα 3.8 απεικονίζεται μια πολλαπλή μη γραμμική παλινδρόμηση πολλαπλών μεταβλητών.

Η πολυπλοκότητα των MLP μπορεί να προσαρμοσθεί ανάλογα με τον αριθμό των κρυμμένων επιπέδων. Διαφορετικός αριθμός κρυμμένων επιπέδων δύναται να μετατρέψει ένα MLP από ένα απλό παραμετρικό μοντέλο σε ένα ευέλικτο μη παραμετρικό μοντέλο [16].



Σχήμα 3.8: Perceptron πολλαπλών επιπέδων με πολλαπλές εισόδους και εξόδους

Ένα MLP αποτελείται συνήθως από αρκετά επίπεδα κόμβων. Το πρώτο ή το χαμηλότερο επίπεδο είναι το επίπεδο εισόδου όπου λαμβάνεται εξωτερική πληροφορία. Το τελευταίο ή υψηλότερο επίπεδο είναι το επίπεδο εξόδου όπου επιτυγχάνεται η λύση του προβλήματος. Το επίπεδο εισόδου και το επίπεδο εξόδου διαχωρίζονται από ένα ή περισσότερα ενδιάμεσα επίπεδα που καλούνται κρυμμένα επίπεδα. Οι κόμβοι στα γειτονικά επίπεδα συνήθως είναι πλήρως συνδεδεμένοι μέσω ακυκλικών τόξων από χαμηλότερα επίπεδα προς υψηλότερα.

Για ένα επεξηγηματικό ή αιτιολογικό πρόβλημα πρόβλεψης, οι είσοδοι στο MLP είναι συνήθως οι ανεξάρτητες μεταβλητές. Η συναρτησιακή σχέση που εκτιμάται από το MLP μπορεί να γραφεί στη μορφή:

$$y = f(x_1, x_2, \dots, x_p),$$

όπου τα x_1, x_2, \dots, x_p είναι οι p ανεξάρτητες μεταβλητές και y είναι η εξαρτημένη μεταβλητή. Υπό αυτή την έννοια, το νευρωνικό δίκτυο όπως προαναφέρθηκε αποτελεί λειτουργικό ισοδύναμο με το μη γραμμικό μοντέλο παλινδρόμησης. Ωστόσο στα προβλήματα πρόβλεψης χρονοσειρών, οι είσοδοι είναι τυπικά οι παρελθοντικές παρατηρήσεις των δεδομένων σειρών και οι έξοδοι οι μελλοντικές τους τιμές. Το MLP εκτελεί την ακόλουθη συνάρτηση:

$$y_{t+1} = f(y_t, y_{t-1}, \dots, y_{t-p}),$$

όπου y_t είναι η παρατήρηση τη χρονική στιγμή t . Συνεπώς το MLP είναι ισοδύναμο με το μη γραμμικό μοντέλο παλινδρόμησης για προβλήματα πρόβλεψης χρονοσειρών.

Πριν ένα οποιοδήποτε ΤΝΔ μπορέσει να χρησιμοποιηθεί για την εκτέλεση κάθε επιθυμητής εργασίας, είναι αναγκαίο να εκπαιδευθεί. Βασικά, εκπαίδευση είναι η διαδικασία του προσδιορισμού των βαρών τα οποία αποτελούν τα θεμελιώδη στοιχεία του ΤΝΔ. Η γνώση που αποκτάται από το δίκτυο αποθηκεύεται στα τόξα (arcs) και στους κόμβους με τη μορφή συναπτικών βαρών και των σταθερών βαρών (nodes biases) των κόμβων. Μέσω των συνδετικών τόξων (arcs) το ΤΝΔ μπορεί να εκτελέσει πολύπλοκες μη γραμμικές συναρτήσεις από τους κόμβους εισόδου προς τους κόμβους εξόδου. Η εκπαίδευση MLP είναι επιβλεπόμενη (supervised) κατά το ότι η επιθυμητή αντίδραση του δικτύου (target value) για κάθε υπόδειγμα εισόδου (παράδειγμα) είναι πάντα διαθέσιμη.

Τα δεδομένα εισόδου εκπαίδευσης βρίσκονται σε μορφή διανυσμάτων των μεταβλητών εισόδου ή των υποδειγμάτων εκπαίδευσης. Σε κάθε στοιχείο του διανύσματος εισόδου αντιστοιχεί ένας κόμβος εισόδου στο επίπεδο εισόδου του δικτύου. Συνεπώς ο αριθμός των κόμβων εισόδου ισούται με τη διάσταση των διανυσμάτων εισόδου. Για ένα τυχαίο πρόβλημα πρόβλεψης, το πλήθος των κόμβων εισόδου είναι ορισμένο επακριβώς και είναι ο αριθμός των ανεξάρτητων μεταβλητών του προβλήματος. Ωστόσο, για ένα πρόβλημα πρόβλεψης χρονοσειρών, ο κατάλληλος αριθμός των κόμβων εισόδου δεν είναι εύκολο να προσδιοριστεί. Τα συνολικά διαθέσιμα δεδομένα συνήθως διανέμονται στο σύνολο εκπαίδευσης (in-sample data) και στο σύνολο δοκιμής (out-of-sample ή hold-out sample). Το σύνολο εκπαίδευσης χρησιμοποιείται για την εκτίμηση των συναπτικών βαρών, ενώ το σύνολο δοκιμής για τη μέτρηση της ικανότητας γενίκευσης του δικτύου.

Η διαδικασία εκπαίδευσης έχει συνήθως ως εξής: Πρώτον, παραδείγματα των συνόλων εκπαίδευσης εισάγονται στους κόμβους εισόδου. Οι τιμές ενεργοποίησης των κόμβων εισόδου σταθμίζονται και συσσωρεύονται σε κάθε κόμβο στο πρώτο κρυμμένο επίπεδο. Έπειτα το σύνολο μορφοποιείται από τη συνάρτηση ενεργοποίησης στις τιμές ενεργοποίησης των κόμβων. Διαδοχικά αναπτύσσεται σε μια είσοδο μέσα στους κόμβους του επόμενου επιπέδου, έως ότου τελικά βρεθούν οι τιμές ενεργοποίησης των εξόδων. Ο αλγόριθμος εκπαίδευσης χρησιμοποιείται για τον προσδιορισμό των βαρών που ελαχιστοποιούν ορισμένα κριτήρια ολικού σφάλματος όπως το άθροισμα των τετραγωνικών σφαλμάτων (sum of squares errors SSE), ή των μέσων τετραγωνικών σφαλμάτων (mean squared errors MSE). Συνεπώς η εκπαίδευση

του δικτύου είναι στην πράξη ένα μη γραμμικό πρόβλημα ελαχιστοποίησης χωρίς περιορισμούς.

Όσον αφορά τα προβλήματα πρόβλεψης χρονοσειρών, το υπόδειγμα εκπαίδευσης αποτελείται από σταθερό αριθμό παρελθοντικών παρατηρήσεων των σειρών. Ας υποθέσουμε ότι έχουμε N παρατηρήσεις y_1, y_2, \dots, y_N στο σύνολο της εκπαίδευσης και χρειαζόμαστε πρόβλεψη ενός βήματος έμπροσθεν (1-step-ahead), τότε χρησιμοποιώντας ένα MLP με n κόμβους εισόδου, έχουμε $N - n$ υποδείγματα εκπαίδευσης. Το πρώτο υπόδειγμα εκπαίδευσης θα αποτελείται από τα y_1, y_2, \dots, y_n ως εισόδους και το y_{n+1} ως την έξοδο που στοχεύουμε. Το δεύτερο υπόδειγμα εκπαίδευσης θα περιέχει τα y_2, y_3, \dots, y_{n+1} ως εισόδους και το y_{n+2} ως την επιθυμητή έξοδο. Τέλος, το τελευταίο υπόδειγμα εκπαίδευσης θα έχει τα $y_{N-n}, y_{N-n+1}, \dots, y_{N-1}$ για εισόδους και το y_N για το στόχο. Τυπικά, μια αντικειμενική συνάρτηση βασισμένη στο SSE ή συνάρτηση κόστους που θα ελαχιστοποιηθεί κατά τη διαδικασία εκπαίδευσης είναι η ακόλουθη:

$$E = \frac{1}{2} \sum_{i=n+1}^N (y_i - a_i)^2,$$

όπου a_i είναι η πραγματική έξοδος του δικτύου και ο συντελεστής $\frac{1}{2}$ περιλαμβάνεται για να απλοποιήσει την παράσταση υπολογισμένων παραγώγων στον αλγόριθμο εκπαίδευσης.

Στα περισσότερα πραγματικά προβλήματα, το νευρωνικό δίκτυο δεν είναι ποτέ 100% σωστό. Τα τεχνητά νευρωνικά δίκτυα είναι προγραμματισμένα να μαθαίνουν έως ένα δεδομένο όριο σφάλματος. Αφού το νευρωνικό δίκτυο μάθει έως το όριο σφάλματος, ο μηχανισμός προσαρμογής των βαρών τερματίζεται και το δίκτυο δοκιμάζεται σε γνωστές περιπτώσεις που δεν έχει ξανασυναντήσει. Η εφαρμογή των νευρωνικών δικτύων σε άγνωστες περιπτώσεις δίδει την πραγματική τιμή σφάλματος. Σε ένα καλά εκπαιδευμένο νευρωνικό δίκτυο το όριο σφάλματος και το πραγματικό σφάλμα πρέπει να ταυτίζονται [17, 18].

3.5 Νευρωνικά δίκτυα υψηλής τάξης

Τα νευρωνικά δίκτυα υψηλής τάξης έχει αποδειχθεί ότι έχουν εκπληκτικές ιδιότητες υπολογισμού, αποθήκευσης και μάθησης. Η απόδοσή τους οφείλεται στο γεγονός ότι η τάξη ή η δομή των νευρωνικών δικτύων υψηλής τάξης, μπορεί να προσαρμοσθεί στην τάξη ή τη δομή του προβλήματος. Συνεπώς το νευρωνικό δίκτυο

που σχεδιάστηκε για συγκεκριμένη κατηγορία προβλημάτων εξειδικεύεται και γίνεται ακόμη πιο αποτελεσματικό στην επίλυση αυτών των προβλημάτων.

Ένα νευρωνικό δίκτυο υψηλής τάξης ορίζεται ως μια υψηλής τάξης λογική μονάδα (High Order Threshold Logic Unit – HOTLU) η οποία περιλαμβάνει όρους με υψηλής τάξης βάρη. Συνήθως, αλλά όχι πάντα, η έξοδος του είναι (1,0) ή (+1,-1). Ένα υψηλής τάξης νευρωνικό δίκτυο στρωμάτων (slab) ορίζεται ως ένα σύνολο λογικών μονάδων υψηλής τάξης (HOTLUs).

Ένα απλό στρώμα νευρωνικού δικτύου υψηλής τάξης περιγράφεται από τη σχέση:

$$y_i = S[net(i)] = S[T_0(i) + T_1(i) + T_2(i) + T_3(i) + \dots + T_k(i)] \quad (1)$$

όπου y_i είναι η έξοδος της i -οστής νευρωνικής μονάδας υψηλής τάξης και S είναι η στιγμοειδής συνάρτηση. $T_n(i)$ είναι ο όρος n -οστής τάξης για την i -οστή μονάδα και k είναι η τάξη της μονάδας.

Ο όρος μηδενικής τάξης είναι ρυθμιζόμενο κατώφλι, που δηλώνεται ως $W_0(i)$. Ο όρος n -οστής τάξης είναι ένα γραμμικό σταθμισμένο άθροισμα από τα n -οστής τάξης γινόμενα των εισόδων, παραδείγματα των οποίων είναι τα:

$$T_1(i) = \sum_j W_1(i, j)x(j), \quad T_2(i) = \sum_j \sum_k W_2(i, j, k)x(j)x(k), \quad (2)$$

όπου $x(j)$ είναι η j -οστή είσοδος στο νευρώνα i -οστής τάξης και $W_n(i, j, \dots)$ είναι ένα ρυθμιζόμενο βάρος που συλλαμβάνει τη n -οστής τάξης συσχέτιση μεταξύ του n -οστής τάξης γινομένου των εισόδων και της εξόδου της μονάδας.

Αρκετά στρώματα μπορούν να συνδεθούν σειριακά τροφοδοτώντας την έξοδο ενός στρώματος ως είσοδο ενός άλλου στρώματος, για την παραγωγή δικτύων πολλαπλών στρωμάτων. Τα sigma-pi νευρωνικά δίκτυα είναι πολυεπίπεδα δίκτυα που μπορούν να έχουν όρους υψηλής τάξης σε κάθε επίπεδο. Ο αλγόριθμος μάθησης για αυτά τα δίκτυα είναι ο γενικευμένος back-propagation. Ωστόσο, οι sigma-pi μονάδες, όπως αρχικά κατασκευάστηκαν, δεν έχουν σταθερούς όρους βαρών, αν και είναι αρκετά απλό να ενσωματωθούν τέτοιες σταθερές στις εν λόγω μονάδες.

Η διαδικασία μάθησης περιλαμβάνει την εφαρμογή μιας καθορισμένης συνάρτησης στο νευρωνικό δίκτυο μέσω επαναληπτικής προσαρμογής των βαρών

βασισμένη σε έναν συγκεκριμένο κανόνα μάθησης και στην αντίδραση του δικτύου στα σύνολα εκπαίδευσης. Η συνάρτηση που θα μαθευτεί αναπαριστάται από ένα σύνολο παραδειγμάτων που αποτελείται από το πιθανό διάνυσμα εισόδου σε συνδυασμό με μία επιθυμητή έξοδο. Το σύνολο εκπαίδευσης είναι υποσύνολο του συνόλου όλων των πιθανών παραδειγμάτων της συνάρτησης. Η υλοποίηση της διαδικασίας μάθησης περιλαμβάνει διαδοχική παρουσίαση παραδειγμάτων σχεδίασης ως ζεύγη εισόδου - εξόδου στο δίκτυο, τα οποία προέρχονται από το σύνολο εκπαίδευσης. Έπειτα από κάθε παρουσίαση, τα βάρη του δικτύου προσαρμόζονται ώστε να συλλαμβάνουν τη συσχέτιση της δομής της σχεδίασης.

Ένας συνηθισμένος κανόνας μάθησης μονού στρώματος είναι ο κανόνας perceptron, ο οποίος για τον κανόνα δεύτερης τάξης ενημέρωσης εκφράζεται ως:

$$W_2'(i, j, k) = W_2(i, j, k) + [t(i) - y(i)]x(j)x(k) \quad (3)$$

Εδώ $t(i)$ είναι η επιθυμητή έξοδος και $y(i)$ είναι η πραγματική έξοδος της i -οστής μονάδας για το x διάνυσμα εισόδου.

Υπάρχουν παρόμοιοι κανόνες μάθησης για τους υπόλοιπους όρους W_i . Εάν το δίκτυο αποδώσει τη σωστή έξοδο για κάθε παράδειγμα εισόδου στο σύνολο εκπαίδευσης, λέμε ότι το δίκτυο συνέκλινε, ή ότι έμαθε το σύνολο εκπαίδευσης. Εάν, έπειτα από τη μάθηση του συνόλου εκπαίδευσης, το δίκτυο δώσει τη σωστή έξοδο για ένα σύνολο παραδειγμάτων του συνόλου εκπαίδευσης, τα οποία δεν έχει δει ακόμη, λέμε ότι το δίκτυο είναι σωστά γενικευμένο.

Ένας ακόμη κανόνας μάθησης για τα νευρωνικά δίκτυα υψηλής τάξης είναι ο κανόνας εξωτερικού γινομένου (outer product rule), που είναι επίσης γνωστός ως κανόνας μάθησης Hebbian. Στην προκειμένη μορφή μάθησης, ο πίνακας συσχέτισης αναπτύσσεται απευθείας μέσω της σχέσης:

$$W_2(i, j, k) = \sum_{s=1}^{Np} [y^s(i) - y(i)] [x^s(j) - x(j)] [x^s(k) - x(k)],$$

$$y(i) = \left[\sum_{s=1}^{Np} y^s(i) \right] / Np, \quad x(j) = \left[\sum_{s=1}^{Np} x^s(j) \right] / Np,$$

όπου το Np δηλώνει τον αριθμό των υποδειγμάτων του συνόλου εκπαίδευσης. Το σύνολο εκπαίδευσης εκφράζεται ως $\{(x^s, y^s), s \in (1, Np)\}$ και τα y και x είναι οι

μέσοι όροι των y^s και x^s στο σύνολο εκπαίδευσης. Στην περίπτωση που το x είναι σχεδόν μηδέν [το οποίο ισχύει όταν οι εισοδοί ανήκουν στο $(+1, -1)$], κερδίζουμε στην απλότητα των παραπάνω σχέσεων με μικρό τίμημα ακρίβειας αν θεωρήσουμε τους όρους x και y μηδενικούς. Αυτό καταδεικνύει την ελάττωση των υπολογισμών που επιτυγχάνεται με την κατάλληλη επιλογή αναπαράστασης των δεδομένων. Ανάλογες σχέσεις ισχύουν και για άλλες τάξεις [19].

Τα νευρωνικά δίκτυα υψηλής τάξης (HONN) αναπτύχθηκαν με την πρόθεση να διευρύνουν τη μη γραμμική περιγραφική ικανότητα των δικτύων feedforward perceptron πολλαπλών επιπέδων. Αυτό επιτυγχάνεται μέσω αύξησης της μη γραμμικής περιγραφικής ικανότητας κάθε νευρώνα ξεχωριστά. Τα νευρωνικά δίκτυα υψηλής τάξης έχουν μονάδες αθροισμάτων (sigma) όπως και γινομένων (pi).

Ένα νευρωνικό δίκτυο υψηλής τάξης κατασκευάζει υψηλής τάξης πολυωνυμικά μοντέλα :

$$P(x) = w_0 + \sum_i w_i x_i + \sum_i \sum_j w_{ij} x_i x_j + \sum_i \sum_j \sum_k w_{ijk} x_i x_j x_k + \dots$$

ή συνοπτικά:

$$P(x) = w_0 + \sum_{i=1}^M w_i \prod_{j=1}^d x_j^r$$

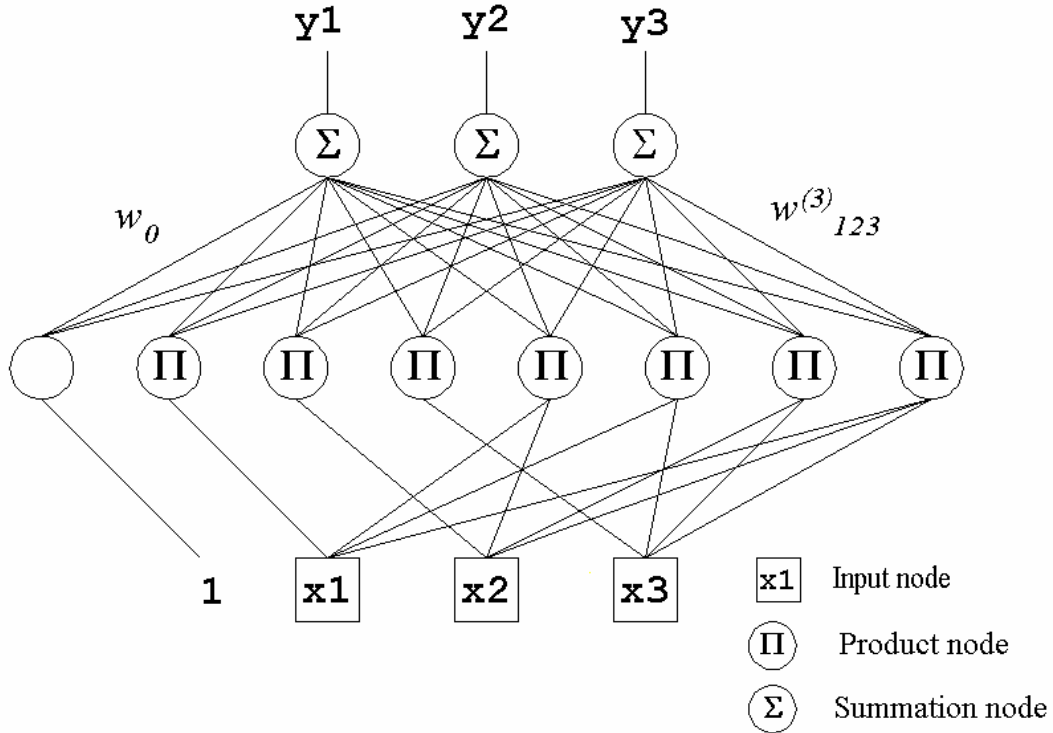
όπου: w_i είναι τα βάρη, M είναι ο μέγιστος αριθμός μονώνυμων, x_i είναι τα χαρακτηριστικά του διανύσματος εισόδου, $r = 0, 1, 2, \dots$ είναι η δύναμη με την οποία το j -οστό χαρακτηριστικό x_j συμμετέχει στο i -οστό μονώνυμο.

Η μη γραμμικότητα των συναρτήσεων δομείται από διαδοχικά επίπεδα τα οποία έχουν μονάδες sigma και pi. Στο σχήμα 3.9 φαίνεται ένα νευρωνικό δίκτυο υψηλής τάξης το οποίο υπολογίζει τη συνάρτηση:

$$P(x) = \sigma(w_0 + \sum_{i=1}^M w_i \sigma(\prod_{j=1}^d x_j^r))$$

όπου x είναι το διάνυσμα εισόδου, w_i είναι τα βάρη, σ είναι η σιγμοειδής συνάρτηση: $\sigma(s) = 1/(1 + e^{-s})$

Το πλήθος των κρυμμένων μονάδων στο πλήρως συνδεδεμένο HONN αυξάνει εκθετικά με το πλήθος των εισόδων (δηλαδή με την αύξηση της διάστασης του διανύσματος εισόδου).



Σχήμα 3.9: Το πλήρως συνδεδεμένο νευρωνικό δίκτυο υψηλής τάξης

3.5.1 Δίκτυα Sigma-Pi

Τα νευρωνικά δίκτυα sigma-pi (SPN) είναι feedforward δίκτυα στα οποία κάθε επίπεδο περιέχει όρους υψηλής τάξης. Τις περισσότερες φορές τα επίπεδα έχουν αθροιστικές μονάδες που τροφοδοτούνται μέσω σταθμικών συνδέσεων από τα ενδιάμεσα αποτελέσματα μονάδων γινομένων.

Τα νευρωνικά δίκτυα sigma-pi αποτελούν σποραδικώς συνδεδεμένα νευρωνικά δίκτυα υψηλής τάξης. Οι ερευνητές περιορίζουν την πολωνυμική τάξη (δηλαδή την τοπολογία του δικτύου) στην κατάλληλη διαμόρφωση ώστε να επιτυγχάνεται ο επιθυμητός βαθμός ακρίβειας χρησιμοποιώντας εκ των προτέρων γνώση για τη δεδομένη εργασία.

Οι μονάδες *sigma* υπολογίζουν το άθροισμα των σταθμικών εισόδων p_j από το χαμηλότερο j -οστό επίπεδο:

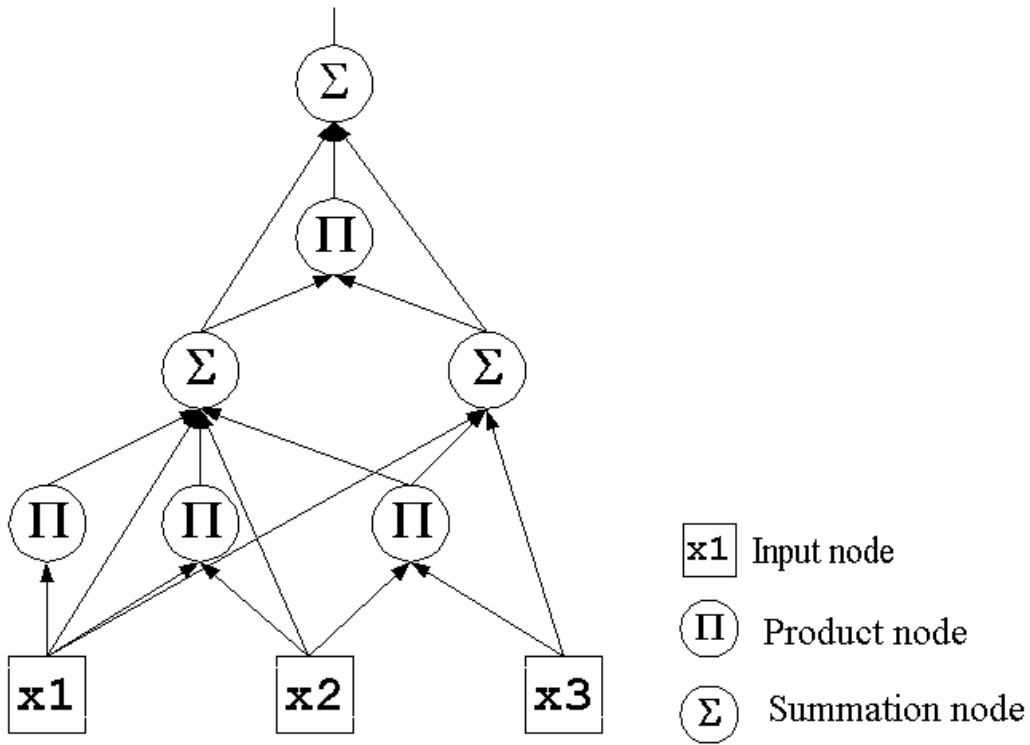
$$s = \sum_j w_j p_j$$

Οι μονάδες *pi* υπολογίζουν το γινόμενο των σταθμικών εισόδων x_i από το χαμηλότερο i -οστό επίπεδο:

$$p_j = \prod_i w_i x_i$$

Η έξοδος των μονάδων διέρχεται μέσω της σιγμοειδούς συνάρτησης:

$$o = \sigma(s) = 1/(1 + e^{-s})$$



Σχήμα 3.10: Το νευρωνικό δίκτυο sigma-pi

Οι αρχές του αλγορίθμου μάθησης backpropagation ισχύουν και για τα sigma-pi δίκτυα. Ο κανόνας Δέλτα για τα βάρη $i \rightarrow j$ στις συνδέσεις των κόμβων στα κρυμμένα επίπεδα έχει ως εξής:

$$\begin{aligned} \partial E / \partial s_j &= \sum_k (\partial E / \partial s_k) (\partial s_k / \partial s_j) \\ &= \sum_k (-\beta_k) \partial s_k / \partial s_j \\ &= \sum_k (-\beta_k) (\partial s_k / \partial o_j) (\partial o_j / \partial s_j) \\ &= \sum_k (-\beta_k) (s'_k) (\partial o_j / \partial s_j) \\ &= \sum_k (-\beta_k) (s'_k) o_j (1 - o_j) \end{aligned}$$

όπου το s_k ορίζεται:

$$s_k = w_0 + \sum_{j1} w_{j1} o_{j1} + \sum_{j1} \sum_{j2} w_{j1j2} o_{j1} o_{j2} + \sum_{j1} \sum_{j2} \sum_{j3} w_{j1j2j3} o_{j1} o_{j2} o_{j3} + \dots$$

η παράγωγος του s_k είναι: $\partial s_k / \partial o_{j2} = s'_k = \sum_{j1} w_{j1j2} o_{j1} + \sum_{j1} \sum_{j3} w_{j1j2j3} o_{j1} o_{j3} + \dots$

Συνεπώς το όφελος (benefit) είναι: $\beta_j = -\partial E / \partial s_j = o_j (1 - o_j) \sum_k \beta_k s'_k$

Αλγόριθμος εκπαίδευσης backpropagation για τα sigma-pi δίκτυα:

Αρχικοποίηση: Παραδείγματα $\{(\mathbf{x}_e, y_e)\}_{e=1}^N$, θέσε τα αρχικά βάρη w_i σε μικρές τυχαίες τιμές, συντελεστής μάθησης $\eta = 0.1$.

Επανάληψη:

για κάθε παράδειγμα εκπαίδευσης (\mathbf{x}, y)

- πολλαπλασίασε τις εξόδους χρησιμοποιώντας τη σιγμοειδή συνάρτηση:

$$o_j = \sigma(s_j) = 1/(1 + e^{-s_j})$$

$$o_k = \sigma(s_k) = 1/(1 + e^{-s_k}), \quad \text{όπου:}$$

$$s_k = w_0 + \sum_{j1} w_{j1} o_{j1} + \sum_{j1} \sum_{j2} w_{j1j2} o_{j1} o_{j2} + \sum_{j1} \sum_{j2} \sum_{j3} w_{j1j2j3} o_{j1} o_{j2} o_{j3} + \dots$$

- πολλαπλασίασε το όφελος β_k στους κόμβους k στο επίπεδο εξόδου:

$$\beta_k = o_k (1 - o_k) [y_k - o_k]$$

- πολλαπλασίασε τις μεταβολές για τα βάρη $j \rightarrow k$ στις συνδέσεις των κόμβων αθροισμάτων στο επίπεδο εξόδου:

$$\Delta w_{j1k} = \eta \beta_k o_{j1} \quad \text{και} \quad \Delta w_{0k} = \eta \beta_k$$

$$\Delta w_{j2k} = \eta \beta_k o_{j1} o_{j2}$$

$$\Delta w_{j3k} = \eta \beta_k o_{j1} o_{j2} o_{j3}$$

- πολλαπλασίασε τα οφέλη β_{j2} για τους κρυμμένους αθροιστικούς κόμβους j σύμφωνα με τον τύπο:

$$\beta_{j2} = o_j (1 - o_j) [\sum_k \beta_k s'_{j2k}] \quad // \text{αποτελέσματα από πολλαπλούς}$$

κόμβους στο επόμενο επίπεδο

$$s'_{j2k} = \sum_{j1} w_{j1j2} o_{j1} + \sum_{j1} \sum_{j3} w_{j1j2j3} o_{j1} o_{j3} + \dots$$

- πολλαπλασίασε τις μεταβολές για τα βάρη $i \rightarrow j$ στις συνδέσεις των κόμβων αθροισμάτων στα κρυμμένα επίπεδα:

$$\Delta w_{ij1} = \eta \beta_{j2} o_{i1} \quad \text{και} \quad \Delta w_{0j} = \eta \beta_{j2}$$

$$\Delta w_{ij2} = \eta \beta_{j2} o_{i1} o_{i2}$$

$$\Delta w_{ij3} = \eta \beta_{j2} o_{i1} o_{i2} o_{i3}$$

- ενημέρωσε τα βάρη με τις μεταβολές που υπολόγισες:

$$w = w + \Delta w$$

ωστόσο να ικανοποιηθεί η συνθήκη τερματισμού [20, 21].

3.5.2 Δίκτυα Pi-Sigma

Τα δίκτυα pi-sigma (PSN) είναι feedforward νευρωνικά δίκτυα με ένα μόνο κρυμμένο επίπεδο γραμμικών μονάδων αθροισμάτων, ενώ έχουν μονάδες γινομένων στο εξωτερικό επίπεδο. Η έννοια “pi-sigma” προέρχεται από το γεγονός ότι τα εν λόγω δίκτυα χρησιμοποιούν τα γινόμενα των αθροισμάτων από τα χαρακτηριστικά εισόδου, αντί για τα αθροίσματα των γινομένων όπως στα δίκτυα “sigma-pi”.

Το νευρωνικό δίκτυο pi-sigma του παρακάτω σχήματος υπολογίζει τη συνάρτηση:

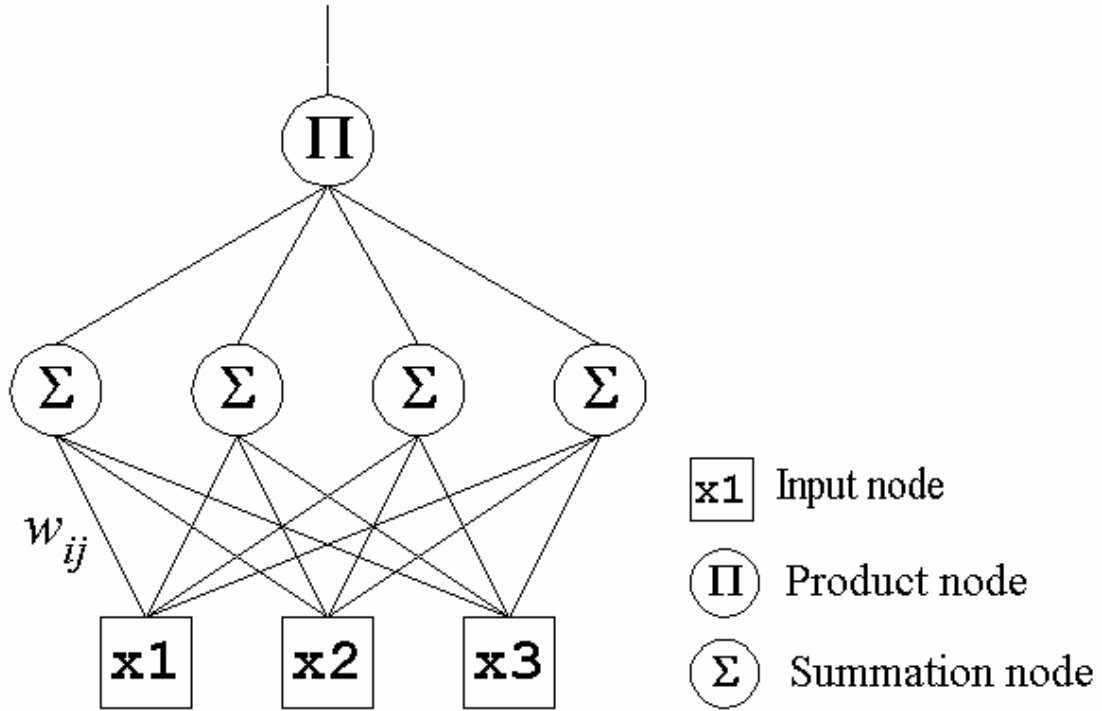
$$P(x) = \sigma\left(\prod_{i=1}^M \left(\sum_{j=1}^d w_j x_j + w_0\right)\right)$$

όπου x είναι το διάνυσμα εισόδου, w_j είναι τα βάρη, M είναι ο αριθμός των νευρώνων στο κρυμμένο επίπεδο και σ είναι η σιγμοειδής συνάρτηση:

$$o = \sigma(s) = 1/(1 + e^{-s})$$

Το πλήθος των βαρών από τις μονάδες εισόδου προς τις κρυμμένες μονάδες είναι γραμμικό στη διάσταση εισόδου. Τα βάρη στις συνδέσεις μεταξύ των κρυμμένων επιπέδων και του επιπέδου εξόδου είναι σταθερά και ίσα με τη μονάδα. Τέτοιου είδους τοπολογία με ένα κρυμμένο επίπεδο δίνει τη δυνατότητα ταχείας μάθησης των βαρών.

Το νευρωνικό δίκτυο pi-sigma διατηρεί την ισχυρή ικανότητα μάθησης των δικτύων πολλαπλών επιπέδων υψηλής τάξης, ενώ παράλληλα αποφεύγει ως ένα βαθμό τη συνδυαστική αύξηση του αριθμού των βαρών και των κρυμμένων επιπέδων, όταν αυξάνεται η διάσταση εισόδου.



Σχήμα 3.11: Το νευρωνικό δίκτυο Pi-Sigma

Ο κανόνας Δέλτα για τα βάρη $i \rightarrow j$ στις συνδέσεις των κόμβων στα κρυμμένα επίπεδα έχει ως εξής:

$$\begin{aligned} \partial E_e / \partial s_j &= (\partial E_e / \partial o_k) (\partial o_k / \partial s_j) \\ &= (y_k - o_k) (\partial o_k / \partial p_j) (\partial p_j / \partial s_j) \\ &= (y_k - o_k) o_k (1 - o_k) (p'_i) \\ &= (y_k - o_k) o_k (1 - o_k) \left(\prod_{i \neq j}^M h_i \right) \end{aligned}$$

όπου το o_k ορίζεται: $o_k = \sigma(p_j) \left(\prod_{i=1}^M h_i \right)$

και η παράγωγος $\partial p_j / \partial s_j = p'_i$ είναι: $p'_j = \prod_{i \neq j}^M h_i$

Ο κανόνας μέγιστης κατάβασης για τα δίκτυα pi-sigma γίνεται:

$$\Delta w_i = \eta (y_k - o_k) o_k (1 - o_k) \left(\prod_{i \neq j}^M h_i \right) x_i$$

Επαυξημένος αλγόριθμος μάθησης Μέγιστης Κατάβασης για τα δίκτυα pi-sigma (Incremental Gradient Descent Learning Algorithm)

Αρχικοποίηση: Παραδείγματα $\{(\mathbf{x}_e, y_e)\}_{e=1}^N$, θέσε τα αρχικά βάρη w_i σε μικρές τυχαίες τιμές, συντελεστής μάθησης $\eta = 0.1$.

Επανάληψη:

για κάθε παράδειγμα εκπαίδευσης (x_e, y_e)

- πολλαπλασίασε την έξοδο: $o = \sigma(\sum_{j=1}^M w_j x_j + w_0)$
- εάν ο perceptron δεν ανταποκριθεί σωστά ενημέρωσε τα βάρη:

$$w_i = w_i + \eta(y_k - o_k) o_k (1 - o_k) (\prod_{i \neq j}^M h_i) x_i$$

ωστότου να ικανοποιηθεί η συνθήκη τερματισμού

Τα νευρωνικά δίκτυα pi-sigma παρέχουν μόνο μια περιορισμένη εκτίμηση των δυναμοσειρών. Εξαιτίας αυτής της μειωμένης ιδιότητας εκτίμησης τα δίκτυα pi-sigma δεν είναι σε θέση να εκτιμήσουν ομοιόμορφα όλες τις συνεχείς συναρτήσεις πολλαπλών μεταβλητών που ορίζονται σε ένα συμπαγές σύνολο (compact set). Ωστόσο, μπορεί να επιτευχθεί γενική εκτίμηση αθροίζοντας τις εξόδους πολλαπλών δικτύων pi-sigma διαφορετικών τάξεων. Το αποτέλεσμα του συνδυασμού των pi-sigma δικτύων καλείται Ridge Polynomial Network (RPN) [22].

3.5.3 Η κατάλληλη αρχιτεκτονική για τα Δίκτυα Υψηλής Τάξης

Το πλήθος των κρυμμένων μονάδων στα νευρωνικά δίκτυα υψηλής τάξης αυξάνεται εκθετικά με τον αριθμό των εισόδων. Ο αντικειμενικός σκοπός για την εύρεση της κατάλληλης αρχιτεκτονικής για τα HONN είναι η εξομάλυνση του προβλήματος της εκθετικής αύξησης και η γενίκευση του δικτύου, δηλαδή η μάθηση συναρτήσεων που παράγουν κατάλληλες αντιδράσεις σε άγνωστα παραδείγματα.

Το πρόβλημα της μάθησης της τοπολογίας και των βαρών των HONN, για τη βελτίωση της ικανότητας γενίκευσης αντιμετωπίζεται με προσεγγίσεις που μπορούν να χωριστούν σε δύο ομάδες:

- προσεγγίσεις κλαδέματος (*pruning approaches*) που ξεκινούν με μεγάλα δίκτυα και κατά την εκπαίδευση διαγράφουν κρυμμένες μονάδες και βάρη
- προσεγγίσεις κατασκευής (*constructive approaches*) που ξεκινούν με μικρά δίκτυα και καθώς αυξάνεται η ακρίβεια προσθέτουν κρυμμένες μονάδες και βάρη

Η εναλλακτική σ' αυτές τις προσεγγίσεις είναι ο εξελικτικός αλγόριθμος που βρίσκει τα βάρη του δικτύου και ταυτόχρονα σμικρύνει και επεκτείνει δυναμικά την τοπολογία του.

3.5.4 Σύγκριση Δικτύων Υψηλής Τάξης με Πολυεπίπεδα Δίκτυα

Τα νευρωνικά δίκτυα υψηλής τάξης χαρακτηρίζονται από διαφορετικές δυνατότητες ταξινόμησης:

- Νευρωνικά Δίκτυα Υψηλής Τάξης (HONN)
 - το *πρώτο επίπεδο* των HONN με τις μονάδες γινομένων δεύτερης τάξης έχει τη δυνατότητα να κατασκευάζει *κοίλες* και *κυρτές περιοχές* (*regions*)
- Νευρωνικά Δίκτυα Πολλαπλών Επιπέδων (MLP)
 - το *πρώτο επίπεδο* μαθαίνει ένα μόνο υπερεπίπεδο
 - το *δεύτερο επίπεδο* κάνει τμηματικά γραμμικές και κυρτές περιοχές
 - το *τρίτο επίπεδο* κάνει κοίλες και κυρτές περιοχές

3.7 Μεθοδολογία κατασκευής ΤΝΔ για πρόβλεψη

Η διαδικασία κατασκευής νευρωνικού δικτύου πρόβλεψης αποτελεί χρονοβόρα και ιδιαίτερα σημαντική εργασία. Ζητήματα μοντελοποίησης που επηρεάζουν την απόδοση των ΤΝΔ πρέπει να ληφθούν υπόψη προσεκτικά. Μια κρίσιμη απόφαση είναι ο προσδιορισμός της κατάλληλης αρχιτεκτονικής, δηλαδή του αριθμού των επιπέδων, των κόμβων κάθε επιπέδου και των τόξων (*arcs*) που συνδέουν τους κόμβους. Άλλες αποφάσεις σχετικά με την αρχιτεκτονική του δικτύου περιλαμβάνουν την επιλογή της συνάρτησης ενεργοποίησης, των κρυμμένων κόμβων και των κόμβων εξόδου, τον αλγόριθμο εκπαίδευσης, τη μετατροπή των δεδομένων ή τις μεθόδους κανονικοποίησης, τα σύνολα εκπαίδευσης και δοκιμής και τα κριτήρια απόδοσης.

1. Η αρχιτεκτονική του δικτύου:

Τα τεχνητά νευρωνικά δίκτυα τυπικά αποτελούνται από επίπεδα και κόμβους. Συγκεκριμένα στο δημοφιλές MLP, όλοι οι κόμβοι εισόδου βρίσκονται σε ένα επίπεδο εισόδου, όλοι οι κόμβοι εξόδου βρίσκονται σε ένα επίπεδο εξόδου και όλοι οι κρυμμένοι κόμβοι κατανέμονται σε ένα ή περισσότερα κρυμμένα επίπεδα ενδιάμεσα. Για το σχεδιασμό ενός MLP πρέπει να προσδιοριστούν οι ακόλουθες μεταβλητές:

- ο αριθμός των κόμβων εισόδου
- ο αριθμός των κρυμμένων επιπέδων και κόμβων
- ο αριθμός των κόμβων εξόδου

Η επιλογή αυτών των παραμέτρων ουσιαστικά εξαρτάται από το πρόβλημα.

Μολονότι υπάρχουν πολλές διαφορετικές προσεγγίσεις για την εύρεση της βέλτιστης αρχιτεκτονικής του ΤΝΔ όπως ο αλγόριθμος κλαδέματος, ο αλγόριθμος πολυωνυμικού χρόνου, η τεχνική κανονικής αποσύνθεσης (canonical decomposition technique) και το κριτήριο πληροφορίας του δικτύου, αυτές οι μέθοδοι είναι συνήθως αρκετά πολύπλοκες εκ φύσεως και δύσκολες στην εφαρμογή. Επιπλέον καμία από αυτές τις μεθόδους δεν μπορεί να εγγυηθεί τη βέλτιστη λύση για όλα τα πραγματικά προβλήματα πρόβλεψης. Μέχρι σήμερα, δεν υπάρχει απλή και σαφής μέθοδος για τον καθορισμό αυτών των παραμέτρων. Οι κατευθυντήριες γραμμές είναι ευρυστικές, είτε βασισμένες σε υποθέσεις που προέρχονται από περιορισμένο αριθμό πειραμάτων. *Συνεπώς η σχεδίαση του ΤΝΔ είναι περισσότερο τέχνη, παρά επιστήμη.*

1.1. Ο αριθμός των κρυμμένων επιπέδων και κόμβων:

Τα κρυμμένα επίπεδα και κόμβοι διαδραματίζουν πολύ σημαντικό ρόλο για σημαντικές εφαρμογές των νευρωνικών δικτύων. Οι κρυμμένοι κόμβοι στο κρυμμένο επίπεδο επιτρέπουν στα νευρωνικά δίκτυα να ανιχνεύουν το χαρακτηριστικό, να συλλαμβάνουν το υπόδειγμα στα δεδομένα και να εκτελούν μη γραμμικές συναρτήσεις μεταξύ των μεταβλητών εισόδου και εξόδου. Είναι προφανές ότι χωρίς τους κρυμμένους κόμβους, τα απλά perceptrons ισοδυναμούν με τα γραμμικά στατιστικά μοντέλα πρόβλεψης.

Επηρεασμένοι από θεωρητικές εργασίες που καταδεικνύουν ότι ένα μόνο κρυμμένο επίπεδο είναι επαρκές ώστε το ΤΝΔ να εκτιμήσει οποιαδήποτε πολύπλοκη μη γραμμική συνάρτηση με την επιθυμητή ακρίβεια, οι περισσότεροι ερευνητές χρησιμοποιούν μόνο ένα κρυμμένο επίπεδο για προγνωστικούς σκοπούς. Ωστόσο, τα δίκτυα με ένα κρυμμένο επίπεδο μπορεί να απαιτούν πολύ μεγάλο πλήθος κρυμμένων κόμβων, το οποίο δημιουργεί αδυναμίες τόσο στο χρόνο εκπαίδευσης όσο και στην ικανότητα γενίκευσης.

Δίκτυα με δύο κρυμμένα επίπεδα είναι πιθανόν να παρέχουν περισσότερα πλεονεκτήματα για ορισμένους τύπους προβλημάτων. Αρκετοί μελετητές

ασχολήθηκαν με τέτοια προβλήματα και θεώρησαν περισσότερα από ένα κρυμμένα επίπεδα (συνήθως δύο κρυμμένα επίπεδα) κατά τη διαδικασία σχεδίασης του δικτύου. Η χρήση δύο κρυμμένων επιπέδων παρατηρήθηκε ότι σε ορισμένες περιπτώσεις είχε ως αποτέλεσμα πιο συμπαγή αρχιτεκτονική η οποία απέφερε μεγαλύτερη αποτελεσματικότητα στη διαδικασία εκπαίδευσης καθώς και την πραγματοποίηση προβλέψεων με μεγαλύτερη ακρίβεια απ' ό,τι τα δίκτυα με ένα κρυμμένο επίπεδο. Ωστόσο στα δίκτυα με περισσότερα από δύο κρυμμένα επίπεδα δεν παρατηρείται κάποια περαιτέρω βελτίωση [9, 23].

Κατά την άποψη μας, ένα κρυμμένο επίπεδο πιθανότατα είναι αρκετό για τα περισσότερα προβλήματα πρόβλεψης. Υπάρχουν αρκετοί λόγοι γιατί πρέπει να χρησιμοποιούμε όσο το δυνατό λιγότερα κρυμμένα επίπεδα:

1. Οι περισσότεροι αλγόριθμοι εκπαίδευσης είναι βαθμωτοί. Το επιπλέον επίπεδο μέσω του οποίου πρέπει να διορθωθούν προς τα πίσω τα σφάλματα προκαλεί αστάθεια στην κατάβαση. Η επιτυχία οποιουδήποτε βαθμωτού αλγόριθμου βελτιστοποίησης εξαρτάται από το βαθμό στον οποίο η κατάβαση παραμένει αμετάβλητη καθώς οι παράμετροι διαφοροποιούνται.
2. Ο αριθμός των τοπικών ελαχίστων αυξάνεται σημαντικά για πολλά κρυμμένα επίπεδα. Οι περισσότεροι βαθμωτοί αλγόριθμοι μπορούν να υπολογίσουν μόνο το τοπικό ελάχιστο, χάνοντας έτσι το γενικό. Ακόμη και αν ο αλγόριθμος είναι σε θέση να βρει το γενικό ελάχιστο, υπάρχει μεγάλη πιθανότητα έπειτα από πολλαπλές και χρονοβόρες επαναλήψεις, το αποτέλεσμα να κολλήσει στο τοπικό ελάχιστο.

Μολαταύτα, η χρήση δύο κρυμμένων επιπέδων μπορεί να αποφέρει καλύτερα αποτελέσματα για συγκεκριμένα προβλήματα, ιδιαίτερα όταν τα δίκτυα ενός επιπέδου αποτυγχάνουν να δώσουν ικανοποιητικό αποτέλεσμα λόγω του υπερβολικού πλήθους κρυμμένων κόμβων [24].

Το ζήτημα του προσδιορισμού του βέλτιστου αριθμού των κρυμμένων κόμβων είναι πολύπλοκο και ιδιαίτερα σημαντικό. Γενικά, τα δίκτυα με λιγότερους κρυμμένους κόμβους προτιμούνται καθώς παρουσιάζουν συνήθως καλύτερη ικανότητα γενίκευσης και λιγότερα προβλήματα υπερεκπαίδευσης. Αλλά τα δίκτυα με πολύ λίγους κρυμμένους κόμβους ίσως να μην έχουν την απαιτούμενη δύναμη μοντελοποίησης και μάθησης των δεδομένων. Δεν υπάρχει θεωρητική βάση για την επιλογή της εν λόγω παραμέτρου, μολονότι έχουν αναφερθεί μερικές συστηματικές

προσεγγίσεις (μέθοδος αποκοπής άχρηστων κρυμμένων κόμβων, μέθοδος πρόσθεσης κρυμμένων κόμβων, έρευνα πλέγματος).

Ο συνηθέστερος τρόπος προσδιορισμού του πλήθους των κρυμμένων κόμβων είναι μέσω πειραμάτων ή με τη μέθοδο δοκιμής και σφάλματος. Αρκετοί εμπειρικοί κανόνες έχουν επίσης προταθεί, όπως ότι ο αριθμός των κρυμμένων κόμβων εξαρτάται από τον αριθμό των υποδειγμάτων εισόδου και κάθε βάρος πρέπει να έχει τουλάχιστον δέκα υποδείγματα εισόδου (μέγεθος δείγματος). Στην προσπάθεια αποφυγής του προβλήματος υπερεκπαίδευσης, ορισμένοι ερευνητές υιοθέτησαν εμπειρικούς κανόνες για τον περιορισμό του αριθμού των κρυμμένων κόμβων. Είναι επίσης γνωστοί ορισμένοι ευριστικοί περιορισμοί σχετικά με τον αριθμό των κρυμμένων κόμβων. Παραδείγματος χάριν στην περίπτωση των δημοφιλών δικτύων με ένα κρυμμένο επίπεδο, υπάρχουν αρκετές κατευθυντήριες γραμμές, όπως οι κανόνες “ $2n + 1$ ”, “ $2n$ ”, “ n ”, “ $n/2$ ”, όπου n ο αριθμός των κόμβων εισόδου. Ωστόσο καμία από αυτές τις ευριστικές επιλογές δε λειτουργεί ικανοποιητικά για όλα τα προβλήματα [9].

1.2. Ο αριθμός των κόμβων εισόδου

Ο αριθμός των κόμβων εισόδου ανταποκρίνεται στον αριθμό των μεταβλητών στο διάνυσμα εισόδου που χρησιμοποιείται για την πρόβλεψη των μελλοντικών τιμών. Για περιπτώσεις συνήθους πρόβλεψης, ο αριθμός των κόμβων εισόδου είναι συχνά προφανής και επιλέγεται σχετικά εύκολα. Στα προβλήματα πρόβλεψης χρονοσειρών, ο αριθμός των κόμβων εισόδου ανταποκρίνεται στον αριθμό των παρελθοντικών παρατηρήσεων που χρησιμοποιούνται για τη σύλληψη των υποκείμενων υποδειγμάτων στις χρονοσειρές και για την πραγματοποίηση προβλέψεων για μελλοντικές τιμές. Ωστόσο, επί του παρόντος δεν υφίσταται προτεινόμενος συστηματικός τρόπος προσδιορισμού του ακριβούς πλήθους. Η επιλογή της εν λόγω παραμέτρου θα πρέπει να συμπεριλαμβάνεται στη διαδικασία κατασκευής του μοντέλου. Για καλύτερα αποτελέσματα, επιθυμούμε έναν μικρό αριθμό αναγκαίων κόμβων που μπορούν να αποκαλύψουν τα μοναδικά χαρακτηριστικά των δεδομένων. Πολύ λίγοι ή υπερβολικά πολλοί κόμβοι εισόδου ίσως να επηρεάσουν είτε την ικανότητα μάθησης είτε την ικανότητα πρόβλεψης του δικτύου.

Οι περισσότεροι μελετητές εκτελούν πειράματα για την επιλογή του αριθμού των κόμβων εισόδου, ενώ άλλοι υιοθετούν κάποιες διαισθητικές ή εμπειρικές ιδέες.

Κατά τη γνώμη μας, ο αριθμός των κόμβων εισόδου είναι πιθανότατα η πλέον κρίσιμη απόφαση σχετικά με τις μεταβλητές του προβλήματος πρόβλεψης χρονοσειρών, αφού περιέχει σημαντικές πληροφορίες σχετικά με την πολύπλοκη (γραμμική και/ ή μη γραμμική) δομή των δεδομένων. Πιστεύουμε ότι αυτή η παράμετρος μπορεί να προσδιοριστεί μέσω θεωρητικής έρευνας στην ανάλυση μη γραμμικών χρονοσειρών, βελτιώνοντας έτσι τη διαδικασία μοντελοποίησης του νευρωνικού δικτύου. Προσφάτως, ιδιαίτερη έμφαση έχει δοθεί στους γενετικούς αλγόριθμους για το βέλτιστο σχεδιασμό του νευρωνικού δικτύου. Οι γενετικοί αλγόριθμοι αποτελούν διαδικασίες βελτιστοποίησης που είναι σε θέση να μιμηθούν τη φυσική επιλογή και τη βιολογική εξέλιξη ώστε να επιτύχουν πιο αποδοτικές διαδικασίες μάθησης για τα ΤΝΔ. Εξαιτίας των μοναδικών τους ιδιοτήτων, οι γενετικοί αλγόριθμοι εφαρμόζονται συχνά σε εμπορικά πακέτα λογισμικού ΤΝΔ [9, 25].

1.3. Ο αριθμός των κόμβων εξόδου

Ο αριθμός των κόμβων εξόδου είναι αρκετά εύκολο να προσδιοριστεί καθώς σχετίζεται άμεσα με το υπό μελέτη πρόβλημα. Για τα προβλήματα πρόβλεψης χρονοσειρών, ο αριθμός των κόμβων εξόδου συχνά ανταποκρίνεται στον ορίζοντα πρόβλεψης. Υπάρχουν δύο τύποι πρόβλεψης: ενός βήματος έμπροσθεν (one-step-ahead) (ο οποίος χρησιμοποιεί έναν κόμβο εξόδου) και η πρόβλεψη πολλαπλών βημάτων έμπροσθεν (multi-step-ahead). Έχουν σημειωθεί δύο τρόποι υλοποίησης πρόβλεψης πολλαπλών βημάτων. Ο πρώτος καλείται επαναληπτική πρόβλεψη καθώς χρησιμοποιείται στο μοντέλο Box-Jenkins στο οποίο οι προγνωστικές τιμές χρησιμοποιούνται επαναληπτικά ως είσοδοι στις επόμενες προβλέψεις. Εν προκειμένω, είναι απαραίτητος μόνο ένας κόμβος εξόδου. Ο δεύτερος τρόπος καλείται άμεση μέθοδος και σύμφωνα με τον οποίο το νευρωνικό δίκτυο έχει αρκετούς κόμβους για άμεση πρόβλεψη σε κάθε βήμα στο μέλλον [9].

Κατά τη γνώμη μας η άμεση πρόβλεψη πολλαπλών περιόδων νευρωνικών δικτύων είναι η καλύτερη μέθοδος (αρκετοί κόμβοι εξόδου) για τους ακόλουθους δύο λόγους. Πρώτον το νευρωνικό δίκτυο μπορεί να κατασκευασθεί άμεσα ώστε να προβλέψει τιμές πολλαπλών βημάτων. Πλεονεκτεί έναντι των επαναληπτικών μεθόδων όπως της Box-Jenkins με την έννοια ότι η επαναληπτική μέθοδος κατασκευάζει μία μόνο συνάρτηση η οποία χρησιμοποιείται για την πρόβλεψη ενός σημείου κάθε φορά και έπειτα επαναλαμβάνει τη συνάρτηση με τις δικές της εξόδους

για να προβλέψει μελλοντικά σημεία. Καθώς οι προβλέψεις τρέχουν, οι παρελθοντικές παρατηρήσεις λιγοστεύουν. Αντίθετα, οι προβλέψεις και όχι οι παρατηρήσεις χρησιμοποιούνται για την πρόγνωση περαιτέρω μελλοντικών στοιχείων. Συνεπώς όσο μεγαλύτερος είναι ο ορίζοντας πρόβλεψης, τόσο πιο ανακριβής είναι η επαναληπτική μέθοδος. Αυτό επίσης εξηγεί γιατί το μοντέλο Box-Jenkins είναι παραδοσιακά πιο κατάλληλο για βραχυπρόθεσμη πρόβλεψη. Δεύτερον, η μεθοδολογία Box-Jenkins βασίζεται σε μεγάλο βαθμό στην αυτοσυσχέτιση μεταξύ των δεδομένων χρονικών στιγμών πίσω. Αξιοσημείωτο είναι ότι η αυτοσυσχέτιση στην ουσία μετρά μόνο τη γραμμική συσχέτιση μεταξύ των παρελθοντικών δεδομένων. Στην πραγματικότητα, η συσχέτιση είναι μη γραμμική και τα μοντέλα Box-Jenkins δεν είναι σε θέση να μοντελοποιήσουν αυτές τις μη γραμμικές σχέσεις. Τα ΤΝΔ συλλαμβάνουν καλύτερα τις μη γραμμικές σχέσεις στα δεδομένα.

1.4. Οι αλληλοσυνδέσεις των κόμβων

Η αρχιτεκτονική του δικτύου χαρακτηρίζεται επίσης από τις αλληλοσυνδέσεις των κόμβων στα διάφορα επίπεδα. Οι συνδέσεις μεταξύ των κόμβων προσδιορίζουν αποφασιστικά τη συμπεριφορά του δικτύου. Για τις περισσότερες εργασίες πρόβλεψης όπως και για τις άλλες εφαρμογές, τα δίκτυα είναι πλήρως συνδεδεμένα, καθώς όλοι οι κόμβοι σε κάθε επίπεδο είναι πλήρως συνδεδεμένοι μόνο με τους κόμβους του επόμενου υψηλότερου επιπέδου, με μόνη εξαίρεση το επίπεδο εξόδου. Ωστόσο, είναι πιθανή η ύπαρξη σποραδικά συνδεδεμένων δικτύων ή δικτύων με άμεσες συνδέσεις από τους κόμβους εισόδου προς τους κόμβους εξόδου. Η πρόσθεση άμεσων συνδετικών στοιχείων μεταξύ των επιπέδων εισόδου και εξόδου μπορεί να αποδειχθεί ωφέλιμη για την ακρίβεια πρόβλεψης, αφού τα εν λόγω στοιχεία δύναται να χρησιμοποιηθούν για τη μοντελοποίηση των γραμμικών δομών των δεδομένων και για την ενίσχυση της δύναμης αναγνώρισης του δικτύου [9].

2. Η συνάρτηση ενεργοποίησης

Η συνάρτηση ενεργοποίησης καλείται επίσης και συνάρτηση μεταφοράς. Προσδιορίζει τη σχέση μεταξύ των εισόδων και εξόδων στους κόμβους και στο δίκτυο. Γενικά, η συνάρτηση ενεργοποίησης εισάγει ένα βαθμό μη γραμμικότητας που είναι πολύτιμος για τις περισσότερες εφαρμογές ΤΝΔ. Θεωρητικά και χωρίς να είμαστε ιδιαίτερα αυστηροί, κάθε διαφορίσιμη συνάρτηση μπορεί να λειτουργήσει ως

συνάρτηση ενεργοποίησης. Στην πράξη, χρησιμοποιείται μόνο ένας μικρός αριθμός από συναρτήσεις ενεργοποίησης με «καλή συμπεριφορά» (πεπερασμένες, μονότονα αύξουσες και διαφορίσιμες). Οι κυριότερες είναι:

1. Η σιγμοειδής (logistic) συνάρτηση:

$$f(x) = (1 + \exp(-x))^{-1}$$

2. Η συνάρτηση υπερβολικής εφαπτομένης (tanh):

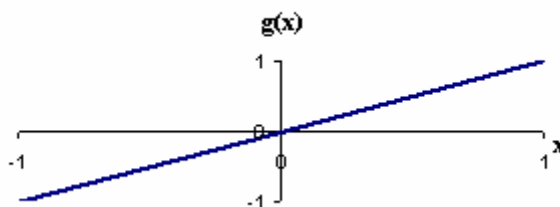
$$f(x) = (\exp(x) - \exp(-x)) / (\exp(x) + \exp(-x))$$

3. Η συνάρτηση ημίτονου ή συνημίτονου:

$$f(x) = \sin(x) \text{ ή } f(x) = \cos(x)$$

4. Η γραμμική συνάρτηση ή ταυτότητα:

$$f(x) = x$$



Σχήμα 3.12: Η γραμμική συνάρτηση ή ταυτότητα

Από τις παραπάνω, η συνάρτηση μεταφοράς logistic είναι η δημοφιλέστερη επιλογή.

Υπάρχουν ορισμένοι ευριστικοί κανόνες για την επιλογή της συνάρτησης ενεργοποίησης. Για παράδειγμα, ορισμένοι ερευνητές προτείνουν τις logistic συναρτήσεις ενεργοποίησης για προβλήματα ταξινόμησης που περιλαμβάνουν μάθηση σχετικά με τη μέση συμπεριφορά και τις συναρτήσεις υπερβολικής εφαπτομένης εάν τα προβλήματα περιλαμβάνουν μάθηση σχετικά με αποκλίσεις από το μέσο όπως τα προβλήματα πρόβλεψης. Ωστόσο, δεν είναι σαφές κατά πόσον διαφορετικές συναρτήσεις ενεργοποίησης επιδρούν σημαντικά στην απόδοση του δικτύου.

Γενικά, ένα δίκτυο μπορεί να έχει διαφορετικές συναρτήσεις ενεργοποίησης για διαφορετικούς κόμβους στο ίδιο ή σε διαφορετικά επίπεδα. Εντούτοις σχεδόν όλα τα δίκτυα χρησιμοποιούν τις ίδιες συναρτήσεις ενεργοποίησης ειδικά για τους κόμβους του ιδίου επιπέδου. Αν και στην πλειοψηφία των εφαρμογών χρησιμοποιείται η logistic συνάρτηση ενεργοποίησης για τους κρυμμένους κόμβους,

δεν υπάρχει απόλυτη συμφωνία για το είδος της συνάρτησης ενεργοποίησης που πρέπει να χρησιμοποιείται στους κόμβους εξόδου. Αξιοσημείωτο είναι ότι όταν χρησιμοποιούνται μη γραμμικές συναρτήσεις στο επίπεδο εξόδου (logistic, υπερβολικής εφαστομένης), οι επιθυμητές τιμές εξόδου συνήθως πρέπει να κανονικοποιηθούν ώστε να εναρμονίζονται με την κλίμακα των πραγματικών εξόδων του δικτύου, αφού ο κόμβος εξόδου με τη συνάρτηση logistic ή υπερβολικής εφαστομένης έχει τυπικό πεδίο $[0,1]$ ή $[-1,1]$ αντίστοιχα.

Συμβατικά, η logistic συνάρτηση ενεργοποίησης δείχνει κατάλληλη για τους κόμβους εξόδου για πολλά προβλήματα ταξινόμησης όπου οι επιθυμητές τιμές είναι συχνά δυαδικές. Ωστόσο, για ένα πρόβλημα πρόβλεψης που περιλαμβάνει συνεχείς επιθυμητές τιμές, είναι λογικό να χρησιμοποιηθεί η γραμμική συνάρτηση ενεργοποίησης για τους κόμβους εξόδου. Αξιοσημείωτο είναι ότι τα feedforward νευρωνικά δίκτυα με γραμμικούς κόμβους εξόδου υπόκεινται στον περιορισμό ότι δεν μπορούν να μοντελοποιήσουν χρονοσειρές που περιέχουν τάση (trend). Συνεπώς, για αυτού του είδους νευρωνικά δίκτυα, απαιτείται προ-διαφοροποίηση ώστε να εξαλειφθούν οι τάσεις. Μέχρι σήμερα δεν έχει διερευνηθεί η συγκριτική απόδοση της χρήσης γραμμικών και μη γραμμικών συναρτήσεων ενεργοποίησης για τους κόμβους εξόδου και δεν έχουν καταγραφεί εμπειρικά αποτελέσματα για την υποστήριξη ιδιαίτερης προτίμησης σε καμία εκ των δύο [9, 26, 27].

3. Ο αλγόριθμος εκπαίδευσης

Η εκπαίδευση του νευρωνικού δικτύου αποτελεί ένα μη γραμμικό πρόβλημα ελαχιστοποίησης χωρίς περιορισμούς, στο οποίο τα συναπτικά βάρη του δικτύου τροποποιούνται επαναληπτικά ώστε να ελαχιστοποιήσουν το ολικό μέσο τετραγωνικό σφάλμα, μεταξύ των επιθυμητών και των πραγματικών τιμών εξόδου, για όλους τους κόμβους εξόδου και όλα τα υποδείγματα εισόδου. Η ύπαρξη πολλών διαφορετικών μεθόδων βελτιστοποίησης παρέχει διάφορες επιλογές για την εκπαίδευση του νευρωνικού δικτύου. Την τρέχουσα περίοδο, δεν υπάρχει αλγόριθμος ο οποίος να εγγυάται τη γενική βέλτιστη λύση για ένα γενικό μη γραμμικό πρόβλημα βελτιστοποίησης σε λογική χρονική διάρκεια. Έτσι όλοι οι αλγόριθμοι βελτιστοποίησης στην πράξη αναπόφευκτα πάσχουν από τα τοπικά προβλήματα βελτιστοποίησης και ό,τι καλύτερο μπορεί να γίνει στον τομέα αυτό είναι η χρήση

των διαθέσιμων μεθόδων βελτιστοποίησης που παρέχουν το “καλύτερο” τοπικό βέλτιστο, εάν η πραγματική γενική λύση είναι αδύνατο να επιτευχθεί.

Η πιο δημοφιλής μέθοδος εκπαίδευσης είναι ο αλγόριθμος backpropagation ο οποίος αποτελεί ουσιαστικά μια μέθοδο μέγιστης κατάβασης (gradient steepest descent method). Για τον αλγόριθμο μέγιστης κατάβασης πρέπει να προσδιοριστεί ένα μέγεθος βήματος, το οποίο καλείται συντελεστής μάθησης (learning rate) στην ορολογία ΤΝΔ. Ο συντελεστής μάθησης είναι αποφασιστικής σημασίας για τον αλγόριθμο μάθησης backpropagation αφού καθορίζει το μέγεθος των μεταβολών στα βάρη. Είναι ευρέως γνωστό ότι η μέθοδος της μέγιστης κατάβασης έχει τα μειονεκτήματα της αργής σύγκλισης, ανεπάρκειας και έλλειψη ευρωστίας. Επιπλέον είναι πολύ ευαίσθητη στην επιλογή του συντελεστή μάθησης. Μικρότεροι συντελεστές τείνουν να επιβραδύνουν τη διαδικασία μάθησης, ενώ μεγαλύτεροι συντελεστές μπορεί να προκαλέσουν ταλάντωση του δικτύου στο χώρο των βαρών. Ένας τρόπος βελτίωσης της πρότυπης μεθόδου μέγιστης κατάβασης είναι να συμπεριληφθεί μια πρόσθετη παράμετρος *ορμής (momentum)* ώστε να είναι εφικτοί μεγαλύτεροι συντελεστές μάθησης με αποτέλεσμα την ταχύτερη σύγκλιση και παράλληλα τη μείωση της τάσης για ταλάντωση. Το σκεπτικό της εισαγωγής της ορμής είναι να πραγματοποιηθεί η επόμενη αλλαγή βάρους στην ίδια κατεύθυνση με την προηγούμενη.

Η πρότυπη τεχνική backpropagation με ορμή χρησιμοποιείται από τους περισσότερους ερευνητές. Εφόσον υπάρχουν λίγοι συστηματικοί τρόποι επιλογής του συντελεστή μάθησης και της ορμής ταυτόχρονα, οι “καλύτερες” τιμές αυτών των παραμέτρων μάθησης επιλέγονται συνήθως μέσω πειραματισμών. Καθώς ο ρυθμός μάθησης και η ορμή είναι δυνατό να πάρουν οποιαδήποτε τιμή μεταξύ 0 και 1, είναι πρακτικά αδύνατο να πραγματοποιηθεί εξαντλητική έρευνα για την εύρεση των καλύτερων συνδυασμών αυτών των παραμέτρων εκπαίδευσης. Μόνο προεπιλεγμένες τιμές μελετώνται από τους ερευνητές. Για παράδειγμα, οι Sharda και Patil (1992) δοκιμάζουν εννέα συνδυασμούς τριών ρυθμών εκπαίδευσης (0.1, 0.5, 0.9) και τριών τιμών ορμής (0.1, 0.5, 0.9). Γενικά οι παράμετροι εκπαίδευσης διαδραματίζουν αποφασιστικό ρόλο στην απόδοση των ΤΝΔ. Ωστόσο, υπάρχουν αντιφατικά συμπεράσματα σχετικά με τις καλύτερες παραμέτρους μάθησης, γεγονός το οποίο κατά τη γνώμη μας οφείλεται στην ανεπάρκεια και στην έλλειψη ευρωστίας του αλγόριθμου μέγιστης κατάβασης.

Εξαιτίας της ανεπάρκειας του συμβατικού αλγορίθμου backpropagation, έχουν προταθεί αρκετές διαφοροποιήσεις ή τροποποιήσεις του backpropagation, όπως η προσαρμοστική μέθοδος, η quickprop και μέθοδοι δεύτερης τάξης. Η τελευταία είναι η πιο αποτελεσματική μη γραμμική μέθοδος βελτιστοποίησης και χρησιμοποιείται στα περισσότερα πακέτα βελτιστοποίησης. Τα χαρακτηριστικά που διαθέτει όπως γρηγορότερη σύγκλιση, ευρωστία και η ικανότητα εύρεσης καλού τοπικού ελαχίστου την καθιστούν δημοφιλή στην εκπαίδευση των ΤΝΔ.

Προσφάτως ερευνητές, πρότειναν τη χρήση ενός μη γραμμικού βελτιστοποιητή γενικού σκοπού, τον GRG2, στην εκπαίδευση των δικτύων. Τα πλεονεκτήματα του GRG2 έχουν προσδιοριστεί για διάφορα προβλήματα ΤΝΔ. Ο GRG2 αποτελεί ένα ευρέως διαθέσιμο λογισμικό βελτιστοποίησης το οποίο επιλύει μη γραμμικά προβλήματα βελτιστοποίησης χρησιμοποιώντας τη γενικευμένη μειωμένη μέθοδο της μέγιστης κατάβασης (reduced gradient method). Με τον GRG2 δεν είναι αναγκαία η επιλογή παραμέτρων μάθησης όπως ο συντελεστής μάθησης και η ορμή. Εν αντιθέσει, ένα διαφορετικό σύνολο παραμέτρων, όπως τα κριτήρια τερματισμού, η διαδικασία αναζήτησης κατεύθυνσης και τα όρια των μεταβλητών, πρέπει να προσδιοριστούν και είναι δυνατό να τεθούν ως προκαθορισμένες τιμές.

Ένα ακόμη σχετικό ζήτημα στην εκπαίδευση των ΤΝΔ είναι ο προσδιορισμός μιας αντικειμενικής συνάρτησης ή συνάρτησης κόστους. Συνήθως χρησιμοποιούνται οι SSE και MSE, αφού ορίζονται σε σχέση με τα σφάλματα. Άλλες αντικειμενικές συναρτήσεις όπως του μεγίστου αποτελέσματος, κέρδους ή ωφέλειας μπορεί να είναι πιο κατάλληλες για προβλήματα οικονομικής πρόβλεψης. Η επιλογή μιας συνάρτησης κόστους μπορεί να επηρεάσει σημαντικά την απόδοση πρόβλεψης του δικτύου εάν ο αλγόριθμος μάθησης (backpropagation) και άλλες παράμετροι του δικτύου είναι αμετάβλητες. Συνεπώς, ένας πιθανός τρόπος άμεσης αντιμετώπισης της τελικής αντικειμενικής συνάρτησης είναι η αλλαγή του αλγορίθμου αναζήτησης από backpropagation σε γενετικούς αλγόριθμους, ή προσομοιωμένη απόσβεση (simulated annealing) και άλλες μεθόδους βελτιστοποίησης οι οποίες επιτρέπουν αναζήτηση σε αυθαίρετες συναρτήσεις χρησιμότητας [9].

4. Κανονικοποίηση των δεδομένων

Οι μη γραμμικές συναρτήσεις ενεργοποίησης όπως η συνάρτηση logistic περιορίζουν την πιθανή έξοδο ενός κόμβου στο διάστημα (0,1) ή (-1,1). Η

κανονικοποίηση των δεδομένων συνήθως εκτελείται πριν την έναρξη της διαδικασίας εκπαίδευσης. Όπως προαναφέρθηκε, όταν μη γραμμικές συναρτήσεις μεταφοράς χρησιμοποιούνται στους κόμβους εξόδου, οι επιθυμητές τιμές εξόδου πρέπει να μετασχηματίζονται στο πεδίο τιμών των πραγματικών εξόδων του δικτύου. Ακόμη κι αν χρησιμοποιείται μια γραμμική συνάρτηση μεταφοράς εξόδου μπορεί να είναι ωφέλιμο να τυποποιηθούν οι εξοδοί όπως και οι εισοδοί για την αποφυγή υπολογιστικών προβλημάτων, την ικανοποίηση των απαιτήσεων του αλγορίθμου και τη διευκόλυνση της μάθησης του δικτύου.

Παρακάτω συνοψίζονται τέσσερις μέθοδοι για την κανονικοποίηση των εισόδων:

1. Κανονικοποίηση along channel: Ως δίοδος (channel) χαρακτηρίζεται ένα σύνολο στοιχείων στην ίδια θέση για όλα τα διανύσματα εισόδου στο σύνολο εκπαίδευσης ή δοκιμής. Δηλαδή, κάθε δίοδος μπορεί να θεωρηθεί ως “ανεξάρτητη” μεταβλητή εισόδου. Η κανονικοποίηση along channel εκτελείται από στήλη προς στήλη εάν τα διανύσματα εισόδου τοποθετηθούν σε μαθηματικό πίνακα. Με άλλα λόγια, κανονικοποιεί κάθε μεταβλητή εισόδου ατομικά.
2. Κανονικοποίηση across channel: Ο εν λόγω τύπος κανονικοποίησης εκτελείται για κάθε διάνυσμα εισόδου ατομικά, δηλαδή, η κανονικοποίηση πραγματοποιείται δια μέσου όλων των στοιχείων στο υπόδειγμα δεδομένων.
3. Κανονικοποίηση mixed channel: Αυτή η μέθοδος όπως υποδηλώνει και η ονομασία της, χρησιμοποιεί συνδυασμό των κανονικοποιήσεων along και across.
4. Εξωτερική κανονικοποίηση: Όλα τα δεδομένα εκπαίδευσης κανονικοποιούνται σε συγκεκριμένο πεδίο τιμών.

Η επιλογή μίας εκ των παραπάνω μεθόδων εξαρτάται συνήθως από τη σύνθεση του διανύσματος εισόδου. Για τα προβλήματα πρόβλεψης χρονοσειρών, η εξωτερική κανονικοποίηση αποτελεί τις περισσότερες φορές τη μόνη κατάλληλη διαδικασία κανονικοποίησης. Οι παρελθοντικές (time lagged) παρατηρήσεις από την ίδια πηγή χρησιμοποιούνται ως μεταβλητές εισόδου και μπορούν να διατηρήσουν τη δομή μεταξύ των διόδων όπως στις αρχικές σειρές. Για τυχαία προβλήματα πρόβλεψης ωστόσο, η μέθοδος που ενδείκνυται είναι η κανονικοποίηση along channel, αφού οι μεταβλητές εισόδου είναι τυπικά οι ανεξάρτητες μεταβλητές που χρησιμοποιούνται για την πρόβλεψη των εξαρτημένων μεταβλητών. Η

κανονικοποίηση across channel για δεδομένα χρονοσειρών μπορεί να προκαλέσει σοβαρό πρόβλημα, αφού τα ίδια δεδομένα σε διαφορετικά υποδείγματα εκπαίδευσης κανονικοποιούνται διαφορετικά και συνεπώς μπορεί να χαθούν πολύτιμες πληροφορίες στην υποκείμενη δομή των αρχικών χρονοσειρών.

Για κάθε μία από τις παραπάνω μεθόδους χρησιμοποιούνται οι ακόλουθοι τύποι:

- γραμμικός μετασχηματισμός στο $[0,1]$: $x_n = (x_0 - x_{\min}) / (x_{\max} - x_{\min})$
- γραμμικός μετασχηματισμός στο $[a,b]$:

$$x_n = (b - a)(x_0 - x_{\min}) / (x_{\max} - x_{\min}) + a$$
- στατιστική κανονικοποίηση: $x_n = (x_0 - \bar{x}) / s$
- απλή κανονικοποίηση: $x_n = x_0 / x_{\max}$

όπου x_n και x_0 είναι τα κανονικοποιημένα και τα αρχικά δεδομένα, x_{\min} , x_{\max} , \bar{x} και s είναι η ελάχιστη, η μέγιστη, η μέση και η τυπική απόκλιση αντίστοιχα κατά μήκος των στηλών ή των σειρών.

Δεν είναι σαφές εάν είναι αναγκαία η κανονικοποίηση των εισόδων, αφού τα συναπτικά βάρη μπορεί να ακυρώσουν την κανονικοποίηση. Υπάρχουν αρκετές μελέτες σχετικά με την επίδραση της κανονικοποίησης των δεδομένων στην μάθηση του δικτύου. Ορισμένοι ερευνητές συμπέραναν ότι γενικά η κανονικοποίηση των δεδομένων είναι ευεργετική, όσον αφορά το ρυθμό ταξινόμησης και το μέσο τετραγωνικό σφάλμα, αλλά τα πλεονεκτήματα ελαττώνονται καθώς αυξάνονται το μέγεθος του δικτύου και του δείγματος. Επιπλέον, η κανονικοποίηση των δεδομένων συνήθως επιβραδύνει τη διαδικασία εκπαίδευσης. Προσφάτως προτάθηκε μια αυτόματη μέθοδο κανονικοποίησης, ο κανόνας gamma learning, που επιτρέπει στο δίκτυο να αυτο-κανονικοποιείται κατά τη διαδικασία μάθησης και εξαλείφει την ανάγκη για κανονικοποίηση των δεδομένων πριν την εκπαίδευση.

Η κανονικοποίηση των τιμών εξόδου (targets) είναι συνήθως ανεξάρτητη από την κανονικοποίηση των εισόδων. Ωστόσο, για προβλήματα πρόβλεψης χρονοσειρών η κανονικοποίηση των τιμών εξόδου εκτελείται τις περισσότερες φορές μαζί με τις εισόδους. Η επιλογή του πεδίου ορισμού στο οποίο οι είσοδοι και οι εξοδοί θα κανονικοποιηθούν εξαρτάται κυρίως από τη συνάρτηση ενεργοποίησης των κόμβων εξόδου, με συνηθέστερα το $[0,1]$ για τη συνάρτηση logistic και το $[-1,1]$ για τη συνάρτηση υπερβολικής εφαπτομένης. Αρκετοί μελετητές κλιμάκωσαν τα δεδομένα

μόνο στο πεδίο ορισμού $[0.1, 0.9]$ ή $[0.2, 0.8]$ βάσει του γεγονότος ότι οι μη γραμμικές συναρτήσεις ενεργοποίησης έχουν συνήθως ασυμπτωτικά όρια (φτάνουν στα όρια μόνο για άπειρες τιμές δεδομένων) και της υπόθεσης ότι οι πιθανές έξοδοι βρίσκονται, για παράδειγμα μόνο στο $[0.1, 0.9]$ ή ακόμη στο $[0.2, 0.8]$ για τη συνάρτηση logistic. Ωστόσο, είναι προφανές ότι κάτι τέτοιο δεν αληθεύει πάντα, αφού η έξοδος από έναν κόμβο logistic μπορεί να είναι τόσο μικρή όπως 0.000045 ή τόσο μεγάλη όσο 0.99995 για την είσοδο -10 ή 10 αντίστοιχα.

Αξίζει να σημειωθεί ότι ως συνέπεια της κανονικοποίησης των τιμών εξόδου, η έξοδος του δικτύου θα ανταποκρίνεται στο κανονικοποιημένο πεδίο ορισμού. Επομένως, για να ερμηνευτούν τα αποτελέσματα του δικτύου, πρέπει να ανακλιμακωθούν οι έξοδοι στο αρχικό πεδίο ορισμού. Από την οπτική του χρήστη, η ακρίβεια που παρέχουν τα ΤΝΔ πρέπει να βασίζεται στο ανακλιμακωμένο σύνολο δεδομένων, όπως επίσης και τα κριτήρια απόδοσης πρέπει να υπολογίζονται βάσει των ανακλιμακωμένων εξόδων. Εντούτοις, μόνο λίγοι μελετητές δηλώνουν καθαρά αν τα μέτρα απόδοσης υπολογίστηκαν στην αρχική ή μετασχηματισμένη κλίμακα [9].

5. Τα δείγματα εκπαίδευσης και δοκιμής

Όπως προαναφέρθηκε, για τη δόμηση ΤΝΔ πρόβλεψης απαιτούνται δείγματα εκπαίδευσης και δοκιμής. Το δείγμα εκπαίδευσης χρησιμοποιείται για την ανάπτυξη του μοντέλου ΤΝΔ και το δείγμα δοκιμής για την εκτίμηση της ικανότητας πρόβλεψης του μοντέλου. Ορισμένες φορές χρησιμοποιείται επίσης ένα τρίτο δείγμα, το δείγμα ελέγχου αξιοπιστίας (validation sample), ώστε να αποφευχθεί το πρόβλημα υπερεκπαίδευσης ή να προσδιοριστεί το σημείο τερματισμού της διαδικασίας εκπαίδευσης. Είναι συνήθης η χρήση ενός συνόλου δοκιμής εξίσου για σκοπούς ελέγχου αξιοπιστίας και δοκιμής, ιδιαίτερα για μικρά σύνολα δεδομένων. Κατά τη γνώμη μας, η επιλογή των δειγμάτων εκπαίδευσης και δοκιμής μπορεί να επηρεάσει την απόδοση των ΤΝΔ.

Το κυρίαρχο ζήτημα εδώ είναι η υποδιαίρεση των δεδομένων στα σύνολα εκπαίδευσης και δοκιμής. Μολονότι δεν υπάρχει γενικός κανόνας λύσης στο εν λόγω πρόβλημα, πρέπει να ληφθούν υπόψη στη απόφαση αρκετοί παράγοντες, όπως τα χαρακτηριστικά του προβλήματος, καθώς και το είδος και το μέγεθος των δεδομένων. Είναι αποφασιστικής σημασίας τόσο το σύνολο εκπαίδευσης όσο και το σύνολο δοκιμής να είναι αντιπροσωπευτικά του δείγματος ή του υποκείμενου μηχανισμού,

ιδιαίτερως για προβλήματα πρόβλεψης χρονοσειρών. Ακατάλληλος διαχωρισμός των συνόλων εκπαίδευσης και δοκιμής θα επηρεάσει την επιλογή της βέλτιστης δομής ΤΝΔ και την εκτίμηση της προγνωστικής απόδοσης του ΤΝΔ.

Οι έρευνες που έχουν διεκπεραιωθεί προσφέρουν ελάχιστη καθοδήγηση στην επιλογή των δειγμάτων εκπαίδευσης και δοκιμής. Οι περισσότεροι μελετητές επιλέγουν βάσει των κανόνων 90% έναντι 10%, 80% έναντι 20% ή 70% έναντι 30% κτλ. Ορισμένοι επιλέγουν βάσει του προβλήματος που μελετάται.

Ένας ακόμη σχετικός παράγοντας είναι το μέγεθος του δείγματος. Δεν υπάρχει συγκεκριμένος κανόνας για τις απαιτήσεις του μεγέθους δείγματος για κάθε πρόβλημα. Η ποσότητα των δεδομένων για την εκπαίδευση του δικτύου εξαρτάται από τη δομή του δικτύου, τη μέθοδο εκπαίδευσης και την πολυπλοκότητα του εκάστοτε προβλήματος ή την ποσότητα του θορύβου στα διαθέσιμα δεδομένα. Γενικά, όπως σε κάθε στατιστική προσέγγιση, το μέγεθος δείγματος σχετίζεται στενά με την απαιτούμενη ακρίβεια του προβλήματος. Όσο μεγαλύτερο το μέγεθος δείγματος, τόσο πιο ακριβή θα είναι τα αποτελέσματα. Έχει καταγραφεί σε ερευνητικές μελέτες ότι καθώς αυξάνεται το μέγεθος, τα ΤΝΔ πρόβλεψης αποδίδουν καλύτερα.

Δεδομένου ενός συγκεκριμένου επιπέδου ακρίβειας, ένα μεγαλύτερο δείγμα απαιτείται, καθώς γίνεται πιο πολύπλοκη η υποκείμενη σχέση μεταξύ των εισόδων και εξόδων ή αυξάνεται ο θόρυβος στα δεδομένα. Ωστόσο, στην πραγματικότητα, το μέγεθος του δείγματος περιορίζεται από τη διαθεσιμότητα των δεδομένων. Η ακρίβεια ενός συγκεκριμένου προβλήματος πρόβλεψης μπορεί επίσης να επηρεάζεται από το μέγεθος του δείγματος που χρησιμοποιείται στο σύνολο εκπαίδευσης και/ ή στο σύνολο δοκιμής.

Αξιοσημείωτο είναι ότι στα πραγματικά προβλήματα υπάρχουν όρια για την ακρίβεια που μπορεί να πετύχει κάθε μοντέλο. Για παράδειγμα, εάν θεωρήσουμε μόνο δύο παράγοντες: το θόρυβο των δεδομένων και το υποκείμενο μοντέλο, τότε το όριο ακριβείας ενός γραμμικού μοντέλου όπως το Box-Jenkins καθορίζεται από το θόρυβο στα δεδομένα και από το βαθμό στον οποίο η υποκείμενη συναρτησιακή μορφή είναι μη γραμμική. Με περισσότερες παρατηρήσεις, η ακρίβεια του μοντέλου δε δύναται να βελτιωθεί εάν δεν υπάρχει μη γραμμική δομή στα δεδομένα. Στα ΤΝΔ ο θόρυβος και μόνο καθορίζει το όριο της ακρίβειας εξαιτίας της ικανότητας εκτίμησης γενικής συνάρτησης που έχουν. Με ένα ικανοποιητικά μεγάλο δείγμα, τα ΤΝΔ είναι σε θέση να μοντελοποιήσουν οποιαδήποτε πολύπλοκη δομή των

δεδομένων. Επομένως, τα ΤΝΔ ωφελούνται περισσότερο από μεγάλα δείγματα απ' ό,τι τα γραμμικά στατιστικά μοντέλα. Ενδιαφέρον έχει ότι τα ΤΝΔ δεν απαιτούν απαραίτητα μεγαλύτερο δείγμα από εκείνο που χρειάζεται για την καλή απόδοση των γραμμικών μοντέλων. Έχει αποδειχθεί ότι τα μοντέλα πρόβλεψης ΤΝΔ αποδίδουν πολύ καλά ακόμη και με μεγέθη δείγματος μικρότερα του 50, ενώ τα μοντέλα Box-Jenkins απαιτούν συνήθως τουλάχιστον 50 σημεία δεδομένων προκειμένου να δώσουν επιτυχημένη πρόβλεψη [9].

6. Κριτήρια απόδοσης

Μολονότι υπάρχουν πολλά κριτήρια απόδοσης για τα ΤΝΔ πρόβλεψης όπως ο χρόνος μοντελοποίησης και ο χρόνος εκπαίδευσης, το βασικό και πιο σημαντικό κριτήριο απόδοσης είναι η ακρίβεια της πρόβλεψης που μπορεί να επιτευχθεί πέραν των δεδομένων εκπαίδευσης. Ωστόσο, ένα κατάλληλο κριτήριο απόδοσης για ένα δεδομένο πρόβλημα δεν είναι πάντα κοινώς αποδεκτό από τους ακαδημαϊκούς και τους επιστήμονες της πρόβλεψης. Ένα κριτήριο απόδοσης συνήθως ορίζεται σε σχέση με το σφάλμα πρόβλεψης, το οποίο είναι η διαφορά μεταξύ της πραγματικής (επιθυμητής) και της προβλέψιμης τιμής. Υπάρχουν αρκετά κριτήρια ακρίβειας και το καθένα έχει πλεονεκτήματα και περιορισμούς. Τα πλέον ευρέως χρησιμοποιούμενα είναι:

- Η μέση απόλυτη απόκλιση (mean absolute deviation MAD) = $\frac{\sum |e_t|}{N}$
- Το άθροισμα των τετραγωνικών σφαλμάτων (sum of squared error SSE) = $\sum (e_t)^2$
- Το άθροισμα των μέσων τετραγωνικών σφαλμάτων (mean squared error MSE) = $\frac{\sum (e_t)^2}{N}$
- Η ρίζα των μέσων τετραγωνικών σφαλμάτων (root mean squared error RMSE) = \sqrt{MSE}
- Το μέσο απόλυτο ποσοστό σφάλματος (mean absolute percentage error MAPE) = $\frac{1}{N} \sum \left| \frac{e_t}{y_t} \right| (100)$

όπου e_i είναι το ατομικό σφάλμα πρόβλεψης, y_i είναι η πραγματική τιμή και N είναι το σύνολο των όρων σφάλματος.

Ενδεικτικά αναφέρουμε ότι στη βιβλιογραφία έχουν χρησιμοποιηθεί και άλλα κριτήρια ακριβείας όπως το μέσο σφάλμα (mean error – ME), η στατιστική μεταβλητή U του Theil (Theil's U-statistic), το μέσο απόλυτο ποσοστό σφάλματος (median absolute percentage error – MdaPE), το γεωμετρικό μέσο σχετικά απόλυτο σφάλμα (geometric mean relative absolute error – GMRAE), η μέση σχετική διακύμανση (average relative variance – ARV) και το κριτήριο πληροφορίας του Bayes (Bayesian information criterion – BIC).

Το MSE είναι το πιο συνηθισμένο μέτρο ακρίβειας, αν και τα πλεονεκτήματα της χρήσης του αμφισβητούνται όταν πρόκειται για την εκτίμηση της σχετικής ακρίβειας μεθόδων πρόβλεψης για διαφορετικά σύνολα δεδομένων. Επιπλέον, το MSE ίσως να μην είναι κατάλληλο για τη δόμηση ενός μοντέλου ΤΝΔ με δείγμα εκπαίδευσης, αφού αγνοεί τις σημαντικές πληροφορίες σχετικά με το νούμερο των παραμέτρων (συναπτικά βάρη), τις οποίες το μοντέλο πρέπει να υπολογίσει. Από τη σκοπιά της στατιστικής, καθώς ο αριθμός των υπολογισμένων παραμέτρων του μοντέλου αυξάνεται, οι βαθμοί ελευθερίας για το ολικό μοντέλο μειώνονται, αυξάνοντας έτσι την πιθανότητα υπερεκπαίδευσης στο δείγμα εκπαίδευσης. Μια βελτιωμένη έκδοση του MSE για το μέρος της εκπαίδευσης είναι το συνολικό άθροισμα των τετραγωνικών σφαλμάτων διαιρεμένο με τους βαθμούς ελευθερίας, που είναι ο αριθμός των παρατηρήσεων μείων τον αριθμό των συναπτικών βαρών και σταθερών βαρών. [9,3]

3.8 Πλεονεκτήματα και Μειονεκτήματα της Χρήσης ΤΝΔ για πρόβλεψη σε Σχέση με Στατιστικές Μεθόδους

Παρά την πρόοδο που έχει σημειωθεί στη θεωρία μη γραμμικής παλινδρόμησης, η πλειοψηφία των εφαρμογών στην οικονομετρία υποθέτει γραμμική σχέση μεταξύ των εξαρτημένων και των ανεξάρτητων μεταβλητών. Η απλότητα των γραμμικών μοντέλων και η πιθανότητα γραμμικής μορφοποίησης μη γραμμικών σχέσεων καθιστά το μοντέλο γραμμικής παλινδρόμησης ελκυστικό και ισχυρό εργαλείο. Συνεπώς θα προσπαθήσουμε να συγκρίνουμε τα νευρωνικά δίκτυα με τα μοντέλα γραμμικής παλινδρόμησης. Στον ακόλουθο πίνακα συνοψίζουμε τα

συγκριτικά πλεονεκτήματα και μειονεκτήματα των νευρωνικών δικτύων που θα αναλύσουμε παρακάτω:

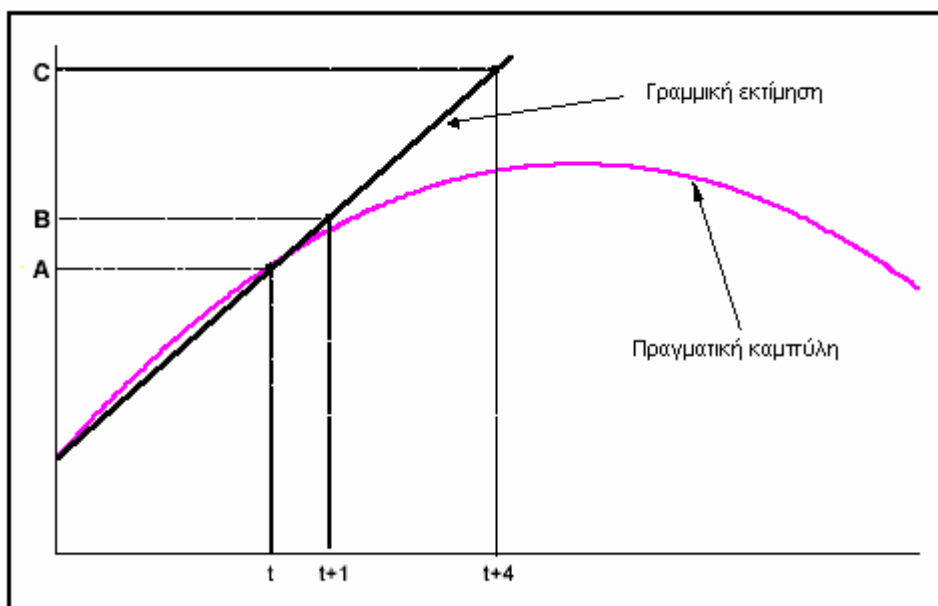
Πλεονεκτήματα	Μειονεκτήματα
Μοντελοποιούν επιτυχώς μη γραμμικές σχέσεις	<i>Δύσκολη ερμηνεία των βαρών που εκτιμήθηκαν από το δίκτυο (πρόβλημα “μαύρου κουτιού”)</i>
Δεν απαιτούν εκ των προτέρων πληροφορία για τη συναρτησιακή μορφή της σχέσης	<i>Απίθανη η εύρεση γενικού ελάχιστου</i>
Η ίδια αρχιτεκτονική είναι πολύ ευέλικτη	<i>Απαίτηση για πολλά δεδομένα και μεγάλο χρόνο εκπαίδευσης</i>

Πίνακας 3.1 : Πλεονεκτήματα – Μειονεκτήματα ΤΝΔ έναντι πολλαπλής γραμμικής παλινδρόμησης

3.8.1 Συγκριτικά πλεονεκτήματα των νευρωνικών δικτύων

1. Μοντελοποιούν επιτυχώς μη γραμμικές σχέσεις

Πρώτον, τα νευρωνικά δίκτυα με μη γραμμικές συναρτήσεις ενεργοποίησης είναι πιο αποτελεσματικά από τα μοντέλα γραμμικής παλινδρόμησης, όταν εμπλέκονται μη γραμμικές σχέσεις. Στο σχήμα 3.13 απεικονίζεται το κλασικό πρόβλημα που προκύπτει κατά την εκτίμηση μη γραμμικών σχέσεων με γραμμικά μοντέλα. Εάν τα δεδομένα από το 0 έως t χρησιμοποιηθούν για την εκτίμηση του γραμμικού μοντέλου, η πρόβλεψη ενός βήματος έμπροσθεν (B) θα είναι πολύ κοντά στην πραγματική μεταβολή της υπό μελέτη μεταβλητής, αφού μια γραμμική συνάρτηση μπορεί να δώσει μια λογική εκτίμηση της τοπικής συμπεριφοράς μιας μη γραμμικής συνάρτησης. Ωστόσο, εάν επιχειρηθούν προβλέψεις αρκετών μελλοντικών περιόδων, η τιμή που εκτιμήθηκε από το γραμμικό μοντέλο (C) ενδεχομένως να απέχει ολοένα και περισσότερο από την πραγματική τιμή της μεταβλητής.



Σχήμα 3.13: Πρόβλεψη μη γραμμικής μεταβλητής με γραμμικό μοντέλο

2. Δεν απαιτούν εκ των προτέρων πληροφορία για τη συναρτησιακή μορφή της σχέσης

Αν και πολλές μη γραμμικές συναρτήσεις δύναται να μορφοποιηθούν γραμμικά χρησιμοποιώντας σχετικά απλούς μαθηματικούς μετασχηματισμούς, κάτι τέτοιο προϋποθέτει ότι ο ερευνητής έχει *a priori* γνώση της φύσης της μη γραμμικότητας η οποία του επιτρέπει να προσδιορίζει τους κατάλληλους μετασχηματισμούς που θα εφαρμοσθούν στα δεδομένα. Είναι ευρέως γνωστό ότι τέτοιες πληροφορίες σπανιότατα είναι διαθέσιμες στον τομέα της μακροοικονομικής πρόβλεψης

Υπάρχει η αντίληψη ότι οι μέθοδοι μη γραμμικής παλινδρόμησης αποδίδουν εξίσου καλά με τα νευρωνικά δίκτυα όταν αντιμετωπίζουν μη γραμμικά φαινόμενα. Θεωρητικά η αντίληψη αυτή αληθεύει. Ωστόσο, στην πράξη πρέπει ο οικονομολόγος που εκτιμά το μοντέλο μη γραμμικής παλινδρόμησης να υποθέσει εκ των προτέρων τη συναρτησιακή μορφή για την υπό μελέτη σχέση. Η επιλογή λανθασμένης συναρτησιακής μορφής οδηγεί σε ανακριβή εκτίμηση των συντελεστών και εσφαλμένη πρόβλεψη. Εν αντιθέσει, κατά την εκτίμηση του νευρωνικού δικτύου ο ερευνητής δε χρειάζεται να ανησυχεί για τη συναρτησιακή μορφή του υπό μελέτη φαινομένου, καθώς η ιδιότητα “γενικής εκτιμήτριας” του δικτύου επιτρέπει τη

μίμηση σχεδόν οποιασδήποτε συναρτησιακής μορφής. Συνεπώς, δεν απαιτείται εκ των προτέρων γνώση για τη λήψη ακριβών προβλέψεων.

3. Η ίδια αρχιτεκτονική είναι πολύ ευέλικτη

Το τρίτο πλεονέκτημα των νευρωνικών δικτύων πηγάζει από την ευελιξία της αρχιτεκτονικής των δικτύων. Όπως είδαμε και παραπάνω, ένα ευρύ φάσμα στατιστικών μεθόδων μπορεί να προσδιορισθεί με μικρές τροποποιήσεις στη συνάρτηση ενεργοποίησης και στη δομή του δικτύου. Συνεπώς η ίδια αρχιτεκτονική είναι ιδιαίτερα ευέλικτη και μπορεί να συμπεριλάβει εξίσου διακριτές και συνεχής εξαρτημένες μεταβλητές.

3.8.1 Μειονεκτήματα και Περιορισμοί των Νευρωνικών Δικτύων

1. Δύσκολη ερμηνεία των βαρών που εκτιμήθηκαν από το δίκτυο (πρόβλημα “μαύρου κουτιού”)

Η πολύπλοκη μη γραμμική συναρτησιακή μορφή του δικτύου καθιστά δύσκολη την ερμηνεία των βαρών που εκτιμήθηκαν. Στα μοντέλα γραμμικής παλινδρόμησης, οι τιμές των συντελεστών παρέχουν ένα άμεσο κριτήριο της συμβολής κάθε μεταβλητής στην έξοδο του μοντέλου. Στην περίπτωση των νευρωνικών δικτύων, είναι ιδιαίτερα πολύπλοκη ο αναλυτικός προσδιορισμός της επίδρασης μιας εισόδου στην εκτιμώμενη τιμή εξόδου. Ακόμη και στην απλούστερη μορφή δικτύου, όπως στο σχήμα 3.3, κάθε είσοδος τροφοδοτείται μέσω μιας μη γραμμικής συνάρτησης ενεργοποίησης και επηρεάζεται επίσης από διαφορετικά βάρη (a_{ij} και b_{ij}). Παρατηρώντας την εξίσωση (10), γίνεται αντιληπτό ότι είναι πολύ δύσκολο να ανακαλύψουμε την επίδραση των X_1 , X_2 ή της Y . Εξαιτίας της εν λόγω δυσκολίας τα νευρωνικά δίκτυα χαρακτηρίζονται ως “μαύρα κουτιά”: το δίκτυο χρησιμοποιεί τις εισόδους για τον υπολογισμό της εξόδου, αλλά ο ερευνητής δεν κατανοεί απόλυτα το λόγο για τον οποίο μια δεδομένη τιμή προβλέπεται. Το πρόβλημα αυτό αντιμετωπίζεται συνήθως με ανάλυση ευαισθησίας.

2. Απίθανη εύρεση γενικού ελαχίστου

Όπως με όλες τις μη γραμμικές μεθόδους, είναι δύσκολη η εύρεση γενικού ελαχίστου της συνάρτησης σφάλματος. Παραδείγματος χάριν, έχει καταγραφεί σε μελέτες ότι ένα πρόβλημα εκτίμησης με 35 βάρη έχει αρκετά δισεκατομμύρια τοπικά ελάχιστα. Ωστόσο, ορισμένα τοπικά ελάχιστα μπορούν να παράγουν ακριβείς προβλέψεις, εάν είναι αρκετά κοντά στο γενικό ελάχιστο.

3. Απαίτηση για πολλά δεδομένα και μεγάλο χρόνο εκπαίδευσης

Ένα επιπλέον μειονέκτημα είναι ότι τα νευρωνικά δίκτυα απαιτούν μεγάλο χρόνο εκπαίδευσης. Υπερβολικές επαναλήψεις χρειάζονται για την εκπαίδευση των νευρωνικών δικτύων. Η μείωση του χρόνου εκπαίδευσης είναι αποφασιστικής σημασίας διότι η κατασκευή ενός συστήματος νευρωνικού δικτύου πρόβλεψης είναι μια διαδικασία δοκιμής και σφάλματος. Συνεπώς, όσο περισσότερα πειράματα μπορεί ένας επιστήμονας να εκτελέσει σε μια ορισμένη χρονική περίοδο, τόσο πιο αξιόπιστο είναι το αποτέλεσμα. Συγκριτικά με τα γραμμικά μοντέλα η όλη διαδικασία σχεδιασμού και εκτίμησης του δικτύου είναι πολύ μεγαλύτερης διάρκειας [10, 14].

ΚΕΦΑΛΑΙΟ 4^ο

Προσέγγιση της Μακροοικονομικής Πρόβλεψης με Τεχνητά Νευρωνικά Δίκτυα (ΤΝΔ)

"Ancora imparo" Ακόμη μαθαίνω, Michelangelo

4.1 Περιγραφή του προβλήματος

Για τη μακροοικονομική πρόβλεψη της Αμερικανικής οικονομίας θα αναπτύξουμε ένα νευρωνικό δίκτυο υψηλής τάξης που εκπαιδεύεται με οπισθόδρομη διάδοση. Για να το κάνουμε αυτό, θα χρησιμοποιήσουμε το προγραμματιστικό περιβάλλον της MATLAB. Τα δεδομένα που θα χρησιμοποιήσουμε είναι όμοια με εκείνα που χρησιμοποιήθηκαν στην προσέγγιση με γραμμικά μοντέλα, δηλαδή οι χρονοσειρές των οικονομικών μεταβλητών από το έτος 1932 έως 1995. Η πρόβλεψη θα είναι μακροπρόθεσμη, δηλαδή θα δουλέψουμε με ετήσια δεδομένα και όχι με τριμηνιαία (βραχυπρόθεσμη πρόβλεψη). Επιλέξαμε τη μακροπρόθεσμη πρόβλεψη εξαιτίας του γεγονότος ότι οι μακροοικονομικές μεταβλητές που χρησιμοποιήσαμε στη μελέτη μας, επηρεάζονται από ενέργειες στον πολιτικό τομέα και καθώς η πολιτική επηρεάζει την οικονομία με σημαντική καθυστέρηση, η μακροπρόθεσμη πρόβλεψη αποδεικνύεται μεγαλύτερης αξίας.

Οι μεταβλητές που θα προβλεφθούν στην παρούσα εργασία είναι:

- πραγματικό προϊόν (Real GDP),
- καθαρές εξαγωγές (Real Net Exports)
- πληθωρισμός (Inflation Rate, cpi)

Οι επεξηγηματικές μεταβλητές που θα εισαχθούν επίσης στο δίκτυο είναι:

- κυβερνητικές δαπάνες (real government exp.)
- exchange rate
- short interest rate (I_Short)
- velocity of M1

Οι παραπάνω μεταβλητές επιλέχθηκαν καθώς χαρακτηρίζονται από αρκετές επιθυμητές ιδιότητες οι σπουδαιότερες από τις οποίες είναι: υπόκεινται σε ελάχιστες αναθεωρήσεις, είναι εγκαίρως διαθέσιμες και μετρώνται με μικρό σφάλμα.

4.2 Εισαγωγή – Προετοιμασία Δεδομένων στη MATLAB

Για την εφαρμογή των νευρωνικών δικτύων στο πρόβλημά μας θα χρησιμοποιήσουμε τη MATLAB 6.5 Release 13. Το νευρωνικό δίκτυο υψηλής τάξης που θα κατασκευάσουμε είναι τύπου sigma-pi και θα φτάσει μέχρι και τρίτης τάξης.

Πριν ξεκινήσουμε την εφαρμογή μας, πρέπει να αποφασίσουμε πώς θα χρησιμοποιήσουμε τα δεδομένα που έχουμε. Θα πρέπει να χρησιμοποιήσουμε ένα μέρος από αυτά για την εκπαίδευση του δικτύου και τα υπόλοιπα για να ελέγξουμε τη σωστή λειτουργία του. Όπως έχουμε αναφέρει, δεν υπάρχει κανόνας για την επιλογή του δείγματος εκπαίδευσης. Ακολουθώντας την εμπειρική αναλογία 80% των δεδομένων για εκπαίδευση και 20% για έλεγχο, θα χρησιμοποιήσουμε τα δεδομένα των ετών 1932 έως 1984 για την εκπαίδευση του δικτύου και τα δεδομένα των ετών 1985 έως 1995 για τον έλεγχο της ορθής λειτουργίας του.

Για την εύκολη επεξεργασία των δεδομένων και την εισαγωγή τους στο δίκτυο, έχουμε κατασκευάσει ένα πρόγραμμα στη MATLAB. Ο κώδικας του προγράμματος φαίνεται παρακάτω. Σημειώνουμε εδώ ότι ολόκληρος ο κώδικας του προγράμματος που χρησιμοποιούμε στην εφαρμογή, παρατίθεται στο ΠΑΡΑΡΤΗΜΑ Γ. Στην παρούσα αναφορά, παραθέτουμε και σχολιάζουμε τον κώδικα τμηματικά.

4.3 Σχολιασμός του κώδικα/ διαδικασία του προβλήματος

Βήμα Πρώτο: Εισαγωγή και Επεξεργασία των Δεδομένων

```
if order==1
    XX=zeros(N,7);
end
if order==2
    XX=zeros(N,56);
end
if order==3
    XX=zeros(N,399);
end
```

Το υψηλής τάξης νευρωνικό δίκτυο τύπου sigma-pi που κατασκευάσαμε όπως είπαμε, μπορεί να είναι πρώτης, δευτέρας ή τρίτης τάξης. Ανάλογα με την τάξη του δικτύου ορίζουμε και αρχικοποιούμε τα δεδομένα εισόδου. Τα δεδομένα εισόδου ορίζονται στη μεταβλητή - πίνακα που ονομάζεται XX, ο οποίος περιέχει τις τιμές των μεταβλητών του προβλήματος

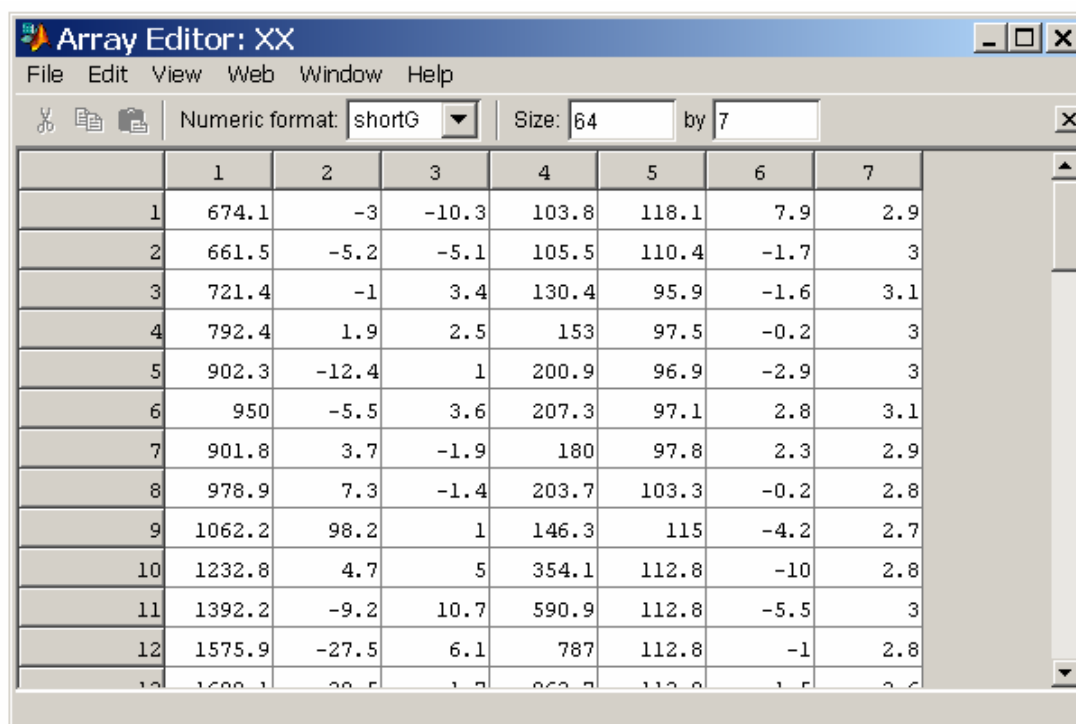
Συγκεκριμένα εάν το δίκτυο είναι πρώτης τάξης, τότε ο πίνακας εισόδου θα είναι διαστάσεων $N \times 7$, όπου N ο αριθμός των παρατηρήσεων μας, δηλαδή 64. Ο αριθμός 7 απεικονίζει το σύνολο των μεταβλητών που χρησιμοποιούμε οι οποίες αποτελούν τους όρους πρώτης τάξης. Εάν το δίκτυο είναι δευτέρας τάξης, τότε ο πίνακας εισόδου θα είναι διαστάσεων $N \times 56$. Ο αριθμός 56 είναι το άθροισμα των όρων πρώτης τάξης (7) και των όρων δευτέρας τάξης (7^2). Τέλος στην περίπτωση που το δίκτυο είναι τρίτης τάξης, ο πίνακας εισόδου θα είναι διαστάσεων $N \times 399$. Ο αριθμός 399 αποτελεί το άθροισμα των όρων πρώτης τάξης (7), των όρων δευτέρας τάξης (7^2) και των όρων τρίτης τάξης (7^3). Και στις τρεις περιπτώσεις, αρχικοποιούμε τους πίνακες εισόδου θέτοντας τους μηδενικούς.

```
for i=1:N,
    XX(i,1:7)=x(i, 1:7);
    if order>1
        xx=x(i,:)'*x(i,:);
        yy=reshape(xx, 49,1);
        XX(i,8:56)=yy';
        if order==3
            yyy=yy*x(i,:);
            yyyy=reshape(yyy, 343,1);
            XX(i,57:399)=yyyy';
        end
    end
end
```

Αφου λοιπόν φτιάξαμε χώρο για τα δεδομένα μας, δημιουργώντας τον μηδενικό πίνακα XX , στη συνέχεια εισάγουμε τις τιμές των δεδομένων στον πίνακα XX . Εάν το δίκτυο είναι πρώτης τάξης, προκύπτει ο πίνακας εισόδου XX διαστάσεων 64×7 , ο οποίος περιέχει τις τιμές των μεταβλητών από το έτος 1932 έως 1995. Εάν το δίκτυο είναι δευτέρας τάξης, προκύπτει ο πίνακας εισόδου XX διαστάσεων 64×56 , ο οποίος περιέχει τις τιμές των μεταβλητών και τους όρους δευτέρας τάξης από το έτος 1932 έως 1995. Τέλος, στην περίπτωση που το δίκτυο είναι τρίτης τάξης, ο πίνακας εισόδου XX περιέχει τις τιμές των μεταβλητών, τους όρους δευτέρας και τους όρους τρίτης τάξης.

Μέρος του πίνακα XX , έτσι όπως τον υπολογίζει η MATLAB για τις δύο πρώτες περιπτώσεις, φαίνεται παρακάτω. Όπως βλέπουμε στον πίνακα εισόδου του δικτύου δευτέρας τάξης έπειτα από τις 7 στήλες με τους όρους πρώτης τάξης ακολουθούν οι υπόλοιπες 49 με τους όρους δευτέρας τάξης, όπως υπολογίστηκαν από τις εντολές του βρόγχου. Ομοίως, στον πίνακα εισόδου του δικτύου τρίτης τάξης, έπειτα από τις 7 στήλες με τους όρους πρώτης τάξης και τις 49 με τους όρους

δευτέρας τάξης, ακολουθούν 343 στήλες με τους όρους τρίτης τάξης, όπως υπολογίστηκαν από τις εντολές του βρόγχου.



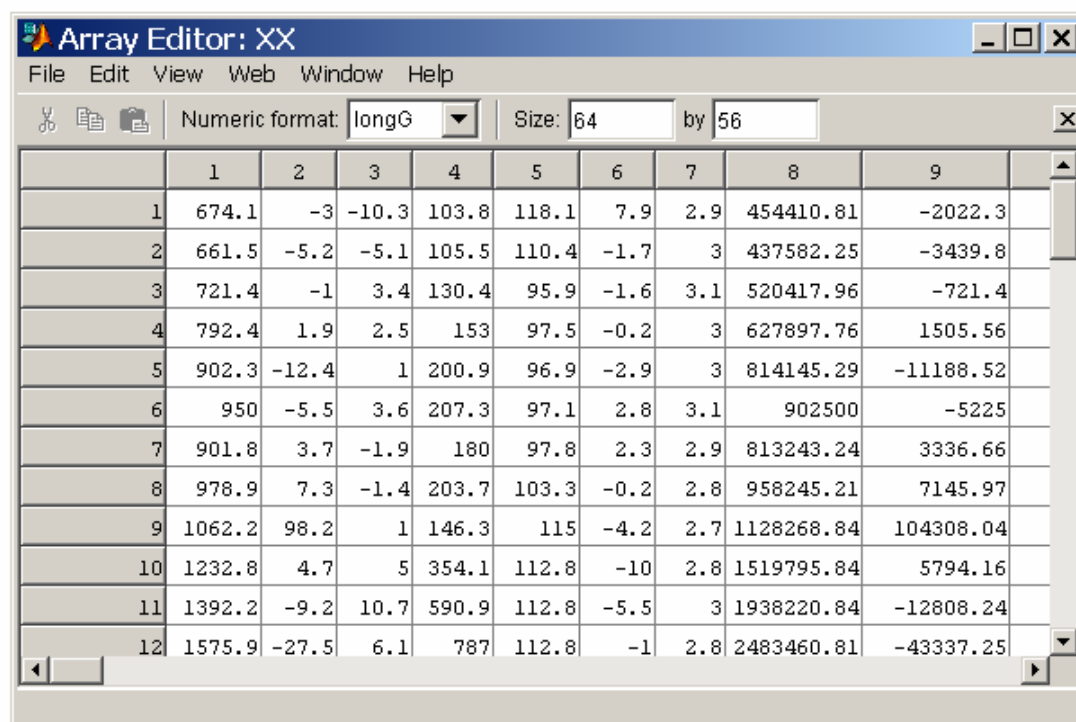
Array Editor: XX

File Edit View Web Window Help

Numeric format: shortG Size: 64 by 7

	1	2	3	4	5	6	7
1	674.1	-3	-10.3	103.8	118.1	7.9	2.9
2	661.5	-5.2	-5.1	105.5	110.4	-1.7	3
3	721.4	-1	3.4	130.4	95.9	-1.6	3.1
4	792.4	1.9	2.5	153	97.5	-0.2	3
5	902.3	-12.4	1	200.9	96.9	-2.9	3
6	950	-5.5	3.6	207.3	97.1	2.8	3.1
7	901.8	3.7	-1.9	180	97.8	2.3	2.9
8	978.9	7.3	-1.4	203.7	103.3	-0.2	2.8
9	1062.2	98.2	1	146.3	115	-4.2	2.7
10	1232.8	4.7	5	354.1	112.8	-10	2.8
11	1392.2	-9.2	10.7	590.9	112.8	-5.5	3
12	1575.9	-27.5	6.1	787	112.8	-1	2.8

Σχήμα 4.1: Ο πίνακας εισόδου XX για το νευρωνικό δίκτυο πρώτης τάξης.



Array Editor: XX

File Edit View Web Window Help

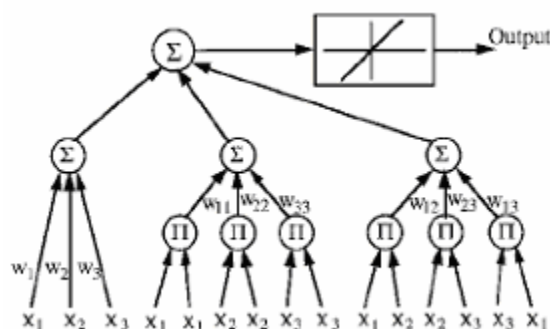
Numeric format: longG Size: 64 by 56

	1	2	3	4	5	6	7	8	9
1	674.1	-3	-10.3	103.8	118.1	7.9	2.9	454410.81	-2022.3
2	661.5	-5.2	-5.1	105.5	110.4	-1.7	3	437582.25	-3439.8
3	721.4	-1	3.4	130.4	95.9	-1.6	3.1	520417.96	-721.4
4	792.4	1.9	2.5	153	97.5	-0.2	3	627897.76	1505.56
5	902.3	-12.4	1	200.9	96.9	-2.9	3	814145.29	-11188.52
6	950	-5.5	3.6	207.3	97.1	2.8	3.1	902500	-5225
7	901.8	3.7	-1.9	180	97.8	2.3	2.9	813243.24	3336.66
8	978.9	7.3	-1.4	203.7	103.3	-0.2	2.8	958245.21	7145.97
9	1062.2	98.2	1	146.3	115	-4.2	2.7	1128268.84	104308.04
10	1232.8	4.7	5	354.1	112.8	-10	2.8	1519795.84	5794.16
11	1392.2	-9.2	10.7	590.9	112.8	-5.5	3	1938220.84	-12808.24
12	1575.9	-27.5	6.1	787	112.8	-1	2.8	2483460.81	-43337.25

Σχήμα 4.2: Ο πίνακας εισόδου XX για το νευρωνικό δίκτυο δεύτερας τάξης.

Βήμα Δεύτερο: Δημιουργία του Δικτύου

Το δίκτυο που θα δημιουργήσουμε είναι ένα δίκτυο υψηλής τάξης τύπου sigma-pi. Το δίκτυο μπορεί να είναι πρώτης, δευτέρας ή τρίτης τάξης. Όπως έχουμε προαναφέρει τα δίκτυα sigma-pi είναι υψηλής τάξης νευρωνικά δίκτυα τα οποία περιέχουν μονάδες sigma και pi. Ο αλγόριθμος μάθησης που θα χρησιμοποιήσουμε είναι ο αλγόριθμος οπισθόδρομης διάδοσης (backpropagation) με παραλλαγή της μεθόδου της μέγιστης κατάβασης (gradient descent method). Η γενική μορφή αυτού του δικτύου είναι:



Σχήμα 4.3: Δίκτυο sigma-pi δευτέρας τάξης.

Το δίκτυο σχεδιάστηκε σύμφωνα με την ακόλουθη αρχιτεκτονική:

- Τρία επίπεδα: Ένα επίπεδο εισόδου, ένα κρυφό επίπεδο και ένα επίπεδο εξόδου
- Επτά νευρώνες στο επίπεδο εισόδου
- Στο κρυμμένο επίπεδο:
 - 7 νευρώνες για το δίκτυο πρώτης τάξης
 - 56 νευρώνες για το δίκτυο δευτέρας τάξης
 - 399 νευρώνες για το δίκτυο τρίτης τάξης
- Τρεις νευρώνες στο επίπεδο εξόδου

Η συνάρτηση ενεργοποίησης υπολογίζει την έξοδο του νευρώνα ανάλογα με την είσοδο από το δίκτυο. Παίρνει σαν όρισμα το διάνυσμα εισόδου και επιστρέφει τις αντίστοιχες τιμές εξόδου. Αναλυτικά οι συναρτήσεις που χρησιμοποιήσαμε είναι:

- Συνάρτηση ενεργοποίησης του νευρώνα εξόδου, επιλέγουμε τη συνάρτηση η οποία κατασκευάσαμε εμείς.
- Συνάρτηση ενεργοποίησης επιλέγουμε τη γραμμική συνάρτηση ή ταυτότητα.

- Συνάρτηση εκπαίδευσης επιλέγουμε την `nnopt` την οποία κατασκευάσαμε εμείς με παραλλαγή του αλγορίθμου μέγιστης κατάβασης. Η συνάρτηση εκπαίδευσης αναπροσαρμόζει τις τιμές των βαρών και των παραμορφώσεων σύμφωνα με τη βαθμιαία κατάβαση. Τα αρχικά βάρη εισάγονται στο δίκτυο από εμάς και είναι μηδενικά διανύσματα. Η εκπαίδευση σταματά όταν το σφάλμα J που υπολογίζει η συνάρτηση `nnopt` συγκλίνει.
- Τα βάρη βελτιώνονται με τη χρήση της συνάρτησης του αθροίσματος των τετραγωνικών σφαλμάτων (SSE – Sum of Squared Error).

Βήμα Τρίτο: Εκπαίδευση του Δικτύου

```
W=fmins('nnopt', W0);
```

Ο κύριος κώδικας που έχουμε αναπτύξει καλεί τη συνάρτηση `nnopt` την οποία έχουμε κατασκευάσει εμείς. Η συνάρτηση `fmins` ουσιαστικά λέει στο δίκτυο να επιστρέψει τον πίνακα W που είναι γενικό ελάχιστο της συνάρτησης `nnopt(W)` κοντά στο W_0 , δηλαδή εφαρμόζουμε τη μέθοδο της μέγιστης κατάβασης. Η συνάρτηση `nnopt` με τη σειρά της, εκπαιδεύει το δίκτυο.

```
function J=nnopt(W)
global XX dx N

J=0;

for i=1:N-10,
    J=J+norm(dx(i,:)'-W*XX(i,:))';
end
J
```

Τα δεδομένα που χρησιμοποιούμε για την εκπαίδευση, όπως βλέπουμε είναι τα πρώτα 54 δεδομένα των χρονοσειρών. Η τιμή του σφάλματος αποθηκεύεται στη μεταβλητή J και αποτελεί το άθροισμα των τετραγώνων των σφαλμάτων: $SSE = \sum (e_i)^2$. Στη MATLAB αυτό εκφράζεται με την εντολή `norm` η οποία επιστρέφει την Ευκλείδεια νόρμα της πραγματικής εξόδου μείον την έξοδο που υπολόγισε το δίκτυο. Ο βρόγχος μεταβάλλει την τιμή του σφάλματος για κάθε επανάληψη και την επιστρέφει ως μεταβλητή J . Η όλη διαδικασία της εκπαίδευσης τερματίζεται μόλις η τιμή του J συγκλίνει. Ο αριθμός των επαναλήψεων καθώς και ο χρόνος εκπαίδευσης που απαιτείται ωστόσο να επέλθει η σύγκλιση δεν είναι γνωστός εκ των προτέρων και ποικίλει ανάλογα με την τάξη του δικτύου.

Παρακάτω παραθέτουμε σε πίνακες τα αποτελέσματα των επαναλήψεων μέχρι τη σύγκλιση του σφάλματος.:

Αριθμός Επαναλήψεων	order=1 J	order=2 J	order=3 J
1	3876.431	3530.31	3031.748
2	3765.581	3325.126	2942.783
3	3727.589	3254.365	2899.521
4	3721.101	3181.749	2886.787
5	3717.642	3028.187	2873.328
6	3717.134	2939.874	2860.632
7	3716.753	2860.421	2825.606
8	3716.539	2557.07	2792.155
9	3716.429	2434.337	
10	3716.329	2402.079	
11	3716.067	2388.446	
12	3716	2376.749	
13	3715.9	2375.277	
14	3715.9	2373.018	
15		2372.52	
16		2372.437	
17		2372.3	
18		2371.985	
19		2370.73	
20		2370.421	
21		2370.2	
22		2370.1	
23		2369.973	
24		2369.9	
25		2369.9	

Ο χρόνος εκπαίδευσης για τα νευρωνικά δίκτυα πρώτης και δευτέρας τάξης δεν ήταν μεγάλος και παρατηρήθηκε σύγκλιση της τιμής του J έπειτα από 14 και 25 επαναλήψεις αντίστοιχα. Με βάσει τα αποτελέσματα αυτά περιμέναμε μικρό χρόνο εκπαίδευσης και για το δίκτυο τρίτης τάξης. Ωστόσο διαψευστήκαμε, καθώς αποδείχθηκε μια εξαιρετικά χρονοβόρα διαδικασία, αφού η κάθε επανάληψη

διαρκούσε 100 ώρες περίπου. Το γεγονός αυτό δικαιολογείται από την αύξηση της πολυπλοκότητας του δικτύου και του αριθμού των κρυμμένων κόμβων.

Βήμα Τέταρτο: Η έξοδος του Νευρωνικού Δικτύου

Η έξοδος του δικτύου που μελετάμε είναι η μεταβολή των μεταβλητών κατάστασης, δηλαδή η μεταβολή των:

- πραγματικό προϊόν (Real GDP)
- καθαρές εξαγωγές (Real Net Exports)
- πληθωρισμός (Inflation Rate, cpi)

```
dx=zeros(N,3);
for i=1:N-1,
    dx(i,1:3)=x(i+1,1:3)-x(i,1:3);
end
```

Αρχικά δημιουργούμε χώρο για τη μεταβλητή – πίνακα dx στην οποία αποθηκεύονται οι μεταβολές των παραπάνω μεταβλητών για τα 64 έτη που μελετάμε, δηλώνοντάς την ως μηδενικό. Έπειτα, μέσω ενός βρόγχου παίρναμε στον πίνακα dx τις πραγματικές μεταβολές των μεταβλητών, όπως υπολογίζονται από τα δεδομένα χρονοσειρών που έχουμε στη διάθεσή μας. Μέρος του πίνακα dx όπως υπολογίστηκε από τη MATLAB, φαίνεται παρακάτω:

	1	2	3
1	-12.6	-2.2	5.2
2	59.9	4.2	8.5
3	71	2.9	-0.9
4	109.9	-14.3	-1.5
5	47.7	6.9	2.6
6	-48.2	9.2	-5.5
7	77.1	3.6	0.5
8	83.3	90.9	2.4
9	170.6	-93.5	4
10	159.4	-13.9	5.7
11	183.7	-18.3	-4.6
12	113.2	-1	-4.4
13	28.5	5.8	8.6

Σχήμα 4.4: Ο πίνακας dx με τις πραγματικές μεταβολές των μεταβλητών

```
dx_est=nn(W);
```

Το δίκτυο τελικά καλεί τη συνάρτηση nn την οποία κατασκευάσαμε εμείς. Η συνάρτηση nn υπολογίζει τις προβλέψεις του δικτύου και τις αποθηκεύει στη μεταβλητή dx_est η οποία είναι ο πίνακας 63 x 3 με τις μεταβολές των μεταβλητών κατάστασης, όπως προβλέφθηκαν από το δίκτυο.

```
function dx_est=nn(W)
global XX dx N

for i=1:N-1
dx_est(i,:)=(W*XX(i,:)')';
end
```

Η συνάρτηση nn που κατασκευάσαμε ουσιαστικά επιστρέφει την έξοδο του νευρωνικού δικτύου υψηλής τάξης:

$$P(x) = w_0 + \sum_i w_i x_i + \sum_i \sum_j w_{ij} x_i x_j + \sum_i \sum_j \sum_k w_{ijk} x_i x_j x_k + \dots$$

ή συνοπτικά:

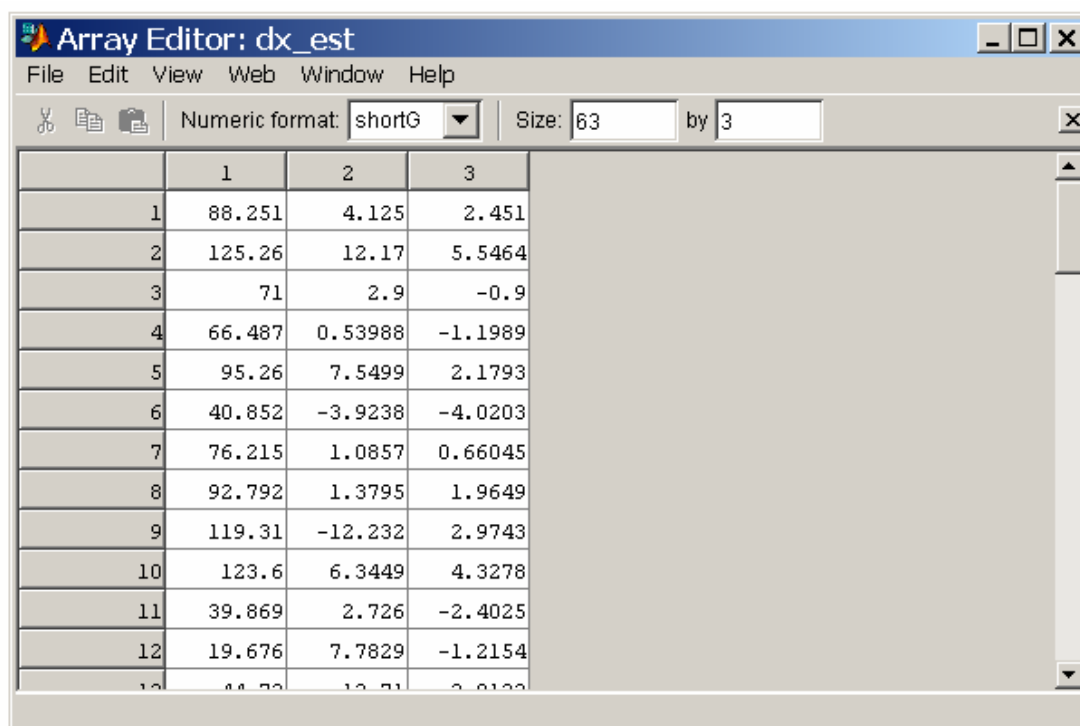
$$P(x) = w_0 + \sum_{i=1}^M w_i \prod_{j=1}^d x_j^r$$

όπου: w_i είναι τα βάρη, M είναι ο μέγιστος αριθμός μονώνυμων, x_i είναι τα χαρακτηριστικά του διανύσματος εισόδου, $r = 0, 1, 2, \dots$ είναι η δύναμη με την οποία το j -οστό χαρακτηριστικό x_i συμμετέχει στο i -οστό μονώνυμο.

Όπως βλέπουμε η συνάρτηση ενεργοποίησης που επιλέξαμε είναι η ταυτότητα: $f(x) = x$.

Για το νευρωνικό δίκτυο πρώτης τάξης βλέπουμε παρακάτω μέρος του πίνακα dx_est όπως υπολογίστηκε στον εργασιακό χώρο της MATLAB, έπειτα από τη σύγκλιση του σφάλματος.

Αντίστοιχα για το νευρωνικό δίκτυο δευτέρας και τρίτης τάξης, βλέπουμε παρακάτω μέρος του πίνακα dx_est όπως υπολογίστηκε στον εργασιακό χώρο της MATLAB έπειτα από τη σύγκλιση του σφάλματος.



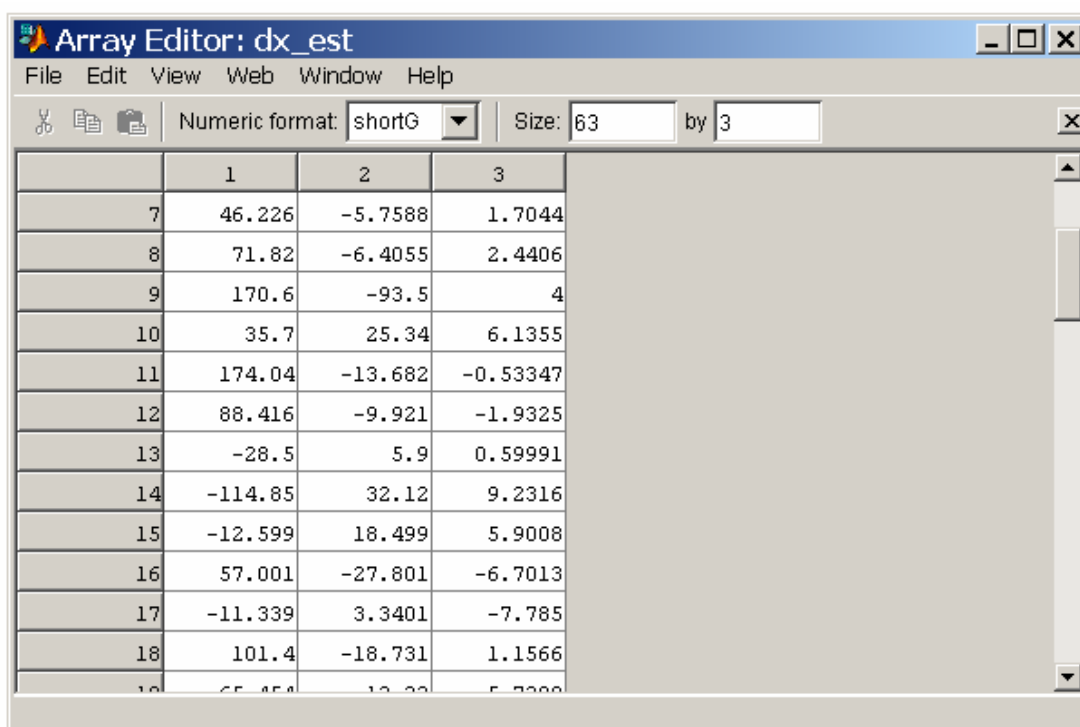
Array Editor: dx_est

File Edit View Web Window Help

Numeric format: shortG Size: 63 by 3

	1	2	3
1	88.251	4.125	2.451
2	125.26	12.17	5.5464
3	71	2.9	-0.9
4	66.487	0.53988	-1.1989
5	95.26	7.5499	2.1793
6	40.852	-3.9238	-4.0203
7	76.215	1.0857	0.66045
8	92.792	1.3795	1.9649
9	119.31	-12.232	2.9743
10	123.6	6.3449	4.3278
11	39.869	2.726	-2.4025
12	19.676	7.7829	-1.2154
13	44.53	12.71	2.8122

Σχήμα 4.5: Ο πίνακας dx_est για το δίκτυο πρώτης τάξης με τις προβλέψεις των μεταβολών των μεταβλητών.



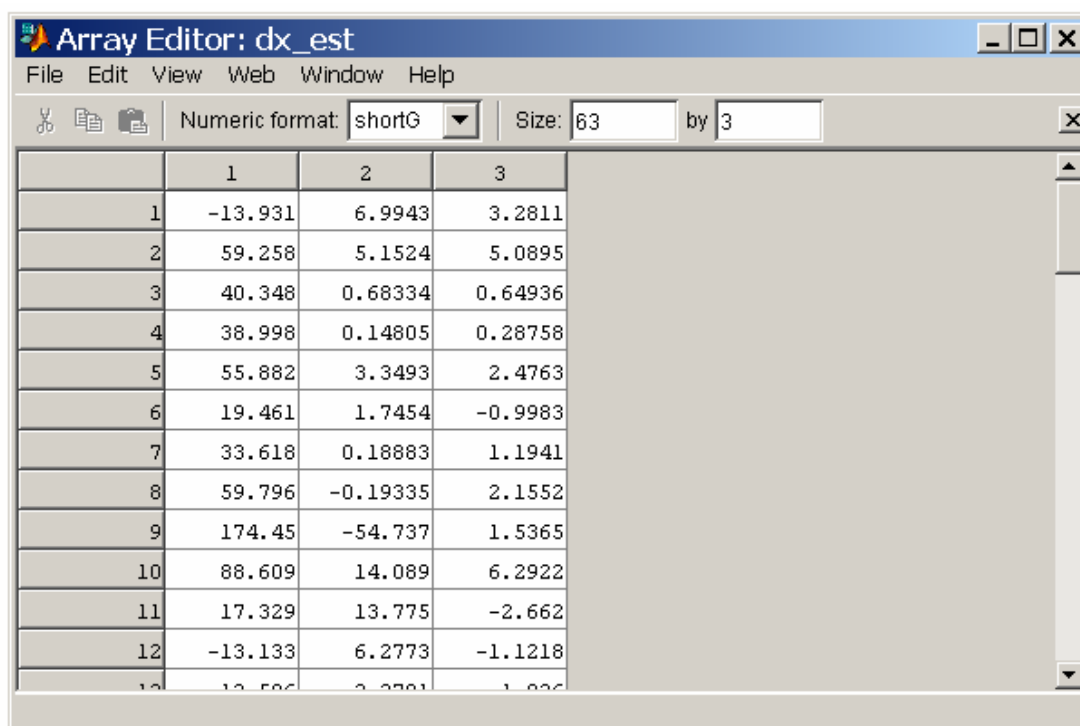
Array Editor: dx_est

File Edit View Web Window Help

Numeric format: shortG Size: 63 by 3

	1	2	3
7	46.226	-5.7588	1.7044
8	71.82	-6.4055	2.4406
9	170.6	-93.5	4
10	35.7	25.34	6.1355
11	174.04	-13.682	-0.53347
12	88.416	-9.921	-1.9325
13	-28.5	5.9	0.59991
14	-114.85	32.12	9.2316
15	-12.599	18.499	5.9008
16	57.001	-27.801	-6.7013
17	-11.339	3.3401	-7.785
18	101.4	-18.731	1.1566
19	65.854	12.22	5.2222

Σχήμα 4.6: Ο πίνακας dx_est για το δίκτυο δευτέρας τάξης με τις προβλέψεις των μεταβολών των μεταβλητών.



	1	2	3
1	-13.931	6.9943	3.2811
2	59.258	5.1524	5.0895
3	40.348	0.68334	0.64936
4	38.998	0.14805	0.28758
5	55.882	3.3493	2.4763
6	19.461	1.7454	-0.9983
7	33.618	0.18883	1.1941
8	59.796	-0.19335	2.1552
9	174.45	-54.737	1.5365
10	88.609	14.089	6.2922
11	17.329	13.775	-2.662
12	-13.133	6.2773	-1.1218
13	13.586	2.2781	1.826

Σχήμα 4.7: Ο πίνακας dx_est για το δίκτυο τρίτης τάξης με τις προβλέψεις των μεταβολών των μεταβλητών.

Βήμα Πέμπτο: Γραφικές Απεικονίσεις

Προκειμένου να συγκρίνουμε την πραγματική μεταβολή των μεταβλητών κατάστασης με εκείνη που πρόβλεψε το νευρωνικό δίκτυο κατασκευάζουμε γραφικές αναπαραστάσεις με το πρόγραμμα figures το οποίο κατασκευάσαμε εμείς. Η πρόβλεψη πραγματοποιείται για τρεις μεταβλητές, συνεπώς περιμένουμε τρία διαγράμματα για κάθε τάξη του δικτύου.

Για το διάγραμμα που συγκρίνει την απόδοση του δικτύου με την πραγματική μεταβολή της μεταβλητής πραγματικό προϊόν (Real GDP) εκτελούνται οι ακόλουθες εντολές:

```
figure(1)
plot(dx(:,1));
hold;
plot(dx_est(:,1), 'r');
hold;
grid;
xlabel('year');
ylabel('\Delta x_1');
axis([1 64 -150 350]);
print -dtiff fig1.tiff
```

Στον άξονα x ο οποίος παίρνει τιμές από 1 έως 64, φαίνονται τα έτη από το έτος 1932 έως το έτος 1995. Στον άξονα y ο οποίος παίρνει τιμές από -150 έως 350, απεικονίζεται η μεταβολή του πραγματικού προϊόντος.

Ομοίως, για το διάγραμμα που συγκρίνει την απόδοση του δικτύου με την πραγματική μεταβολή της μεταβλητής καθαρές εξαγωγές (Real Net Exports) εκτελούνται οι ακόλουθες εντολές:

```
figure(2)
plot(dx(:,2));
hold;
plot(dx_est(:,2), 'r');
hold;
grid
xlabel('year')
ylabel('\Delta x_2')
axis([1 64 -150 100]);
print -dtiff fig2.tiff
```

Στον άξονα x, ο οποίος παίρνει τιμές από 1 έως 64, φαίνονται τα έτη από το έτος 1932 έως το έτος 1995. Στον άξονα y ο οποίος παίρνει τιμές από -150 έως 100, απεικονίζεται η μεταβολή των καθαρών εξαγωγών.

Τέλος για το διάγραμμα που συγκρίνει την απόδοση του δικτύου με την πραγματική μεταβολή της μεταβλητής πληθωρισμός (Inflation Rate), εκτελούνται οι ακόλουθες εντολές:

```
figure(3)
plot(dx(:,3));
hold;
plot(dx_est(:,3), 'r');
hold;
grid
xlabel('year')
ylabel('\Delta x_3')
axis([1 64 -30 40]);
print -dtiff fig3.tiff
```

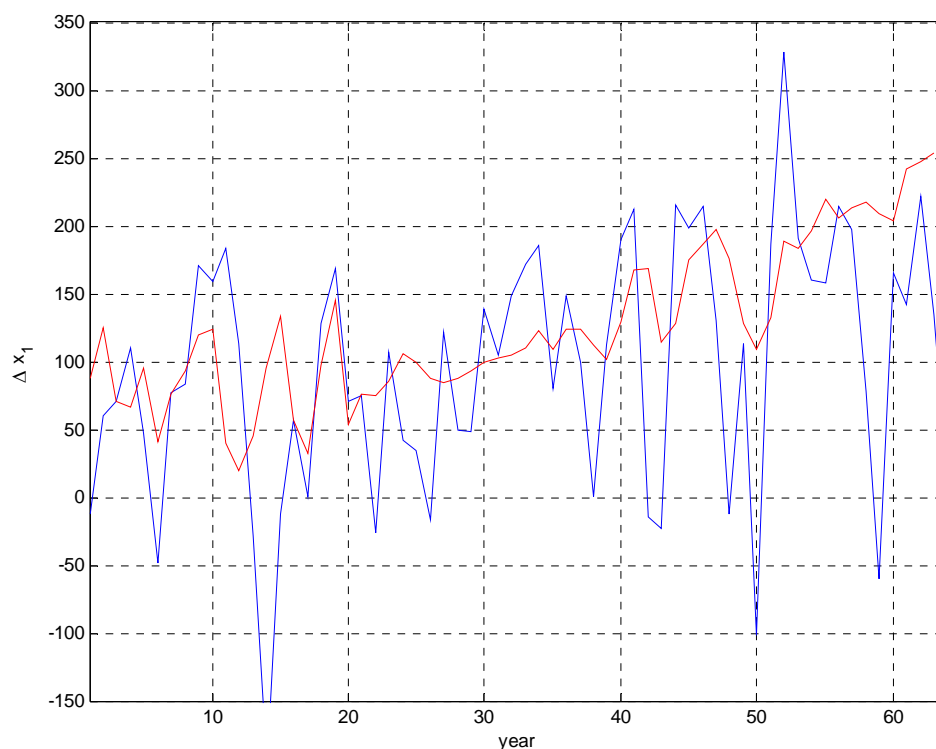
Στον άξονα x, ο οποίος παίρνει τιμές από 1 έως 64, φαίνονται τα έτη από το έτος 1932 έως το έτος 1995. Στον άξονα y ο οποίος παίρνει τιμές από -30 έως 40, απεικονίζεται η μεταβολή του πληθωρισμού.

4.4 Σχολιασμός των Αποτελεσμάτων

Στην παράγραφο αυτή θα δούμε κάποιες γραφικές παραστάσεις με τα δεδομένα της εκπαίδευσης, αλλά και της εφαρμογής (πρόβλεψης). Αντιπαραβάλλοντας τις πραγματικές μεταβολές των μεταβλητών κατάστασης και τις μεταβολές αυτές όπως προβλέφθηκαν από τα δίκτυα πρώτης, δευτέρας και τρίτης τάξης οπτικοποιημένες, μπορούμε να εξάγουμε συμπεράσματα για την απόδοση των δικτύων και την ικανότητα πρόβλεψης με τα εν λόγω δεδομένα.

Νευρωνικό Δίκτυο Πρώτης Τάξης

Συγκεκριμένα, για το δίκτυο πρώτης τάξης, τρέχουμε το πρόγραμμα που δημιουργήσαμε για την έκδοση των γραφημάτων στον εργασιακό χώρο της MATLAB, στον οποίο έχει πλέον συγκλίνει η τιμή του σφάλματος στην τιμή $J=3715.9$ έπειτα από 14 επαναλήψεις και μικρό χρόνο εκπαίδευσης. Έτσι για τη μεταβολή της πρώτης μεταβλητής κατάστασης, που είναι το πραγματικό προϊόν (Real GDP) ακολουθεί το αντίστοιχο διάγραμμα. Με μπλε χρώμα φαίνεται η πραγματική μεταβολή της τιμής της μεταβλητής και με κόκκινο η μεταβολή της τιμής που προέβλεψε το νευρωνικό δίκτυο πρώτης τάξης. Παρατηρούμε ότι υπάρχουν μεγάλες διαφορές στα αποτελέσματα, ενώ οι τιμές ταυτίζονται σε λίγες περιπτώσεις.

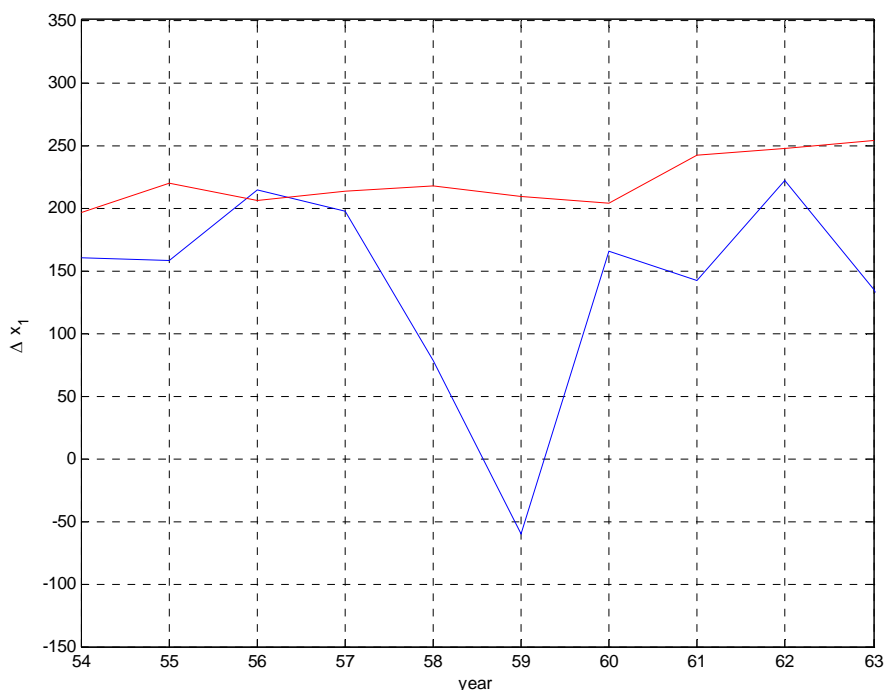


Διάγραμμα 4.1: Πρόβλεψη του GDP με το νευρωνικό δίκτυο πρώτης τάξης.

Το δίκτυο, όπως προαναφέραμε εκπαιδεύτηκε μέχρι το έτος 1985. Συνεπώς τα δεδομένα των χρονοσειρών των 10 τελευταία έτη χρησιμοποιήθηκαν ως δείγμα ελέγχου. Για να εξασφαλίσουμε το διάγραμμα για τα δεδομένα ελέγχου, στο πρόγραμμα figures που κατασκευάσαμε για την εξαγωγή των διαγραμμάτων, πραγματοποιήσαμε την ακόλουθη μεταβολή:

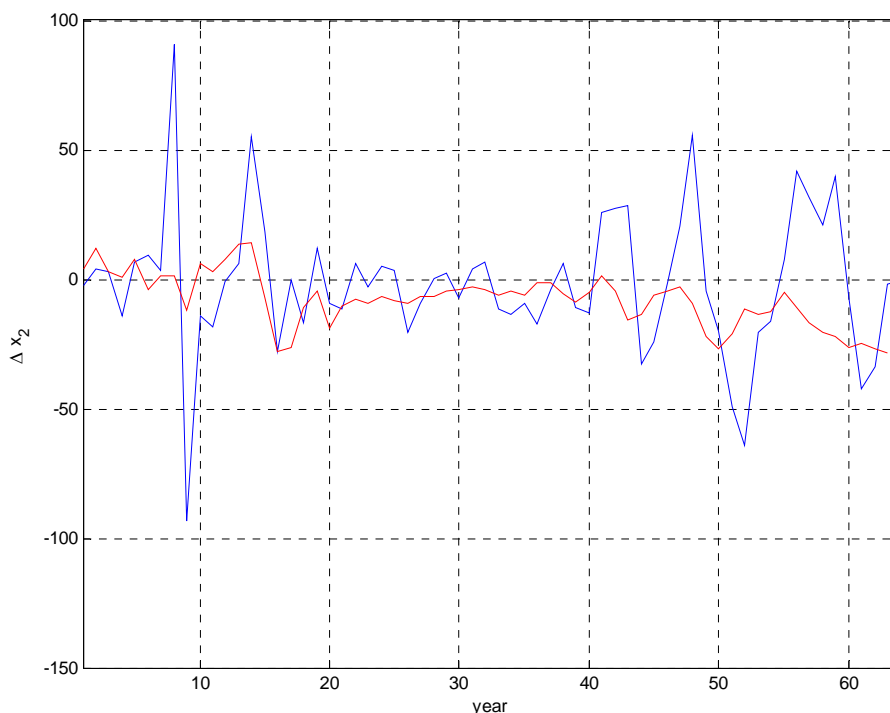
```
figure(1)
plot(dx(:,1));
hold;
plot(dx_est(:,1), 'r');
hold;
grid;
xlabel('year');
print -dtiff fig1_new.tiff
ylabel('\Delta x_1');
axis([54 63 -150 350]);
```

Στο παρακάτω διάγραμμα λοιπόν απεικονίζεται ευκρινέστερα η απόδοση του δικτύου για τα τελευταία δέκα έτη. Μόνο για το έτος 1987 η πραγματική τιμή και η τιμή που προβλέφθηκε είναι αρκετά κοντά, ενώ ειδικά για το έτος 1990 η πραγματική τιμή απέχει σημαντικά από εκείνη της πρόβλεψης. Αξιοσημείωτο είναι ότι το 1990 παρουσιάστηκε οικονομική ύφεση στις ΗΠΑ, γεγονός που δικαιολογεί την πολύ χαμηλότερη πραγματική τιμή του GDP, από εκείνη που προέβλεψε το δίκτυο. Η ύφεση διήρκησε έως το 1993, όπου και παρατηρείται ανάκαμψη της τιμής του πραγματικού προϊόντος. Τα αποτελέσματα δεν είναι ικανοποιητικά.



Διάγραμμα 4.2: Πρόβλεψη του GDP για τα δεδομένα ελέγχου.

Για τη μεταβολή της τιμής της δεύτερης μεταβλητής, δηλαδή των καθαρών εξαγωγών (Real Net Exports), έχουμε:



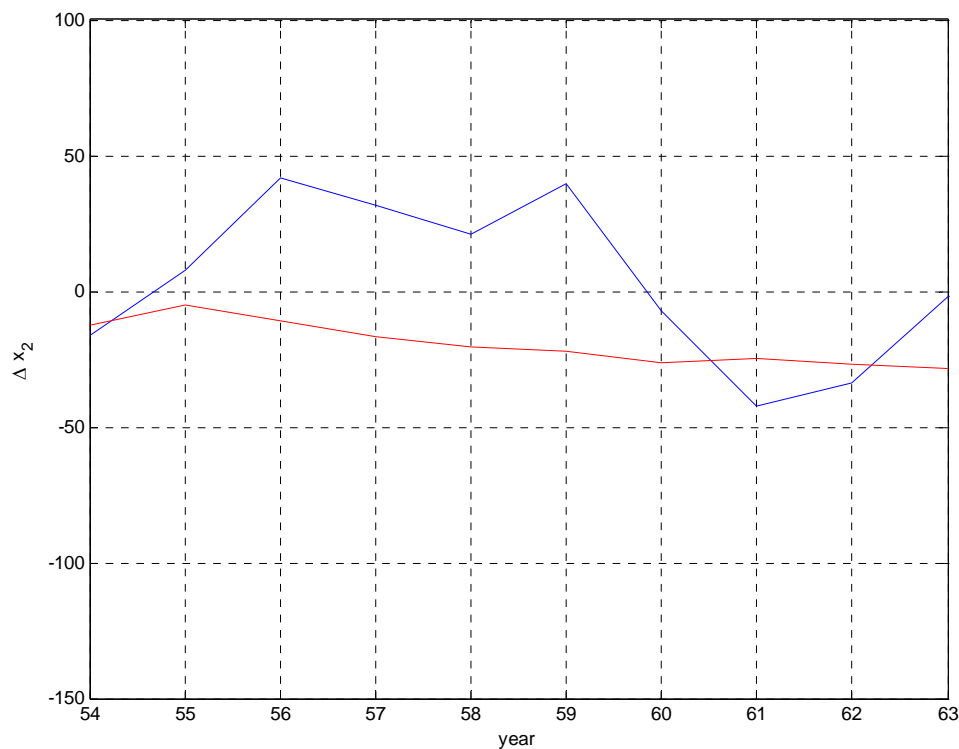
Διάγραμμα 4.3: Πρόβλεψη των καθαρών εξαγωγών με το νευρωνικό δίκτυο πρώτης τάξης.

Με μπλε χρώμα φαίνεται η πραγματική μεταβολή της τιμής της μεταβλητής και με κόκκινο η μεταβολή της τιμής που προέβλεψε το νευρωνικό δίκτυο πρώτης τάξης. Τα αποτελέσματα καταδεικνύουν σημαντικές αποκλίσεις. Οι τιμές ταυτίζονται σε λίγες περιπτώσεις, ενώ είναι χαρακτηριστικό ότι το δίκτυο δεν κατορθώνει να προσεγγίσει τις ακραίες πραγματικές τιμές, δηλαδή είτε πολύ μεγάλες, είτε πολύ μικρές.

Το διάγραμμα με τα δεδομένα ελέγχου εξασφαλίστηκε από το πρόγραμμα figures, αφού πραγματοποιήσαμε την ακόλουθη μεταβολή:

```
figure(2)
plot(dx(:,2));
hold;
plot(dx_est(:,2), 'r');
hold;
grid
xlabel('year')
ylabel('\Delta x_2')
axis([54 63 -150 100]);
print -dtiff fig2_new.tiff
```

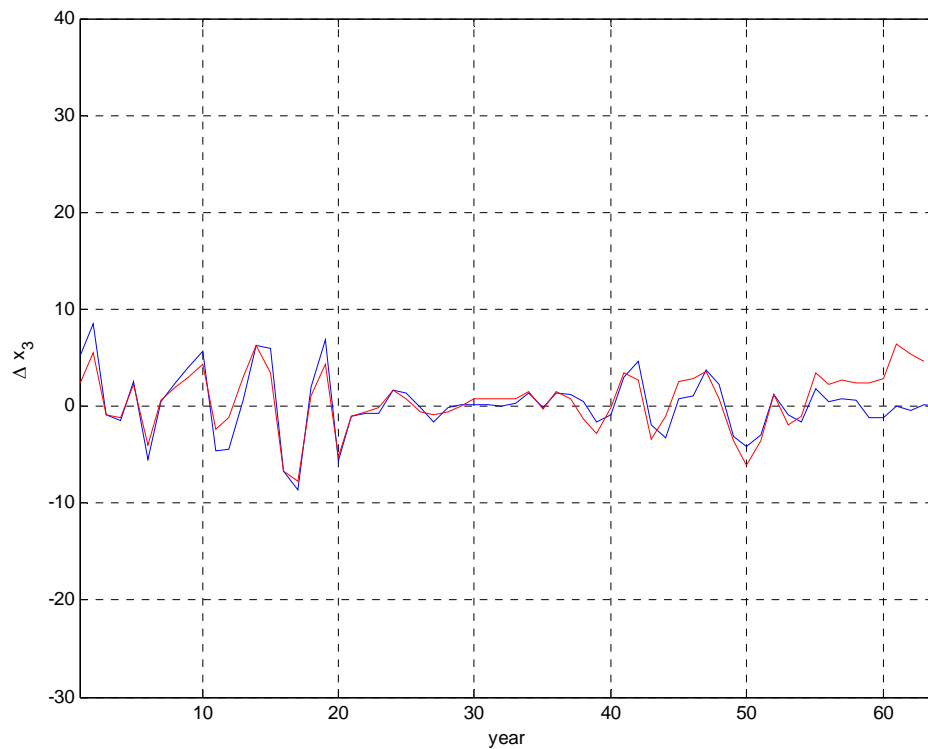
Παρακάτω βλέπουμε το διάγραμμα με την απόδοση του δικτύου για τα δεδομένα ελέγχου:



Διάγραμμα 4.4: Πρόβλεψη των καθαρών εξαγωγών για τα δεδομένα ελέγχου.

Είναι φανερό πως οι πραγματικές τιμές και οι τιμές που προβλέφθηκαν δε συμπίπτουν για κανένα έτος, ενώ οι αποκλίσεις είναι σημαντικές με τις μεγαλύτερες να παρουσιάζονται κατά τα έτη 1987 και 1990. Συνεπώς τα αποτελέσματα της πρόβλεψης για τη μεταβολή της τιμής των καθαρών εξαγωγών δεν είναι ικανοποιητικά.

Για τη μεταβολή της τιμής της τρίτης μεταβλητής, δηλαδή του πληθωρισμού έχουμε:



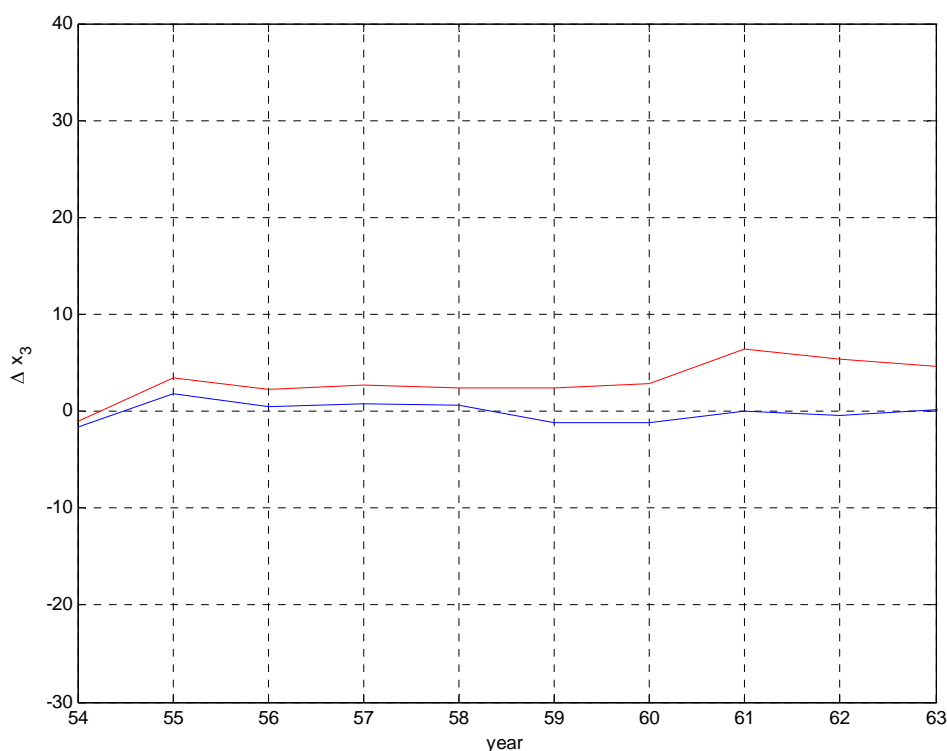
Διάγραμμα 4.5: Πρόβλεψη του πληθωρισμού με το νευρωνικό δίκτυο πρώτης τάξης.

Το δίκτυο φαίνεται να έχει εκπαιδευθεί αρκετά καλά, αφού για τα δεδομένα εκπαίδευσης η πραγματική τιμή της μεταβολής του πληθωρισμού και η τιμή που προβλέφθηκε από το δίκτυο έχουν ελάχιστες αποκλίσεις. Ωστόσο προς τα τελευταία έτη, στα δεδομένα ελέγχου δηλαδή, παρατηρούμε μείωση της απόδοσης του δικτύου.

Το διάγραμμα με τα δεδομένα ελέγχου εξασφαλίστηκε από το πρόγραμμα figures, έπειτα από την ακόλουθη μεταβολή:

```
figure(3)
plot(dx(:,3));
hold;
plot(dx_est(:,3), 'r');
hold;
grid
xlabel('year')
ylabel('\Delta x_3')
axis([54 63 -30 40]);
print -dtiff fig3_new.tiff
```

Έτσι προκύπτει το παρακάτω γράφημα:



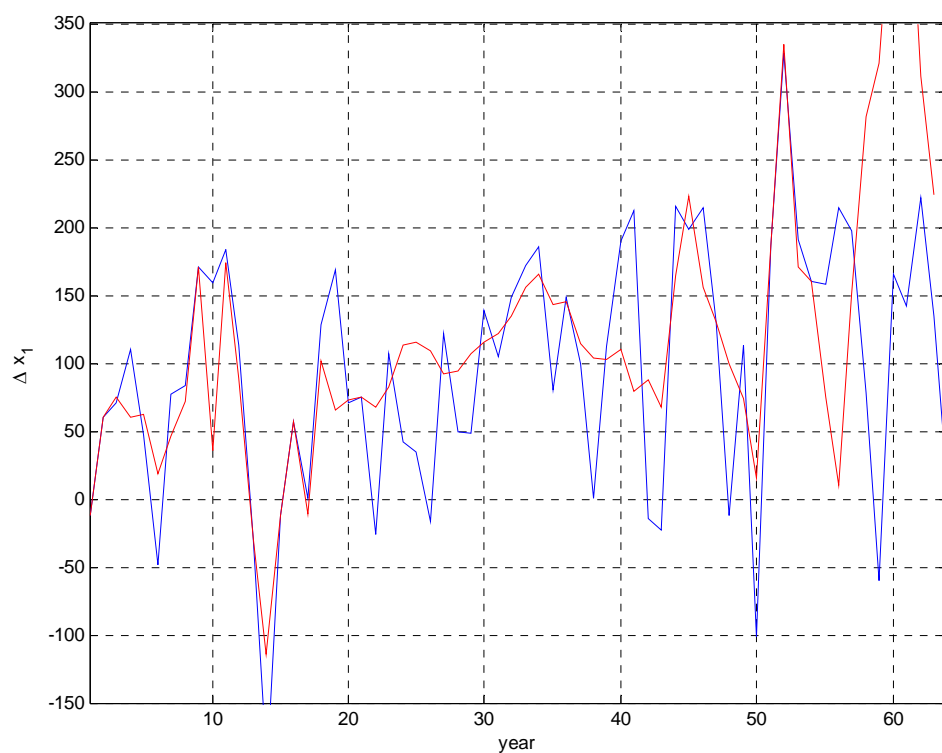
Διάγραμμα 4.6: Πρόβλεψη του πληθωρισμού για τα δεδομένα ελέγχου.

Παρατηρούμε πως για κανένα έτος δεν ταυτίζονται οι πραγματικές τιμές με εκείνες που προέβλεψε το δίκτυο, οι αποκλίσεις ωστόσο δεν είναι πολύ μεγάλες. Η μεγαλύτερη σημειώνεται για το έτος 1992. Η ικανότητα πρόβλεψης του δικτύου λοιπόν για τη μεταβολή της τιμής του πληθωρισμού δεν είναι η επιθυμητή, ωστόσο είναι καλύτερη από την ικανότητα πρόβλεψης του δικτύου για τις προηγούμενες μεταβλητές.

Νευρωνικό Δίκτυο Δευτέρας Τάξης

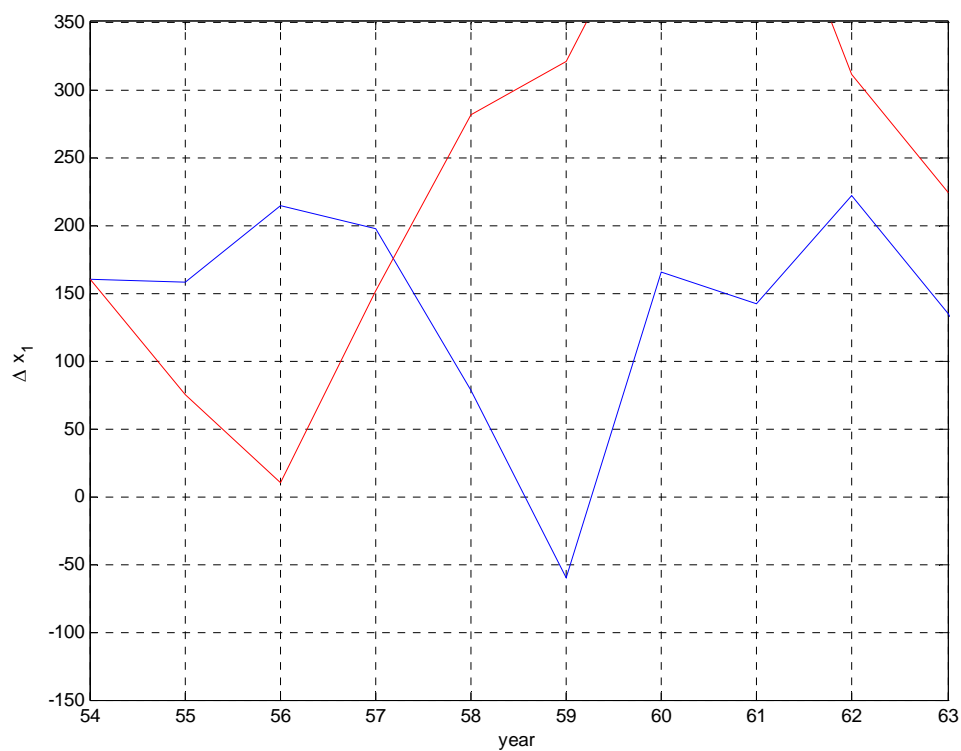
Στον εργασιακό χώρο της MATLAB στον οποίο συνέκλινε η μεταβλητή του σφάλματος στην τιμή $J = 2369.9$ έπειτα από 25 επαναλήψεις και μικρό χρόνο εκπαίδευσης, ωστόσο μεγαλύτερο από εκείνον του δικτύου πρώτης τάξης, τρέχουμε το πρόγραμμα για την έκδοση των γραφημάτων.

Το παρακάτω γράφημα απεικονίζει τη μεταβολή του πραγματικού προϊόντος. Παρατηρούμε μεγάλες αποκλίσεις ανάμεσα στην πραγματική τιμή της μεταβολής του GDP και στην τιμή που προβλέφθηκε και για τα δεδομένα εκπαίδευσης, αλλά ακόμη πιο έντονα για τα δεδομένα ελέγχου.



Διάγραμμα 4.7: Πρόβλεψη του GDP με το νευρωνικό δίκτυο δευτέρας τάξης.

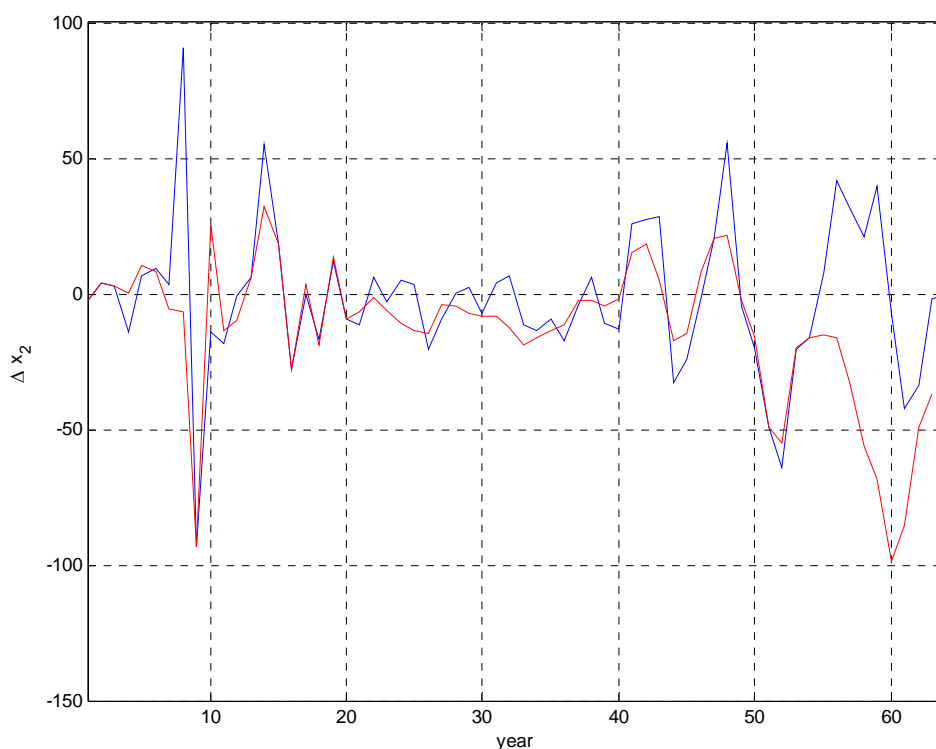
Ακολουθεί το διάγραμμα με την απόδοση του δικτύου για τα δεδομένα ελέγχου:



Διάγραμμα 4.8: Πρόβλεψη του GDP για τα δεδομένα ελέγχου.

Είναι προφανείς οι σημαντικές διαφορές των πραγματικών και των προβλεπόμενων τιμών. Ωστόσο για τα έτη 1985 και 1988 το δίκτυο πραγματοποίησε έγκυρες προβλέψεις. Δεν μπορούμε παρόλα αυτά να παραβλέψουμε τις υπερβολικές αποκλίσεις όλων των υπόλοιπων ετών. Και πάλι για το έτος 1990 έως και το 1993, παρατηρούμε πως η οικονομική ύφεση και τα αποτελέσματά της στη μεταβολή της μεταβλητής, ήταν αδύνατο να προβλεφθούν από το δίκτυο.

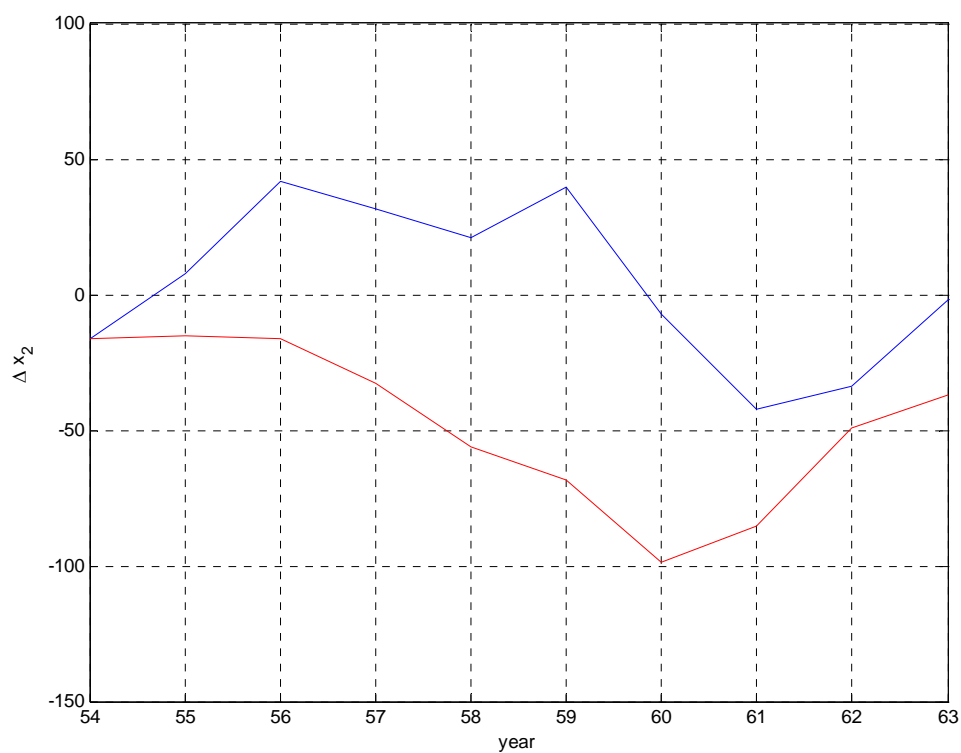
Για τη μεταβολή της τιμής της δεύτερης μεταβλητής, δηλαδή των καθαρών εξαγωγών (Real Net Exports), έχουμε:



Διάγραμμα 4.9: Πρόβλεψη των καθαρών εξαγωγών με το νευρωνικό δίκτυο δευτέρας τάξης.

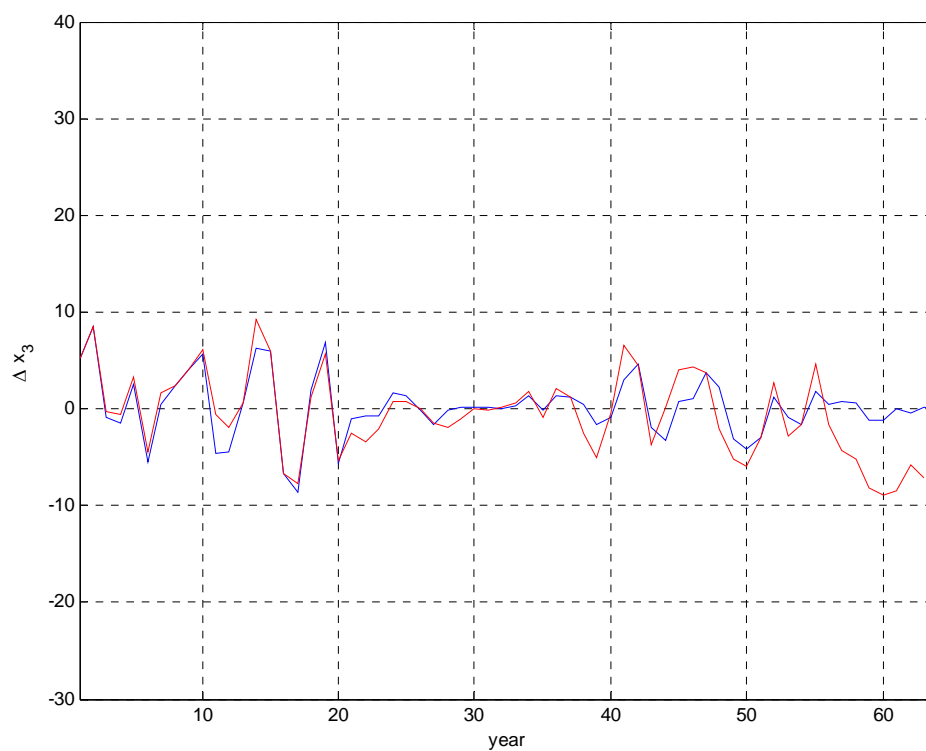
Παρατηρούμε μικρές αποκλίσεις από την πραγματική τιμή και την τιμή που πρόβλεψε το δίκτυο για τα δεδομένα εκπαίδευσης, ενώ ακόμη και όταν αυτές εμφανίζονται η τάση των αποτελεσμάτων είναι ίδια. Ωστόσο για τα τελευταία έτη, που αποτελούν τα δεδομένα ελέγχου, οι διαφορές αυξάνονται ραγδαία.

Ακολουθεί η γραφική παράσταση για τα δεδομένα ελέγχου. Και πάλι η τάση των αποτελεσμάτων είναι η ίδια, οι διαφορές όμως κάθε άλλο παρά αμελητέες είναι. Συγκεκριμένα, πραγματικές και προβλεπόμενες τιμές δε συμπίπτουν για κανένα έτος, ενώ η μεγαλύτερη απόκλιση καταγράφεται και πάλι για το έτος 1990.



Διάγραμμα 4.10: Πρόβλεψη των καθαρών εξαγωγών για τα δεδομένα ελέγχου.

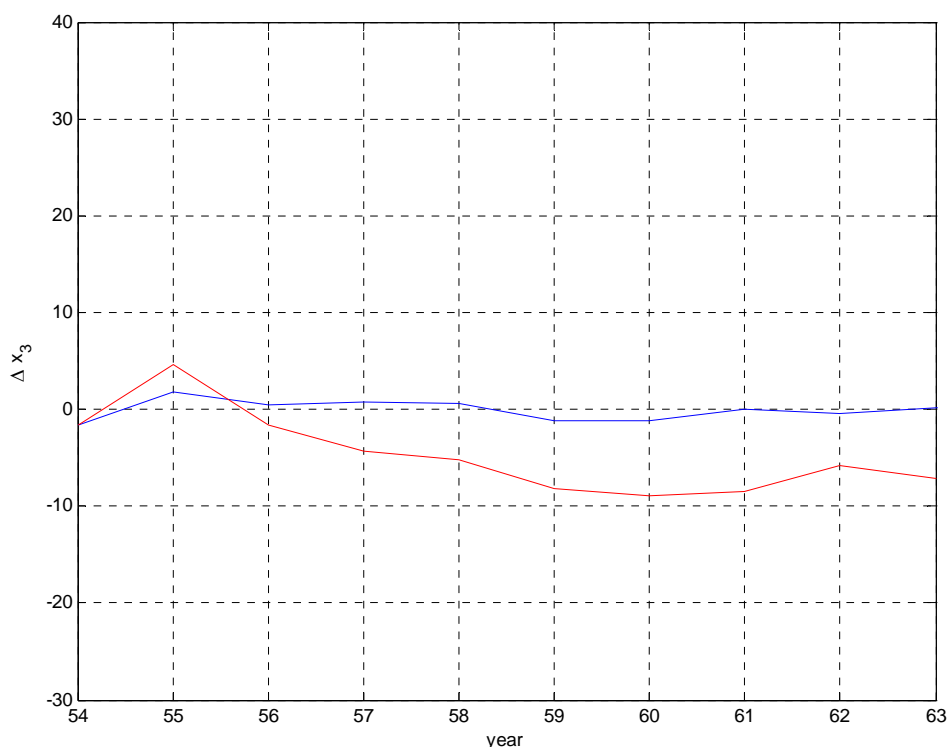
Για τη μεταβολή της τιμής της τρίτης μεταβλητής, δηλαδή του πληθωρισμού έχουμε:



Διάγραμμα 4.11: Πρόβλεψη του πληθωρισμού με το νευρωνικό δίκτυο δευτέρας τάξης.

Όπως βλέπουμε για τα πρώτα έτη τα αποτελέσματα απέχουν ελάχιστα. Πραγματικές και προβλεπόμενες τιμές συμπίπτουν στα περισσότερα σημεία και ακόμη και όταν απέχουν, οι διαφορές κρίνονται αμελητέες. Ωστόσο τα τελευταία έτη οι αποκλίσεις αυξάνονται.

Στο παρακάτω διάγραμμα φαίνεται η απόδοση του δικτύου για τα δεδομένα ελέγχου:



Διάγραμμα 4.12: Πρόβλεψη του πληθωρισμού για τα δεδομένα ελέγχου.

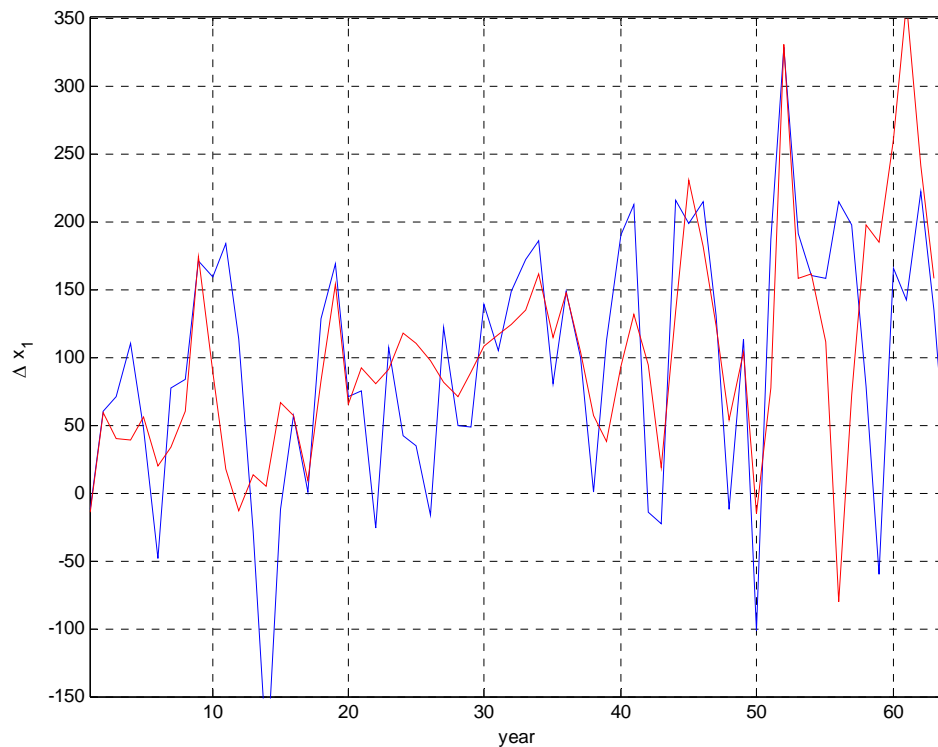
Το δίκτυο τείνει να υπολογίζει μικρότερη τιμή της μεταβολής του πληθωρισμού από την πραγματική. Για κανένα έτος δεν υπάρχει ταύτιση των δύο τιμών, ενώ το μέγεθος των αποκλίσεων κυμαίνεται περίπου στα ίδια επίπεδα. Ωστόσο και πάλι οι μεγαλύτερες αποκλίσεις παρουσιάζονται για την περίοδο 1990 έως 1993.

Νευρωνικό Δίκτυο Τρίτης Τάξης

Το νευρωνικό δίκτυο τρίτης τάξης, εξαιτίας του αυξημένου αριθμού των κόμβων στο κρυμμένο επίπεδο και της πολυπλοκότητας των υπολογισμών, απαιτεί ιδιαίτερα μεγάλο χρόνο εκπαίδευσης. Ενδεικτικά σημειώνουμε ότι κάθε επανάληψη απαιτεί τουλάχιστον 100 ώρες. Μέχρι τη συγγραφή της παρούσας εργασίας η τιμή

του σφάλματος δεν έχει συγκλίνει, θα παρουσιάσουμε όμως τα αποτελέσματά μας για την τελευταία τιμή του σφάλματος $J = 2792.155$, η οποία προέκυψε έπειτα από 8 επαναλήψεις. Στον αντίστοιχο λοιπόν εργασιακό χώρο της MATLAB, τρέχουμε το πρόγραμμα figures για την εξαγωγή των παρακάτω γραφικών παραστάσεων.

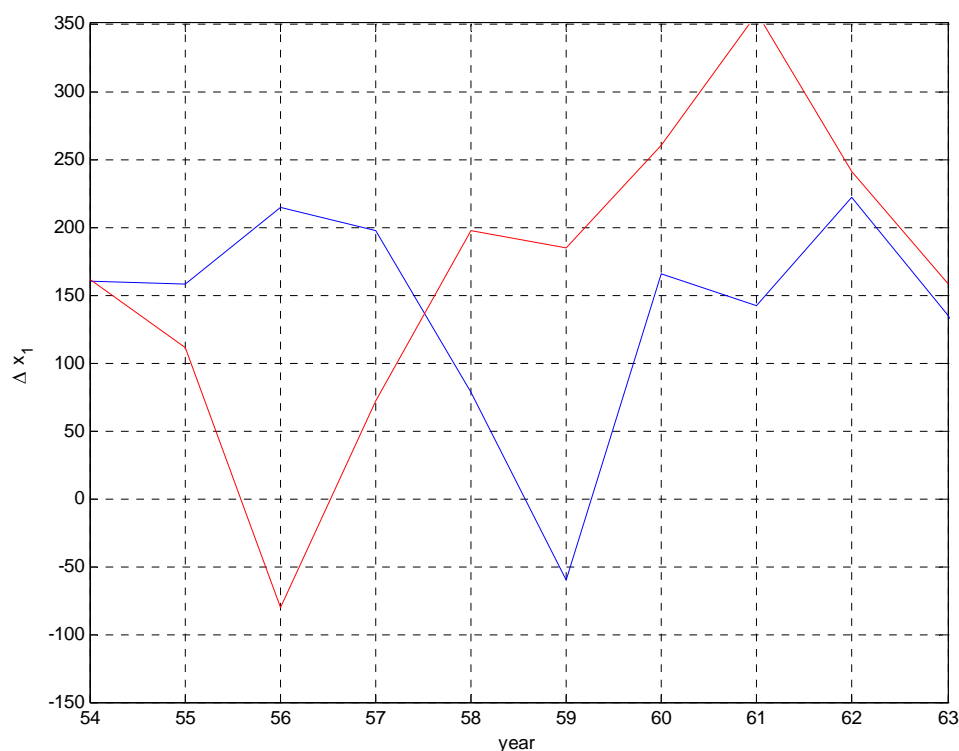
Για τη μεταβολή του πραγματικού προϊόντος έχουμε:



Διάγραμμα 4.13: Πρόβλεψη του GDP με το νευρωνικό δίκτυο τρίτης τάξης.

Το δίκτυο προσεγγίζει σε ορισμένα σημεία την πραγματική μεταβολή του προϊόντος, ωστόσο σημειώνονται και σημαντικές αποκλίσεις, ειδικά για τα δεδομένα ελέγχου.

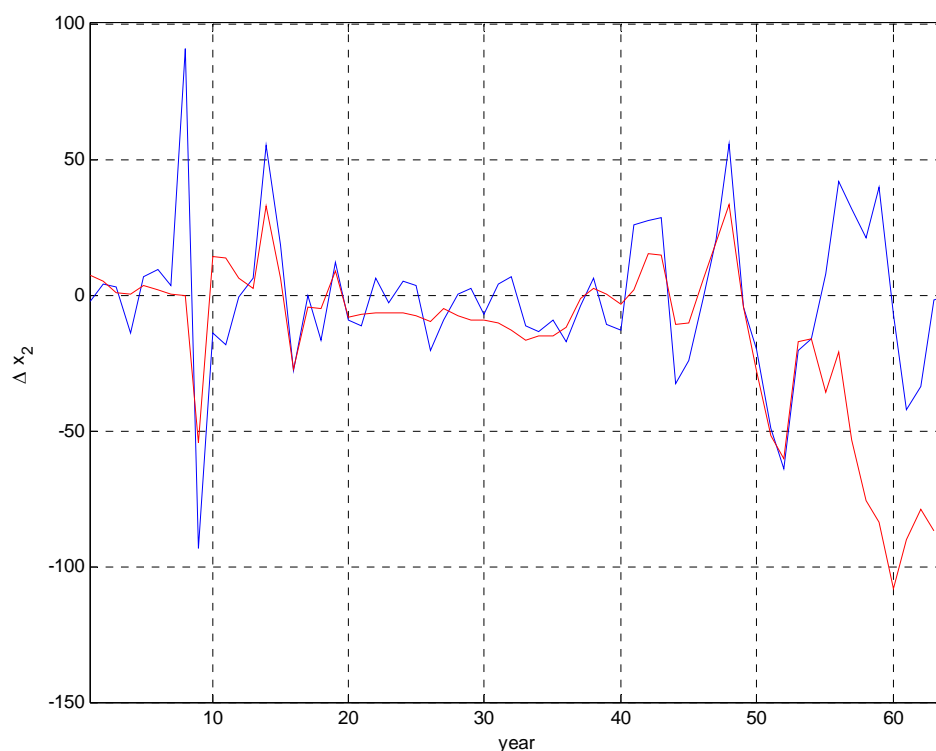
Στο παρακάτω διάγραμμα φαίνεται η απόδοση του δικτύου για τα δεδομένα ελέγχου:



Διάγραμμα 4.14: Πρόβλεψη του GDP για τα δεδομένα ελέγχου.

Η μοναδική περίπτωση που το δίκτυο προσεγγίζει με ακρίβεια την πραγματική μεταβολή του GDP είναι για το έτος 1985, ενώ αρκετά καλή είναι και η πρόβλεψη για το έτος 1994. Για όλα τα υπόλοιπα έτη, οι διαφορές των αποτελεσμάτων είναι σημαντικές με τη μεγαλύτερη να εμφανίζεται για το έτος 1990.

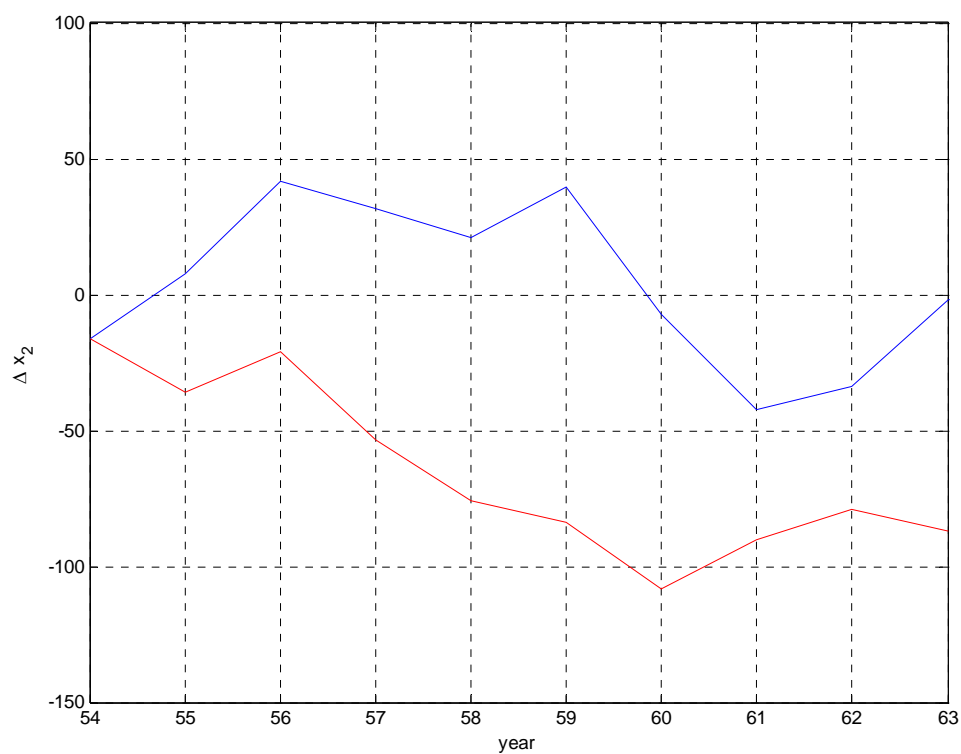
Για τη μεταβολή της τιμής της δεύτερης μεταβλητής, δηλαδή των καθαρών εξαγωγών (Real Net Exports), έχουμε:



Διάγραμμα 4.15: Πρόβλεψη των καθαρών εξαγωγών με το νευρωνικό δίκτυο τρίτης τάξης.

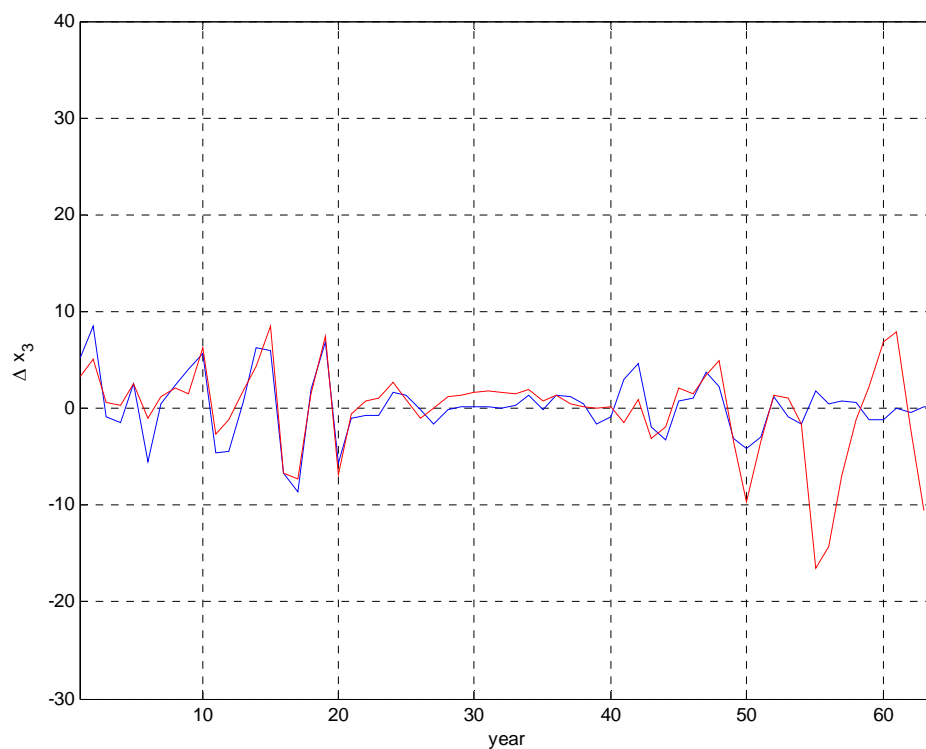
Παρατηρούμε σχετικά μικρές αποκλίσεις, μεταξύ των πραγματικών και των προβλεπόμενων τιμών, αλλά στα δεδομένα ελέγχου οι αποκλίσεις είναι παραπάνω από εμφανείς.

Συγκεκριμένα στο παρακάτω διάγραμμα φαίνεται ευκρινέστερα η απόδοση του δικτύου για τα δεδομένα ελέγχου. Όπως βλέπουμε, τα αποτελέσματα ταυτίζονται μόνο για το έτος 1984, ενώ για όλα τα υπόλοιπα έτη οι διαφορές είναι μεγάλες. Το έτος 1990 παρατηρείται η σημαντικότερη απόκλιση, γεγονός που πιθανότατα οφείλεται στην αναπάντεχη οικονομική ύφεση.



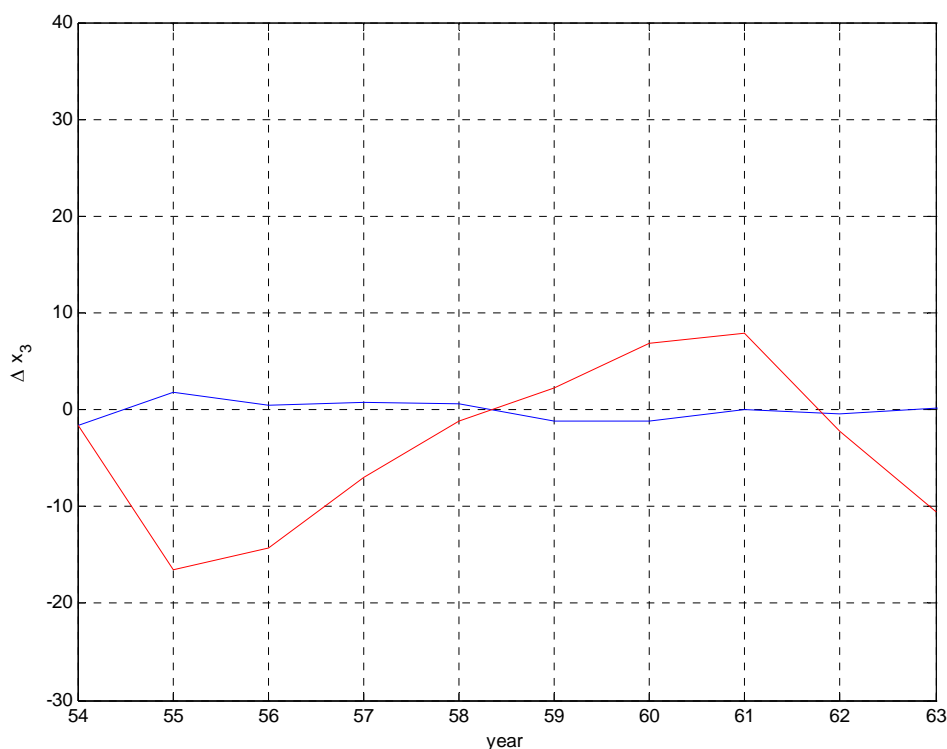
Διάγραμμα 4.16: Πρόβλεψη των καθαρών εξαγωγών για τα δεδομένα ελέγχου.

Για τη μεταβολή της τιμής της τρίτης μεταβλητής, δηλαδή του πληθωρισμού έχουμε:



Διάγραμμα 4.17: Πρόβλεψη του πληθωρισμού με το νευρωνικό δίκτυο τρίτης τάξης.

Παρατηρούμε πως για τα δεδομένα εκπαίδευσης το δίκτυο δίδει έγκυρες προβλέψεις, αφού η διαφορά στα αποτελέσματα μοιάζει πολύ μικρή. Τα δεδομένα ελέγχου, ωστόσο δίνουν διαφορετική εικόνα για την απόδοση του δικτύου. Συγκεκριμένα έχουμε:



Διάγραμμα 4.18: Πρόβλεψη του πληθωρισμού για τα δεδομένα ελέγχου.

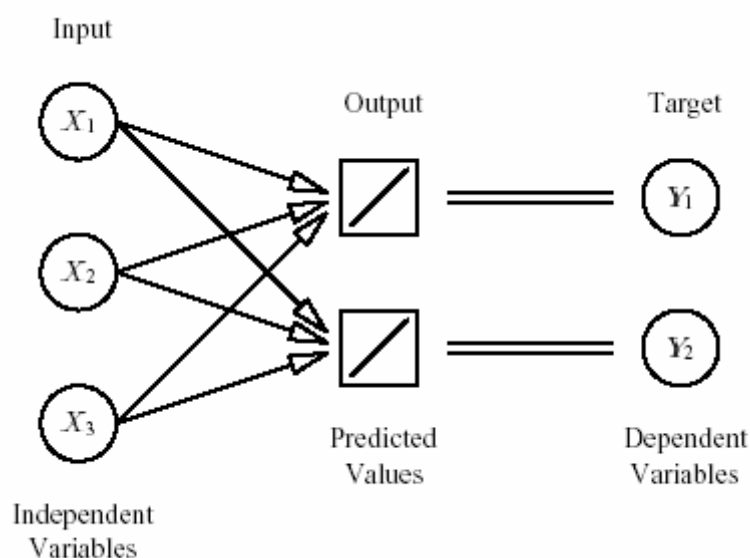
Για τα έτη 1985, 1989 και 1993 οι προβλέψεις του δικτύου είναι έγκυρες. Για όλα τα υπόλοιπα έτη παρατηρούμε σημαντικές αποκλίσεις στα αποτελέσματα, αφού το δίκτυο τείνει είτε να υποτιμά είτε να υπερεκτιμά τη μεταβολή του πληθωρισμού.

ΚΕΦΑΛΑΙΟ 5^ο

Συμπεράσματα

Στην παρούσα εργασία μελετήθηκε η παραδοσιακή μέθοδος της πολλαπλής γραμμικής παλινδρόμησης και η καινοτόμος μέθοδος των νευρωνικών δικτύων υψηλής τάξης, για την πρόβλεψη της μεταβολής μακροοικονομικών μεταβλητών, με σκοπό την πρόβλεψη της εθνικής οικονομίας των ΗΠΑ. Καταλήξαμε στα εξής συμπεράσματα:

- Το μοντέλο της πολλαπλής γραμμικής παλινδρόμησης ταυτίζεται πλήρως με το νευρωνικό δίκτυο πρώτης τάξης, όπως εξάλλου ήταν αναμενόμενο. Οι στατιστικές μέθοδοι και τα νευρωνικά δίκτυα δεν αποτελούν ανταγωνιστικές τεχνικές για την ανάλυση δεδομένων και την πρόβλεψη. Υπάρχουν κοινά χαρακτηριστικά και στα δύο μοντέλα.



Σχήμα 5.1: Δίκτυο πρώτης τάξης = Πολλαπλή γραμμική παλινδρόμηση

- Τα αποτελέσματα της πολλαπλής γραμμικής παλινδρόμησης, αλλά και των υψηλής τάξης νευρωνικών δικτύων εξαρτώνται από τα δεδομένα. Συγκεκριμένα για τη μακροοικονομική πρόβλεψη τα δεδομένα χαρακτηρίζονται από μη γραμμικότητα και από απρόβλεπτα ακραία γεγονότα.

Στα διαγράμματα έγινε φανερό πως για περιόδους οικονομικής ύφεσης οι προβλέψεις δεν ήταν έγκυρες. Η ικανότητα πρόβλεψης και γενίκευσης των συγκεκριμένων δικτύων είναι ιδιαίτερα υψηλή, ωστόσο με τα συγκεκριμένα δεδομένα, τα αποτελέσματα δεν ήταν τα επιθυμητά.

- Τα μοντέλα πρόβλεψης νευρωνικών δικτύων καθώς και τα γραμμικά μοντέλα θα πρέπει να χρησιμοποιούνται σε συνδυασμό με επιπλέον εργαλεία για τη μακροοικονομική πρόβλεψη, η οποία αποτελεί πρόκληση εξαιτίας των κριτικών και εξωγενών παραγόντων που την επηρεάζουν.
- Ο κώδικας που αναπτύχθηκε είναι δυνατόν να εφαρμοσθεί με επιτυχία σε επιπλέον τομείς πρόβλεψης, πέραν της μακροοικονομικής που μελετήθηκε στην παρούσα εργασία.

Παράρτημα Ι

Δεδομένα Χρονοσειρών

	real gdp (x1)	real net exports (x2)	inflation rate (x3)	real government exp (x4) (u1)	exchange rate (x4) (u2)	short interest rate (x5) (u3)	M1 (x7) (z)
1932	674.1	-3.0	-10.3	103.8	118.1	7.9	2.9
1933	661.5	-5.2	-5.1	105.5	110.4	-1.7	3.0
1934	721.4	-1.0	3.4	130.4	95.9	-1.6	3.1
1935	792.4	1.9	2.5	153.0	97.5	-0.2	3.0
1936	902.3	-12.4	1.0	200.9	96.9	-2.9	3.0
1937	950.0	-5.5	3.6	207.3	97.1	2.8	3.1
1938	901.8	3.7	-1.9	180.0	97.8	2.3	2.9
1939	978.9	7.3	-1.4	203.7	103.3	-0.2	2.8
1940	1062.2	98.2	1.0	146.3	115.0	-4.2	2.7
1941	1232.8	4.7	5.0	354.1	112.8	-10.0	2.8
1942	1392.2	-9.2	10.7	590.9	112.8	-5.5	3.0
1943	1575.9	-27.5	6.1	787.0	112.8	-1.0	2.8
1944	1689.1	-28.5	1.7	862.7	112.8	-1.5	2.6
1945	1660.6	-22.6	2.3	751.6	112.8	-7.8	2.3
1946	1461.4	32.3	8.5	339.5	112.8	-13.5	2.1
1947	1448.8	50.8	14.4	263.2	112.8	-6.7	2.2
1948	1505.8	23.0	7.7	294.8	112.8	2.4	2.4
1949	1506.2	22.5	-0.9	317.2	114.4	0.5	2.5
1950	1634.8	5.6	1.0	340.1	127.1	-6.5	2.6
1951	1803.9	17.5	7.9	475.1	128.5	-0.1	2.9
1952	1874.3	8.3	2.3	546.7	130.3	1.0	2.9
1953	1949.4	-3.2	1.3	583.9	128.4	1.9	3.0

1954	1923.5	3.0	0.6	527.0	128.4	1.8	3.0
1955	2030.3	0.0	-0.2	515.1	121.1	0.7	3.2
1956	2072.1	5.2	1.5	516.7	116.2	0.4	3.3
1957	2106.7	8.4	2.9	534.7	115.0	1.1	3.5
1958	2090.6	-12.3	2.7	543.3	115.5	1.4	3.5
1959	2212.6	-21.8	1.0	557.8	114.7	3.1	3.6
1960	2261.7	-21.3	0.9	579.8	115.2	2.8	3.8
1961	2309.8	-19.1	1.1	602.2	116.6	1.8	3.8
1962	2449.1	-26.5	1.2	643.3	119.0	2.0	4.0
1963	2554.1	-22.7	1.3	662.1	119.6	2.2	4.1
1964	2702.9	-15.9	1.3	681.9	119.7	2.4	4.2
1965	2874.8	-27.3	1.6	710.6	119.7	1.4	4.4
1966	3060.2	-40.9	3.0	775.6	119.6	2.8	4.6
1967	3140.2	-50.1	2.8	824.9	120.0	0.9	4.7
1968	3288.6	-67.2	4.2	855.7	122.1	0.5	4.8
1969	3388.0	-71.3	5.4	857.4	122.4	1.9	4.9
1970	3388.2	-65.0	5.9	835.8	121.1	3.5	5.0
1971	3500.1	-75.8	4.2	829.0	117.8	1.8	5.0
1972	3690.3	-88.9	3.3	841.3	109.1	-1.5	5.2
1973	3902.3	-63.0	6.3	849.2	99.1	-2.9	5.4
1974	3888.2	-35.6	11.0	864.6	101.4	0.7	5.6
1975	3865.1	-7.2	9.1	861.8	98.5	0.5	5.8
1976	4081.0	-39.9	5.8	870.1	105.6	-1.1	6.1
1977	4279.3	-64.2	6.5	886.6	103.3	-2.0	6.3
1978	4493.7	-65.6	7.6	921.7	92.4	-3.3	6.6
1979	4624.0	-45.3	11.3	944.6	88.1	-2.6	6.9
1980	4611.9	10.1	13.5	965.9	87.4	1.9	7.0
1981	4724.9	5.6	10.4	983.3	103.3	8.6	7.3
1982	4623.6	-14.1	6.2	965.9	116.5	8.7	7.2
1983	4810.1	-63.3	3.2	985.0	125.3	4.5	7.0
1984	5138.2	-127.3	4.4	1026.3	138.3	6.6	7.2

1985	5329.5	-147.9	3.5	1081.7	143.2	6.1	7.1
1986	5489.9	-163.9	1.9	1131.4	112.3	2.7	6.6
1987	5648.4	-156.2	3.7	1161.8	96.9	2.8	6.3
1988	5862.9	-114.4	4.1	1178.6	92.8	2.8	6.5
1989	6060.3	-82.7	4.8	1216.6	98.7	3.4	7.0
1990	6138.7	-61.9	5.4	1251.1	89.0	3.8	7.1
1991	6078.9	-22.3	4.2	1257.7	89.5	4.8	6.9
1992	6244.4	-29.5	3.0	1263.8	86.6	0.9	6.5
1993	6386.4	-72.0	3.0	1261.6	93.2	0.7	6.1
1994	6608.7	-105.7	2.6	1261.6	93.2	1.8	6.1
1995	6742.9	-107.6	2.8	1262.5	93.2	3.1	6.4

Παράρτημα II

Κώδικας

Neural.m

```

clear all
close all

global XX dx N

load matlab

% o xristis orizei tin taxsi toy diktiou
order=1;
N=max(size(x));
if order==1
    XX=zeros(N,7);
end
if order==2
    XX=zeros(N,56);
end
if order==3
    XX=zeros(N,399);
end

for i=1:N,
    XX(i,1:7)=x(i, 1:7);
if order>1
    xx=x(i,:)*x(i,:);
    yy=reshape(xx, 49,1);
    XX(i,8:56)=yy';
    if order==3
        yyy=yy*x(i,:);
        yyyy=reshape(yyy, 343,1);
        XX(i,57:399)=yyyy';
    end
end
end

dx=zeros(N,3);
for i=1:N-1,
    dx(i,1:3)=x(i+1,1:3)-x(i,1:3);
end
if order==1

```

```

    load w1.dat
    W0=w1;
end
if order==2
    load w2.dat
    W0=w2;
end
if order==3
    load w3.dat
    W0=w3;
end
W=fmins('nnopt', W0);

```

```

if order==1
    save W1.dat W -ascii
end
if order==2
    save W2.dat W -ascii
end
if order==3
    save W3.dat W -ascii
end

dx_est=nn(W);

```

nnopt.m

```

function J=nnopt(W)
global XX dx N

J=0;

for i=1:N-10,
    J=J+norm(dx(i,:)-W*XX(i,:));
end
J

```

nn.m

```

function dx_est=nn(W)
global XX dx N

for i=1:N-1
    dx_est(i,:)=(W*XX(i,:))';
end

```

figures.m

```
figure(1)
plot(dx(:,1));
hold;
plot(dx_est(:,1), 'r');
hold;
grid;
xlabel('year');
ylabel('\Delta x_1');
axis([1 64 -150 350]);
print -dtiff fig1.tiff
```

```
figure(2)
plot(dx(:,2));
hold;
plot(dx_est(:,2), 'r');
hold;
grid;
xlabel('year')
ylabel('\Delta x_2')
axis([1 64 -150 100]);
print -dtiff fig2.tiff
```

```
figure(3)
plot(dx(:,3));
hold;
plot(dx_est(:,3), 'r');
hold;
grid;
xlabel('year')
ylabel('\Delta x_3')
axis([1 64 -30 40]);
print -dtiff fig3.tiff
```

figures_control.m

```
figure(1)
plot(dx(:,1));
hold;
plot(dx_est(:,1), 'r');
hold;
grid;
xlabel('year');
ylabel('\Delta x_1');
axis([54 63 -150 350]);
print -dtiff fig1_new.tiff
```

```
figure(2)
plot(dx(:,2));
hold;
plot(dx_est(:,2), 'r');
hold;
grid;
xlabel('year')
ylabel('\Delta x_2')
axis([54 63 -150 100]);
```

```
print -dtiff fig2_new.tiff
```

```
figure(3)
plot(dx(:,3));
hold;
plot(dx_est(:,3), 'r');
hold;
grid;
xlabel('year')
ylabel('\Delta x_3')
axis([54 63 -30 40]);
```

```
print -dtiff fig3_new.tiff
```

Βιβλιογραφία

- [1] Hendry, D.F., Clements M.P., “Economic Forecasting: Some Lessons from Recent Research”, *European Central Bank working series*, 2001
- [2] Hendry, D.F., “How economists forecast”, *MIT Press*, 2001
- [3] Makridakis, S., Wheelwright, S., Hyndman, R., “Forecasting, Methods and Applications”, *3rd Edition, Wiley*, 1998
- [4] Martin, M., “Best Practices in Macroeconomic Forecasting: Key Issues for Discussion”, *Dept Relief International*, 1999
- [5] Stedman, B., “Professional Role A: An Introduction to the Business of Design”
- [6] Rubinfeld, D.L., “Reference Guide on Multiple Regression”
- [7] Boutsikas M.V., “Σημειώσεις μαθήματος “Στατιστικά Προγράμματα””, Τμήμα Στατ. & Ασφ. Επιστήμης, *Πανεπιστήμιο Πειραιώς*, 2004
- [8] Shachmurove, Y., “Applying Artificial Neural Networks to Business, Economics and Finance”, *The City College of the City University of New York and the University of Pennsylvania*
- [9] Zhang, G.B., Patuwo, E., Hu, M.Y., “Forecasting with Artificial Neural Networks: The state of the art”, *International Journal of forecasting*, 1998
- [10] Gonzalez, S., “Neural Networks for Macroeconomic Forecasting: A Complementary Approach to Linear Regression Models”, Working Paper 2000-07, *Department of Finance, Canada*
- [11] Kuan, C.M., White, H., “Artificial Neural Networks: An Econometric Perspective”, *Econometric Reviews*, 1994
- [12] Funahashi, K., “On the Approximate Realization of Continuous Mappings by Neural Networks”, *Neural Networks*, 1989
- [13] Hornik K., Stinchcombe M., White H., “Multilayer Feedforward Networks are Universal Approximations”, *Neural Networks*, 1989
- [14] Swanson N.R., White H., “A Model Selection Approach to Real Time Macroeconomic Forecasting Using Linear Models and Artificial Neural Networks”, *Review of Economic and Statistics*, 1997
- [15] Sarle, W.S., “Neural Networks and Statistical Models”, *Proceedings of the Nineteenth Annual SAS Users Group International Conference*, 1994

- [16] Cheng, B., Titterington, D.M., “Neural Networks: A review from a statistical perspective”, *Statistical Science*, 1994
- [17] Fine, T.L., “Feed-forward Neural Network Methology”, *NY: Springer*, 1999
- [18] Lippmann, R.P., “An Introduction to Computing with Neural Networks”, *IEEE ASSP Magazine*, 1987
- [19] Giles, C.L., Maxwell, T., “Learning, Invariance and Generalization in High-Order Neural Networks”, *Applied Optics*, 1987
- [20] Heywood, M., Noakes, P., “A Framework for Improved Training of Sigma-Pi Networks”, *IEEE Transaction on Neural Networks*, 1995
- [21] Kosmatopoulos, E.B., Polycarpou, M.M., Christodoulou, M.A., Ioannou, P.A., “Higher Order Neural Network Structures for Identification of Dynamical Systems”, *IEEE Transactions on Neural Networks*, 1995
- [22] Shin, Y., Ghosh, J., “The Pi-Sigma Network: An Efficient Higher-Order Network for Pattenm Classification and Function Approximation”, *Proc. Int.Joint Conference on Neural Networks IJCNN*, 1991
- [23] Zhang, X., “Time Series Analysis and prediction by neural networks”, *Optimization Methods and Software 4*, 1994
- [24] Happel, B.L.M., Murre, J.M.J., “The design and evolution of modular neural networks architectures”, *Neural Networks 7*, 1994
- [25] Cottrell, M., Girard, B., Girard, Y., Mangeas, M., Muller, C., “Neural Modelling for time series: a statistical stepwise method for weight elimeination”, *IEEE Transaction on Neural Networks 6*, 1995