



Τμήμα Μηχανικών Παραγωγής & Διοίκησης

Διπλωματική εργασία με θέμα:

«Πρόβλεψη τιμών πολύτιμων μετάλλων με τη χρήση
νευρο-ασαφών μεθόδων (ANFIS)»

Παπαδόπουλος Ιωάννης

Επιβλέπων καθηγητής: Ατσαλάκης Γεώργιος

Χανιά, ΦΕΒΡΟΥΑΡΙΟΣ 2015

ΠΕΡΙΕΧΟΜΕΝΑ

ΚΕΦΑΛΑΙΟ 1:Εισαγωγή.....	3
ΚΕΦΑΛΑΙΟ 2:Σχετική βιβλιογραφία.....	4
ΚΕΦΑΛΑΙΟ 3:Θεωρητική προσέγγιση των μοντέλων.....	8
3.1 Προσδιορισμός συστήματος.....	8
3.2 Εκτιμητής ελαχίστων τετραγώνων (Least-squares estimator).....	10
3.3 Οπισθοδρόμηση για πρωσοτροφοδοτούμενα δίκτυα.....	15
3.4 Νευρωνικά δίκτυα.....	26
3.4.1 Perceptrons. Αρχιτεκτονική και κανόνας εκμάθησης.....	26
3.4.2 Οπισθοδρομούμενα πολυεπίπεδα δίκτυα perceptron.....	30
3.4.3 Κανόνας εκμάθησης οπισθοδρόμησης.....	32
3.5 Υβριδικός αλγόριθμος εκμάθησης: Συνδυασμός βαθμωτής ελαχιστοποίησης και εκτιμητή ελαχίστων τετραγώνων.....	35
3.6 Θεωρητικό υπόβαθρο του ANFIS.....	40
ΚΕΦΑΛΑΙΟ 4: Ανάλυση και εφαρμογή των μοντέλων.....	45
4.1 Εισαγωγή.....	45
4.2 Αρχιτεκτονική του ANFIS.....	45
4.3 Περιορισμοί του ANFIS.....	51
4.4 Εκπαίδευση του μοντέλου ANFIS.....	52
ΚΕΦΑΛΑΙΟ 5: Αποτελέσματα.....	52
5.1 Αποτελέσματα του μοντέλου ANFIS.....	53
5.2 Αποτελέσματα του νευρωνικού δικτύου.....	69
5.3 Συμπεράσματα.....	73
ΚΕΦΑΛΑΙΟ 6: Βιβλιογραφία.....	75

ΚΕΦΑΛΑΙΟ 1: Εισαγωγή

Ο κύριος σκοπός αυτής της εργασίας είναι να προβλεφθεί η τιμή πολύτιμων μετάλλων χρησιμοποιώντας το ιστορικό των τιμών που επηρεάζονται από πολλούς οικονομικούς παράγοντες. (Ismail et al). Συνήθως ο κύριος στόχος πολλών επενδυτών είναι να πάρουν μια αξιόπιστη μέθοδο πρόβλεψης για τη χρηματιστηριακή αγορά, αυτός ο στόχος μπορεί να επιτευχθεί με τη βοήθεια οικονομετρικών εννοιών, στατιστικών μεθόδων και τεχνικών ανάλυσης το οποίο οδήγησε πολλούς ερευνητές να επινοήσουν νέα μοντέλα και μεθοδολογίες για την πρόβλεψη. Κατά τις τελευταίες δεκαετίες, η τεχνητή νοημοσύνη βρέθηκε να παρέχει πολύτιμα αποτελέσματα σε αυτά προβλήματα. Πολλά πολύτιμα μοντέλα εφαρμόζονται για να λύσουν αυτό το πρόβλημα στην επιστήμη των υπολογιστών και επιχειρησιακής έρευνας, από τα τεχνητά νευρωνικά δίκτυα (ΤΝΔ), την ασαφή λογική και εξελικτικούς αλγορίθμους όπως τους γενετικούς αλγόριθμους και διαφορική εξέλιξη αλλά η κάθε μεθοδολογία έχει κάποια πλεονεκτήματα και μειονεκτήματα. (Qinghua Wen et al) Ένα νέο μοντέλο που προτείνεται το οποίο συνδυάζει τη γενετική αλγορίθμου και νευρο-ασαφές συστήματος με την κβαντική πληροφορική και έδειξε τη δύναμη της χρήσης των κβαντικών γενετικών αλγορίθμων για τη βελτιστοποίηση είναι το ANFIS, ένα μοντέλο με μια πολύ καλή απόδοση και σχετικά πολύ καλά αποτελέσματα . (Osman Hegazy et al).

ΚΕΦΑΛΑΙΟ 2: Σχετική βιβλιογραφία

Το 1998, οι Garcia-Ferrer και Queralt έδειξαν ότι η παράγωγος της τάσης του μοντέλου τους μπορεί να χρησιμοποιηθεί ως δείκτης της ελλοχεύουσας (underlying) αύξησης καθώς επίσης και σαν ένα εργαλείο για την πρόβλεψη των σημείων αλλαγής σε εποχιακές οικονομικές χρονοσειρές. Το μοντέλο χρησιμοποίησε δεδομένα από το 1975 έως το 1993 για τέσσερις μηνιαίους δείκτες τις Ισπανικής οικονομίας. Ο στόχος ήταν να αξιολογηθεί η ικανότητα του μοντέλου μετά την πτώση των δεικτών κατά το έτος 1993 να προβλέψει την αλλαγή της τάσης και την άνοδό τους κατά το έτος 1994. Στη συνέχεια τα αποτελέσματα που προέκυψαν από τη χρήση του μοντέλου συγκρίθηκαν με τα αποτελέσματα από άλλα μοντέλα όπως το ARIMA των Box και Jenkins (1970) και το Βασικό Δομικό Μοντέλο των Harvey και Todd (1983) καθώς και με τα αποτελέσματα δύο ανεξάρτητων αναφορών πρόβλεψης για την Ισπανική οικονομία: το 'Forecasting and Monitoring Service for the Spanish Economy' (Treadway, 1994) και το σχέδιο ERISTE (Garcia-Ferrer et al). Τα αποτελέσματα δείχνουν ότι μοντέλο τους προβλέπει καλύτερα και πιο αποτελεσματικά την αλλαγή της τάσης για τους τέσσερις δείκτες το έτος 1994 σε σχέση με τα υπόλοιπα μοντέλα.

Το 2003, οι Canova και Ciccarelli χρησιμοποίησαν ένα Μπαγεσιανό Var μοντέλο (Bayesian panel Var) με στόχο την πρόβλεψη των σημείων αλλαγής της τάσης. Το μοντέλο χρησιμοποιήθηκε για την πρόβλεψη των σημείων αλλαγής τάσης στους ρυθμούς ανάπτυξης των επτά πιο πλούσιων κρατών της γης. Χρησιμοποιήθηκαν τριμηνιαία δεδομένα από το 1973 έως το 1993 για τρεις εθνικούς δείκτες και ένα παγκόσμιο δείκτη. Τα αποτελέσματα από τη χρήση του μοντέλου συγκρίνονται με τα αποτελέσματα από το ARIMA, τα μη περιορισμένα (unrestricted) Var μοντέλα και άλλα δομικά μοντέλα. Τα αποτελέσματα δείχνουν ότι το μοντέλο τους έχει μεγαλύτερη πιθανότητα να προβλέψει τα σημεία αλλαγής τάσης σε σχέση με τα άλλα μοντέλα.

Οι Kholodilin και Yao το 2005 ανέπτυξαν ένα δυναμικό μοντέλο δύο παραγόντων (bi-factor) με αλλαγές Markov (Markov-switching). Το

μοντέλο χρησιμοποιεί δεδομένα για οκτώ δείκτες της Αμερικανικής οικονομίας από το 1967 έως το 2004. Στη συνέχεια τα αποτελέσματα συγκρίθηκαν με τα αποτελέσματα από το Εθνικό Γραφείο Οικονομικής Έρευνας (National Bureau of Economic Research). Το μοντέλο προβλέπει καλύτερα τις αλλαγές τάσης της Αμερικανικής οικονομίας. Συγκεκριμένα, στα δεδομένα στα οποία έχει εκπαιδευτεί, το μοντέλο τους παρουσιάζει καλύτερα αποτελέσματα, ενώ στα δεδομένα στα οποία δεν έχει εκπαιδευτεί, τα αποτελέσματα είναι παρεμφερή με αυτά του Εθνικού Γραφείου Οικονομικής Έρευνας.

Το 2002, ο Andersson ανέπτυξε ένα μη-παραμετρικό μοντέλο που χρησιμοποιεί μόνο περιορισμούς μονοτονίας για την ανίχνευση των σημείων αλλαγής τάσης. Η αξιολόγηση του μοντέλου γίνεται με μηνιαία δεδομένα για τον Σουηδικό βιομηχανικό δείκτη παραγωγής από το 1976 έως το 1993. Το αποτέλεσμα δείχνει ότι το μοντέλο ανιχνεύει τα σημεία αλλαγής τάσης με αρκετή επιτυχία.

Το 2005, ο Andersson, απέδειξε ότι η πιθανότητα να εντοπίσει ένα μοντέλο ανίχνευσης τα σημεία αλλαγής τάσης αυξάνει όταν αυξάνεται η άνοδος της αναμενόμενης τιμής μετά την αλλαγή.

Στην παρούσα έρευνα χρησιμοποιείται το σύστημα ANFIS και ένα νευρωνικό δίκτυο βασισμένο σε γενετικούς αλγόριθμους για να προβλέψουν τις τιμές 4 πολύτιμων μετάλλων. Τόσο το νευρωνικό δίκτυο όσο και το μοντέλο ANFIS δίνουν αρκετά ικανοποιητικά αποτελέσματα.

2.1 Takagi Sugeno Type fuzzy rule based system

Είναι ένα ασαφές μοντέλο που βασίζεται σε κανόνες, όπου ο προηγούμενος αποτελείται από γλωσσικές μεταβλητές και ο επόμενος θα αντιπροσωπεύεται από μια συνάρτηση των μεταβλητών εισόδου.

2.2 The fuzzy subtractive clustering

Όταν ο αριθμός των συστάδων είναι άγνωστος για ένα δεδομένο σύνολο από δεδομένα, (Stephen L. Chiu et al) η αφαιρετική ομαδοποίηση, είναι ένας γρήγορος, με ένα πέρασμα αλγόριθμος για τον υπολογισμό του αριθμού των συστάδων και μπορεί να χρησιμοποιηθεί για να προετοιμαστεί η επαναληπτική βελτιστοποίηση με βάση την ομαδοποίηση και το μοντέλο ταυτοποίησης (όπως ANFIS).

2.3 ANFIS Structure

Τα νευρωνικά δίκτυα και η ασαφής λογική είναι δύο συμπληρωματικές τεχνικές για τη δημιουργία ευφυών συστημάτων. Ενώ τα νευρωνικά δίκτυα είναι μια υπολογιστική δομή που εκτελούν καλά όταν ασχολούνται με τα ανεπεξέργαστα δεδομένα, η ασαφής λογική ασχολείται με τη λογική χρήση γλωσσικών πληροφοριών που παρέχονται από την ανάλυση του συστήματος. Τα ασαφή συστήματα δεν μπορούν να μάθουν και να προσαρμοστούν στο νέο κλίμα. Από την άλλη πλευρά, τα νευρωνικά δίκτυα μπορούν να μάθουν, αλλά είναι δυσδιάκριτα στο χρήστη. Το νευρο - ασαφές σύστημα είναι ένα νευρωνικό δίκτυο που λειτουργεί σαν ένα ασαφές μοντέλο. Μπορεί να εκπαιδευτεί για να αναπτύξει τους ασαφείς κανόνες και να καθορίσει συναρτήσεις συμμετοχής για την είσοδο και έξοδο των μεταβλητών στο σύστημα. Το ANFIS είναι ένας τύπος προσαρμοστικών δικτύων που λειτουργεί σαν ασαφές σύστημα.

2.4 Differential Evolution

Η Διαφορική Εξέλιξη "αλγόριθμος" είναι μια μέθοδος που βελτιστοποιεί ένα πρόβλημα για την ελαχιστοποίηση στην αντικειμενική συνάρτηση το οποίο θα μπορέσει να υλοποιήσει τους στόχους του προβλήματος χωρίς οποιαδήποτε αλλαγή των περιορισμών. Ο αλγόριθμος αυτός είναι σαν τους γενετικούς αλγόριθμους. Η κύρια διαφορά στην εξεύρεση καλύτερων λύσεων είναι ότι οι γενετικοί αλγόριθμοι εξαρτώνται από τη διασταύρωση ενώ η Διαφορική Εξέλιξη βασίζεται στη λειτουργία της μετάλλαξης. Αυτή η διαδικασία βασίζεται στις διαφορές μεταξύ των ζευγών στον πληθυσμό όταν επιλεγούν τυχαία.

ΚΕΦΑΛΑΙΟ 3: Θεωρητική προσέγγιση των μοντέλων

3.1 Προσδιορισμός συστήματος

Το πρόβλημα προσδιορισμού ενός μαθηματικού μοντέλου για ένα άγνωστο σύστημα (το οποίο αναφέρεται επίσης και ως σύστημα στόχος) παρατηρώντας τα ζευγάρια δεδομένων εισόδου-εξόδου του αναφέρεται γενικά ως αναγνώριση συστήματος. Οι στόχοι του προσδιορισμού συστήματος είναι πολλαπλοί:

1. Για να προβλέψει μια συμπεριφορά του συστήματος, όπως στην πρόβλεψη χρονοσειρών και στην πρόβλεψη καιρού.
2. Για να εξηγήσει τις αλληλεπιδράσεις και τις σχέσεις μεταξύ των εισόδων και των εξόδων ενός συστήματος. Για παράδειγμα, ένα μαθηματικό μοντέλο μπορεί να χρησιμοποιηθεί για να εξετάσει εάν η ζήτηση ποικίλλει πράγματι ανάλογα με την προσφορά σε ένα οικονομικό σύστημα.
3. Για να σχεδιάσει έναν ελεγκτή βασισμένο στο μοντέλο ενός συστήματος, όπως στον έλεγχο αεροσκαφών και σκαφών. Επίσης για να γίνει η προσομοίωση των υπολογιστών του συστήματος υπό έλεγχο, χρειάζεται ένα μοντέλο του συστήματος.

Ο προσδιορισμός συστημάτων περιλαμβάνει γενικά δύο βήματα:

Προσδιορισμός δομής: Σε αυτό το βήμα, πρέπει να εφαρμοστεί η *a priori* γνώση για το σύστημα στόχο για να καθοριστεί μια κατηγορία μοντέλων μέσα στην οποία πρόκειται να διεξαχθεί η αναζήτηση του καταλληλότερου μοντέλου. Συνήθως αυτή η κατηγορία μοντέλων δείχνεται από μια παραμετρική συνάρτηση $y=f(u;\theta)$, όπου y είναι η έξοδος του μοντέλου, u είναι το διάνυσμα εισόδου και θ είναι το διάνυσμα παραμέτρου. Ο προσδιορισμός της συνάρτησης f είναι εξαρτώμενο πρόβλημα και η συνάρτηση βασίζεται στην εμπειρία και τη διαίσθηση του σχεδιαστή και τους νόμους της φύσης που κυβερνούν το σύστημα στόχο.

Προσδιορισμός παραμέτρου: Στο δεύτερο βήμα, η δομή του μοντέλου είναι γνωστή και το μόνο που χρειάζεται να γίνει είναι να εφαρμοστούν

τεχνικές βελτιστοποίησης για να καθοριστεί το διάνυσμα παραμέτρου $\theta = \hat{\theta}$ τέτοιο ώστε το προκύπτον μοντέλο $\hat{y}=f(u; \hat{\theta})$ να μπορεί να περιγράψει το σύστημα κατάλληλα.

Εάν δεν υπάρχει οποιαδήποτε a priori γνώση για το σύστημα στόχο, τότε ο προσδιορισμός δομής γίνεται ένα δύσκολο πρόβλημα και πρέπει να επιλεγεί η δομή με δοκιμή και σφάλμα. Ευτυχώς, υπάρχει αρκετή γνώση για τις δομές των περισσότερων συστημάτων εφαρμοσμένης μηχανικής και βιομηχανικών διαδικασιών και έτσι είναι συνήθως δυνατό να παραχθεί μια συγκεκριμένη κατηγορία μοντέλων συγκεκριμένα, μια παραμετρική συνάρτηση που μπορεί καλύτερα να περιγράψει το σύστημα στόχο. Συνεπώς, το πρόβλημα προσδιορισμού συστημάτων περιορίζεται συνήθως σε αυτό του προσδιορισμού παραμέτρου.

Γενικά, ο προσδιορισμός συστημάτων δεν είναι μια διαδικασία ενός περάσματος αλλά πρέπει να γίνει προσδιορισμός δομής και παραμέτρου επανειλημμένα έως ότου να βρεθεί ένα ικανοποιητικό πρότυπο ως εξής:

- Καθορίζεται και παραμετροποιείται μια κατηγορία μαθηματικών μοντέλων που αντιπροσωπεύουν το σύστημα που προσδιορίζεται.
- Εκτελείται ο προσδιορισμός παραμέτρου για να επιλεγθούν οι παράμετροι που ταιριάζουν καλύτερα στο σύνολο δεδομένων εκπαίδευσης.
- Πραγματοποιούνται δοκιμές επικύρωσης για να φανεί εάν το μοντέλο που προσδιορίζεται ανταποκρίνεται σωστά σε ένα άγνωστο σύνολο δεδομένων. (Αυτό το σύνολο στοιχείων είναι χωρισμένο από το σύνολο δεδομένων εκπαίδευσης και αναφέρεται ως **σύνολο δεδομένων δοκιμής, επικύρωσης ή ελέγχου.**)
- Ολοκληρώνεται η διαδικασία μόλις τα αποτελέσματα της δοκιμής επικύρωσης είναι ικανοποιητικά. Διαφορετικά, μια άλλη κατηγορία μοντέλων επιλέγεται και επαναλαμβάνονται τα βήματα από το δεύτερο μέχρι το τέταρτο.

Το ANFIS χρησιμοποιεί έναν υβριδικό αλγόριθμο εκμάθησης για να προσδιορίσει τις παραμέτρους των ασαφών συστημάτων τύπου Sugeno. Εφαρμόζει έναν συνδυασμό της μεθόδου ελαχίστων τετραγώνων (least-squares) και της μεθόδου οπισθόδρομης βαθμωτής ελαχιστοποίησης (backpropagation gradient descent) για την εκπαίδευση παραμέτρων των συναρτήσεων συμμετοχής τύπου FIS ώστε να μιμηθεί ένα δοσμένο σύνολο δεδομένων εκπαίδευσης. Στη συνέχεια, θα περιγραφούν η μέθοδος ελαχίστων τετραγώνων και η μέθοδος οπισθόδρομης βαθμωτής ελαχιστοποίησης και έπειτα θα παρουσιαστεί ο τρόπος με τον οποίο το ANFIS συνδυάζει αυτές τις δύο μεθόδους προκειμένου να βρεθούν γρηγορότερα και καλύτερα αποτελέσματα.

3.2 Εκτιμητής ελαχίστων τετραγώνων (Least-squares estimator)

Στο γενικό πρόβλημα ελαχίστων τετραγώνων, η έξοδος του γραμμικού μοντέλου y δίνεται από τη γραμμικά παραμετρικοποιημένη έκφραση

$$y = \theta_1 f_1(\mathbf{u}) + \theta_2 f_2(\mathbf{u}) + \dots + \theta_n f_n(\mathbf{u}), \quad (2.2.1)$$

όπου $\mathbf{u}=[u_1, \dots, u_p]^T$ είναι το διάνυσμα εισόδου του μοντέλου, f_1, \dots, f_n είναι γνωστές συναρτήσεις του \mathbf{u} , και $\theta_1, \dots, \theta_n$ είναι άγνωστες παράμετροι που θα υπολογιστούν. Στη στατιστική, το να βάλει κανείς δεδομένα χρησιμοποιώντας ένα γραμμικό μοντέλο αναφέρεται σαν γραμμική παλινδρόμηση. Κατά συνέπεια η εξίσωση (2.2.1) καλείται επίσης συνάρτηση παλινδρόμησης, και τα θ_i ονομάζονται συντελεστές παλινδρόμησης.

Για να προσδιοριστούν οι άγνωστες παράμετροι θ_i , συνήθως πρέπει να εκτελεστούν πειράματα για να βρεθεί ένα σύνολο δεδομένων εκπαίδευσης που αποτελείται από τα ζευγάρια δεδομένων $\{(\mathbf{u}_i, y_i), i=1, \dots, m\}$ τα οποία αντιπροσωπεύουν τα επιθυμητά ζευγάρια εισόδου-εξόδου του συστήματος στόχου που θα μοντελοποιηθεί. Η αντικατάσταση κάθε ζευγαριού στοιχείων στην εξίσωση (2.2.1) παράγει ένα σύνολο γραμμικών εξισώσεων m :

$$\begin{aligned} f_1(\mathbf{u}_1)\theta_1 + f_2(\mathbf{u}_1)\theta_2 + \dots + f_n(\mathbf{u}_1)\theta_n &= y_1 \\ f_1(\mathbf{u}_2)\theta_1 + f_2(\mathbf{u}_2)\theta_2 + \dots + f_n(\mathbf{u}_2)\theta_n &= y_2 \\ &\vdots \\ f_1(\mathbf{u}_m)\theta_1 + f_2(\mathbf{u}_m)\theta_2 + \dots + f_n(\mathbf{u}_m)\theta_n &= y_m \end{aligned} \quad (2.2.2)$$

Χρησιμοποιώντας τη μορφή πινάκων, μπορούν να ξαναγραφτούν οι προηγούμενες εξισώσεις σε μια συνοπτική μορφή:

$$\mathbf{A}\boldsymbol{\theta} = \mathbf{y} \quad (2.2.3)$$

Όπου \mathbf{A} είναι ένας $m \times n$ πίνακας (μερικές φορές λέγεται πίνακας σχεδιασμού):

$$\mathbf{A} = \begin{bmatrix} f_1(\mathbf{u}_1) & \dots & f_n(\mathbf{u}_1) \\ \vdots & \vdots & \vdots \\ f_1(\mathbf{u}_m) & \dots & f_n(\mathbf{u}_m) \end{bmatrix},$$

θ είναι ένα $n \times 1$ διάνυσμα άγνωστων παραμέτρων:

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_n \end{bmatrix},$$

Και y είναι ένα $m \times 1$ διάνυσμα εξόδου:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}.$$

Η i -οστή σειρά του ενωμένου πίνακα δεδομένων $[\mathbf{A} : \mathbf{y}]$, που δηλώνεται με $[\mathbf{a}_i^T : y_i]$, σχετίζεται με το i -οστό ζευγάρι δεδομένων $(\mathbf{u}_i ; y_i)$ μέσω της

$$\mathbf{a}_i^T = [f_1(\mathbf{u}_i), \dots, f_n(\mathbf{u}_i)].$$

Εφόσον το μεγαλύτερο μέρος των υπολογισμών βασίζεται στους πίνακες \mathbf{A} και \mathbf{y} , μερικές φορές γίνεται αναφορά στο $(\mathbf{a}_i^T ; y_i)$ σαν το i -οστό ζευγάρι δεδομένων του συνόλου δεδομένων εκπαίδευσης.

Για να προσδιοριστεί μεμονωμένα το άγνωστο διάνυσμα $\boldsymbol{\theta}$, είναι απαραίτητο να ισχύει $m \geq n$. Αν ο \mathbf{A} είναι τετραγωνικός ($m=n$) και

αντιστρέψιμος, τότε η εξίσωση (2.2.3) μπορεί να λυθεί ως προς τον άγνωστο \mathbf{x} και γίνεται:

$$\boldsymbol{\theta} = \mathbf{A}^{-1} \mathbf{y} \quad . \quad (2.2.4)$$

Εντούτοις, ο m είναι συνήθως μεγαλύτερος από τον n , που σημαίνει ότι υπάρχουν περισσότερα ζευγάρια στοιχείων από τις παραμέτρους. Σε αυτήν την περίπτωση, μια ακριβής λύση που να ικανοποιεί όλες τις m εξισώσεις δεν είναι πάντα δυνατή, δεδομένου ότι τα στοιχεία μπορεί να μολυνθούν από θόρυβο, ή το μοντέλο μπορεί να μην είναι κατάλληλο για την περιγραφή του συστήματος στόχου. Κατά συνέπεια η εξίσωση (2.2.3) πρέπει να τροποποιηθεί με την ενσωμάτωση ενός διανύσματος λάθους \mathbf{e} για να αποτελέσει το τυχαίο λάθος θορύβου ή το τυχαίο λάθος διαμόρφωσης ως εξής:

$$\mathbf{A}\boldsymbol{\theta} + \mathbf{e} = \mathbf{y} \quad . \quad (2.2.5)$$

Τώρα, αντί της εύρεσης της ακριβούς λύσης στην εξίσωση (2.2.3), πρέπει να βρεθεί το $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ που ελαχιστοποιεί το άθροισμα του τετραγωνικού σφάλματος και ορίζεται ως

$$E(\boldsymbol{\theta}) = \sum_{i=1}^m (y_i - \mathbf{a}_i^T \boldsymbol{\theta})^2 = \mathbf{e}^T \mathbf{e} = (\mathbf{y} - \mathbf{A}\boldsymbol{\theta})^T (\mathbf{y} - \mathbf{A}\boldsymbol{\theta}) \quad (2.2.6)$$

όπου $\mathbf{e} = \mathbf{y} - \mathbf{A}\boldsymbol{\theta}$ είναι το διάνυσμα λάθους που παράγεται από μια συγκεκριμένη επιλογή του $\boldsymbol{\theta}$. Πρέπει να σημειωθεί ότι το $E(\boldsymbol{\theta})$ είναι σε τετραγωνική μορφή και έχει ένα μοναδικό ελάχιστο $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ □ Το ακόλουθο θεώρημα δηλώνει έναν απαραίτητο όρο που ικανοποιείται από τον εκτιμητή ελαχίστων τετραγώνων $\hat{\boldsymbol{\theta}}$.

Θεώρημα 2.2.1: Εκτιμητής ελαχίστων τετραγώνων

Το τετραγωνικό σφάλμα στην εξίσωση (2.2.6) ελαχιστοποιείται όταν $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, οποίος καλείται εκτιμητής ελαχίστων τετραγώνων (LSE για συντομία) και ο οποίος ικανοποιεί την κανονική εξίσωση

$$\mathbf{A}^T \mathbf{A} \hat{\boldsymbol{\theta}} = \mathbf{A}^T \mathbf{y} \quad (2.2.7)$$

Αν ο $\mathbf{A}^T \mathbf{A}$ είναι αντιστρέψιμος, ο $\hat{\boldsymbol{\theta}}$ είναι μοναδικός και δίνεται από την

$$\hat{\boldsymbol{\theta}} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y} \quad (2.2.8)$$

3.3 Οπισθοδρόμηση για πρωσοτροφοδοτούμενα δίκτυα

Η ενότητα αυτή παρουσιάζει ένα βασικό κανόνα εκμάθησης για προσαρμόσιμα δίκτυα, που είναι στην ουσία η πιο απλή μέθοδος βαθμωτής ελαχιστοποίησης. Το κεντρικό μέρος αυτού του κανόνα εκμάθησης αφορά στο πώς να επιλεγεί επαναληπτικά ένα διάνυσμα κλίσης στο οποίο κάθε στοιχείο ορίζεται ως την παράγωγο ενός μέτρου σφάλματος ως προς μια παράμετρο. Αυτό γίνεται με τη βοήθεια του κανόνα αλυσίδας, ενός βασικού τύπου για το διαφορισμό σύνθετων συναρτήσεων ο οποίος αναλύεται σε κάθε εγχειρίδιο μαθηματικών. Η διαδικασία εύρεσης ενός διανύσματος κλίσης σε ένα δίκτυο αναφέρεται γενικά σαν οπισθοδρόμηση (backpropagation) επειδή το διάνυσμα κλίσης υπολογίζεται σε κατεύθυνση αντίθετη από τη ροή της εξόδου κάθε κόμβου. Μόλις επιλεγεί η κλίση, διάφορες τεχνικές βελτιστοποίησης και παλινδρόμησης βασισμένες στις παραγώγους είναι διαθέσιμες για την ενημέρωση των παραμέτρων. Ειδικότερα, εάν χρησιμοποιούμε το διάνυσμα κλίσης σε μια απλή μέθοδο βαθμωτής ελαχιστοποίησης, το προκύπτον παράδειγμα εκμάθησης αναφέρεται συχνά ως κανόνας οπισθόδρομης εκμάθησης. Αυτός ο κανόνας εκμάθησης θα αναλυθεί στη συνέχεια αυτής της ενότητας.

Ας υποθεθεί ότι δεδομένο πρωσοτροφοδοτούμενο προσαρμόσιμο δίκτυο στην αναπαράστασή του σε επίπεδα έχει L επίπεδα και το επίπεδο l (όπου $l=0,1,\dots,L;l=0$ αντιπροσωπεύει το επίπεδο εισαγωγής) έχει $N(l)$ κόμβους. Τότε η έξοδος και η συνάρτηση του κόμβου i [$i=1,\dots,N(l)$] στο επίπεδο l μπορούν να αναπαρασταθούν σαν $x_{l,i}$, και $f_{l,i}$, αντίστοιχα. Χωρίς απώλεια γενικότητας, ας υποθεθεί ότι δεν υπάρχει καμία σύνδεση μεταξύ μη συνεχόμενων στρωμάτων. Δεδομένου ότι η έξοδος ενός κόμβου εξαρτάται από τα εισερχόμενα σήματα και το σύνολο παραμέτρων του κόμβου, προκύπτει η ακόλουθη γενική έκφραση για

τη συνάρτηση των κόμβων $f_{l,i}$:

$$x_{l,i} = f_{l,i}(x_{l-1,1}, \dots, x_{l-1,N(l-1)}, a, \beta, \gamma, \dots), \quad (2.3.1)$$

όπου a, β, γ κλπ. είναι οι παράμετροι αυτού του κόμβου.

Υποθέτοντας ότι το δοσμένο σύνολο δεδομένων εκπαίδευσης έχει P καταχωρήσεις, μπορεί να οριστεί ένα μέτρο σφάλματος για την p -οστή ($1 \leq p \leq P$) καταχώρηση των δεδομένων εκπαίδευσης σαν το άθροισμα των τετραγωνικών σφαλμάτων:

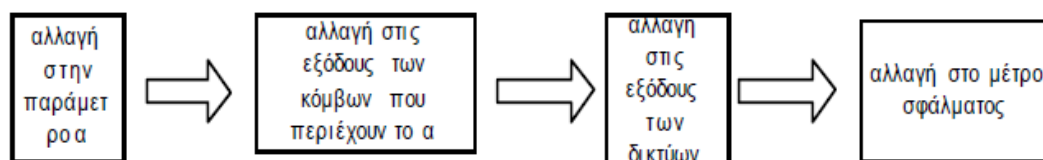
$$E_p = \sum_{k=1}^{N(l)} (d_k - x_{L,k})^2, \quad (2.3.2)$$

όπου d_k είναι το k -οστό συστατικό του p -οστού επιθυμητού διανύσματος εξόδου και $x_{L,k}$, είναι το k -οστό συστατικό του πραγματικού διανύσματος εξόδου που παράγεται με την παρουσίαση του p -οστού διανύσματος εισόδου στο δίκτυο. (Για σημειογραφική απλότητα, παραλείπεται ο δείκτης p και από το d_k και από το $x_{L,k}$). Προφανώς, όταν το E_p είναι ίσο με το μηδέν, το δίκτυο είναι ικανό να αναπαράγει ακριβώς το επιθυμητό διάνυσμα εξόδου στο p -οστό ζευγάρι δεδομένων εκπαίδευσης. Κατά συνέπεια ο στόχος εδώ είναι να ελαχιστοποιηθεί ένα συνολικό μέτρο σφάλματος, που ορίζεται ως

$$E = \sum_{p=1}^P E_p.$$

Υπενθυμίζεται ότι ο ορισμός του E_p στην εξίσωση (2.3.2) δεν είναι καθολικός αλλά είναι δυνατός και άλλος ορισμός για το E_p για συγκεκριμένες καταστάσεις ή εφαρμογές. Επομένως, θα αποφευχθεί μια ρητή έκφραση για το μέτρο σφάλματος E_p για να δοθεί έμφαση στη γενικότητα. Επιπλέον, ας υποτεθεί ότι το E_p εξαρτάται μόνο από τους κόμβους εξόδου.

Για να χρησιμοποιηθεί η βαθμωτή ελαχιστοποίηση ώστε να ελαχιστοποιηθεί το μέτρο σφάλματος, πρέπει πρώτα να βρεθεί το διάνυσμα κλίσης. Πριν υπολογιστεί το διάνυσμα κλίσης, πρέπει να παρατηρηθούν οι ακόλουθες αιτιώδεις σχέσεις:



όπου τα βέλη \Rightarrow δείχνουν τις αιτιώδεις σχέσεις. Με άλλα λόγια, μια μικρή αλλαγή σε μια παράμετρο a θα επηρεάσει την έξοδο του κόμβου που περιέχει το a . Αυτό με τη σειρά του θα επηρεάσει την έξοδο του τελευταίου επιπέδου και συνεπώς το μέτρο σφάλματος. Επομένως, η βασική αρχή στον υπολογισμό του διανύσματος κλίσης είναι να περαστούν μια σειρά από πληροφορίες παραγώγων ξεκινώντας από το επίπεδο εξόδου και πηγαίνοντας ανάποδα από επίπεδο σε επίπεδο έως ότου καταλήξει η διαδικασία στο επίπεδο εισόδου.

Το σήμα σφάλματος $\epsilon_{l,i}$ ορίζεται σαν τη παράγωγο του μέτρου σφάλματος E_p ως προς την έξοδο του κόμβου i στο επίπεδο l , λαμβάνοντας υπόψη και τις άμεσες και τις έμμεσες πορείες. Σε σύμβολα,

$$\epsilon_{l,i} = \frac{\partial^+ E_p}{\partial x_{l,i}} \quad (2.3.3)$$

Η έκφραση αυτή ονομάστηκε διατεταγμένη παράγωγος από τον Werbos. Η διαφορά μεταξύ της διατεταγμένης παραγώγου και της συνηθισμένης μερικής παραγώγου βρίσκεται στον τρόπο με τον οποίο η συνάρτηση διαφορίζεται. Για την έξοδο ενός εσωτερικού κόμβου $x_{l,i}$

όπου ($l \neq L$),

η μερική παράγωγος $\frac{\partial E_p}{\partial x_{l,i}}$ είναι ίση με το μηδέν, εφόσον το E_p δεν εξαρτάται από το $x_{l,i}$ έμμεσα, αφού μια αλλαγή στο $x_{l,i}$ θα διαδοθεί μέσω έμμεσων πορειών στο επίπεδο εξόδου και συνεπώς θα προκαλέσει μια αντίστοιχη αλλαγή στην τιμή του E_p . Επομένως, το $\epsilon_{l,i}$ μπορεί να υπολογιστεί ως μια αναλογία αυτών των δύο αλλαγών όταν αυτές γίνονται απειροελάχιστες.

Το σήμα σφάλματος για τον i -οστό κόμβο εξόδου (στο επίπεδο L) μπορεί να υπολογιστεί απευθείας:

$$\epsilon_{L,i} = \frac{\partial^+ E_p}{\partial x_{L,i}} = \frac{\partial E_p}{\partial x_{L,i}} \quad (2.3.4)$$

Αυτό ισούται με $\epsilon_{L,i} = -2(d_i - x_{L,i})$ αν το E_p ορίζεται όπως στην εξίσωση (2.3.2). Για τον εσωτερικό κόμβο στην i -οστή θέση του επιπέδου l , το σήμα σφάλματος μπορεί να βρεθεί από τον κανόνα αλυσίδας:

$$\epsilon_{l,i} = \underbrace{\frac{\partial^+ E_p}{\partial x_{l,i}}}_{\text{error signal at layer } l} = \sum_{m=1}^{N(l+1)} \underbrace{\frac{\partial^+ E_p}{\partial x_{l+1,m}}}_{\text{error signal at layer } l+1} \frac{\partial f_{l+1,m}}{\partial x_{l,i}} = \sum_{m=1}^{N(l+1)} \epsilon_{l+1,m} \frac{\partial f_{l+1,m}}{\partial x_{l,i}}, \quad (2.3.5)$$

όπου $0 \leq l \leq L-1$. Δηλαδή το σήμα σφάλματος ενός εσωτερικού κόμβου στο στρώμα l μπορεί να εκφραστεί ως γραμμικός συνδυασμός των κόμβων

στο στρώμα $l+1$. Επομένως, για οποιαδήποτε l και i [και $1 \leq i \leq N(l)$],

$$\epsilon_{l,i} = \frac{\partial^+ E_p}{\partial x_{l,i}}$$

μπορούν να βρεθούν τα

εφαρμόζοντας πρώτα την εξίσωση (2.3.4) μία φορά για να υπολογιστούν τα σήματα σφάλματος στο επίπεδο εξόδου, και έπειτα εφαρμόζοντας την εξίσωση (2.3.5) επαναληπτικά έως ότου καταλήξει η διαδικασία στο επιθυμητό επίπεδο l . Η διαδικασία αυτή καλείται οπισθοδρόμηση δεδομένου ότι τα σήματα σφάλματος λαμβάνονται διαδοχικά από το επίπεδο εξόδου προς το επίπεδο εισόδου.

Το διάνυσμα κλίσης ορίζεται ως την παράγωγο του μέτρου σφάλματος ως προς κάθε παράμετρο, έτσι πρέπει να εφαρμοστεί ο κανόνας αλυσίδας ξανά για να βρεθεί το διάνυσμα κλίσης. Εάν a είναι μια παράμετρος του i -οστού κόμβου στο επίπεδο l , τότε ισχύει:

$$\frac{\partial^+ E_p}{\partial a} = \frac{\partial^+ E_p}{\partial x_{l,i}} \frac{\partial f_{l,i}}{\partial a} = \epsilon_{l,i} \frac{\partial f_{l,i}}{\partial a}. \quad (2.3.6)$$

Ας σημειωθεί ότι αν επιτραπεί στην παράμετρο a να μοιράζεται μεταξύ διαφορετικών κόμβων, τότε η εξίσωση (2.3.6) πρέπει να αλλαχθεί σε μια πιο γενική μορφή:

$$\frac{\partial^+ E_p}{\partial a} = \sum_{x^* \in S} \frac{\partial^+ E_p}{\partial x^*} \frac{\partial f^*}{\partial a}, \quad (2.3.7)$$

όπου S είναι το σύνολο των κόμβων που περιέχουν το a σαν παράμετρο ενώ x^* και f^* είναι η έξοδος και η συνάρτηση, αντίστοιχα, ενός γενικού κόμβου στο S .

Η παράγωγος του γενικού μέτρου σφάλματος E ως προς το a είναι

$$\frac{\partial^+ E}{\partial a} = \sum_{p=1}^P \frac{\partial^+ E_p}{\partial a} \quad . \quad (2.3.8)$$

Συνεπώς, για την απλούστερη βαθμωτή ελαχιστοποίηση χωρίς ελαχιστοποίηση γραμμών, ο τύπος για τη γενική παράμετρο a είναι

$$\Delta a = -\eta \frac{\partial^+ E_p}{\partial a}, \quad (2.3.9)$$

όπου το η είναι ο ρυθμός εκμάθησης, το οποίο μπορεί να εκφραστεί περαιτέρω ως

$$\eta = \frac{\kappa}{\sqrt{\sum_{\alpha} \left(\frac{\partial E}{\partial \alpha} \right)^2}}, \quad (2.3.10)$$

όπου το κ είναι το μέγεθος βήματος, το μήκος δηλαδή κάθε μετάβασης κατά μήκος της κατεύθυνσης κλίσης στο διάστημα παραμέτρου. Συνήθως το μέγεθος βήματος μπορεί να αλλαχθεί για να μεταβληθεί η ταχύτητα της σύγκλισης.

Όταν ένα πρωσοτροφοδοτούμενο δίκτυο n -κόμβων αναπαριστάται στην τοπολογική του διάταξη, μπορεί να υπολογιστεί το μέτρο σφάλματος E_p σαν την έξοδο ενός επιπλέον κόμβου με ένδειξη $n+1$, του

οποίου η συνάρτηση κόμβου f_{n+1} μπορεί να οριστεί από τις εξόδους κάθε κόμβου με μικρότερη ένδειξη. (Επομένως, το E_p μπορεί να εξαρτάται άμεσα από οποιουσδήποτε κόμβους.) Εφαρμόζοντας πάλι τον κανόνα αλυσίδας, ισχύει ο ακόλουθος συνοπτικός τύπος για τον υπολογισμό του σήματος σφάλματος

$$\epsilon_i = \frac{\partial E_p}{\partial x_i} :$$

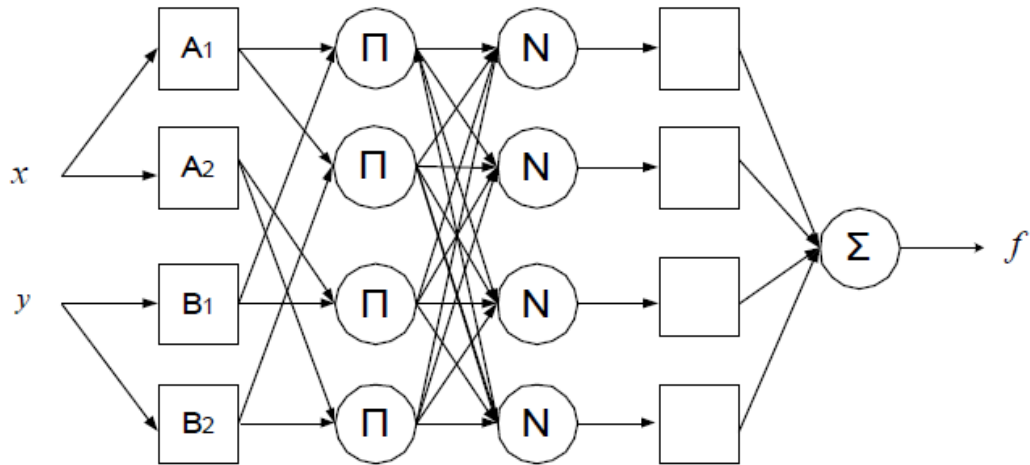
$$\frac{\partial^+ E_p}{\partial x_i} = \frac{\partial f_{n+1}}{\partial x_i} + \sum_{i < j \leq n} \epsilon_j \frac{\partial^+ E_p}{\partial x_j} \frac{\partial f_j}{\partial x_i} , \quad (2.3.11)$$

ή

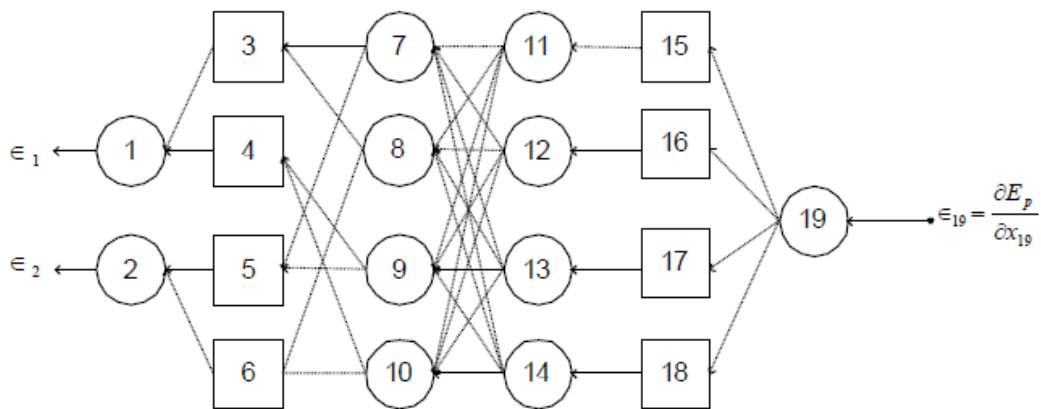
$$\epsilon_i = \frac{\partial f_{n+1}}{\partial x_i} + \sum_{i < j \leq n} \epsilon_j \frac{\partial f_j}{\partial x_i} , \quad (2.3.12)$$

όπου ο πρώτος όρος δείχνει μια άμεση επίδραση του x_i στο E_p μέσω της άμεσης διαδρομής από τον κόμβο i στον κόμβο $n+1$ και κάθε όρος παραγώγου στο άθροισμα δείχνει την έμμεση επίδραση του x_i στο E_p . Μόλις βρεθεί το σήμα σφάλματος για κάθε κόμβο, τότε το διάνυσμα κλίσης για τις παραμέτρους παράγεται όπως πριν.

Ένας άλλος συστηματικός τρόπος να υπολογιστούν τα σήματα σφάλματος είναι μέσω της αναπαράστασης του δικτύου διάδοσης σφάλματος (ή του μοντέλου ευαισθησίας), το οποίο λαμβάνεται από το αρχικό προσαρμόσιμο δίκτυο με την αντιστροφή των συνδέσεων και την παροχή των σημάτων σφάλματος στο επίπεδο εξόδου ως είσοδοι στο νέο δίκτυο. Το ακόλουθο σχήμα δείχνει το μοντέλο ANFIS και το δίκτυο διάδοσης σφάλματος του:



Σχήμα 2.1(a): Το προσαρμόσιμο δίκτυο



Σχήμα 2.1(b): Το δίκτυο διάδοσης σφάλματος

Τώρα θα υπολογιστούν τα σήματα σφάλματος στους εσωτερικούς κόμβους. Χρησιμοποιούνται τα f_i και x_i για το συμβολισμό της συνάρτησης και της εξόδου του κόμβου i . Η έξοδος του κόμβου i είναι το σήμα σφάλματος αυτού του κόμβου στο πραγματικό προσαρμόσιμο δίκτυο. Σε σύμβολα, αν επιλεγθεί το τετραγωνικό μέτρο σφάλματος για E_p , τότε ισχύει το ακόλουθο:

$$\epsilon_{19} = -2(d_{19} - x_{19}),$$

Αυτό είναι επειδή ο κόμβος 19 είναι μόνο ένας κόμβος προσωρινής αποθήκευσης στο δίκτυο διάδοσης σφάλματος. Για τους κόμβους 15, 16, 17 και 18 ισχύουν:

$$\epsilon_{18} = \frac{\partial^+ E_p}{\partial x_{18}} = \frac{\partial^+ E_p}{\partial x_{19}} \frac{\partial f_{19}}{\partial x_{18}} = \epsilon_{19} \frac{\partial f_{19}}{\partial x_{18}},$$

$$\epsilon_{17} = \frac{\partial^+ E_p}{\partial x_{17}} = \frac{\partial^+ E_p}{\partial x_{19}} \frac{\partial f_{19}}{\partial x_{17}} = \epsilon_{19} \frac{\partial f_{19}}{\partial x_{17}},$$

$$\epsilon_{16} = \frac{\partial^+ E_p}{\partial x_{16}} = \frac{\partial^+ E_p}{\partial x_{19}} \frac{\partial f_{19}}{\partial x_{16}} = \epsilon_{19} \frac{\partial f_{19}}{\partial x_{16}},$$

$$\epsilon_{15} = \frac{\partial^+ E_p}{\partial x_{15}} = \frac{\partial^+ E_p}{\partial x_{19}} \frac{\partial f_{19}}{\partial x_{15}} = \epsilon_{19} \frac{\partial f_{19}}{\partial x_{15}}.$$

Αυτό είναι επειδή όλοι αυτοί οι κόμβοι εξαρτώνται από τον κόμβο 19. Αυτό ισχύει και για τους κόμβους 11, 12, 13 και 14 με τη διαφορά ότι κάθε ένας από αυτούς εξαρτάται από διαφορετικό κόμβο:

$$\epsilon_{14} = \frac{\partial^+ E_p}{\partial x_{14}} = \frac{\partial^+ E_p}{\partial x_{18}} \frac{\partial f_{18}}{\partial x_{14}} = \epsilon_{18} \frac{\partial f_{18}}{\partial x_{14}},$$

$$\epsilon_{13} = \frac{\partial^+ E_p}{\partial x_{13}} = \frac{\partial^+ E_p}{\partial x_{17}} \frac{\partial f_{17}}{\partial x_{13}} = \epsilon_{17} \frac{\partial f_{17}}{\partial x_{13}},$$

$$\epsilon_{12} = \frac{\partial^+ E_p}{\partial x_{12}} = \frac{\partial^+ E_p}{\partial x_{16}} \frac{\partial f_{16}}{\partial x_{12}} = \epsilon_{16} \frac{\partial f_{16}}{\partial x_{12}},$$

$$\epsilon_{11} = \frac{\partial^+ E_p}{\partial x_{11}} = \frac{\partial^+ E_p}{\partial x_{15}} \frac{\partial f_{15}}{\partial x_{11}} = \epsilon_{15} \frac{\partial f_{15}}{\partial x_{11}}.$$

Αντιθέτως, οι κόμβοι 7, 8, 9 και 10 εξαρτώνται από τέσσερις διαφορετικούς κόμβους όπως φαίνεται στο σχήμα 2.1. Έτσι, ισχύει:

$$\epsilon_{10} = \frac{\partial^+ E_p}{\partial x_{10}} = \epsilon_{14} \frac{\partial f_{14}}{\partial x_{10}} + \epsilon_{13} \frac{\partial f_{13}}{\partial x_{10}} + \epsilon_{12} \frac{\partial f_{12}}{\partial x_{10}} + \epsilon_{11} \frac{\partial f_{11}}{\partial x_{10}},$$

$$\epsilon_9 = \frac{\partial^+ E_p}{\partial x_9} = \epsilon_{14} \frac{\partial f_{14}}{\partial x_9} + \epsilon_{13} \frac{\partial f_{13}}{\partial x_9} + \epsilon_{12} \frac{\partial f_{12}}{\partial x_9} + \epsilon_{11} \frac{\partial f_{11}}{\partial x_9},$$

$$\epsilon_8 = \frac{\partial^+ E_p}{\partial x_8} = \epsilon_{14} \frac{\partial f_{14}}{\partial x_8} + \epsilon_{13} \frac{\partial f_{13}}{\partial x_8} + \epsilon_{12} \frac{\partial f_{12}}{\partial x_8} + \epsilon_{11} \frac{\partial f_{11}}{\partial x_8},$$

$$\epsilon_7 = \frac{\partial^+ E_p}{\partial x_7} = \epsilon_{14} \frac{\partial f_{14}}{\partial x_7} + \epsilon_{13} \frac{\partial f_{13}}{\partial x_7} + \epsilon_{12} \frac{\partial f_{12}}{\partial x_7} + \epsilon_{11} \frac{\partial f_{11}}{\partial x_7}.$$

Οι κόμβοι 3, 4, 5 και 6 εξαρτώνται από δύο διαφορετικούς κόμβους και σαν αποτέλεσμα προκύπτουν οι ακόλουθες σχέσεις:

$$\epsilon_6 = \frac{\partial^+ E_p}{\partial x_6} = \frac{\partial^+ E_p}{\partial x_{10}} \frac{\partial f_{10}}{\partial x_6} + \frac{\partial^+ E_p}{\partial x_8} \frac{\partial f_8}{\partial x_6} = \epsilon_{10} \frac{\partial f_{10}}{\partial x_6} + \epsilon_8 \frac{\partial f_8}{\partial x_6},$$

$$\epsilon_5 = \frac{\partial^+ E_p}{\partial x_5} = \frac{\partial^+ E_p}{\partial x_9} \frac{\partial f_9}{\partial x_5} + \frac{\partial^+ E_p}{\partial x_7} \frac{\partial f_7}{\partial x_5} = \epsilon_9 \frac{\partial f_9}{\partial x_5} + \epsilon_7 \frac{\partial f_7}{\partial x_5},$$

$$\epsilon_4 = \frac{\partial^+ E_p}{\partial x_4} = \frac{\partial^+ E_p}{\partial x_{10}} \frac{\partial f_{10}}{\partial x_4} + \frac{\partial^+ E_p}{\partial x_9} \frac{\partial f_9}{\partial x_4} = \epsilon_{10} \frac{\partial f_{10}}{\partial x_4} + \epsilon_9 \frac{\partial f_9}{\partial x_4},$$

$$\epsilon_3 = \frac{\partial^+ E_p}{\partial x_3} = \frac{\partial^+ E_p}{\partial x_8} \frac{\partial f_8}{\partial x_3} + \frac{\partial^+ E_p}{\partial x_7} \frac{\partial f_7}{\partial x_3} = \epsilon_8 \frac{\partial f_8}{\partial x_3} + \epsilon_7 \frac{\partial f_7}{\partial x_3}.$$

Τέλος, οι κόμβοι 1 και 2 εξαρτώνται από δύο κόμβους και έτσι ισχύει:

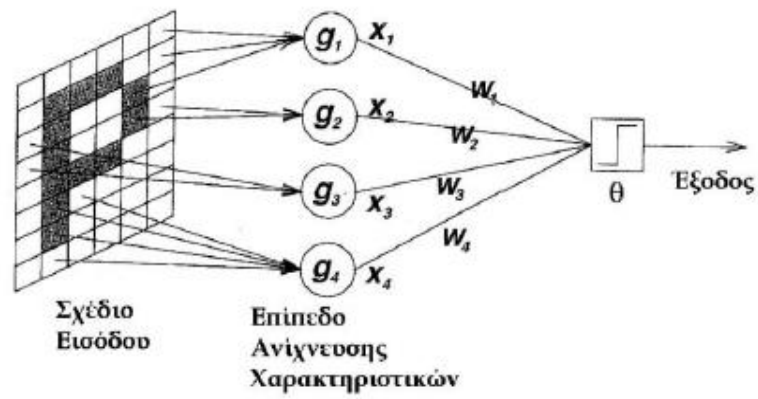
$$\epsilon_2 = \frac{\partial^+ E_p}{\partial x_2} = \frac{\partial^+ E_p}{\partial x_6} \frac{\partial f_6}{\partial x_2} + \frac{\partial^+ E_p}{\partial x_5} \frac{\partial f_5}{\partial x_2} = \epsilon_6 \frac{\partial f_6}{\partial x_2} + \epsilon_5 \frac{\partial f_5}{\partial x_2},$$

$$\epsilon_1 = \frac{\partial^+ E_p}{\partial x_1} = \frac{\partial^+ E_p}{\partial x_4} \frac{\partial f_4}{\partial x_1} + \frac{\partial^+ E_p}{\partial x_3} \frac{\partial f_3}{\partial x_1} = \epsilon_4 \frac{\partial f_4}{\partial x_1} + \epsilon_3 \frac{\partial f_3}{\partial x_1}$$

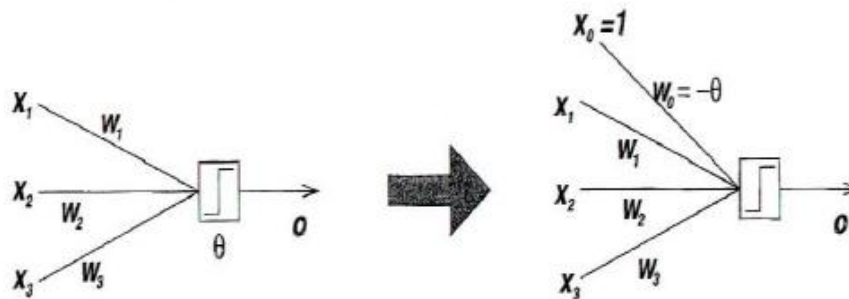
3.4 Νευρωνικά δίκτυα

3.4.1 Perceptrons. Αρχιτεκτονική και κανόνες εκμάθησης

Το δίκτυο perceptron αντιπροσωπεύει μία από τις πρώτες προσπάθειες για να κατασκευαστούν έξυπνα και αυτοεκπαιδευόμενα συστήματα χρησιμοποιώντας απλά συστατικά. Προήλθε από το βιολογικό μοντέλο εγκεφαλικού νευρώνα που παρουσιάστηκε από τους McCulloch και Pitts το 1943. Αργότερα ο Rosenblatt σχεδίασε το δίκτυο perceptron με τη σκέψη να εξηγήσει και να μοντελοποιήσει τις δυνατότητες αναγνώρισης σχημάτων των βιολογικών συστημάτων όρασης. Παρόλο που ο στόχος είναι φιλόδοξος, το παράδειγμα είναι απλό. Το σχήμα 2.2 είναι μια τυπική διάταξη δικτύου perceptron για εφαρμογές αναγνώρισης σχημάτων, στα οποία τα οπτικά σχήματα αναπαριστώνται σαν πίνακες με στοιχεία μεταξύ 0 και 1. Το πρώτο επίπεδο του δικτύου perceptron λειτουργεί σαν ένα σετ «ανιχνευτών χαρακτηριστικών» που δέχονται τα σήματα εισόδου για να ανιχνεύσουν συγκεκριμένα χαρακτηριστικά. Το δεύτερο επίπεδο (έξοδος) παίρνει τις εξόδους των «ανιχνευτών χαρακτηριστικών» του πρώτου επιπέδου και ταξινομεί το δεδομένο σχέδιο εισόδου. Η εκπαίδευση ξεκινά με την πραγματοποίηση ρυθμίσεων στα σχετικά βάρη συνδέσεων (π.χ. βάρη w_i) και μία τιμή κατωφλίου (threshold) θ . Για ένα πρόβλημα δύο τάξεων (για παράδειγμα, να καθοριστεί αν το σχέδιο του σχήματος 2.2 είναι P ή όχι), το επίπεδο εξόδου έχει συνήθως μόνο ένα κόμβο. Για ένα πρόβλημα n τάξεων με το n να είναι μεγαλύτερο ή ίσο με 3, το επίπεδο εξόδου έχει συνήθως n κόμβους, καθένας από τους οποίους αντιστοιχεί σε μία τάξη και ο κόμβος εξόδου με την μεγαλύτερη τιμή δείχνει σε ποια τάξη ανήκει το διάνυσμα εισόδου.



Σχήμα 2.2: Το δίκτυο perceptron



Σχήμα 2.3: Εισαγωγή τεχνητού βάρους σύνδεσης

Κάθε συνάρτηση g_i στο επίπεδο 1 είναι μια σταθερή συνάρτηση που πρέπει να επιλεγεί εκ των προτέρων. Η συνάρτηση χαρτογραφεί όλο ή ένα μέρος του σχεδίου (pattern) εισόδου σε μία δυαδική τιμή $x_i \in \{-1, 1\}$ ή σε μία διπολική τιμή $x_i \in \{0, 1\}$. Ο όρος x_i αναφέρεται ως ενεργή τιμή αν η τιμή είναι 1, ανενεργή αν η τιμή είναι 0 και απαγορευτική αν η τιμή είναι -1. Η μονάδα εξόδου είναι ένα γραμμικό στοιχείο κατωφλιού με μία τιμή κατωφλιού θ :

$$\begin{aligned} o &= f\left(\sum_{i=1}^n w_i x_i - \theta\right), \\ &= f\left(\sum_{i=1}^n w_i x_i + w_0\right), w_0 \equiv -\theta, \\ &= f\left(\sum_{i=0}^n w_i x_i\right), x_0 \equiv 1. \end{aligned} \tag{2.5.1}$$

όπου w_i είναι ένα μεταβαλλόμενο βάρος που σχετίζεται με ένα εισερχόμενο σήμα x_i και w_0 ($= -\theta$) είναι ο όρος θορύβου (bias). Η εξίσωση (2.5.1) δείχνει ότι το κατώφλι μπορεί να θεωρηθεί σαν ένα βάρος σύνδεσης μεταξύ της μονάδας εξόδου και ενός «τεχνητού» εισερχόμενου σήματος x_0 που είναι πάντα 1, όπως φαίνεται στο σχήμα 2.3. Στην εξίσωση (2.5.1), η είναι η $f(\cdot)$ συνάρτηση ενεργοποίησης του δικτύου perceptron και είναι τυπικά ή σιγμοειδής (sigmoid) συνάρτηση $\text{sgn}(x)$ ή βηματική συνάρτηση $\text{step}(x)$:

$$\begin{aligned} \text{sgn}(x) &= \begin{cases} 1 & \text{αν } x > 0, \\ -1 & \text{αλλιώς,} \end{cases} \\ \text{step}(x) &= \begin{cases} 1 & \text{αν } x > 0, \\ 0 & \text{αλλιώς.} \end{cases} \end{aligned}$$

Πρέπει να σημειωθεί ότι ο «ανιχνευτής χαρακτηριστικών» g_i μπορεί να είναι οποιαδήποτε συνάρτηση του σχεδίου εισόδου αλλά η διαδικασία εκμάθησης μόνο ρυθμίζει τα βάρη σύνδεσης στη μονάδα εξόδου (στο τελευταίο επίπεδο). Επειδή μπορούν να μεταβληθούν μόνο τα βάρη που οδηγούν στο τελευταίο επίπεδο, το δίκτυο perceptron του σχήματος 2.2 θεωρείται σαν δίκτυο perceptron 39 ενός επιπέδου (single-layer perceptron). Ξεκινώντας με ένα σετ τυχαίων βαρών σύνδεσης, ο βασικός αλγόριθμος εκμάθησης για ένα δίκτυο perceptron ενός επιπέδου επαναλαμβάνει τα ακόλουθα βήματα έως ότου τα βάρη συγκλίνουν:

Βήμα 1: Επιλογή ενός διανύσματος εισόδου \mathbf{x} από τα δεδομένα εισόδου.

Βήμα 2: Αν το δίκτυο perceptron δίνει λάθος απόκριση, γίνεται μετατροπή όλων των βαρών σύνδεσης w_i σύμφωνα με τη σχέση

$$\Delta w_i = \eta t_i x_i$$

όπου t_i είναι η έξοδος στόχος και η είναι ο ρυθμός εκμάθησης.

Ο προηγούμενος κανόνας εκμάθησης μπορεί να εφαρμοστεί επίσης για την ενημέρωση του κατωφλιού θ ($= -w_0$) σύμφωνα με τη σχέση 2.5.1. Η τιμή για τον ρυθμό εκμάθησης η μπορεί να είναι σταθερή σε όλη τη διάρκεια της εκπαίδευσης ή να είναι μια μεταβλητή ποσότητα ανάλογη με το σφάλμα. Ένα η που είναι ανάλογο με το σφάλμα συνήθως οδηγεί σε γρηγορότερη σύγκλιση αλλά μπορεί να προκαλέσει ασταθή εκμάθηση.

Ο προηγούμενος αλγόριθμος εκμάθησης βασίζεται περίπου στη βαθμωτή ελαχιστοποίηση. Ο Rosenblatt απέδειξε ότι υπάρχει μία μέθοδος για τη ρύθμιση των βαρών που διασφαλίζει ότι θα συγκλίνει για να παράγει την απαιτούμενη έξοδο αν και μόνον αν υπάρχει τέτοιο σετ βαρών. Αυτό ονομάζεται θεώρημα σύγκλισης δικτύου perceptron. Επιπλέον, τα δίκτυα perceptron μπορούν να ομαδοποιηθούν σε διαφορετικές οικογένειες βάση των συναρτήσεων g_i .

Στις αρχές της δεκαετίας του 1960, τα δίκτυα perceptron δημιούργησαν ενδιαφέρον και αισιοδοξία για την κατασκευή αυτοεκπαιδευόμενων έξυπνων συστημάτων. Όμως ο αρχικός ενθουσιασμός ελαττώθηκε με τη δημοσίευση των Minsky και Papert *Perceptrons* το 1969, στην οποία ανέλυσαν τα δίκτυα perceptron διεξοδικά και συμπέραναν ότι τα δίκτυα perceptron ενός επιπέδου μπορούν μόνο να χρησιμοποιηθούν για πάρα πολύ απλά προβλήματα.

3.4.2 Οπισθοδρομούμενα πολυεπίπεδα δίκτυα perceptron

Το δίκτυο perceptron ενός επιπέδου είναι ένα βασικό συστατικό των νευρωνικών δικτύων και παρέχει τις βάσεις για την κατανόηση τους και τις περισσότερες εφαρμογές τους. Όμως επειδή οι συναρτήσεις ενεργοποίησης δεν είναι διαφορήσιμες, οι στρατηγικές εκπαίδευσης των πρώτων πολυεπίπεδων δικτύων perceptron με σιγμοειδείς (signum) ή βηματικές συναρτήσεις ενεργοποίησης δεν είναι εμφανείς εκτός και αν χρησιμοποιούνται συνεχείς συναρτήσεις ενεργοποίησης.

Ένα οπισθοδρομούμενο πολυεπίπεδο δίκτυο perceptron είναι ένα προσαρμόσιμο δίκτυο του οποίου οι κόμβοι (ή νευρώνες) εκτελούν την ίδια συνάρτηση στα εισερχόμενα σήματα. Αυτή η συνάρτηση του κόμβου είναι συνήθως μία σύνθεση του σταθμισμένου αθροίσματος και της διαφορήσιμης μη γραμμικής συνάρτησης ενεργοποίησης που είναι γνωστή και ως συνάρτηση μεταφοράς. Οι τρεις πιο συχνά χρησιμοποιούμενες συναρτήσεις ενεργοποίησης είναι οι:

$$\text{Logistic function : } f(x) = \frac{1}{1 + e^{-x}}$$

$$\text{Hyperbolic tangent function : } f(x) = \tanh(x/2) = \frac{1 - e^{-x}}{1 + e^{-x}}$$

$$\text{Identity function : } f(x) = x$$

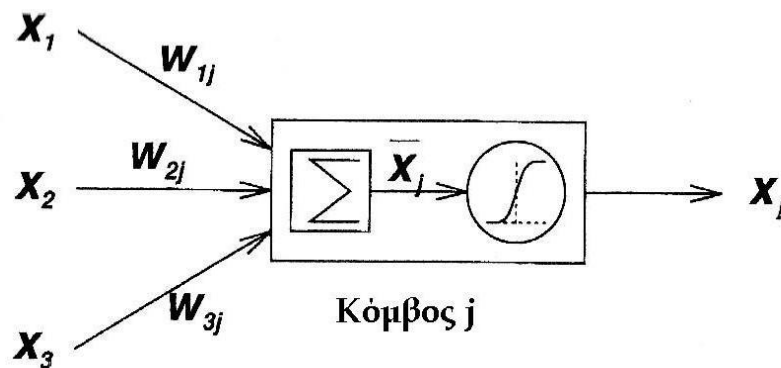
Η υπερβολική εφαπτομένη (hyperbolic tangent) και η λογιστική προσεγγίζουν τη σιγμοειδή (sigmoid) και τη βηματική συνάρτηση, αντίστοιχα. Μερικές φορές αυτές οι δύο συναρτήσεις ενεργοποίησης αναφέρονται ως squashing συναρτήσεις επειδή οι είσοδοι για αυτές τις συναρτήσεις «στριμώνονται» στο πεδίο $[0,1]$ ή $[-1,1]$. Λέγονται επίσης σιγμοειδείς συναρτήσεις επειδή οι καμπύλες τους τύπου S παρουσιάζουν ομαλότητα και ασυμπτωτικές ιδιότητες. Και οι δύο συναρτήσεις χρησιμοποιούνται συνήθως σε προβλήματα παλινδρόμησης και ταξινόμησης.

Για να μπορέσει ένα νευρωνικό δίκτυο να προσεγγίσει μία συνεχή συνάρτηση που δεν περιορίζεται στο διάστημα $[0,1]$ ή στο $[-1,1]$, επιλέγουμε τη συνάρτηση του κόμβου για το επίπεδο εξόδου να είναι ένα σταθμισμένο άθροισμα χωρίς squashing συναρτήσεις. Αυτό είναι ανάλογο με την περίπτωση στην οποία η συνάρτηση ενεργοποίησης είναι μια συνάρτηση ταυτοποίησης (identity) και οι κόμβοι εξόδου αυτού του τύπου ονομάζονται συχνά γραμμικοί κόμβοι.

Τα οπισθοδρομούμενα πολυεπίπεδα δίκτυα perceptron είναι με διαφορά οι πιο συχνά χρησιμοποιούμενες δομές νευρωνικών δικτύων με πολλές εφαρμογές, όπως η αναγνώριση σχεδίων, η επεξεργασία σημάτων, η συμπίεση δεδομένων (data compression) και ο αυτόματος έλεγχος.

3.4.3 Κανόνας εκμάθησης οπισθοδρόμησης

Το οπισθοδρομούμενο πολυεπίπεδο δίκτυο perceptron θα θεωρηθεί για ευκολία ότι χρησιμοποιεί τη λογιστική συνάρτηση σαν συνάρτηση ενεργοποίησης.

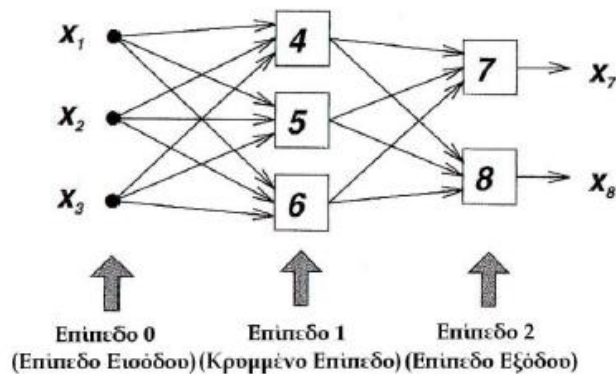


Σχήμα 2.4: Κόμβος j ενός οπισθοδρομούμενου πολυεπίπεδου δικτύου perceptron

Η είσοδος δικτύου (net input) \bar{x} ενός κόμβου ορίζεται σαν το σταθμισμένο άθροισμα των εισερχόμενων σημάτων συν ένα όρο θορύβου (bias). Για παράδειγμα, η είσοδος δικτύου και η έξοδος του κόμβου j στο σχήμα 2.4 είναι

$$\begin{aligned}\bar{x}_j &= \sum_i w_{ij} x_i + w_j, \\ x_j &= f(\bar{x}_j) = \frac{1}{1 + \exp(-\bar{x}_j)},\end{aligned}\tag{2.5.2}$$

όπου x_i είναι η έξοδος του κόμβου i που βρίσκεται σε οποιοδήποτε από τα προηγούμενα επίπεδα, w_{ij} είναι το βάρος της σύνδεσης του i με το j και w_j είναι ο θόρυβος (bias) του κόμβου j . Επειδή τα βάρη w_{ij} είναι στην πραγματικότητα ενδιάμεσες παράμετροι που σχετίζονται με κάθε κόμβο j , η αλλαγή των βαρών ενός κόμβου θα αλλάξει τη συμπεριφορά του κόμβου και αυτό με τη σειρά του θα αλλάξει τη συμπεριφορά όλου του οπισθοδρομούμενου πολυεπίπεδου δικτύου perceptron. Το σχήμα 2.5 δείχνει ένα οπισθοδρομούμενο πολυεπίπεδο δίκτυο perceptron 2 επιπέδων με 3 εισόδους στο επίπεδο εισόδου, 3 νευρώνες στο κρυμμένο επίπεδο και 2 νευρώνες εξόδου στο επίπεδο εξόδου.



Σχήμα 2.5: Οπισθοδρομούμενο πολυεπίπεδο δίκτυο perceptron 3-3-2

Η οπισθοδρόμηση ή γενικευμένος κανόνας δέλτα εξηγείται στη συνέχεια. Αρχικά, ένα τετραγωνικό μέτρο σφάλματος για το p -οστό ζεύγος εισόδου εξόδου ορίζεται ως:

$$E_p = \sum_k (d_k - x_k)^2, \quad (2.5.3)$$

όπου d_k είναι η επιθυμητή έξοδος για τον κόμβο k , και x_k είναι η πραγματική έξοδος για τον κόμβο k . Για να βρεθεί το βαθμωτό διάνυσμα (gradient vector), ορίζεται ένας όρος σφάλματος $\bar{\epsilon}_i$ για τον κόμβο i ως

$$\bar{\epsilon}_i = \frac{\partial^+ E_p}{\partial \bar{x}_i}. \quad (2.5.4)$$

Από τον κανόνα αλυσίδας, η σχέση για το $\bar{\epsilon}_i$ μπορεί να γραφεί ως

$$\bar{\epsilon}_i = \begin{cases} -2(d_i - x_i) \frac{\partial x_i}{\partial \bar{x}_i} = -2(d_i - x_i)x_i(1 - x_i) & \text{αν ο κόμβος } i \text{ είναι κόμβος εξόδου,} \\ \frac{\partial x_i}{\partial \bar{x}_i} = \sum_{j:i < j} \frac{\partial^+ E_p}{\partial \bar{x}_j} \frac{\partial \bar{x}_j}{\partial x_i} = x_i(1 - x_i) \sum_{j:i < j} \bar{\epsilon}_j w_{ij} & \text{αλλιώς,} \end{cases} \quad (2.5.5)$$

όπου w_{ij} είναι το βάρος σύνδεσης από τον κόμβο i στον κόμβο j και w_{ij} είναι μηδέν αν δεν υπάρχει απευθείας σύνδεση. Τότε η ενημέρωση του βάρους w_{ki} για απευθείας (on-line) εκπαίδευση (από σχέδιο σε σχέδιο) είναι

$$\Delta w_{ki} = -\eta \frac{\partial^+ E_p}{\partial w_{ki}} = -\eta \frac{\partial^+ E_p}{\partial \bar{x}_i} \frac{\partial \bar{x}_i}{\partial w_{ki}} = -\eta \bar{\epsilon}_i x_k, \quad (2.5.6)$$

όπου η είναι ο ρυθμός εκμάθησης που επηρεάζει την ταχύτητα σύγκλισης και τη σταθερότητα των βαρών κατά τη διάρκεια της εκμάθησης.

Για εκτός γραμμής (off-line) εκμάθηση, το βάρος σύνδεσης w_{ki} ανανεώνεται μόνο μετά την παρουσίαση ολόκληρου του σετ δεδομένων, ή μόνο μετά από μία εποχή:

$$\Delta w_{ki} = -\eta \frac{\partial^+ E}{\partial w_{ki}} = -\eta \sum_p \frac{\partial^+ E_p}{\partial w_{ki}}, \quad (2.5.7)$$

ή σε μορφή διανύσματος,

$$\Delta \mathbf{w} = -\eta \frac{\partial^+ E}{\partial \mathbf{w}} = -\nabla_{\mathbf{w}} E, \quad (2.5.8)$$

όπου $E = \sum_p E_p$.

Αυτό αντιστοιχεί σε ένα τρόπο χρήσης της πραγματικής βαθμωτής κατεύθυνσης που βασίζεται σε ολόκληρο το σετ δεδομένων.

3.5 Υβριδικός αλγόριθμος εκμάθησης: Συνδυασμός βαθμωτής ελαχιστοποίησης και εκτιμητή ελαχίστων τετραγώνων

Αν και μπορεί να εφαρμοστεί η οπισθοδρόμηση ή η βαθμωτή ελαχιστοποίηση για να προσδιοριστούν οι παράμετροι σε ένα προσαρμόσιμο δίκτυο, αυτή η απλή μέθοδος βελτιστοποίησης παίρνει συνήθως πολύ χρόνο προτού να συγκλίνει. Μπορεί να παρατηρηθεί, όμως, ότι η έξοδος ενός προσαρμόσιμου δικτύου είναι γραμμική σε μερικές από τις παραμέτρους του δικτύου. Έτσι μπορούν να προσδιοριστούν αυτές οι γραμμικές παράμετροι με τη γραμμική μέθοδο ελαχίστων τετραγώνων που περιγράφηκε παραπάνω.

Αυτή η προσέγγιση οδηγεί σε ένα υβριδικό κανόνα εκμάθησης που συνδυάζει τη βαθμωτή ελαχιστοποίηση (SD) και τον εκτιμητή ελαχίστων τετραγώνων (LSE) για γρήγορο προσδιορισμό των παραμέτρων. Στη συνέχεια θα παρουσιαστεί πως αυτός λειτουργεί.

Το προσαρμόσιμο δίκτυο έχει μία έξοδο που αναπαριστάται ως

$$o = F(i, S), \quad (2.4.1)$$

όπου i είναι το διάνυσμα των μεταβλητών εισόδου, S είναι το σύνολο των παραμέτρων, και F είναι η συνολική συνάρτηση που εφαρμόζεται από το προσαρμόσιμο δίκτυο. Αν υπάρχει μια συνάρτηση H τέτοια ώστε η σύνθετη συνάρτηση $H \circ F$ είναι γραμμική σε μερικά από τα στοιχεία του S , τότε αυτά τα στοιχεία μπορούν να προσδιοριστούν από τη μέθοδο ελαχίστων τετραγώνων. Πιο συγκεκριμένα, αν το σύνολο παραμέτρων S μπορεί να χωριστεί σε δύο σύνολα

$$S = S_1 \oplus S_2, \quad (2.4.2)$$

(όπου \oplus αντιπροσωπεύει άμεσο άθροισμα) τέτοια ώστε η $H \circ F$ είναι γραμμική στα στοιχεία του S_2 , τότε εφαρμόζοντας την H στην εξίσωση (2.4.1), προκύπτει ότι

$$H(o) = H \circ F(Bi, S), \quad (2.4.3)$$

που είναι γραμμική στα στοιχεία του S_2 . Τώρα με δεδομένα στοιχεία του S_I , μπορούν να χρησιμοποιηθούν P δεδομένα εκπαίδευσης στην εξίσωση (2.4.2) και να προκύψει μια εξίσωση πινάκων:

$$\mathbf{A}\boldsymbol{\theta} = \mathbf{y} \quad (2.4.4)$$

όπου $\boldsymbol{\theta}$ είναι ένα άγνωστο διάνυσμα του οποίου τα στοιχεία είναι παράμετροι στο S_2 . Αυτό είναι ένα κλασσικό γραμμικό πρόβλημα ελαχίστων τετραγώνων, και η καλύτερη λύση για το $\boldsymbol{\theta}$, που ελαχιστοποιεί το, $\|\mathbf{A}\boldsymbol{\theta} - \mathbf{y}\|^2$ είναι ο εκτιμητής ελαχίστων τετραγώνων (LSE) $\boldsymbol{\theta}^*$:

$$\boldsymbol{\theta}^* = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y}, \quad (2.4.5)$$

όπου \mathbf{A}^T είναι ο ανάστροφος του \mathbf{A} και $(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T$ είναι ο ψευδοαντίστροφος του \mathbf{A} αν $\mathbf{A}^T \mathbf{A}$ είναι αντιστρέψιμος.

Τώρα μπορεί να συνδυαστεί η βαθμωτή ελαχιστοποίηση και ο εκτιμητής ελαχίστων τετραγώνων για να ενημερωθούν οι παράμετροι σε ένα προσαρμόσιμο δίκτυο. Για να εφαρμοστεί υβριδική εκμάθηση σε μια παρτίδα επεξεργασίας (batch mode), κάθε εποχή αποτελείται από ένα πέρασμα προς τα μπρος και ένα πέρασμα προς τα πίσω. Στο πέρασμα προς τα μπρος, αφού παρουσιάζεται ένα διάνυσμα εισόδου, υπολογίζονται οι έξοδοι των κόμβων στο δίκτυο από επίπεδο σε επίπεδο μέχρι να βρεθεί μια αντίστοιχη σειρά στους πίνακες \mathbf{A} και \mathbf{y} στην εξίσωση (2.4.4). Αυτή η διαδικασία επαναλαμβάνεται για όλα τα ζευγάρια δεδομένων εκπαίδευσης για να σχηματιστούν οι

ολοκληρωμένοι \mathbf{A} και \mathbf{y} . Έπειτα οι παράμετροι στο S_2 προσδιορίζονται από τον ψευδοαντίστροφο τύπο στην εξίσωση (2.4.5). Αφού προσδιορίζονται οι παράμετροι στο S_2 , μπορεί να υπολογιστεί το μέτρο σφάλματος για κάθε ζευγάρι δεδομένων εκπαίδευσης. Στο πέρασμα προς τα πίσω, τα σήματα σφάλματος [η παράγωγος του μέτρου σφάλματος ως προς την έξοδο κάθε κόμβου, βλέπε εξισώσεις (2.3.4) και (2.3.5)] διαδίδονται από το άκρο της εξόδου ως το άκρο της εισόδου. Το διάνυσμα κλίσης αυξάνεται για κάθε είσοδο δεδομένων εκπαίδευσης. Στο τέλος του περάσματος προς τα πίσω για όλα τα δεδομένα εκπαίδευσης, οι παράμετροι στο S_1 ενημερώνονται από τη βαθμωτή ελαχιστοποίηση στην εξίσωση (2.3.9).

Για δεδομένες σταθερές τιμές των παραμέτρων στο S_1 , οι παράμετροι στο S_2 που βρίσκονται είναι εγγυημένο ότι είναι στο ολικό βέλτιστο σημείο του διαστήματος παραμέτρων του S_2 εξαιτίας της επιλογής του τετραγωνικού μέτρου σφάλματος. Αυτός ο υβριδικός κανόνας εκμάθησης μπορεί όχι μόνο να μειώσει τη διάσταση του διαστήματος εξερεύνησης που εξερευνείται από την πραγματική μέθοδο βαθμωτής ελαχιστοποίησης, αλλά, γενικά, θα μειώσει σημαντικά το χρόνο που απαιτείται για να φθάσει στη σύγκλιση.

Από την αρχιτεκτονική του ANFIS που παρουσιάστηκε προηγουμένως, παρατηρείται ότι όταν οι τιμές των αρχικών παραμέτρων (premise parameters) είναι σταθερές, η συνολική έξοδος μπορεί να εκφραστεί σαν γραμμικός συνδυασμός των επακόλουθων παραμέτρων (consequent parameters). Σε σύμβολα, η έξοδος f μπορεί να ξαναγραφτεί ως:

$$\begin{aligned}
f &= \frac{w_1}{w_1 + w_2 + w_3 + w_4} f_1 + \frac{w_2}{w_1 + w_2 + w_3 + w_4} f_2 + \frac{w_3}{w_1 + w_2 + w_3 + w_4} f_3 + \frac{w_4}{w_1 + w_2 + w_3 + w_4} f_4 \\
&= \bar{w}_1(p_1x + q_1y + r_1) + \bar{w}_2(p_2x + q_2y + r_2) + \bar{w}_3(p_3x + q_3y + r_3) + \bar{w}_4(p_4x + q_4y + r_4) \\
&= (\bar{w}_1x)p_1 + (\bar{w}_1y)q_1 + (\bar{w}_1)r_1 + (\bar{w}_2x)p_2 + (\bar{w}_2y)q_2 + (\bar{w}_2)r_2 + (\bar{w}_3x)p_3 + (\bar{w}_3y)q_3 + (\bar{w}_3)r_3 \\
&\quad + (\bar{w}_4x)p_4 + (\bar{w}_4y)q_4 + (\bar{w}_4)r_4
\end{aligned}$$

που είναι γραμμική στις επακόλουθες παραμέτρους (consequent parameters)

$p_1, q_1, r_1, p_2, q_2, r_2, p_3, q_3, r_3, p_4, q_4$ και r_4 . Από την παρατήρηση αυτή προκύπτει

S = σύνολο συνολικών παραμέτρων,

S_1 = σύνολο αρχικών παραμέτρων (μη γραμμικών),

S_2 = σύνολο επακόλουθων παραμέτρων (γραμμικών)

στην εξίσωση (2.4.2) ενώ $H(\cdot)$ και $F(\cdot, \cdot)$ είναι η μοναδιαία συνάρτηση και η συνάρτηση του ασαφούς συστήματος συμπερασμού (fuzzy inference system), αντίστοιχα στην εξίσωση (2.4.3). Επομένως, ο υβριδικός αλγόριθμος εκμάθησης που αναπτύσσεται παραπάνω μπορεί να εφαρμοστεί άμεσα. Πιο συγκεκριμένα, στο πέρασμα προς τα μπρος του υβριδικού αλγορίθμου, οι έξοδοι κόμβων πηγαίνουν προς τα εμπρός μέχρι το επίπεδο 4 και οι επακόλουθες παράμετροι προσδιορίζονται με τη μέθοδο ελαχίστων τετραγώνων. Στο πέρασμα προς τα πίσω, τα σήματα σφάλματος διαδίδονται προς τα πίσω και οι αρχικές παράμετροι ενημερώνονται από την βαθμωτή ελαχιστοποίηση. Ο πίνακας 2.1 συνοψίζει τις δραστηριότητες που πραγματοποιούνται σε κάθε πέρασμα.

	Πέρασμα προς τα μπρος	Πέρασμα προς τα πίσω
Αρχικές παράμετροι	Σταθερές	Βαθμωτή ελαχιστοποίηση
Επακόλουθες παράμετροι	Εκτιμητής ελαχίστων τετραγώνων	Σταθερές
Σήματα	Έξοδοι κόμβων	Σήματα σφάλματος

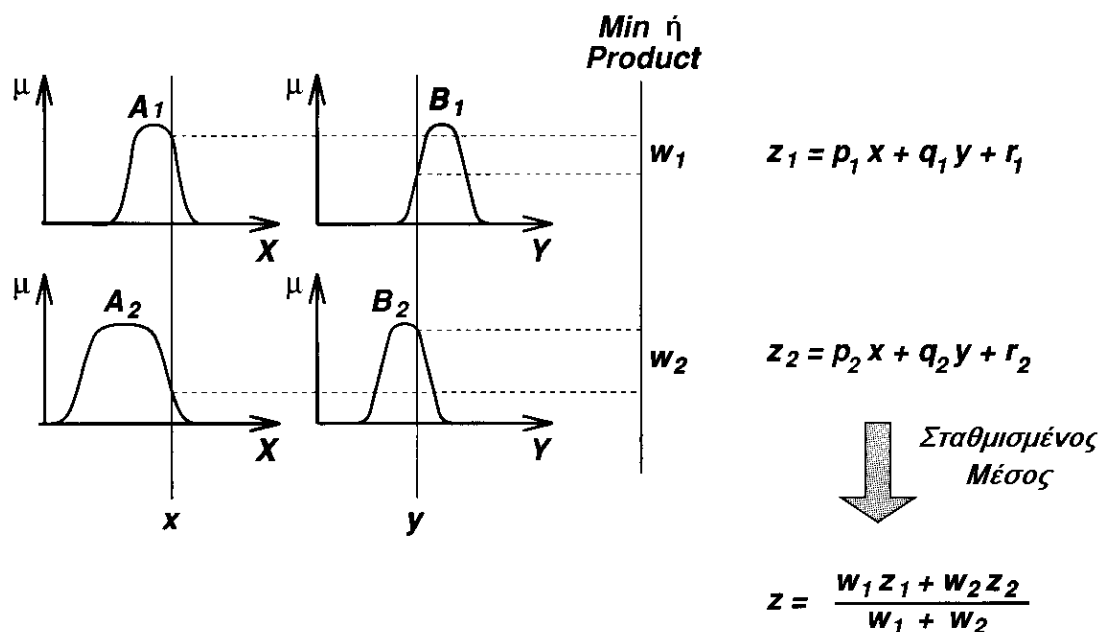
Πίνακας 2.1: Δύο περάσματα στην υβριδική διαδικασία εκμάθησης για το ANFIS

3.6 Θεωρητικό υπόβαθρο του ANFIS.

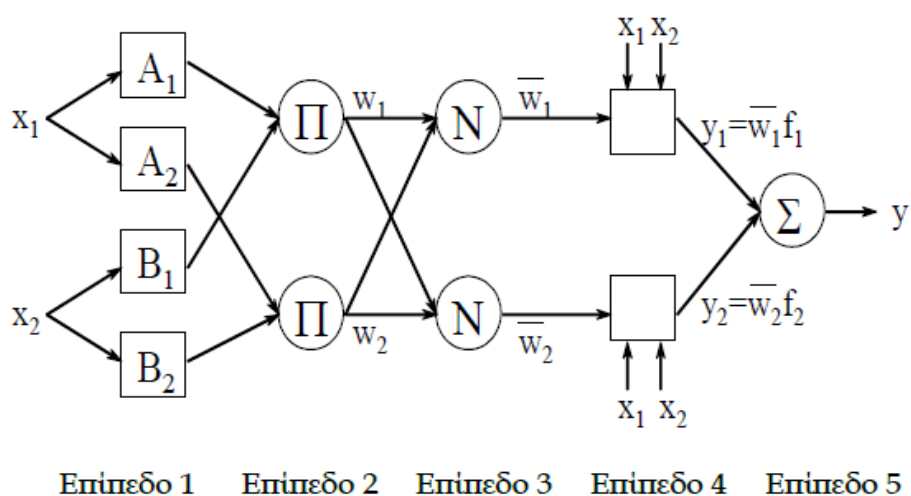
Ένα νευρο-ασαφές σύστημα ορίζεται ως ένας συνδυασμός τεχνητών νευρωνικών δικτύων (ANN) και Ασαφούς Συστήματος Συμπερασμού (Fuzzy Inference System, FIS) κατά τέτοιο τρόπο ώστε ο αλγόριθμος εκμάθησης νευρωνικών δικτύων χρησιμοποιείται για να καθορίσει τις παραμέτρους του FIS. Το Προσαρμοστικό Νευροασαφές σύστημα συμπερασμού (Adaptive Neural Fuzzy Inference System, ANFIS) είναι ένα σύστημα που ανήκει στην νευρο-ασαφή κατηγορία.

Λειτουργικά, δεν υπάρχει σχεδόν κανένας περιορισμός στις συναρτήσεις κόμβων ενός προσαρμόσιμου (adaptive) δικτύου εκτός από τμήματα που μπορούν να διαφοριστούν (piecewise differentiability). Δομικά, ο μόνος περιορισμός της διαμόρφωσης δικτύων είναι ότι πρέπει να είναι τύπου πρόσθιας τροφοδότησης. Λόγω αυτού του μικρού περιορισμού, οι εφαρμογές του προσαρμόσιμου δικτύου είναι άμεσες και απέραντες σε διάφορους τομείς. Σε αυτό το τμήμα, προτείνεται μια κατηγορία προσαρμόσιμων δικτύων, τα οποία είναι λειτουργικά ισοδύναμα με τα ασαφή συστήματα συμπερασμού.

Ο ασαφής μηχανισμός συμπερασμού παρουσιάζεται στο επόμενο σχήμα:



Σχήμα 1.1: Ασαφής συλλογιστική (Jang 1995)



Σχήμα 1.2: Αρχιτεκτονική του ANFIS (Jang 1995 [20])

Για απλούστευση, γίνεται η υπόθεση ότι το υπό εξέταση ασαφές σύστημα συμπερασμού έχει δύο εισόδους x και y και μία έξοδο z καθώς επίσης και ότι η βάση κανόνων περιέχει δύο ασαφείς κανόνες τύπου Takagi και Sugeno:

Κανόνας 1:

$$\text{Αν } x \text{ είναι } A_1 \text{ και } y \text{ είναι } B_1 \text{ τότε } f_1 = p_1 \cdot x + q_1 \cdot y + r_1$$

Κανόνας 2:

$$\text{Αν } x \text{ είναι } A_2 \text{ και } y \text{ είναι } B_2 \text{ τότε } f_2 = p_2 \cdot x + q_2 \cdot y + r_2$$

Η αρχιτεκτονική του ANFIS παρουσιάζεται πιο αναλυτικά στο κεφάλαιο 4.2

Ο συγγραφέας A. Abraham, παρουσιάζει στον Πίνακα 1.1, μία συγκριτική ανάλυση ορισμένων μοντέλων. Για να συγκρίνει τα αποτελέσματα της πρόβλεψης κάθε μοντέλου, χρησιμοποίησε ως δεδομένα την γνωστή χαοτική χρονολογική σειρά Mackey-Glass. Ως μέτρο σύγκρισης χρησιμοποίησε τη ρίζα του μέσου τετραγωνικού σφάλματος (RMSE). Αποδεικνύει ότι το ANFIS είναι καλύτερο μοντέλο καθώς δίνει την χαμηλότερη τιμή στην ρίζα του μέσου τετραγωνικού σφάλματος.

Πίνακας 1.1: Απόδοση των νευρο-ασαφών συστημάτων (Abraham, 2005 [1])

Σύστημα	Επαναλήψεις (epochs)	RMSE
ANFIS	75	0,0017
NEFPROX	216	0,0332
EFuNN	1	0,0140
dmEFuNN	1	0,0042
SONFIN	1	0,0180

Όμως σε υπολογιστικό χρόνο το ANFIS δεν υπερτερεί καθώς απαιτούνται 75 επαναλήψεις για να εκπαιδευτεί. Τα συστήματα που χρησιμοποιούν Sugeno μηχανισμό συμπερασμού είναι περισσότερο ακριβή αλλά απαιτούν περισσότερο υπολογιστικό χρόνο. Ενώ αυτά που χρησιμοποιούν Mamdani ερμηνεύονται περισσότερο και απαιτούν λιγότερο υπολογιστικό χρόνο αλλά έχουν λιγότερη ακρίβεια.

Στο ίδιο συμπέρασμα καταλήγει και ο συγγραφέας Jang (Jang, 1997) ο οποίος συγκρίνει το σύστημα ANFIS με τρία άλλα μοντέλα αποδεικνύοντας ότι το ANFIS δίνει το μικρότερο RMSE. Τα αποτελέσματα παρουσιάζονται στον Πίνακα 1.2.

Πίνακας 1.2: Σύγκριση του ANFIS με άλλα μοντέλα (Jang, 1997 [21])

Μοντέλο	Σφάλμα εκπαίδευσης (training)	Σφάλμα ελέγχου (checking)	Αριθμός παραμέτρων	Μέγεθος δειγματος εκπαίδευσης	Μέγεθος δειγματος ελέγχου
ANFIS	0,043%	1,066%	50	216	125
GMDH	4,7%	5,7%	-	20	20
Fuzzy 1	1,5%	2,1%	22	20	20
Fuzzy 2	0,59%	3.4%	32	20	20

Για τους λόγους αυτούς απεφασίσθη στην παρούσα έρευνα να χρησιμοποιηθεί το ANFIS. Αρκετοί συγγραφείς έχουν χρησιμοποιήσει το ANFIS για επεξεργασία χρονολογικών σειρών [Atsalakis, 2006, a, b, c, d, e, f.] [Atsalakis 2005, a, b, c, d] [Ucenic, 2003, a,b] [Ucenic, 2005, a, b]. [M. Rezaeianzadeh, H. Tabari, A. Arabi Yazdi, S. Isik, L. Kalin (2013)] [M. Mohandes, S. Rehman, S.M. Rahman 2011][F. Deregeh, M. Karimian, and H. Nezamabadi-Pour 2013][Azadeh, S.M. Asadzadeh, M. Saberi, V. Nadimi, A. Tajvidi, M. Sheikalishahi 2011][K. Ramesh, A. P. Kesarkar, J. Bhate, M. Venkat Ratnam, and A. Jayaraman 2014]

ΚΕΦΑΛΑΙΟ 4: Ανάλυση και εφαρμογή των μοντέλων

4.1 Εισαγωγή

Στο κεφάλαιο αυτό θα γίνει μια περιγραφή του μοντέλου ANFIS και θα παρουσιαστεί η δομή καθώς και οι περιορισμοί του. Τέλος, θα γίνει μια περιγραφή του νευρωνικού δικτύου που χρησιμοποιήθηκε.

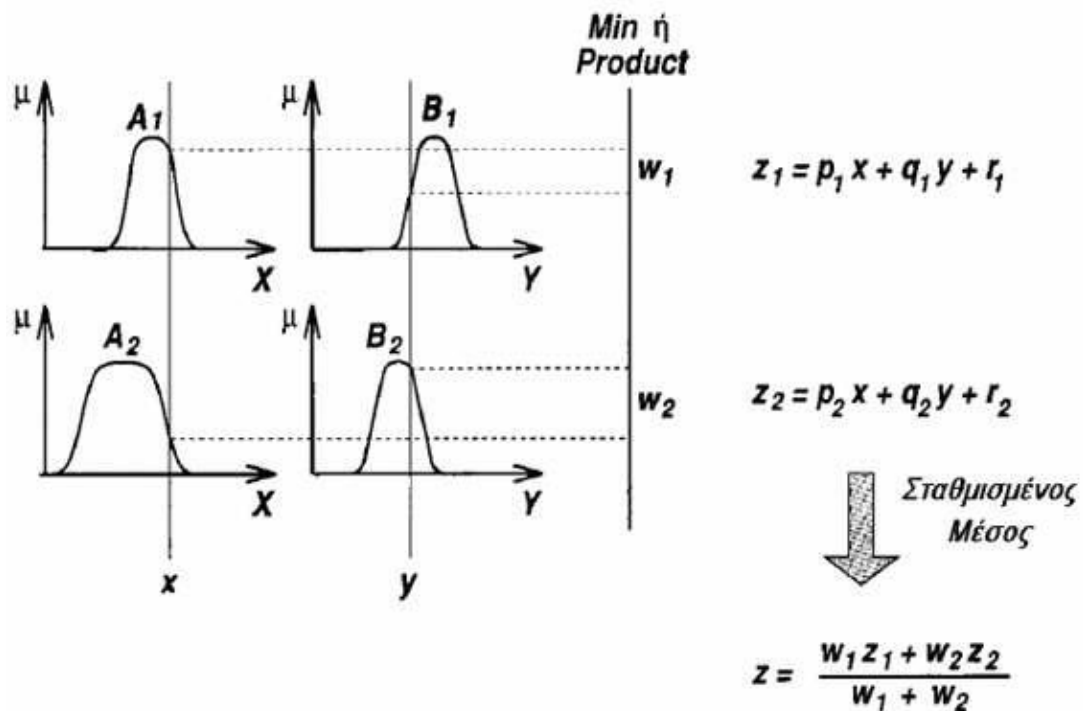
4.2 Αρχιτεκτονική του ANFIS

Ο αλγόριθμος ANFIS είναι από τους κυριότερους αλλά και ταυτόχρονα από τους πρώτους που εφαρμόστηκαν στο πεδίο της νευρο-ασαφούς προσέγγισης προβλημάτων. Παρακάτω περιγράφεται το δίκτυο, υποθέτοντας ότι το πρόβλημα το οποίο θα αναλυθεί έχει δύο εισόδους x και y και μία έξοδο z . Υποθέτοντας ότι για ένα πρώτης τάξης μοντέλο Sugeno, μία τυπική βάση κανόνων (rule base) θα μπορούσε να είναι και η εξής:

Κανόνας 1: Εάν x είναι A_1 και y είναι B_1 τότε $f_1 = p_1 \times x + q_1 \times y + r_1$

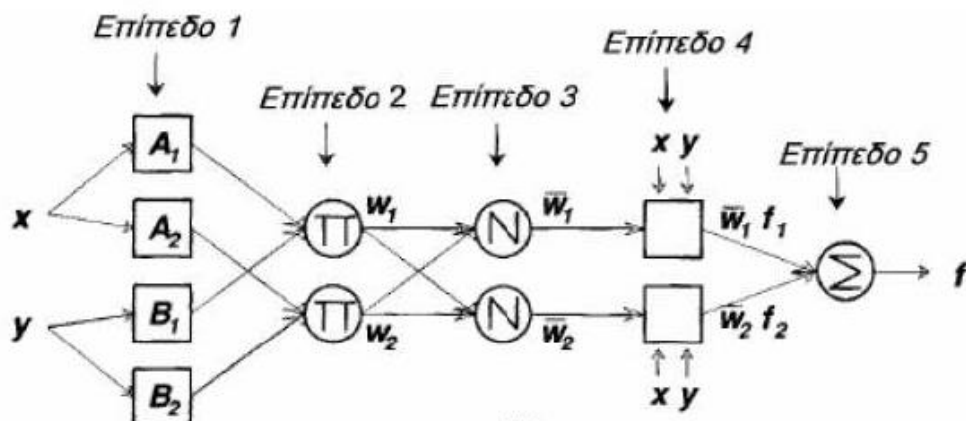
Κανόνας 2: Εάν x είναι A_2 και y είναι B_2 τότε $f_2 = p_2 \times x + q_2 \times y + r_2$

Το παρακάτω Σχήμα 3.1 δείχνει με απλό τρόπο τη διαδικασία συμπερασμού (inference procedure) του μοντέλου Sugeno, στην περίπτωση όπου για τοperator έχει επιλεγεί η τομή των δύο ασαφών συνόλων (A,B), οπότε $\mu A \cap B(x) = \min[\mu A(x), \mu B(x)]$.



Σχήμα 3.1: Ασαφής Συλλογιστική (fuzzy reasoning)

Όπως γίνεται φανερό, η έξοδος z του πρωτοβάθμιου μοντέλου Sugeno είναι ένας σταθμικός μέσος όρος. Η αντίστοιχη αναπαράσταση του δικτύου ANFIS παρουσιάζεται στο παρακάτω Σχήμα 3.2 .



Σχήμα 3.2: Η αρχιτεκτονική του ANFIS

Το παραπάνω σχήμα απεικονίζει τον συλλογιστικό μηχανισμό (reasoning) για αυτό το μοντέλο Sugeno και η αντίστοιχη ισοδύναμη αρχιτεκτονική του ANFIS όπου οι κόμβοι του ίδιου επιπέδου έχουν παρόμοιες συναρτήσεις. Παρακάτω παρουσιάζεται πιο αναλυτικά η διεργασία που εκτελείται σε κάθε επίπεδο.

Επίπεδο 1: Κάθε κόμβος i σε αυτό το επίπεδο είναι ένας προσαρμόσιμος (adaptive) κόμβος με μια συνάρτηση κόμβου:

$$\begin{aligned} O_{1,i} &= \mu_{A_i}(x) \quad \text{για } i = 1,2 \text{ ή} \\ O_{1,i} &= \mu_{B_{i-2}}(y) \quad \text{για } i = 3,4 \text{ ή} \\ O_{1,i} &= \mu_{C_j}(y) \quad \text{για } i = j+4 \text{ και } j=1,2, \end{aligned}$$

όπου x (ή y ή z)- η είσοδος στον κόμβο i , A_i (ή B_{i-2} ή C_j)- η γλωσσική μεταβλητή (small, large, κλπ.) που σχετίζεται με αυτή τη συνάρτηση του κόμβου. Με άλλα λόγια, το $O_{1,i}$ είναι ο βαθμός συμμετοχής του A ($= A_1, A_2, B_1, B_2, C_1$ ή C_2) και καθορίζει το βαθμό στον οποίο η είσοδος x (ή y ή z) ικανοποιεί τον ποσοτικοποίηση A . Εδώ η συνάρτηση συμμετοχής για το A μπορεί να είναι οποιαδήποτε κατάλληλη παραμετρική συνάρτηση συμμετοχής όπως η καμπανοειδής

για παράδειγμα:
$$\mu_{A_i}(x) = e^{-\frac{1}{2} \left(\frac{x-c_i}{a_i} \right)^2}$$

όπου a_i, c_i είναι το σύνολο των παραμέτρων. Καθώς οι τιμές αυτών των παραμέτρων αλλάζουν, οι συναρτήσεις ποικίλλουν ανάλογα, παρουσιάζοντας έτσι διάφορες μορφές της συνάρτησης συμμετοχής για το ασαφές σύνολο A . Οι παράμετροι σε αυτό το επίπεδο αναφέρονται ως αρχικοί παράμετροι (premise parameters).

Επίπεδο 2: Κάθε κόμβος σε αυτό το επίπεδο είναι ένας σταθερός (fixed) κόμβος Π, του οποίου η έξοδος είναι το γινόμενο όλων των εισερχόμενων σημάτων:

$$O_{2,i} = w_i = \mu_{A_i}(x) * \mu_{B_i}(y) * \mu_{C_i}(z), \text{ για } i = 1,2$$

$$O_{2,i} = w_i = \mu_{A_i}(x) * \mu_{B_{i-2}}(y) * \mu_{C_2}(z), \text{ για } i = 3,4$$

$$O_{2,i} = w_i = \mu_{A_2}(x) * \mu_{B_i}(y) * \mu_{C_j}(z), \text{ για } i = j+4, j=1,2$$

$$O_{2,i} = w_i = \mu_{A_2}(x) * \mu_{B_2}(y) * \mu_{C_{i-2}}(z), \text{ για } i = j+4, j=3,4$$

Επίπεδο 3: Κάθε κόμβος σε αυτό το επίπεδο είναι ένας σταθερός κόμβος Ν. Ο i-ιστός κόμβος υπολογίζει το λόγο της βαθμού ενεργοποίησης (firing strength) του i-οστού κανόνα στο άθροισμα των βαθμών ενεργοποίησης όλων των κανόνων:

$$O_{3,i} = \bar{w}_i = \frac{w_i}{w_1 + w_2 + w_3 + w_4 + w_5 + w_6 + w_7 + w_8}, \quad i=1,2,3,4,5,6,7,8.$$

Για ευκολία, οι έξοδοι αυτού του επιπέδου ονομάζονται κανονικοποιημένοι βαθμοί ενεργοποίησης (normalized firing strengths).

Επίπεδο 4: Κάθε κόμβος i σε αυτό το επίπεδο είναι ένας προσαρμόσιμος κόμβος με μία συνάρτηση κόμβου.

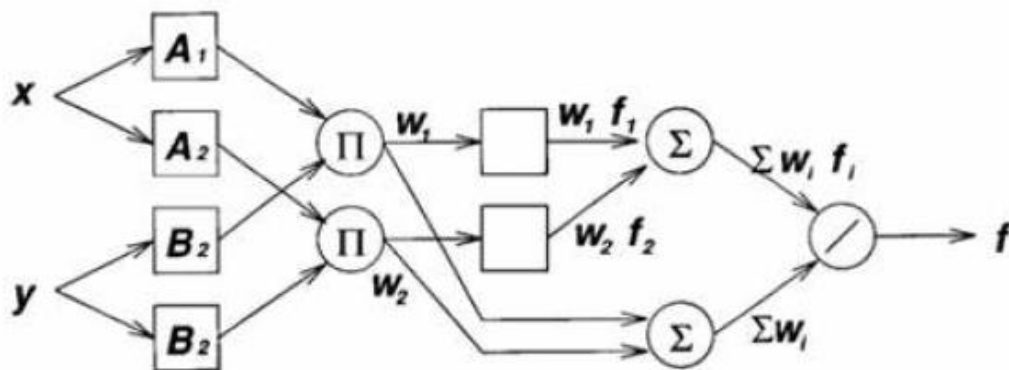
$$O_{4,i} = \bar{w}_i * f_i = \bar{w}_i * (p_i * x + q_i * y + s_i * z + r_i)$$

όπου: w_i - η έξοδος του επιπέδου 3, $\{p_i, q_i, s_i, r_i\}$ - το σύνολο των παραμέτρων. Οι παράμετροι σε αυτό το επίπεδο αναφέρονται ως επακόλουθοι (consequent parameters).

Επίπεδο 5: Ο μοναδικός κόμβος σε αυτό το επίπεδο είναι ένας σταθερός κόμβος Σ που υπολογίζει τη συνολική έξοδο σαν το ολικό άθροισμα όλων των εισερχόμενων σημάτων:

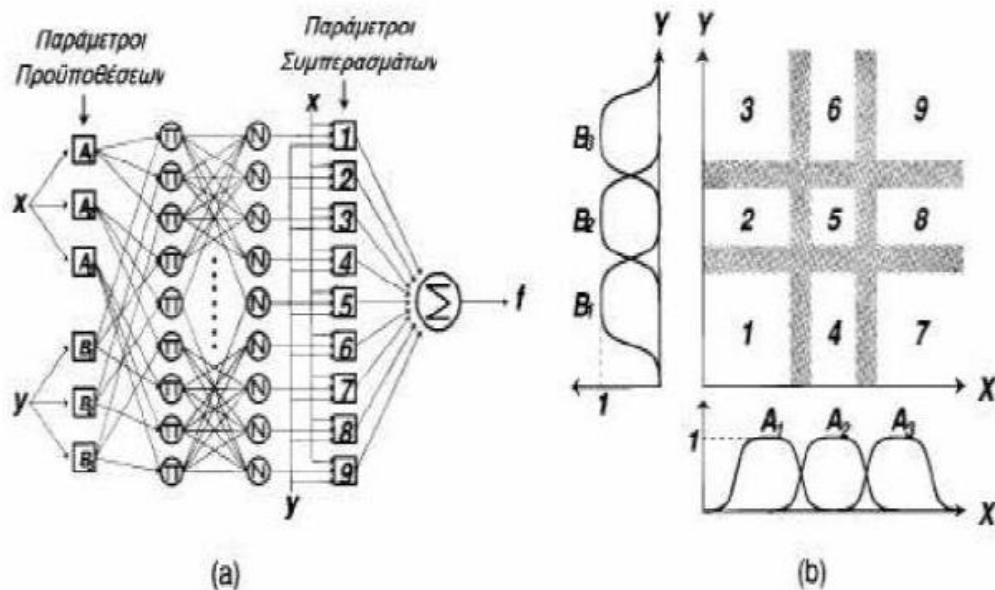
$$\text{overall output} = O_{5,i} = \sum_i \bar{w}_i * f_i = \frac{\sum_i w_i * f_i}{\sum_i w_i}$$

Αυτό το προσαρμοστικό δίκτυο είναι λειτουργικά ισοδύναμο με το ασαφές μοντέλο Sugeno. Μπορούμε να συνδυάσουμε τα επίπεδα 3 και 4 για να αποκτήσουμε ένα ισοδύναμο δίκτυο με τέσσερα μόνο επίπεδα. Με το ίδιο δείγμα μπορούμε να πραγματοποιήσουμε την κανονικοποίηση των βαρών στο τελευταίο επίπεδο. Το σχήμα 3.3 απεικονίζει ένα ANFIS αυτού του τύπου. Στην ακραία περίπτωση μπορούμε να συρρικνώσουμε ακόμα και όλο το δίκτυο, σε έναν μόνο προσαρμοστικό κόμβο με το ίδιο σύνολο παραμέτρων. Η ανάθεση συναρτήσεων κόμβων και η σύνθεση του δικτύου είναι αυθαίρετες, εφόσον κάθε κόμβος και κάθε επίπεδο πραγματοποιούν λειτουργίες που είναι σημαντικές και έχουν δυνατότητα να αποτελούνται από επιμέρους τμήματα.



Σχήμα 3.3: Η αρχιτεκτονική του ANFIS για το ασαφές σύνολο Sugeno, όπου η κανονικοποίηση βαρών πραγματοποιείται στο τελευταίο επίπεδο.

Εκτός από το μηχανισμό συμπερασμού του Sugeno, μπορούμε να κατασκευάσουμε ANFIS και με το μοντέλο Mamdani. Το Sugeno χρησιμοποιείται περισσότερο συχνά, μιας και διακρίνεται για την διαφάνεια και την αποτελεσματικότητά του. Στο σχήμα 3.4 (a) βλέπουμε μια αρχιτεκτονική ANFIS που είναι ισοδύναμη με ένα ασαφές μοντέλο Sugeno πρώτου βαθμού δυο εισόδων και εννέα κανόνων, σε κάθε είσοδο θεωρούμε ότι αντιστοιχούν τρεις συναρτήσεις συμμετοχής. Το σχήμα 3.4 (b) απεικονίζει πως δύο διαστάσεων χώρος εισόδου είναι χωρισμένος σε εννέα υπέρθετες (overlapping) ασαφείς περιοχές όπου κάθε μια ελέγχεται από ένα ασαφή κανόνα if-then. Αυτό σημαίνει ότι το μέρος των προϋποθέσεων ενός κανόνα προσδιορίζει μια ασαφή περιοχή, ενώ το μέρος των συμπερασμάτων προσδιορίζει την έξοδο μέσα στην περιοχή.



Σχήμα 3.4(a) Η αρχιτεκτονική του ANFIS για το ασαφές μοντέλο Sugeno με δύο εισόδους και εννέα κανόνες (b) ο χώρος εισόδου, χωρισμένος σε εννέα ασαφείς περιοχές.

4.3 Περιορισμοί του ANFIS

Οι κυριότεροι περιορισμοί του ANFIS που αναφέρονται στην διεθνή βιβλιογραφία είναι:

- Βασίζεται σε ένα FIS τύπου Sugeno.
- Έχει μια έξοδο που λαμβάνεται με τη μέθοδο της αποασαφοποίησης σταθμισμένου μέσου.
- Όλες οι συναρτήσεις συμμετοχής εξόδου πρέπει να είναι ίδιου τύπου, είτε γραμμικές είτε σταθερές.
- Δεν μπορεί να γίνεται κοινή χρήση κανόνων. Διαφορετικοί κανόνες δεν μπορούν να έχουν την ίδια συνάρτηση συμμετοχής εξόδου.
- Πρέπει να υπάρχουν βάρη σε κάθε κανόνα.
- Δεν μπορεί να χρησιμοποιηθεί οποιαδήποτε συνάρτηση συμμετοχής, παρά μόνο προκαθορισμένες που επιβάλλουν οι περιορισμοί του ANFIS.

4.4 Εκπαίδευση του μοντέλου ANFIS

Σε αυτή τη μελέτη θα χρησιμοποιηθεί ένα Προσαρμοστικό Νευροασαφές Σύστημα Συμπερασμού (Adaptive Neural Fuzzy Inference, ANFIS) μοντέλο για την πρόβλεψη των σημείων αλλαγής τάσης στο χρηματιστήριο. Τα ασαφή συστήματα συμπερασμού (Fuzzy inference systems) που χρησιμοποιούν νευρωνικά δίκτυα προτάθηκαν ώστε να μπορούν να αποφευχθούν τα αδύναμα σημεία της ασαφούς λογικής. Το μεγαλύτερο πλεονέκτημα είναι ότι μπορούν να χρησιμοποιήσουν την ικανότητα εκμάθησης των νευρωνικών δικτύων και ότι μπορούν να αποφύγουν τον χρόνο ταιριάσματος κανόνων μιας μηχανής συμπεράσματος στο παραδοσιακό σύστημα ασαφούς λογικής. Το μοντέλο έχει τρεις εισόδους και μια έξοδο και η τιμή πρόβλεψης δίνεται από την ακόλουθη εξίσωση:

$$y(t+1) = f(y(t), y(t-1), y(t-2))$$

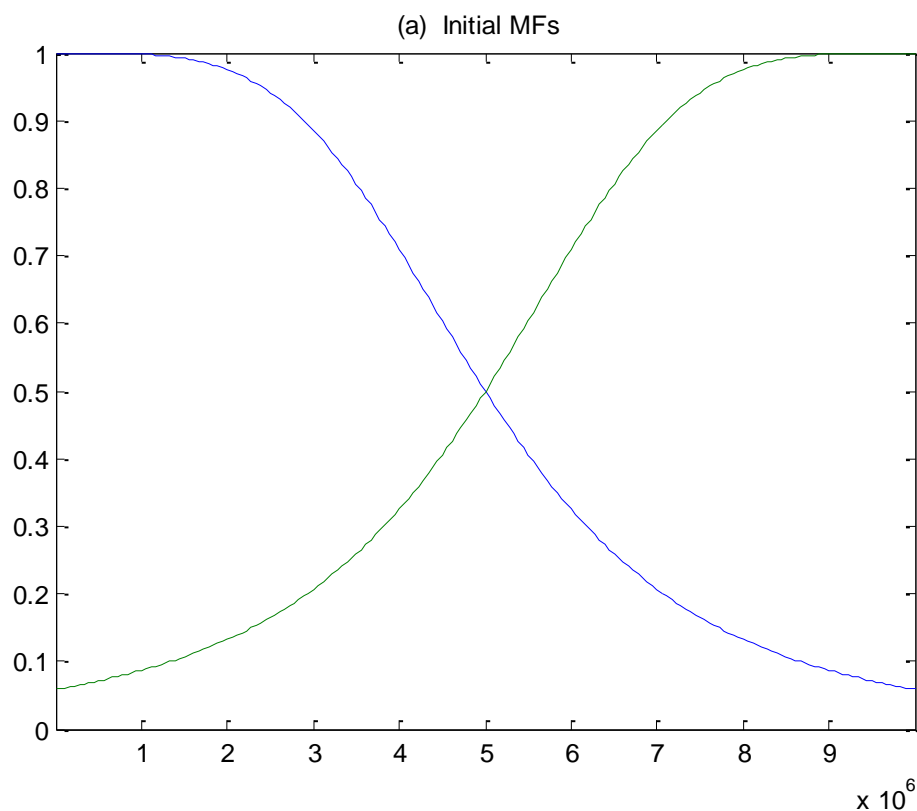
Τα δεδομένα που χρησιμοποιήθηκαν αφορούν ημερήσιες τιμές κλεισίματος πολύτιμων μετάλλων (gold, silver, platinum, palladium). Τα δεδομένα αφορούν μια περίοδο 16 ετών, από το 1999 έως 2014.

ΚΕΦΑΛΑΙΟ 5: Αποτελέσματα

Στην παράγραφο αυτή θα παρουσιαστούν τα αποτελέσματα από τη χρήση του μοντέλου μας. Στα επόμενα σχήματα παρουσιάζονται τα αποτελέσματα για δεδομένα στα οποία έχει εκπαιδευτεί το μοντέλο (εντός δείγματος) και για δεδομένα στα οποία δεν έχει εκπαιδευτεί το μοντέλο (εκτός δείγματος).

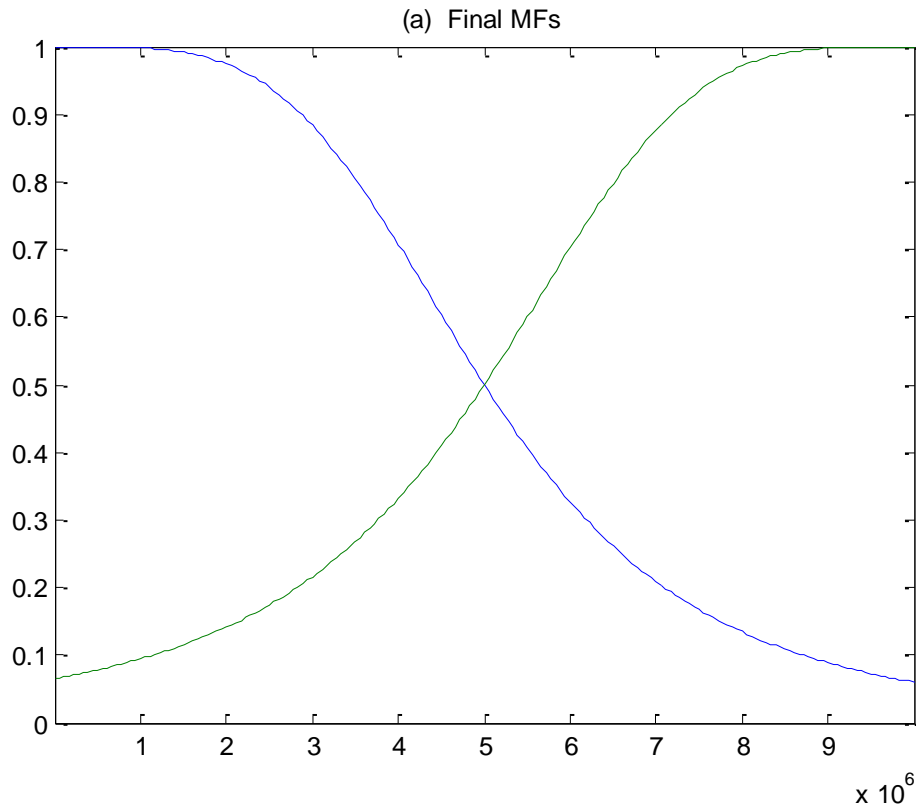
5.1 Αποτελέσματα του μοντέλου ANFIS

Τα αποτελέσματα του μοντέλου ANFIS για Gold , Palladium , Platinum και Silver παρατίθενται στα παρακάτω αντίστοιχα διαγράμματα.



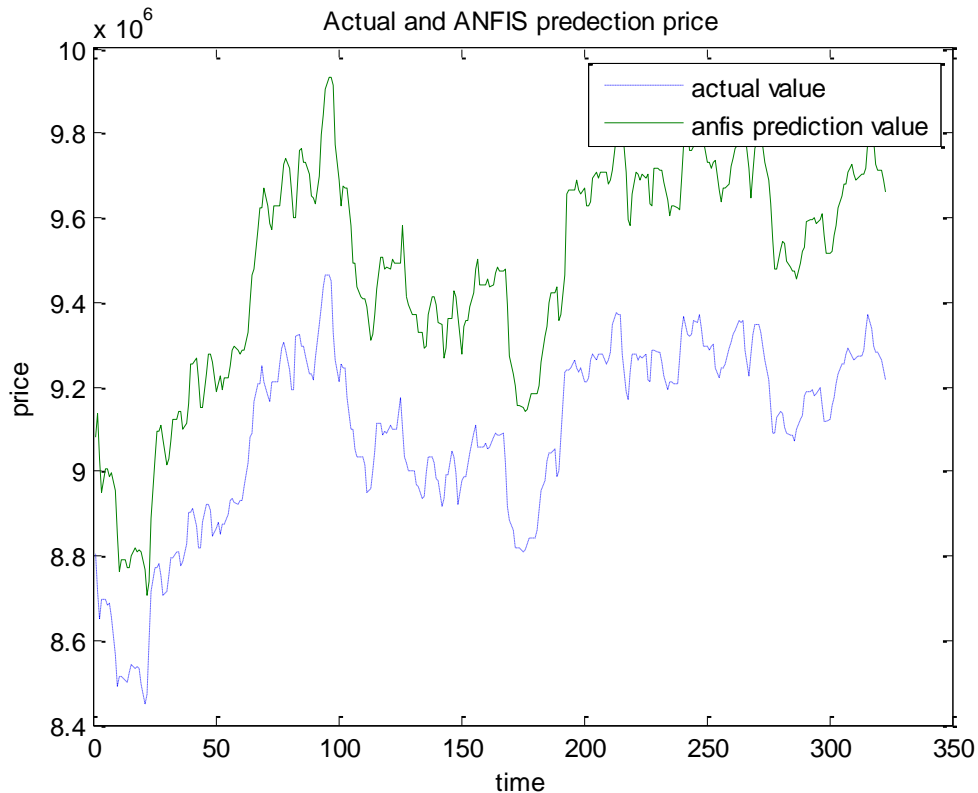
Γράφημα 1: Gold – Αρχικές συναρτήσεις συμμετοχής

Στο παραπάνω γράφημα παρουσιάζονται οι αρχικές συναρτήσεις συμμετοχής πριν την εκπαίδευση των παραμέτρων του ANFIS.



Γράφημα 2: Gold – Τελικές συναρτήσεις συμμετοχής

Στο παραπάνω γράφημα παρουσιάζονται οι τελικές συναρτήσεις συμμετοχής μετά την εκπαίδευση των παραμέτρων του ANFIS.



Γράφημα 3: Gold - Σύγκριση προβλέψεων του μοντέλου ANFIS έναντι των επιθυμητών τιμών.

Στο παραπάνω γράφημα αξιολογούμε το ANFIS συγκρίνοντας τα αποτελέσματα του ως προς τις πραγματικές τιμές και οπτικοποιούμε το αποτέλεσμα με κοινή τους παρουσίαση.
(μπλε γραμμή → πραγματικές τιμές, πράσινη γραμμή → πρόβλεψη ANFIS)

Παρατηρούμε εδώ ότι υπάρχει μια σημαντική διαφορά ανάμεσα στις πραγματικές τιμές και στην πρόβλεψη του ANFIS. Θα μπορούσαμε να πούμε δηλαδή ότι η πρόβλεψη του ANFIS για Gold δεν είναι και πάρα πολύ καλή.

Σφάλματα:

Ρίζα μέσου τετραγωνικού σφάλματος:

(Root Mean Squared Error)

$$RMSE = \sqrt{\frac{\sum_{t=1}^N e_t^2}{N}}$$

Μέσο τετραγωνικό σφάλμα:

(Mean Square Error)

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2.$$

Μέσο απόλυτο σφάλμα:

(Mean Absolute Error)

$$MAE = \frac{1}{N} \cdot \sum_{t=1}^N |e_t|$$

Μέσο απόλυτο ποσοστιαίο σφάλμα:

(Mean Absolute Percentage Error)

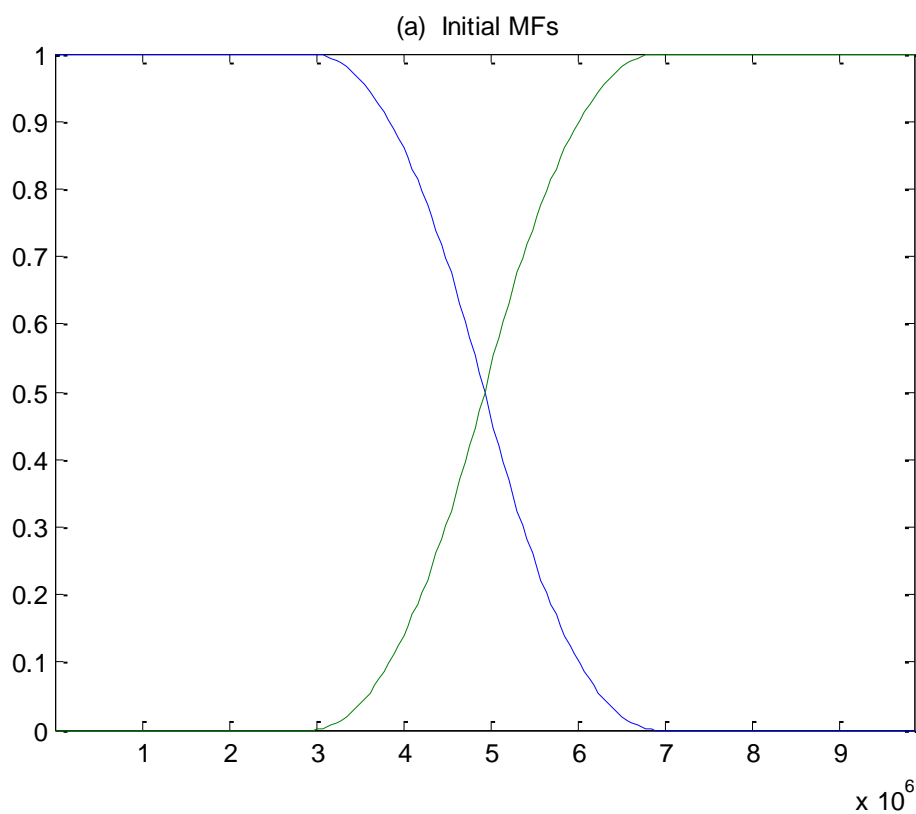
$$MAPE = \frac{100}{N} \cdot \sum_{t=1}^N \left| \frac{F_t - A_t}{A_t} \right|$$

Σε αυτό το σημείο ακολουθεί περαιτέρω αξιολόγηση του μοντέλου ANFIS με τα μοντέλα AR, και ARMA με τη βοήθεια του System Identification Toolbox της Matlab, από τα οποία εξάγουμε τους ίδιους τύπους σφαλμάτων όπως και στο ANFIS.

Πίνακας 1: Gold - Συγκριτικός πίνακας σφαλμάτων
(MSE/RMSE/MAE/MAPE)

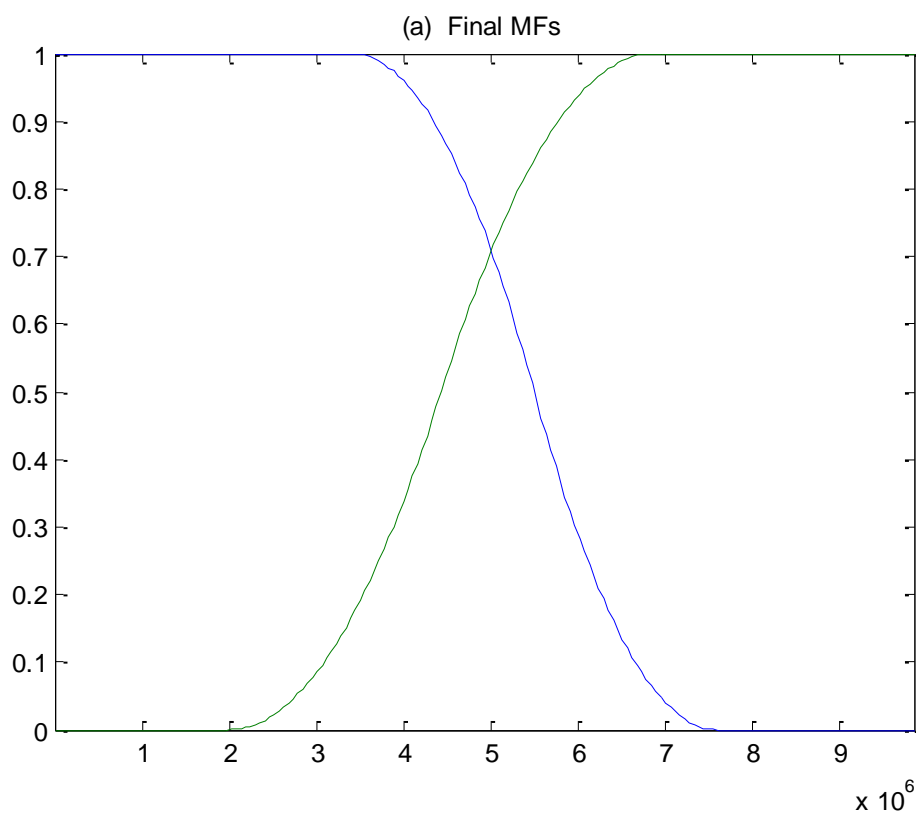
GOLD	MSE	RMSE	MAE	MAPE
AR	1,0477	0,000000747	0,000000774	0,000000477
ARMA	1,0438	0,000000843	0,000000834	0,000000438
ANFIS	0,0191	0,000000019	0,000000091	0,000000011

Όπως προκύπτει από την σύγκριση των μοντέλων, για το μέσο τετραγωνικό σφάλμα (MSE), μικρότερο σφάλμα παρουσίασε το ANFIS, στη συνέχεια το μοντέλο AR και έπεται το ARMA. Για το μέσο απόλυτο σφάλμα (MAE) και πάλι το ANFIS είχε μικρότερο σφάλμα. Τα μοντέλα AR, ARMA ακολουθούν. Για το μέσο απόλυτο ποσοστιαίο σφάλμα (MAPE) το ANFIS παρουσίασε και εδώ την χαμηλότερη τιμή από τα υπόλοιπα μοντέλα. Τέλος, για το σφάλμα της ρίζας του μέσου τετραγωνικού σφάλματος (RMSE) το ANFIS παρουσίασε την μικρότερη τιμή σφάλματος έναντι των μοντέλων AR και ARMA. Συμπερασματικά το ANFIS παρουσίασε καλύτερα αποτελέσματα έναντι των μοντέλων AR, και ARMA.



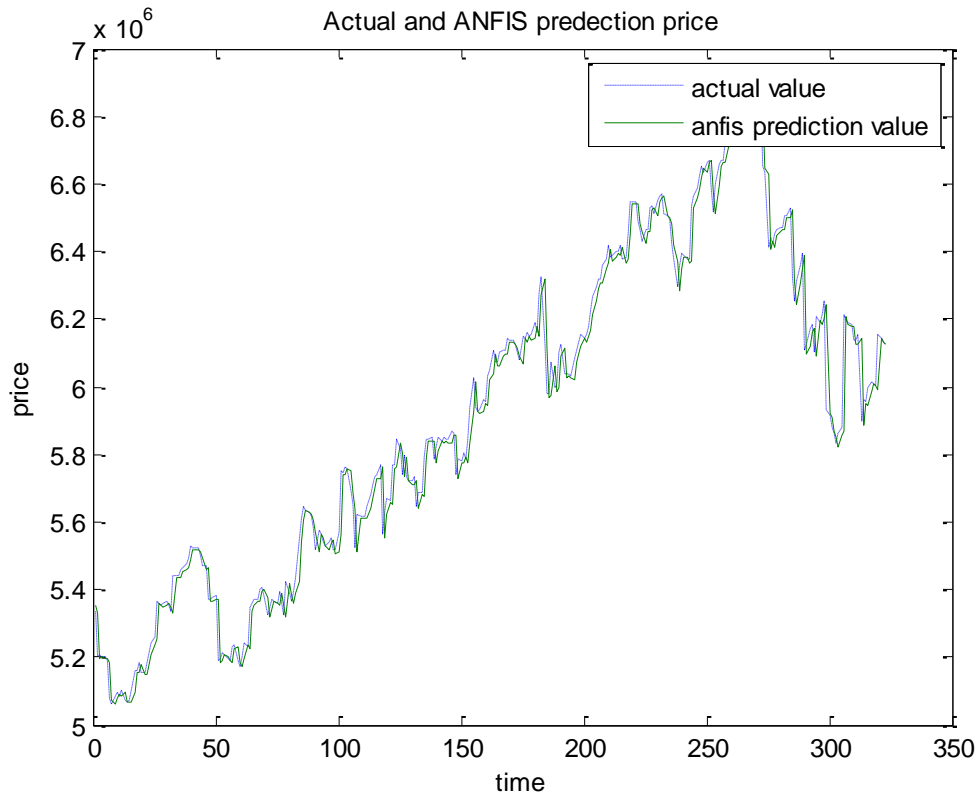
Γράφημα 4: Palladium – Αρχικές συναρτήσεις συμμετοχής

Στο παραπάνω γράφημα παρουσιάζονται οι αρχικές συναρτήσεις συμμετοχής πριν την εκπαίδευση των παραμέτρων του ANFIS.



Γράφημα 5: Palladium – Τελικές συναρτήσεις συμμετοχής

Στο παραπάνω γράφημα παρουσιάζονται οι τελικές συναρτήσεις συμμετοχής μετά την εκπαίδευση των παραμέτρων του ANFIS.



Γράφημα 6: Palladium - Σύγκριση προβλέψεων του μοντέλου ANFIS έναντι των επιθυμητών τιμών.

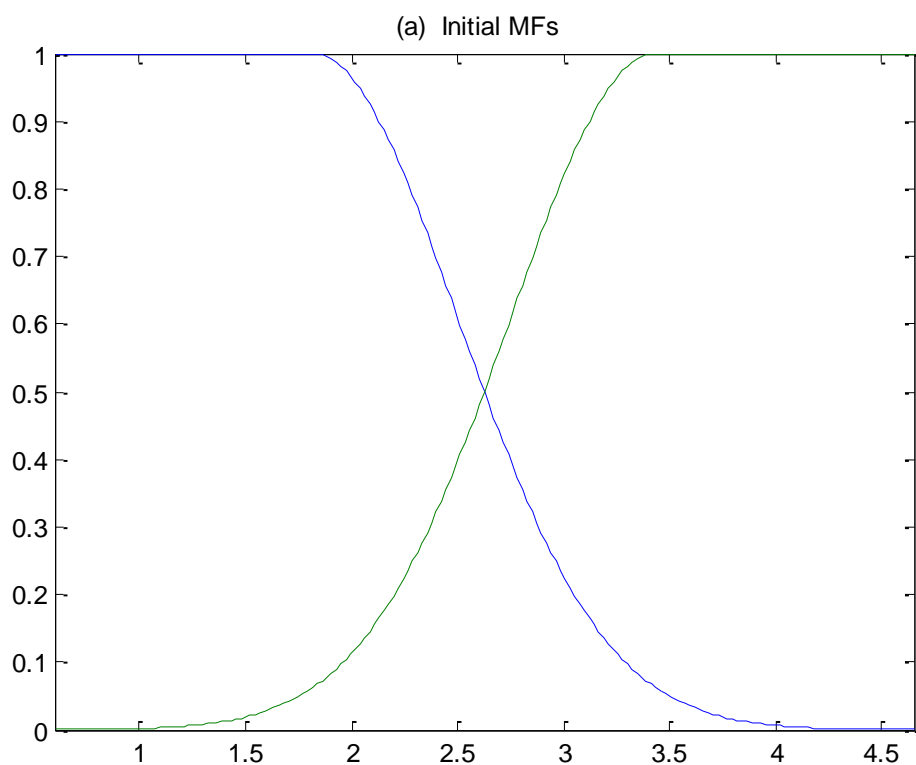
Στο παραπάνω γράφημα αξιολογούμε το ANFIS συγκρίνοντας τα αποτελέσματα του ως προς τις πραγματικές τιμές και οπτικοποιούμε το αποτέλεσμα με κοινή τους παρουσίαση.
(μπλε γραμμή → πραγματικές τιμές, πράσινη γραμμή → πρόβλεψη ANFIS)

Παρατηρούμε εδώ ότι υπάρχει μια σχετική ταύτιση ανάμεσα στις πραγματικές τιμές και στην πρόβλεψη του ANFIS. Θα μπορούσαμε λοιπόν να πούμε ότι η πρόβλεψη του ANFIS για Palladium ότι είναι πάρα πολύ καλή.

Πίνακας 2: Palladium - Συγκριτικός πίνακας σφαλμάτων
(MSE/RMSE/MAE/MAPE)

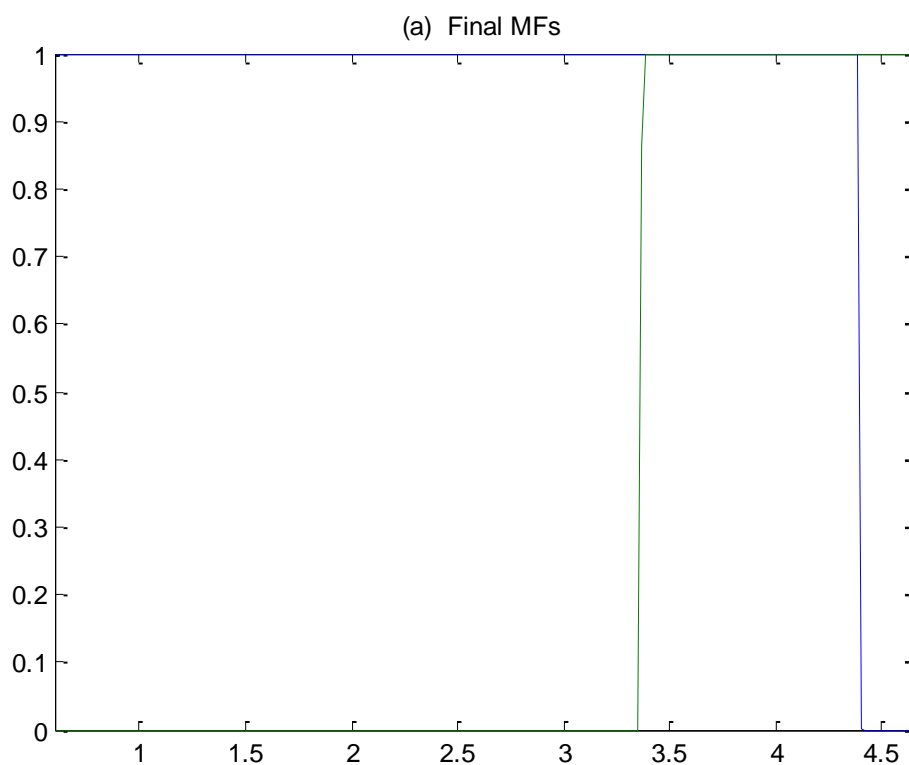
PALLADIUM	MSE	RMSE	MAE	MAPE
AR	5,6529	0,000000529	0,000000629	0,000000695
ARMA	5,6645	0,000000645	0,000000654	0,000000664
ANFIS	0,0413	0,000000013	0,000000034	0,000000031

Όπως προκύπτει από την σύγκριση των μοντέλων, για το μέσο τετραγωνικό σφάλμα (MSE), μικρότερο σφάλμα παρουσίασε το ANFIS, στη συνέχεια το μοντέλο AR και έπεται το ARMA. Για το μέσο απόλυτο σφάλμα (MAE) και πάλι το ANFIS είχε μικρότερο σφάλμα. Τα μοντέλα AR, ARMA ακολουθούν. Για το μέσο απόλυτο ποσοστιαίο σφάλμα (MAPE) το ANFIS παρουσίασε και εδώ την χαμηλότερη τιμή από τα υπόλοιπα μοντέλα. Τέλος, για το σφάλμα της ρίζας του μέσου τετραγωνικού σφάλματος (RMSE) το ANFIS παρουσίασε την μικρότερη τιμή σφάλματος έναντι των μοντέλων AR και ARMA. Συμπερασματικά το ANFIS παρουσίασε καλύτερα αποτελέσματα έναντι των μοντέλων AR, και ARMA.



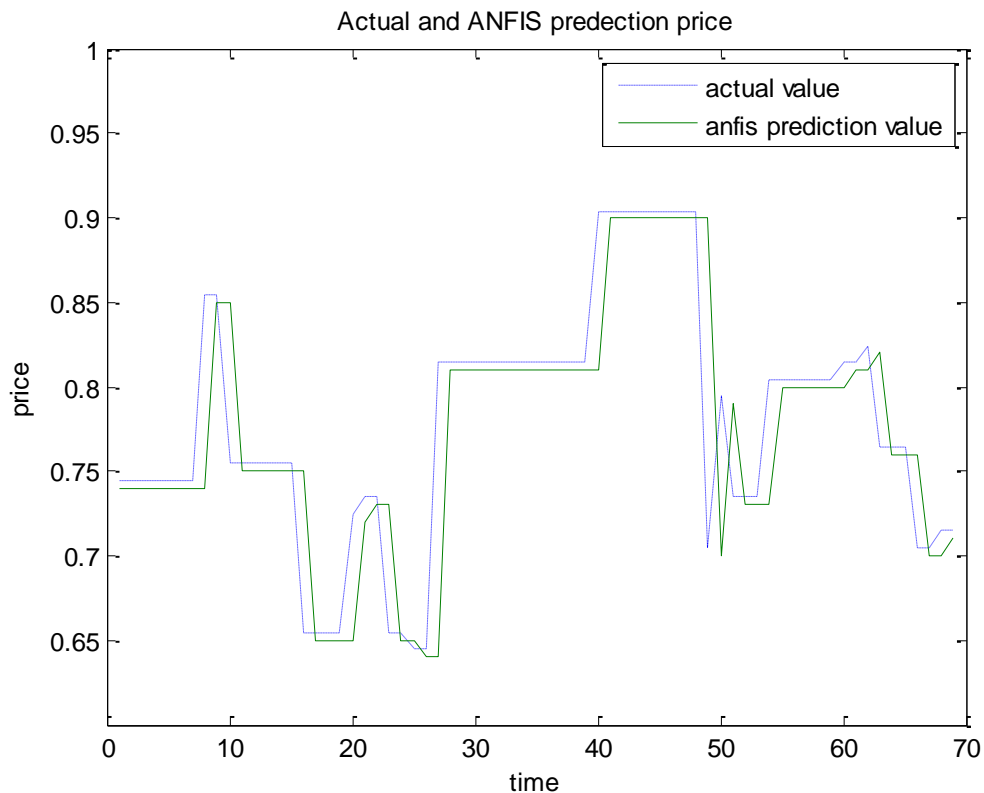
Γράφημα 7: Platinum – Αρχικές συναρτήσεις συμμετοχής

Στο παραπάνω γράφημα παρουσιάζονται οι αρχικές συναρτήσεις συμμετοχής πριν την εκπαίδευση των παραμέτρων του ANFIS.



Γράφημα 8: Platinum – Τελικές συναρτήσεις συμμετοχής

Στο παραπάνω γράφημα παρουσιάζονται οι τελικές συναρτήσεις συμμετοχής μετά την εκπαίδευση των παραμέτρων του ANFIS.



Γράφημα 9: Platinum - Σύγκριση προβλέψεων του μοντέλου ANFIS έναντι των επιθυμητών τιμών.

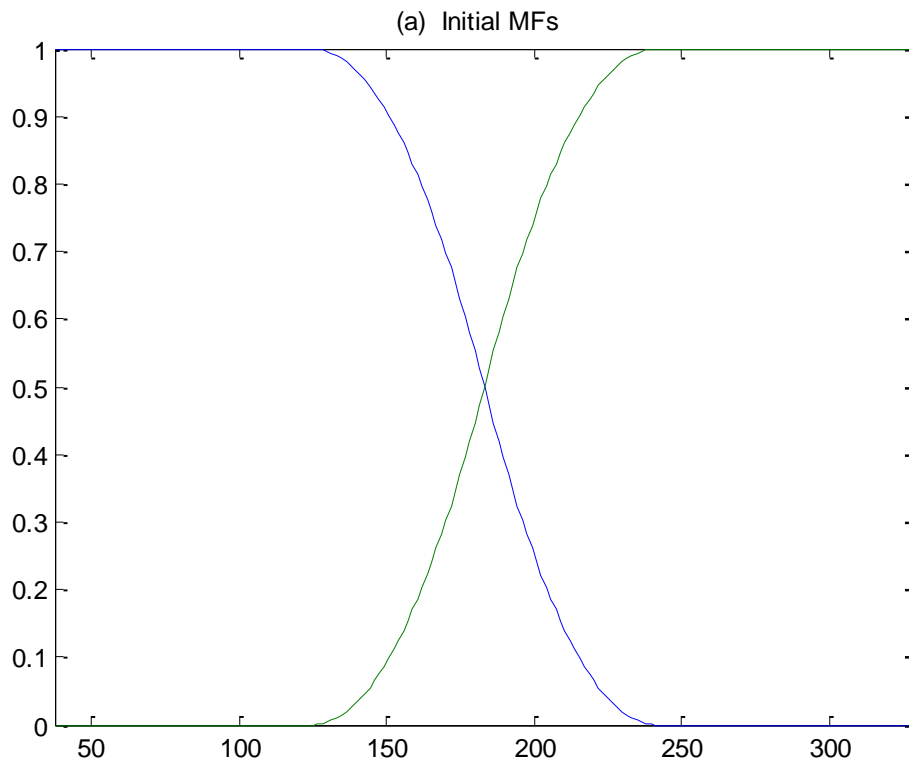
Στο παραπάνω γράφημα αξιολογούμε το ANFIS συγκρίνοντας τα αποτελέσματα του ως προς τις πραγματικές τιμές και οπτικοποιούμε το αποτέλεσμα με κοινή τους παρουσίαση.
(μπλε γραμμή → πραγματικές τιμές, πράσινη γραμμή → πρόβλεψη ANFIS)

Παρατηρούμε εδώ ότι υπάρχει μια πολύ μικρή διαφορά ανάμεσα στις πραγματικές τιμές και στην πρόβλεψη του ANFIS. Θα μπορούσαμε να πούμε δηλαδή ότι η πρόβλεψη του ANFIS για Platinum ότι είναι ικανοποιητική.

Πίνακας 3: Platinum - Συγκριτικός πίνακας σφαλμάτων
(MSE/RMSE/MAE/MAPE)

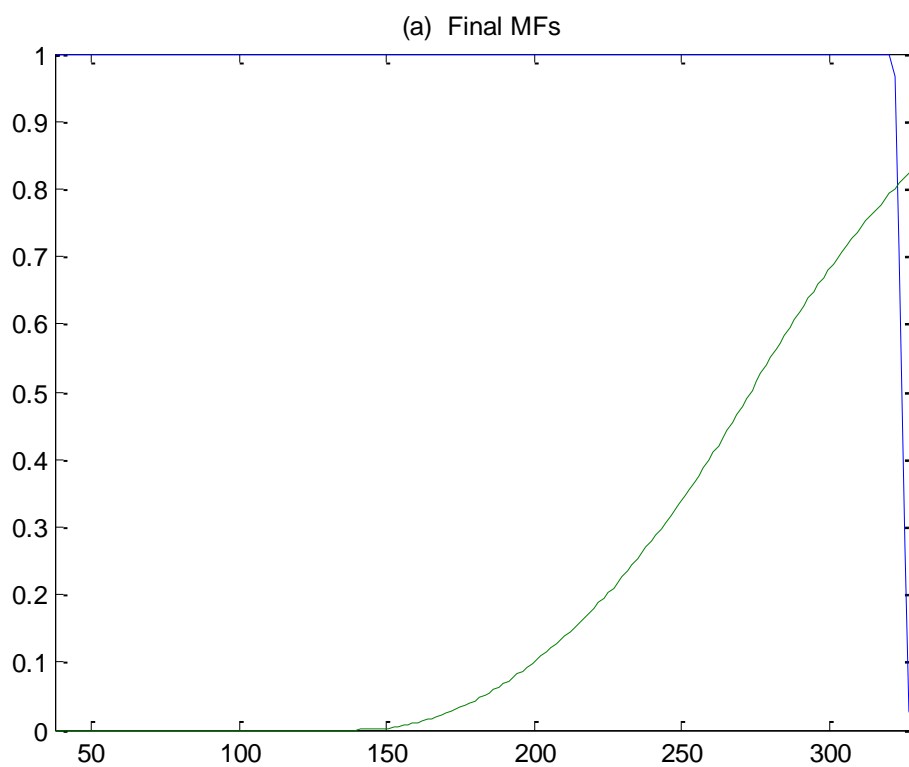
PLATINUM	MSE	RMSE	MAE	MAPE
AR	0,0063	0,0795	0,0620	7,7944
ARMA	0,0063	0,0794	0,0620	7,7935
ANFIS	0,0021	0,0459	0,0224	2,9601

Όπως προκύπτει από την σύγκριση των μοντέλων, για το μέσο τετραγωνικό σφάλμα (MSE), μικρότερο σφάλμα παρουσίασε το ANFIS, στη συνέχεια το μοντέλο AR και έπεται το ARMA. Για το μέσο απόλυτο σφάλμα (MAE) και πάλι το ANFIS είχε μικρότερο σφάλμα. Τα μοντέλα AR, ARMA ακολουθούν. Για το μέσο απόλυτο ποσοστιαίο σφάλμα (MAPE) το ANFIS παρουσίασε και εδώ την χαμηλότερη τιμή από τα υπόλοιπα μοντέλα. Τέλος, για το σφάλμα της ρίζας του μέσου τετραγωνικού σφάλματος (RMSE) το ANFIS παρουσίασε την μικρότερη τιμή σφάλματος έναντι των μοντέλων AR και ARMA. Συμπερασματικά το ANFIS παρουσίασε καλύτερα αποτελέσματα έναντι των μοντέλων AR, και ARMA.



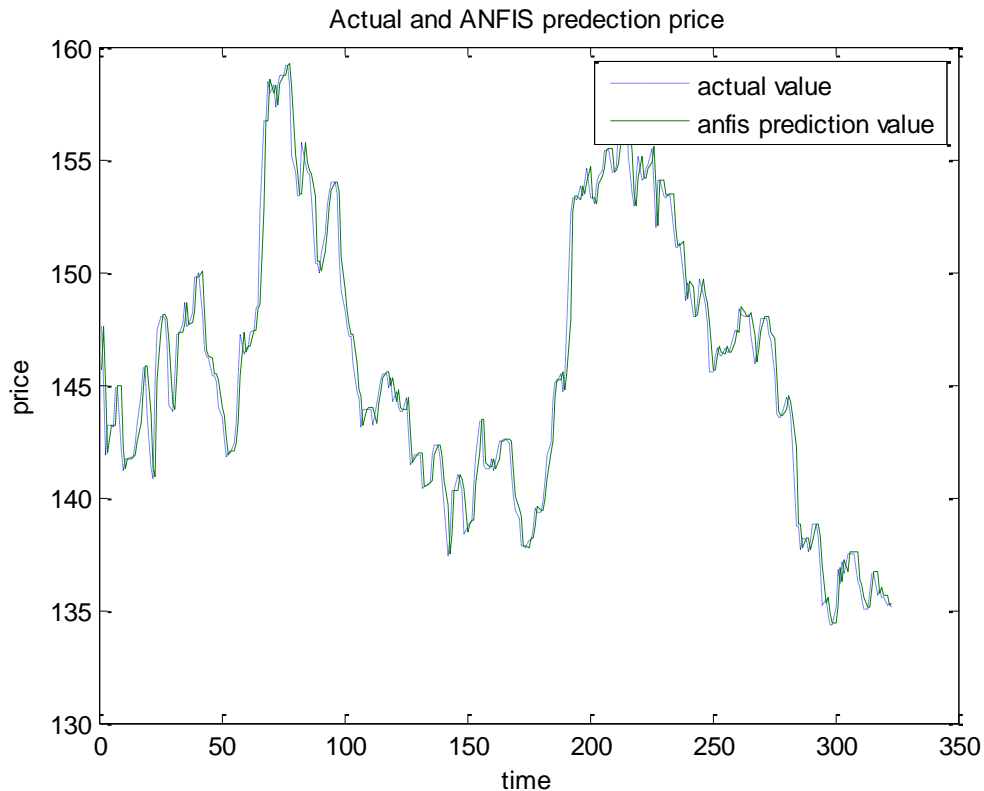
Γράφημα 10: Silver – Αρχικές συναρτήσεις συμμετοχής

Στο παραπάνω γράφημα παρουσιάζονται οι αρχικές συναρτήσεις συμμετοχής πριν την εκπαίδευση των παραμέτρων του ANFIS.



Γράφημα 11: Silver – Τελικές συναρτήσεις συμμετοχής

Στο παραπάνω γράφημα παρουσιάζονται οι τελικές συναρτήσεις συμμετοχής μετά την εκπαίδευση των παραμέτρων του ANFIS.



Γράφημα 12: Silver - Σύγκριση προβλέψεων του μοντέλου ANFIS έναντι των επιθυμητών τιμών.

Στο παραπάνω γράφημα αξιολογούμε το ANFIS συγκρίνοντας τα αποτελέσματα του ως προς τις πραγματικές τιμές και οπτικοποιούμε το αποτέλεσμα με κοινή τους παρουσίαση.
(μπλε γραμμή → πραγματικές τιμές, πράσινη γραμμή → πρόβλεψη ANFIS)

Παρατηρούμε εδώ ότι υπάρχει μια απόλυτη ταύτιση ανάμεσα στις πραγματικές τιμές και στην πρόβλεψη του ANFIS. Θα μπορούσαμε λοιπόν να πούμε ότι η πρόβλεψη του ANFIS για Silver ότι είναι η πλέον ικανοποιητική.

Πίνακας 4: Silver - Συγκριτικός πίνακας σφαλμάτων
(MSE/RMSE/MAE/MAPE)

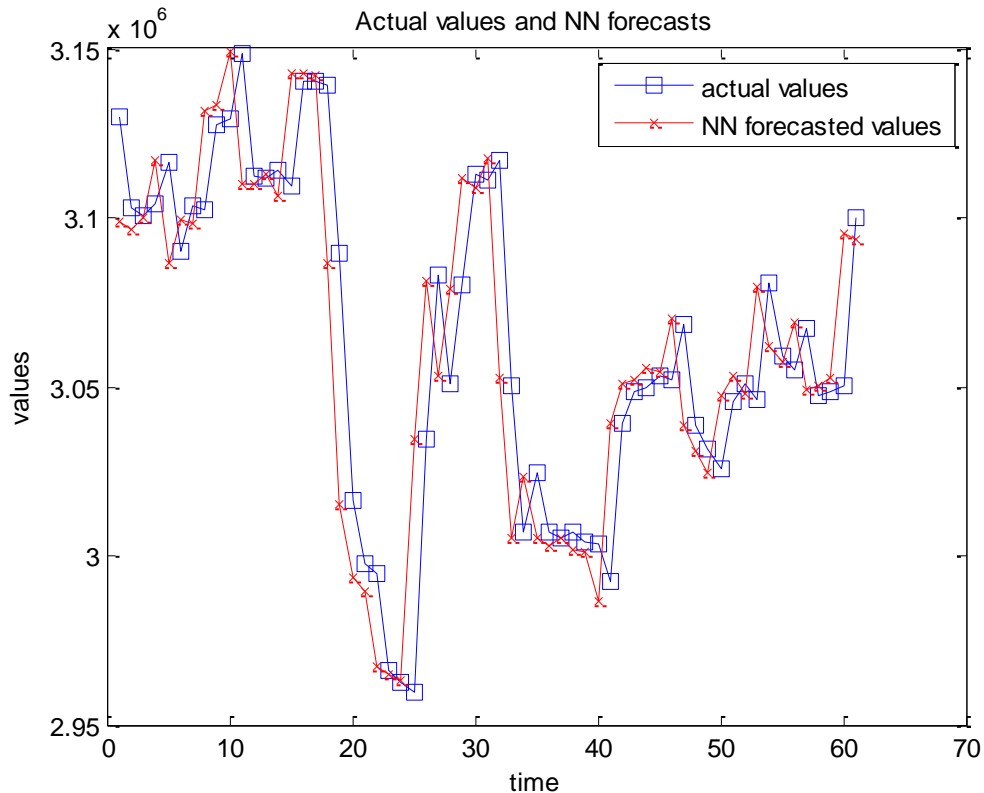
SILVER	MSE	RMSE	MAE	MAPE
AR	40,2868	6,3472	5,2957	3,6678
ARMA	41,4779	6,4403	5,3906	3,7397
ANFIS	1,1810	1,0867	0,7473	0,5110

Όπως προκύπτει από την σύγκριση των μοντέλων, για το μέσο τετραγωνικό σφάλμα (MSE), μικρότερο σφάλμα παρουσίασε το ANFIS, στη συνέχεια το μοντέλο AR και έπεται το ARMA. Για το μέσο απόλυτο σφάλμα (MAE) και πάλι το ANFIS είχε μικρότερο σφάλμα. Τα μοντέλα AR, ARMA ακολουθούν. Για το μέσο απόλυτο ποσοστιαίο σφάλμα (MAPE) το ANFIS παρουσίασε και εδώ την χαμηλότερη τιμή από τα υπόλοιπα μοντέλα. Τέλος, για το σφάλμα της ρίζας του μέσου τετραγωνικού σφάλματος (RMSE) το ANFIS παρουσίασε την μικρότερη τιμή σφάλματος έναντι των μοντέλων AR και ARMA. Συμπερασματικά το ANFIS παρουσίασε καλύτερα αποτελέσματα έναντι των μοντέλων AR, και ARMA.

5.2 Αποτελέσματα του νευρωνικού δικτύου

Έγινε επίσης η χρήση ενός νευρωνικού δικτύου στο οποίο χρησιμοποιήθηκαν τα ίδια ημερήσια δεδομένα για Gold. Τα αποτελέσματα παρατίθενται στα παρακάτω διαγράμματα. Εν συνεχεία, θα γίνει μια σύγκριση μεταξύ των αποτελεσμάτων από το μοντέλο ANFIS και αυτά του νευρωνικού δικτύου.

Για συντομότερη και ευκολότερη αξιολόγηση και αναπαράσταση των αποτελεσμάτων χρησιμοποιήθηκαν μόνο δεδομένα για Gold.



Γράφημα 13: Gold - Σύγκριση προβλέψεων του νευρωνικού δικτύου έναντι των επιθυμητών τιμών.

Στο παραπάνω γράφημα αξιολογούμε το νευρωνικό δίκτυο συγκρίνοντας τα αποτελέσματα του ως προς τις πραγματικές τιμές και οπτικοποιούμε το αποτέλεσμα με κοινή τους παρουσίαση. (μπλε γραμμή → πραγματικές τιμές, κόκκινη γραμμή → πρόβλεψη νευρωνικού δικτύου)

Παρατηρούμε εδώ ότι υπάρχει μια μικρή απόκλιση ανάμεσα στις πραγματικές τιμές και στην πρόβλεψη του νευρωνικού δικτύου. Θα μπορούσαμε λοιπόν να πούμε ότι η πρόβλεψη του νευρωνικού δικτύου για Gold ότι είναι η σχετικά ικανοποιητική.

Πίνακας 5: Gold - Συγκριτικός πίνακας σφαλμάτων
(MSE/RMSE/MAE/MAPE)

GOLD	MSE	RMSE	MAE	MAPE
AR	1,6349	1,0866	0,7444	0,5100
ARMA	1,6337	1,0066	0,7065	0,5060
NN	0,0692	0,0096	0,0062	0,0060

Όπως προκύπτει από την σύγκριση των μοντέλων, για το μέσο τετραγωνικό σφάλμα (MSE), μικρότερο σφάλμα παρουσίασε το νευρωνικό δίκτυο, στη συνέχεια το μοντέλο AR και έπεται το ARMA. Για το μέσο απόλυτο σφάλμα (MAE) και πάλι το νευρωνικό δίκτυο είχε μικρότερο σφάλμα. Τα μοντέλα AR, ARMA ακολουθούν. Για το μέσο απόλυτο ποσοστιαίο σφάλμα (MAPE) το νευρωνικό δίκτυο παρουσίασε και εδώ την χαμηλότερη τιμή από τα υπόλοιπα μοντέλα. Τέλος, για το σφάλμα της ρίζας του μέσου τετραγωνικού σφάλματος (RMSE) το νευρωνικό δίκτυο παρουσίασε την μικρότερη τιμή σφάλματος έναντι των μοντέλων AR και ARMA. Συμπερασματικά το νευρωνικό δίκτυο παρουσίασε καλύτερα αποτελέσματα έναντι των μοντέλων AR, και ARMA όπως φυσικά έγινε και με το ANFIS στους προηγούμενους συγκριτικούς πίνακες.

Ας δούμε όμως παρακάτω τι θα προκύψει εάν συγκρίνουμε τα αποτελέσματα του ANFIS με αυτά του νευρωνικού δικτύου.

Πίνακας 6: Gold - Συγκριτικός πίνακας σφαλμάτων
(MSE/RMSE/MAE/MAPE)

Εδώ συγκρίνεται το ANFIS με το νευρωνικό δίκτυο NN.

GOLD	MSE	RMSE	MAE	MAPE
ANFIS	0,0191	0,000000019	0,000000091	0,000000011
NN	0,0692	0,0096	0,0062	0,0060
AR	1,0477	0,000000747	0,000000774	0,000000477
ARMA	1,0438	0,000000843	0,000000834	0,000000438

Όπως προκύπτει από την σύγκριση των μοντέλων, για το μέσο τετραγωνικό σφάλμα (MSE), μικρότερο σφάλμα παρουσίασε το ANFIS, μετά το νευρωνικό δίκτυο και στη συνέχεια τα μοντέλα AR και ARMA. Για το μέσο απόλυτο σφάλμα (MAE) και πάλι το ANFIS είχε μικρότερο σφάλμα. Τα μοντέλα AR, ARMA καθώς και το νευρωνικό δίκτυο ακολουθούν. Για το μέσο απόλυτο ποσοστιαίο σφάλμα (MAPE) το ANFIS παρουσίασε και εδώ την χαμηλότερη τιμή από τα υπόλοιπα μοντέλα. Τέλος, για το σφάλμα της ρίζας του μέσου τετραγωνικού σφάλματος (RMSE) το ANFIS παρουσίασε την μικρότερη τιμή σφάλματος έναντι των μοντέλων AR, ARMA και του νευρωνικού δικτύου.

Συμπερασματικά, το ANFIS παρουσίασε καλύτερα αποτελέσματα συνολικά έναντι όλων μοντέλων, AR, ARMA και του νευρωνικού δικτύου.

5.3 Συμπεράσματα.

Ένα ANFIS για να μπορέσει να μελετηθεί, απαιτεί από τον χρήστη να έχει ένα ανεπτυγμένο γνωστικό υπόβαθρο, το οποίο θα περιλαμβάνει τις στατιστικές μεθόδους, την αριθμητική ανάλυση, τα αλγεβρικά συστήματα και τους πίνακες. Επίσης να γνωρίζει τι είναι τα νευρωνικά δίκτυα και πως μπορούν να εκπαιδευτούν και να γίνουν προσαρμοστικά και τη θεωρία της ασαφούς λογικής. Έτσι, μπορούμε να πούμε ότι αυτή η πολυπλοκότητα του δίνει τη δύναμη να μπορεί να προσαρμόζεται καλύτερα στις διάφορες απαιτήσεις των όποιων προβλημάτων. Δίνει έτσι μικρότερα σφάλματα, άρα και καλύτερα αποτελέσματα.

Η εκπαίδευση ενός τέτοιου συστήματος μπορεί να γίνει εύκολα, καθώς μπορούμε να αντιληφθούμε άμεσα τυχόν προβλήματα μέσα από τους κανόνες και τα γραφήματα που εξάγουμε. Αντίθετα, ενώ η πολυπλοκότητα του σημαίνει και καλύτερες δυνατότητες προσαρμογής, αυτόματα σημαίνει και μεγαλύτερες ανάγκες για επεξεργαστική ισχύ. δηλ. χρειάζεται περισσότερο χρόνο για να ολοκληρώσει την εξαγωγή αποτελεσμάτων σε σχέση με ένα απλούστερο παραδοσιακό μοντέλο καθώς επίσης απαιτεί μεγαλύτερες επενδύσεις σε τεχνολογικό εξοπλισμό.

Ένα ακόμα μειονέκτημα αποτελεί και το γεγονός ότι χρειαζόμαστε πολλά δεδομένα για την εκπαίδευση του συστήματος, σε αντίθετη περίπτωση δηλ. εάν τα δεδομένα είναι λίγα παραμονεύει ο κίνδυνος της μη γενίκευσης του μοντέλου. Παρόλα αυτά, τα νευρωνικά-δίκτυα θα είναι ένα επιπλέον αξιόλογο μοντέλο για την πρόβλεψη τιμών πολύτιμων μετάλλων.

Συμπερασματικά, όπως γίνεται εύκολα αντιληπτό, από τα διαγράμματα και τους συγκριτικούς πίνακες στο προηγούμενο κεφάλαιο, τα αποτελέσματα του ANFIS είναι πάρα πολύ κοντά με τις πραγματικές τιμές και συγκριτικά καλύτερα από άλλα μοντέλα όπως το AR και το ARMA. Το ίδιο βέβαια παρατηρούμε και με τα αποτελέσματα του νευρωνικού δικτύου στη συνέχεια. Στον συγκριτικό πίνακα που βλέπουμε στο τέλος του προηγούμενου κεφαλαίου το ANFIS εμφανίζεται καλύτερο από το νευρωνικό δίκτυο στη σύγκριση μεταξύ τους.

Θα μπορούσαμε μελλοντικά να κάνουμε και άλλες συγκρίσεις εκτός των σφαλμάτων με αυτά τα μοντέλα:

- Το ποσοστό επιτυχίας → υποδηλώνει τον αριθμό των σωστών προβλέψεων του αποθέματος προς τον αριθμό των δοκιμών.
- Σύγκριση κατά την αγορά και τη ζήτηση → ο επενδυτής επενδύει ένα χρηματικό ποσό και κατέχει την επένδυση μέχρι το τέλος του χρονικού ορίζοντα προσομοίωσης. Χρησιμοποιώντας το προτεινόμενο σύστημα, ο επενδυτής διαθέτει περιουσιακά στοιχεία στο χρηματιστήριο.
- Αξιολόγηση από τη σύγκριση με 13 άλλα μοντέλα → Μια συγκριτική μελέτη για την αξιολόγηση με ποσοστά πρόβλεψης για την ορθότητα των τάσεων των τιμών των μετοχών. Η μελέτη συγκρίνει νευρο-ασαφή συστήματα (παρόμοια ή και πολύ σχετικά). Ωστόσο, αυτή η μελέτη στοχεύει στην ανάδειξη πρόβλεψης στα επίπεδα ακρίβειας των διαφόρων τεχνικών, και ιδίως το υψηλότερο επίπεδο.

ΚΕΦΑΛΑΙΟ 6: Βιβλιογραφία

Abraham A., (2005). Adaptation of Fuzzy Inference System Using neural Learning, Fuzzy System Engineering: Theory and Practice, Nadia Nedjah et al. (Eds), Studies in Fuzziness and Soft Computing, Springer Verlag Germany, pp. 53-83.

Andersson, E. (2005). On-line detection of turning points using non-parametric surveillance: The effect of the growth after the turn. Statistics & Probability Letters, 73, 433–439.

Andersson, E., (2002). Monitoring cyclical processes. A non-parametric approach. Journal of Applied Statistics, Vol. 29, No. 7, 973-990.

Atsalakis, G., and Ucenic, C., (2005, a). Time Series Prediction of Water Consumption Using Neuro-fuzzy (ANFIS) Approach. International Conference on Water Economics, Statistics and Finance, Greece.

Atsalakis, G., May (2006 b). Neuro-fuzzy and Technical Analysis for stock prediction. Working paper, Technical University of Crete.

Atsalakis, G., Skiadas C. and Braimis I., (2006,f). Probability of trend prediction of exchange rate by neuro-fuzzy techniques, International Symposium on Applied Stochastic Models and Data Analysis, USA.

Atsalakis, G., Ucenic, C. and Skiadas C., (2005, d). Time series prediction of the Greek manufacturing index for the non-metallic minerals sector using a Neuro-fuzzy approach (ANFIS), International Symposium on Applied Stochastic Models and Data Analysis, France, Brest.

Atsalakis, G., Ucenic, C., (2005, c). Forecasting the profit for the Greek food sector using a neuro-fuzzy approach, International Conference on NHIBE, p.p.686-694, Corfu, Greece.

Atsalakis, G., Ucenic, C., (2006, c). Electric load forecasting by Neuro-fuzzy approach. WSEAS International Conference on Energy and Environmental Systems, Chalkida Evia, Greece.

Atsalakis, G., Ucenic, C., (2006, d). Forecasting the Electricity Demand Using a Neuro-fuzzy Approach Versus Traditions Methods. Journal of WSEAS Transactions on Business and Economics, issue 1, vol. 3, pp. 9-17.

Atsalakis, G., Ucenic, C., (2006, e). Forecasting the wind energy production using a Neuro-fuzzy model. Journal of WSEAS Transactions on Environment and Development, issue 6, vol. 2, pp. 823-829.

Atsalakis, G., Ucenic, C., Plokamakis G., (2005, b). Forecasting of Electricity Demand Using Neuro-fuzzy (ANFIS) Approach, International Conference on NHIBE, Corfu, Greece.

Atsalakis, G., Valavanis K., September (2006, a). A Neuro-Fuzzy controller for Stock market forecasting. Working paper August, Technical University of Crete.

Avagaddi Prasad, Durga Prasad B, Ravindra Babu M, Manohar K and Siva Mallikarjuna Rao N, "A Wavelet -Based Adaptive Neuro Fuzzy Inference System Approach To Fault Classification For High Speed Protective Relaying", International Journal of Recent Advances in Engineering & Technology (IJRAET), 2347-2812, Volume-1, Issue -1, 2013

Balwant Kumar and Dr. Dharmender Kumar, "A review on Artificial Bee Colony algorithm", International Journal of Engineering and Technology, 2(3) 175-186, 2013.

Bengoechea, P., Camacho, M. Perez-Quiros, G. (2006). A useful tool for forecasting the Euro-area business cycle phases. International Journal of Forecasting.

Canova, F., Ciccarelli, M. (2004). Forecasting and turning point predictions in a Bayesian panel VAR model. Journal of Econometrics, 120, 327 – 359.

Chen-Sen Ouyang, Wan-Jui Lee, and Shie-Jue Lee, "A TSK-Type Neurofuzzy Network Approach to System Modeling Problems", IEEE transactions on systems, man, and cybernetics—part B: cybernetics, vol. 35, No. 4, 2005.

Dervis Karaboga, and Seluk kdem, "A Simple and Global Optimization Algorithm for Engineering Problems: Differential Evolution Algorithm", Turkish Journal of Electrical Engineering & Computer

Garcia-Ferrer, A., Queralt, R. A. (1998). Can univariate models forecast turning points in seasonal economic time series? International Journal of Forecasting, 14, 433–446.

Harding, D., Pagan, A. (2006). Synchronization of cycles. Journal of Econometrics, 132, 59–79.

Harvey, A. C., Trimbur, T. M., Van Dijk, H. K.. Trends and cycles in economic time series: A Bayesian approach. Journal of Econometrics.

Holland, J. H. (1975). Adaptation in natural and artificial systems. University of Michigan Press, Michigan.

Ismail Z, Yahya A and Shabri A, "Forecasting Gold Prices Using Multiple Linear Regression Method", American Journal of Applied Sciences 6: 1509 -1514, 2009

Jang, J.-S.R, C.-T. Sun, (1995). Neuro-fuzzy Modeling and Control, Proceedings of the IEEE, Vol. 83, No. 3, pp. 378-406. 91

Jang, J-S.R., C-T.E. Sun and E. Mizutani, (1997). Neuro-fuzzy and soft computing : a computational approach to learning and machine intelligence, Prentice Hall.

Jinn-Tsong Tsai, Jyh-Horng Chou and Wen-Hsien Ho, "Improved Quantum-Inspired Evolutionary Algorithm for Engineering Design Optimization", Mathematical Problems in Engineering, 2012.

Kholodilin, K. A., Yao V. W. (2005). Measuring and predicting turning points using a dynamic bi-factor model. International Journal of Forecasting, 21, 525– 537.

Luong Trung Tuan and Suet Peng Yong, "A Fuzzy Neural Based Data Classification System", DMIN : 383 -388, 2006.

Makridakis, S. & Hibon, M. (1979). Accuracy of forecasting: An empirical investigation. *Journal of the Royal Statistical Society, A*, 142, 97-145.

McCulloch, W. S. & Pitts, W. (1943). A logical calculus of the ideas imminent in neural nets. *Bulletin of Mathematical Biophysics*, 5, 115-137.

Minsky, M. & Papert, S. (1969). *Perceptrons*. MIT Press, Cambridge, M.A.

Omar Kettani, Faycal Ramdani and Benaissa Tadili, "A Quantum Differential Evolutionary Algorithm for the Independent Set Problem", *International Journal of Computer Applications* (0975 –8887) , Volume 58–No.14, 2012

Osman Hegazy, Omar S. Soliman and Ahmed A. Toony, "Hybrid of neuro -fuzzy inference system and quantum genetic algorithm for prediction in stock market", *Issues in Business Management and Economics (IBME)* Vol.2 (6), pp. 094-102, June 2014

Peilin Liu, Wenhao Leng and Wei Fang, "Training ANFIS Model with an Improved Quantum -Behaved Particle Swarm Optimization Algorithm", *Mathematical Problems in Engineering*, 2013.

Qinghua Wen, Zehong Yang and Yixu Song, "Hybrid Approaches for Stock Price Prediction", *New Trends in Information Science and Service Science (NISS)*, 4th International Conference, 2010

Rosenblatt, F. (1962). Principles of neurodynamics: perceptrons and the theory of brain mechanisms. Spartan, New York. *Sciences*, Vol. 12 Issue 1, p53, 2004.

Stephen L. Chiu, "Fuzzy Model Identification Based on Cluster Estimation", *Journal of Intelligent & Fuzzy Systems*, Vol. 2, No. 3, 1994

Toha S F and Tokhi M O, "ANFIS modelling of a twin rotor system using particle swarm optimisation and RLS", *Cybernetic Intelligent Systems (CIS)*. IEEE 9th International Conference, 2010.

Ucenic, C. and Atsalakis, G., (2005, a). Time series prediction of the Greek manufacturing index for the basic metals sector using a Neuro-fuzzy approach (ANFIS). Management of Technological changes, Proceeding of 4th International Conference on the Management of Technological Changes, TUC, Chania, Greece.

Ucenic, C. and Atsalakis, G., (2005, b). Forecasting the profit for the Greek non-metallic sector using a Neuro-fuzzy approach (ANFIS). Management of Technological changes, Proceeding of 4th International Conference on the Management of Technological Changes, Technical University Crete, Chania, Greece.

Ucenic, C., and Atsalakis, G., (2003, b). Trade integration of goods and some technical sources of growth. Management of Technological changes, Proceeding of 3rd International Conference on the Management of Technological Changes, Technical University Crete, Chania, Greece.

Ucenic, C., and G. Atsalakis, (2003, a). Study the influences of the economic environment on the gross domestic product using a fuzzy model. Management of Technological changes, Proceeding of 3rd International Conference on the Management of Technological Changes, TUC, Chania, Greece.

Van Nieuwerburgh, S., Veldkamp, L. (2006). Learning asymmetries in real business cycles. *Journal of Monetary Economics*, 53, 753–772.

Vasan A and Srinivasa Raju K, "Optimal Reservoir Operation using Differential Evolution", Proceedings of International Conference on Hydraulic Engineering: Research and Practice (ICON-HERP 2004), Indian Institute of Technology, Roorkee, October 26-28, 2004.

Ευχαριστίες

Καταρχάς, θα ήθελα να ευχαριστήσω τον καθηγητή κ. Γιώργο Ατσαλάκη με την πολύτιμη καθοδήγηση του οποίου κατάφερα να ολοκληρώσω με επιτυχία την παρούσα εργασία. Με την ολοκλήρωση της παρούσας διπλωματικής εργασίας μου, ολοκληρώνεται μια πολύ σημαντική περίοδος της μέχρι τώρα ζωής μου. Ένα μεγάλο ευχαριστώ στους ανθρώπους εκείνους που με στήριξαν και με ώθησαν να φτάσω ως εδώ. Ένα πολύ μεγάλο ευχαριστώ στην οικογένεια μου και τους φίλους μου για την αγάπη τους και τη συμπαράσταση τους.

Σας ευχαριστώ...