

ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΑΡΑΓΩΓΗΣ
& ΔΙΟΙΚΗΣΗΣ



ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ ΜΕ ΤΙΤΛΟ:
ΠΡΟΒΛΕΨΗ ΙΣΟΤΙΜΙΑΣ ΕΥΡΩ/ΔΟΛΛΑΡΙΟΥ
ΜΕ ΧΡΗΣΗ ΝΕΥΡΟ-ΑΣΑΦΩΝ ΤΕΧΝΙΚΩΝ.

Επιβλέπων Καθηγητής: Χρήστος Σκιαδάς
Επιμέλεια: Ηλίας Μπραΐμης

XANIA 2005

ΕΥΧΑΡΙΣΤΙΕΣ

Στο σημείο αυτό θα ήθελα να ευχαριστήσω όλους εκείνους που με βοήθησαν και μου συμπαραστάθηκαν στην προσπάθειά μου να ολοκληρώσω την παρούσα διπλωματική εργασία.

Αρχικά, θα ήθελα να ευχαριστήσω τον κ. Σκιαδά Χρήστο που μου έδωσε την ευκαιρία να ασχοληθώ με ένα τόσο ενδιαφέρον θέμα.

Επίσης, θα ήθελα να ευχαριστήσω τον κ. Ατσαλάκη Γεώργιο, ΕΕΔΙΠ ΙΙ του Πολυτεχνείου Κρήτης, χωρίς την σωστή καθοδήγηση και βοήθεια του οποίου θα ήταν αδύνατο να ολοκληρώσω την παρούσα διπλωματική εργασία.

Θα ήθελα ακόμα να ευχαριστήσω τους φίλους και συναδέλφους μου: Κατσαμάνη Γιώργο, Λεντζάκη Αντώνη, Πρωτονοτάριο Βασίλη, Στρατουδάκη Μάνο και Φούσια Χρήστο για τη στήριξη που μου παρείχαν όλα αυτά τα χρόνια.

Τέλος, το μεγαλύτερο ευχαριστώ το οφείλω στους γονείς μου για όλα όσα μου έχουν προσφέρει μέχρι σήμερα.

Μπραΐμης Ηλίας
Σεπτέμβριος 2005

ΠΕΡΙΕΧΟΜΕΝΑ

ΠΕΡΙΛΗΨΗ.....	4
ΚΕΦΑΛΑΙΟ 1: Περιγραφή των μοντέλων.....	6
1.1 Εισαγωγή	6
1.2 Μοντέλο AR	7
1.3 Μοντέλο ARMA.....	9
1.4 ANFIS	10
ΚΕΦΑΛΑΙΟ 2: Ανάλυση του μοντέλου ANFIS.....	15
2.1 Εισαγωγή	15
2.2 Περιγραφή του μοντέλου.....	15
2.3 Αρχιτεκτονική του ANFIS.....	19
ΚΕΦΑΛΑΙΟ 3: Μέθοδοι για προσδιορισμό συστήματος	30
3.1 Προσδιορισμός συστήματος: Μια εισαγωγή	30
3.2 Εκτιμητής ελαχίστων τετραγώνων (Least-squares estimator).....	32
3.3 Οπισθοδρόμηση για πρωσοτροφοδοτούμενα δίκτυα.....	37
3.4 Υβριδικός αλγόριθμος εκμάθησης: Συνδυασμός βαθμωτής ελαχιστοποίησης και εκτιμητή ελαχίστων τετραγώνων.....	46
ΚΕΦΑΛΑΙΟ 4: Αποτελέσματα από τη χρήση του μοντέλου ANFIS.....	51
4.1 Εισαγωγή	51
4.2 Περιγραφή του μοντέλου	51
4.3 Αποτελέσματα.....	54
4.4 Συμπεράσματα	61
ΒΙΒΛΙΟΓΡΑΦΙΑ	62

ΠΕΡΙΛΗΨΗ

Κατά τη διάρκεια των τελευταίων δεκαετιών έχουν χρησιμοποιηθεί πολλές μέθοδοι προκειμένου να γίνει δυνατή η πρόβλεψη της συναλλαγματικής ισοτιμίας νομισμάτων. Όμως τα αποτελέσματα που προκύπτουν από τη χρήση των μεθόδων αυτών δεν είναι ικανοποιητικά. Λόγω της υψηλής αστάθειας, της πολυπλοκότητας και του θορυβώδους περιβάλλοντος αγοράς όλο και περισσότεροι ερευνητές στράφηκαν στη χρήση των τεχνικών νευρωνικών δικτύων για να επιτύχουν καλύτερα αποτελέσματα.

Σκοπός της παρούσης διπλωματικής εργασίας είναι να δείξει ότι η χρήση των νευρωνικών δικτύων για την πρόβλεψη της συναλλαγματικής ισοτιμίας νομισμάτων είναι η καλύτερη επιλογή. Για το λόγο αυτό κατασκευάστηκε ένα μοντέλο ANFIS, του οποίου τα αποτελέσματα συγκρίνονται με τα αποτελέσματα από τη χρήση άλλων συμβατικών μεθόδων.

Η παρούσα διπλωματική εργασία περιλαμβάνει, ανά κεφάλαιο, τα εξής:

Στο **πρώτο** κεφάλαιο γίνεται μια σύντομη ιστορική αναδρομή και περιγράφονται τα μοντέλα AR και ARMA με τα αποτελέσματα των οποίων θα γίνει η σύγκριση των αποτελεσμάτων του μοντέλου μας. Επίσης, γίνεται μια γενική περιγραφή του μοντέλου ANFIS.

Στο **δεύτερο** κεφάλαιο γίνεται μια λεπτομερής ανάλυση του μοντέλου ANFIS και παρουσιάζεται ένα αριθμητικό παράδειγμα για την καλύτερη κατανόηση του τρόπου λειτουργίας του.

Στο **τρίτο** κεφάλαιο γίνεται παρουσίαση και ανάλυση των μεθόδων που χρησιμοποιεί το ANFIS για να προσδιορίσει τις παραμέτρους ενός άγνωστου συστήματος από τα δεδομένα εισόδου-εξόδου.

Τέλος, στο **τέταρτο** κεφάλαιο γίνεται η παρουσίαση των αποτελεσμάτων του μοντέλου ANFIS και η σύγκρισή τους με τα αποτελέσματα των μοντέλων AR και ARMA καθώς επίσης και τα συμπεράσματα που προκύπτουν.

ΚΕΦΑΛΑΙΟ 1: Περιγραφή των μοντέλων

1.1 Εισαγωγή

Κατά τη διάρκεια των τελευταίων δεκαετιών πολλές μέθοδοι χρησιμοποιήθηκαν προκειμένου να προβλεφθεί η συναλλαγματική ισοτιμία νομισμάτων. Στη δεκαετία του '60, οι δημοφιλέστερες μέθοδοι για τη βραχυπρόθεσμη πρόβλεψη στις αγορές συναλλάγματος ήταν ευρετικές μέθοδοι, όπως του κινούμενου μέσου όρου (Brown, 1963) της εκθετικής εξομάλυνσης και της προσαρμοστικής εκθετικής εξομάλυνσης (Triggs, 1967). Εντούτοις, αυτές οι μέθοδοι δεν κατείχαν σταθερά υπόβαθρα στη στατιστική θεωρία. Στη δεκαετία του '70, πολλοί ερευνητές άρχισαν να χρησιμοποιούν τις μεθόδους πρόβλεψης ARIMA (Box-Jenkins, 1970) επειδή ήταν στατιστικά πιο ελκυστικές και χρησιμοποιήθηκαν ευρέως από την αρχή της δεκαετίας του '80.

Αργότερα, στα μέσα της δεκαετίας του '80, η αστάθεια των συναλλαγματικών ισοτιμιών προκάλεσε μεγάλη ανησυχία στις αγορές συναλλάγματος. Προκειμένου να προβλεφθεί η βραχυπρόθεσμη αστάθεια, η αυτοπαλινδρόμηση υποθετική ετεροσκεδαστικότητα (Autoregressive Conditional Heteroskedasticity, ARCH) προτάθηκε από τον Engle (1982) και έγινε πολύ δημοφιλής προς το τέλος της δεκαετίας του '80 και στις αρχές της δεκαετίας του '90. Μια άλλη μέθοδος που έγινε πολύ δημοφιλής στη βραχυπρόθεσμη πρόβλεψη στις αγορές συναλλάγματος είναι το γενικό μοντέλο ARCH (GARCH), το οποίο είναι γραμμικό στο μέσο όρο και μη γραμμικό στη διασπορά και έχει χρησιμοποιηθεί από πολλούς ερευνητές προκειμένου να συλληφθεί η ετεροσκεδαστικότητα των καθημερινών συναλλαγματικών ισοτιμιών. Μια άλλη μέθοδος πρόβλεψης που χρησιμοποιείται στην πρόβλεψη των συναλλαγματικών ισοτιμιών είναι το διάλυμα αυτοπαλινδρόμησης (VAR). Τέλος, οι Chien και Leung (2003) ανέπτυξαν ένα μοντέλο διόρθωσης σφάλματος Μπαγιεσιανού διανύσματος (Bayesian vector error correction model) για την πρόβλεψη των συναλλαγματικών ισοτιμιών.

Εντούτοις, οι παραδοσιακές στατιστικές τεχνικές για την πρόβλεψη συναλλαγματικών ισοτιμιών νομισμάτων δεν έχουν ικανοποιητικά αποτελέσματα. Οι Yao και Tan (2000) υποστήριξαν ότι η κλασσική ανάλυση χρονοσειρών, που βασίζεται στη θεωρία των στάσιμων στοχαστικών διαδικασιών, δεν αποδίδει ικανοποιητικά στις οικονομικές χρονοσειρές. Τα οικονομικά δεδομένα δεν είναι απλές διαδικασίες ARIMA, δεν μπορούν να περιγραφούν από απλές γραμμικές δομές και δεν είναι απλός λευκός θόρυβος ή τυχαίοι περίπατοι. Λόγω της υψηλής αστάθειας, της πολυπλοκότητας και του θορυβώδους περιβάλλοντος αγοράς η χρήση των τεχνικών νευρωνικών δικτύων για λόγους πρόβλεψης θεωρήθηκε η καλύτερη επιλογή. Η τεχνολογία νευρωνικών δικτύων χρησιμοποιείται αυτήν την περίοδο σε πολλούς τομείς εφαρμογής στις επιχειρήσεις, ειδικά όταν το πρόβλημα περιλαμβάνει κατά ένα μεγάλο μέρος ταξινόμηση, αναγνώριση και πρόβλεψη. Μερικοί ερευνητές έχουν παρουσιάσει σημαντικές μελέτες σχετικά με τη πρόβλεψη συναλλαγματικών ισοτιμιών νομισμάτων χρησιμοποιώντας τα νευρωνικά δίκτυα (El Shazly & El Shazly, 1997, Lisi & Schiavo, 1999, Leung et al, 2000, Qi & Wu, 2003, El Shazly & El Shazly, 1999, Chen & Leung, 2004, Poddig & Rehkugler, 1996).

Στη συνέχεια του κεφαλαίου θα κάνουμε μια περιγραφή των μοντέλων AR και ARMA που χρησιμοποιούνται παρακάτω καθώς επίσης και μια γενική περιγραφή του μοντέλου ANFIS.

1.2 Μοντέλο AR

Ένα στοχαστικό μοντέλο που μπορεί να είναι εξαιρετικά χρήσιμο στην αναπαράσταση ορισμένων πρακτικά εμφανιζόμενων σειρών είναι το αποκαλούμενο αυτοπαλινδρόμενο μοντέλο. Στο μοντέλο αυτό, η τρέχουσα τιμή της διαδικασίας εκφράζεται σαν ένα πεπερασμένο, γραμμικό άθροισμα των προηγούμενων τιμών της διαδικασίας και μια διαταραχή a_t . Ας θεωρήσουμε τις τιμές της διαδικασίας σε ίσα χρονικά διαστήματα

$t, t-1, t-2, \dots$ ως $z_t, z_{t-1}, z_{t-2}, \dots$. Επίσης έστω ότι $\tilde{z}_t, \tilde{z}_{t-1}, \tilde{z}_{t-2}, \dots$ είναι αποκλίσεις από το μ (για παράδειγμα $\tilde{z}_t = z_t - \mu$). Τότε το

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \phi_2 \tilde{z}_{t-2} + \dots + \phi_p \tilde{z}_{t-p} + \alpha_t \quad (1.2.1)$$

ονομάζεται αυτοπαλινδρόμενη (AR) διαδικασία τάξης p . Ο λόγος για το όνομα αυτό είναι ότι ένα γραμμικό μοντέλο

$$\tilde{z} = \phi_1 \tilde{x}_1 + \phi_2 \tilde{x}_2 + \dots + \phi_p \tilde{x}_p + a$$

που σχετίζει μια εξαρτημένη μεταβλητή z με ένα σετ ανεξάρτητων μεταβλητών x_1, x_2, \dots, x_p , συν ένα σφάλμα a , αναφέρεται συχνά σαν ένα παλινδρόμενο μοντέλο, και το z λέγεται ότι παλινδρομείται στα x_1, x_2, \dots, x_p . Στην (1.2.1) η μεταβλητή z παλινδρομείται σε προηγούμενες τιμές του εαυτού της, συνεπώς το μοντέλο είναι αυτοπαλινδρόμενο. Αν ορίσουμε ένα αυτόπαλινδρόμενο τελεστή τάξης p ως

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$$

τότε το αυτοπαλινδρόμενο μοντέλο μπορεί να γραφτεί οικονομικά ως

$$\phi(B)\tilde{z}_t = \alpha_t$$

Το μοντέλο περιλαμβάνει $p+2$ άγνωστες παραμέτρους $\mu, \phi_1, \phi_2, \dots, \phi_p, \sigma_\alpha^2$ που στην πράξη πρέπει να υπολογιστούν από τα δεδομένα. Η επιπλέον παράμετρος σ_α^2 είναι η διασπορά της διαδικασίας λευκού θορύβου α_t .

Δεν είναι δύσκολο να δούμε ότι το αυτοπαλινδρόμενο μοντέλο είναι μια ειδική περίπτωση του γραμμικού μοντέλου φίλτρου (linear filter model) του (1.2.1). Για

παράδειγμα μπορούμε να απαλείψουμε τον \tilde{z}_{t-1} από το δεξιό μέρος της (1.2.1) μετασχηματίζοντάς τη ως εξής

$$\tilde{z}_{t-1} = \phi_1 \tilde{z}_{t-2} + \phi_2 \tilde{z}_{t-3} + \dots + \phi_p \tilde{z}_{t-p-1} + a_{t-1}$$

Ομοίως μπορούμε να μετασχηματίσουμε για τον \tilde{z}_{t-2} κλπ.

Συμβολικά έχουμε ότι το

$$\phi(B)\tilde{z}_t = a_t$$

είναι ισοδύναμο με το

$$\tilde{z}_t = \psi(B)a_t$$

με

$$\psi(B) = \phi^{-1}(B)$$

Οι αυτοπαλινδρόμενες διαδικασίες μπορεί να είναι στάσιμες και μη στάσιμες. Για να είναι στάσιμες πρέπει τα ϕ να επιλέγονται έτσι ώστε τα βάρη ψ_1, ψ_2, \dots στην $\psi(B) = \phi^{-1}(B)$ να σχηματίζουν συγκλίνουσα σειρά.

1.3 Μοντέλο ARMA

Για να επιτύχουμε μεγαλύτερη ευκαμψία στην προσαρμογή κανονικών χρονοσειρών, είναι μερικές φορές καλύτερο να χρησιμοποιούμε όρους και αυτοπαλινδρόμενους και κινούμενου μέσου στο μοντέλο. Αυτό οδηγεί στο μοντέλο ARMA (mixed autoregressive-moving average)

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \phi_2 \tilde{z}_{t-2} + \dots + \phi_p \tilde{z}_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad (1.3.1)$$

ή

$$\phi(B)\tilde{z}_t = \theta(B)\alpha_t$$

που έχει $p + q + 2$ άγνωστες παραμέτρους $\mu; \phi_1, \dots, \phi_p; \theta_1, \dots, \theta_q; \sigma_\alpha^2$, οι οποίες υπολογίζονται από τα δεδομένα.

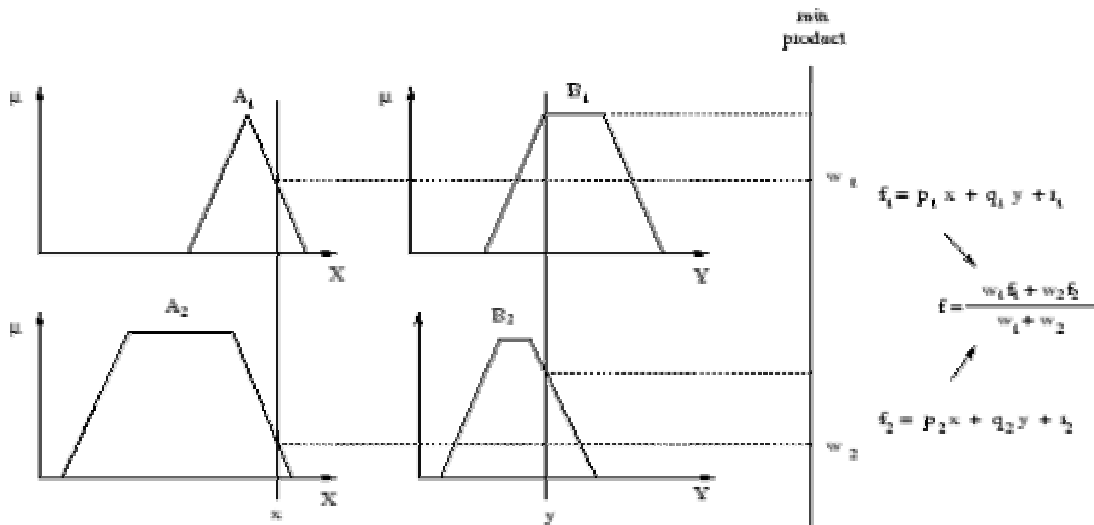
Στην πράξη είναι συνήθως αλήθεια ότι επαρκής αναπαράσταση των πραγματικά εμφανιζόμενων στάσιμων χρονοσειρών μπορεί να παρθεί με αυτοπαλινδρόμενα, κινούμενου μέσου, ή μεικτών μοντέλων, στα οποία τα p και q δεν είναι μεγαλύτερα από 2 και συχνά είναι μικρότερα από 2.

1.4 ANFIS

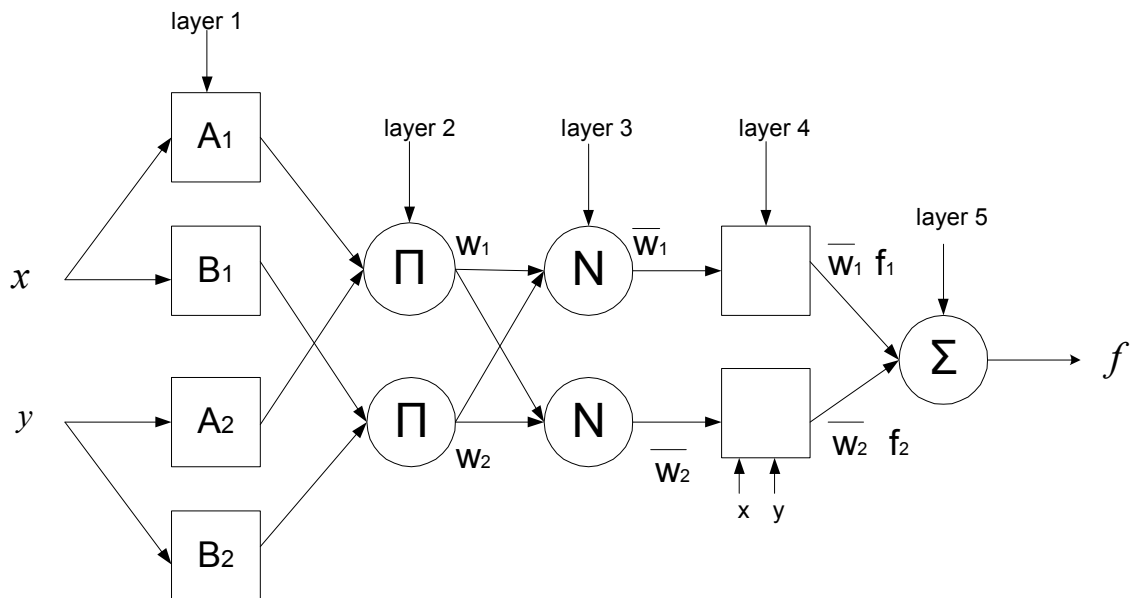
Ένα νευρο-ασαφές σύστημα ορίζεται ως ένας συνδυασμός τεχνητών νευρωνικών δικτύων (ANN) και Ασαφούς Συστήματος Συμπερασμού (Fuzzy Inference System, FIS) κατά τέτοιο τρόπο ώστε ο αλγόριθμος εκμάθησης νευρωνικών δικτύων χρησιμοποιείται για να καθορίσει τις παραμέτρους του FIS [1]. Το Προσαρμοστικό Νευροασαφές 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, προτείνουμε μια κατηγορία προσαρμόσιμων δικτύων, τα οποία είναι λειτουργικά ισοδύναμα με τα ασαφή συστήματα συμπερασμού.

Ο ασαφής μηχανισμός συμπερασμού παρουσιάζεται στο επόμενο σχήμα:



Σχήμα 1: Ασαφής μηχανισμός συμπερασμού



Σχήμα 2: Αρχιτεκτονική του ANFIS

Για ευκολία, υποθέτουμε ότι το ασαφές σύστημα συμπερασμού υπό εξέταση έχει δύο εισόδους x και y και μία έξοδο z . Υποθέτουμε 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 στο σχήμα 2. Οι συναρτήσεις κόμβων στο ίδιο επίπεδο είναι της ίδιας οικογένειας συναρτήσεων όπως περιγράφουμε παρακάτω:

Επίπεδο 1 Κάθε κόμβος i σε αυτό το επίπεδο είναι ένας προσαρμόσιμος κόμβος με μία συνάρτηση κόμβου.

$$O_i^1(x) = \mu_{A_i}(x)$$

όπου x – η είσοδος στον κόμβο i

A_i - η γλωσσική ταμπέλα (small, large, κλπ.) που σχετίζεται με αυτή τη συνάρτηση του κόμβου.

Με άλλα λόγια, το O_i^1 είναι ο βαθμός συμμετοχής του A_i και καθορίζει το βαθμό στον οποίο η είσοδος x ικανοποιεί τον ποσοτικοποιητή A_i . Συνήθως επιλέγεται το $\mu_{A_i}(x)$ σε σχήμα καμπάνας με το μέγιστο ίσο με 1 και το ελάχιστο ίσο με 0, όπως η γενικευμένη καμπανοειδής συνάρτηση

$$\mu_{A_i}(x) = \frac{1}{1 + \left[\left(\frac{x - c_i}{a_i} \right)^2 \right]^{b_i}}$$

ή η γκαουσιανή συνάρτηση

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

όπου a_i, b_i, c_i είναι το σύνολο παραμέτρων.

Καθώς οι τιμές αυτών των παραμέτρων αλλάζουν, οι συναρτήσεις ποικίλλουν ανάλογα, παρουσιάζοντας έτσι διάφορες μορφές της συνάρτησης συμμετοχής για τη γλωσσική ταμπέλα A_i . 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:

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

$$\bar{w}_i = \frac{w_i}{w_1 + w_2}, \quad i=1,2$$

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.

$$O_{4,i} = \bar{w}_i * f_i = \bar{w}_i * (p_1 * x + q_i * y + r_i)$$

όπου: \bar{w}_i - η έξοδος του επιπέδου 3

$\{p_i, q_i, r_i\}$ - το σύνολο 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:

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

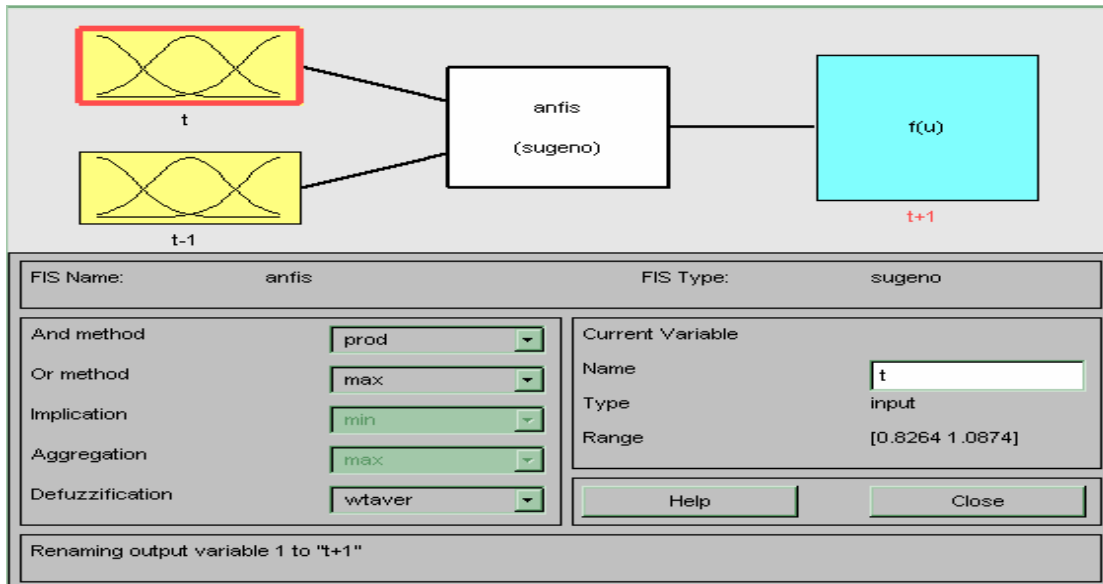
ΚΕΦΑΛΑΙΟ 2: Ανάλυση του μοντέλου ANFIS

2.1 Εισαγωγή

Στο κεφάλαιο αυτό θα ασχοληθούμε με την ανάλυση του μοντέλου ANFIS που χρησιμοποιήσαμε στην έρευνά μας. Αρχικά θα κάνουμε μια περιγραφή του μοντέλου και θα παρουσιάσουμε τη δομή του. Στη συνέχεια θα εξετάσουμε κάθε επίπεδο του μοντέλου ξεχωριστά και θα παρουσιάσουμε τι είδους υπολογισμοί γίνονται σε κάθε επίπεδο. Τέλος, θα παρουσιάσουμε ένα αριθμητικό παράδειγμα όπου θα φαίνονται αναλυτικά όλοι οι υπολογισμοί που πραγματοποιεί το ANFIS σε κάθε επίπεδο, έτσι ώστε να γίνει πλήρως κατανοητή η λειτουργία του μοντέλου.

2.2 Περιγραφή του μοντέλου

Το μοντέλο μας αποτελείται από δύο εισόδους και μια έξοδο. Σαν εισόδους παίρνουμε την ισοτιμία ευρώ/δολαρίου μιας τυχαίας ημέρας (t) και την ισοτιμία της ακριβώς προηγούμενης ημέρας ($t-1$), ενώ σαν έξοδο παίρνουμε την πρόβλεψη του ANFIS για την ισοτιμία ευρώ/δολαρίου της επόμενης από την τυχαία ημέρα ($t+1$). Η δομή του μοντέλου παρουσιάζεται στο ακόλουθο σχήμα:



Σχήμα 1: Δομή του μοντέλου

Από το σχήμα βλέπουμε ότι το ANFIS είναι τύπου Sugeno. Το ασαφές μοντέλο τύπου Sugeno προτάθηκε από τους Tagaki, Sugeno και Kang σε μια προσπάθεια να αναπτύξουν μια συστηματική προσέγγιση που θα τους επιτρέψει την παραγωγή ασαφών κανόνων από ένα συγκεκριμένο σύνολο δεδομένων εισόδου-εξόδου. Ένας τυπικός ασαφής κανόνας για ένα ασαφές μοντέλο τύπου Sugeno έχει την ακόλουθη μορφή:

$$\text{αν } x \text{ είναι } A \text{ και } y \text{ είναι } B \text{ τότε } z = f(x,y)$$

όπου τα A και B είναι ασαφή σύνολα στα αίτια (antecedent), ενώ το $z = f(x,y)$ είναι μια σαφή (crisp) συνάρτηση στο συμπέρασμα (consequent). Συνήθως το $f(x,y)$ είναι ένα πολυώνυμο των μεταβλητών εισόδου x και y , αλλά μπορεί να είναι οποιαδήποτε συναρτηση αρκεί να μπορεί να περιγράψει κατάλληλα την έξοδο του μοντέλου εντός της ασαφούς περιοχής που καθορίζεται από τα αίτια του κανόνα. Εμείς έχουμε ένα πρώτου βαθμού ασαφές μοντέλο τύπου Sugeno το οποίο σημαίνει ότι το $f(x,y)$ είναι ένα πρώτου βαθμού πολυώνυμο. Στο μοντέλο μας ισχύουν οι παρακάτω κανόνες:

Αν (t) είναι small και (t-1) είναι small τότε (t+1) είναι very small

Αν (t) είναι small και (t-1) είναι large τότε (t+1) είναι small

Αν (t) είναι large και (t-1) είναι small τότε (t+1) είναι large

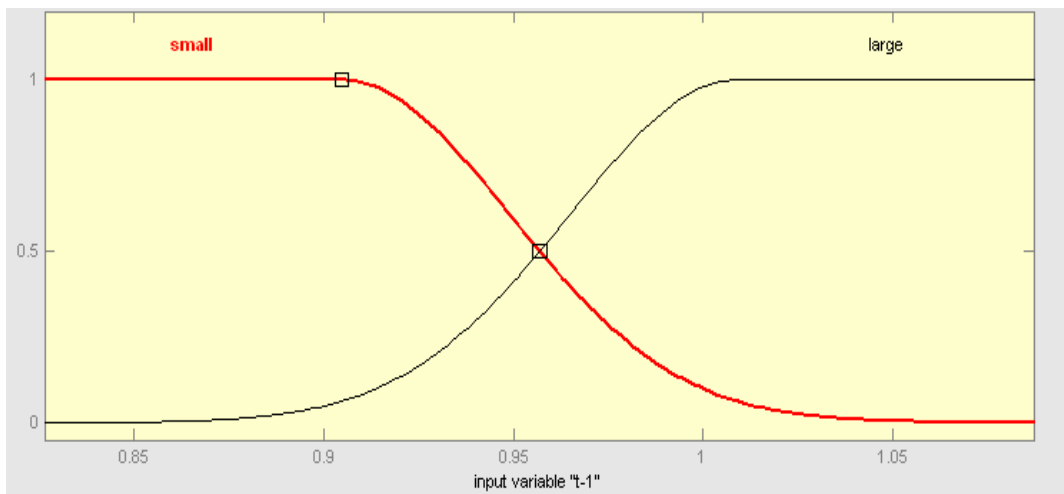
Αν (t) είναι large και (t-1) είναι large τότε (t+1) είναι very large

Θα δούμε πως χρησιμοποιεί το μοντέλο τους κανόνες και πως γίνονται οι υπολογισμοί στην ενότητα που θα αναφερθούμε στην αρχιτεκτονική του ANFIS.

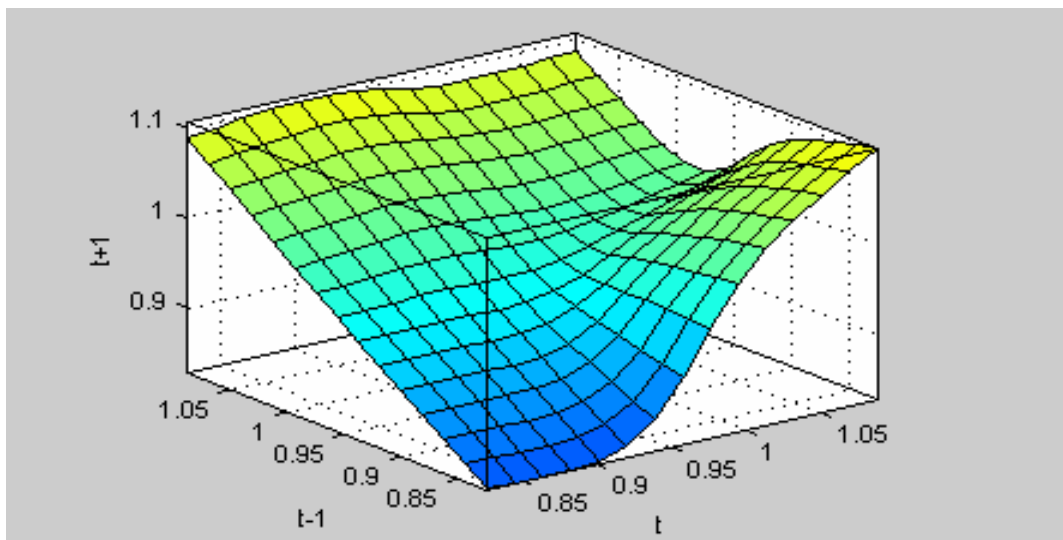
Οι συναρτήσεις συμμετοχής των εισόδων (t) και (t-1) είναι οι ίδιες, αφού και οι δύο συναρτήσεις είναι τύπου gauss2mf. Η εξίσωση των συναρτήσεων τύπου gauss2mf είναι:

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

όπου τα a_i, b_i, c_i είναι το σύνολο των παραμέτρων. Στο σχήμα 2 μπορούμε να δούμε τη συνάρτηση συμμετοχής των εισόδων (t) και (t-1), ενώ στο σχήμα 3 είναι η προκύπτουσα επιφάνεια εισόδου-εξόδου:

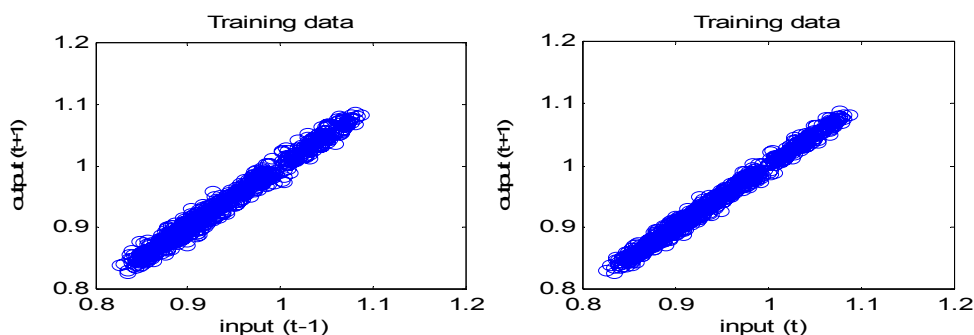


Σχήμα 2: Συνάρτηση συμμετοχής

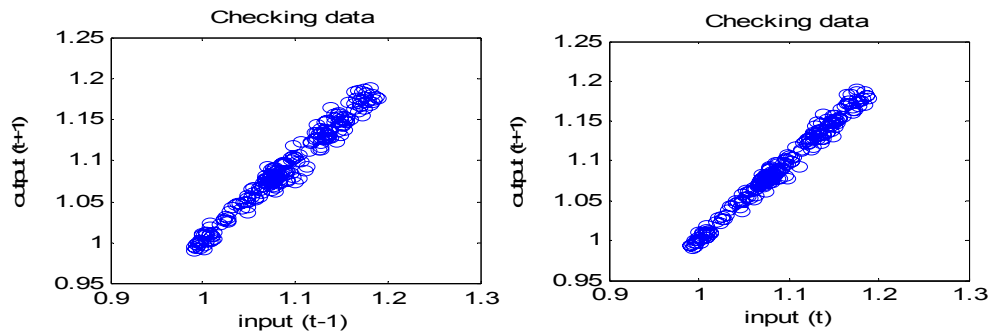


Σχήμα 3: Συνολική επιφάνεια εισόδου-εξόδου

Στο σχήμα 3 έχουμε μια τρισδιάστατη επιφάνεια του ασαφούς μοντέλου. Όπως μπορούμε να δούμε η επιφάνεια είναι αρκετά ομαλή, πράγμα που σημαίνει ότι έχουμε πάρει ένα ικανοποιητικό αριθμό δεδομένων και έτσι τα αποτελέσματα που θα πάρουμε από τη χρήση του μοντέλου που κατασκευάσαμε θα είναι αρκετά καλά. Εντούτοις, αυτό δεν φαίνεται να συμβαίνει στη κάτω δεξιά γωνία. Αυτή η ανωμαλία μπορεί να εξηγηθεί με την εξέταση των γραφημάτων διασποράς στα επόμενα σχήματα:



Σχήμα 4: Γραφήματα διασποράς δεδομένων εκπαίδευσης



Σχήμα 5: Γραφήματα διασποράς δεδομένων ελέγχου

Όπως βλέπουμε από τα προηγούμενα σχήματα, ο λόγος για αυτήν την ανωμαλία είναι η έλλειψη ικανοποιητικού αριθμού δεδομένων στη συγκεκριμένη περιοχή των γραφημάτων διασποράς.

2.3 Αρχιτεκτονική του ANFIS

Σε αυτή την παράγραφο θα εξετάσουμε πως είναι δομημένο το μοντέλο ANFIS. Επίσης θα δούμε σε ποια επίπεδα χωρίζεται το μοντέλο και ποιοι υπολογισμοί γίνονται σε κάθε επίπεδο ούτως ώστε να πάρουμε το τελικό αποτέλεσμα.

Όπως αναφέρθηκε στην προηγούμενη ενότητα, το μοντέλο μας είναι ένα ασαφές μοντέλο τύπου Sugeno πρώτου βαθμού το οποίο έχει δύο εισόδους και μία έξοδο. Επίσης, στο μοντέλο ισχύουν τέσσερις ασαφείς κανόνες οι οποίοι είναι οι εξής:

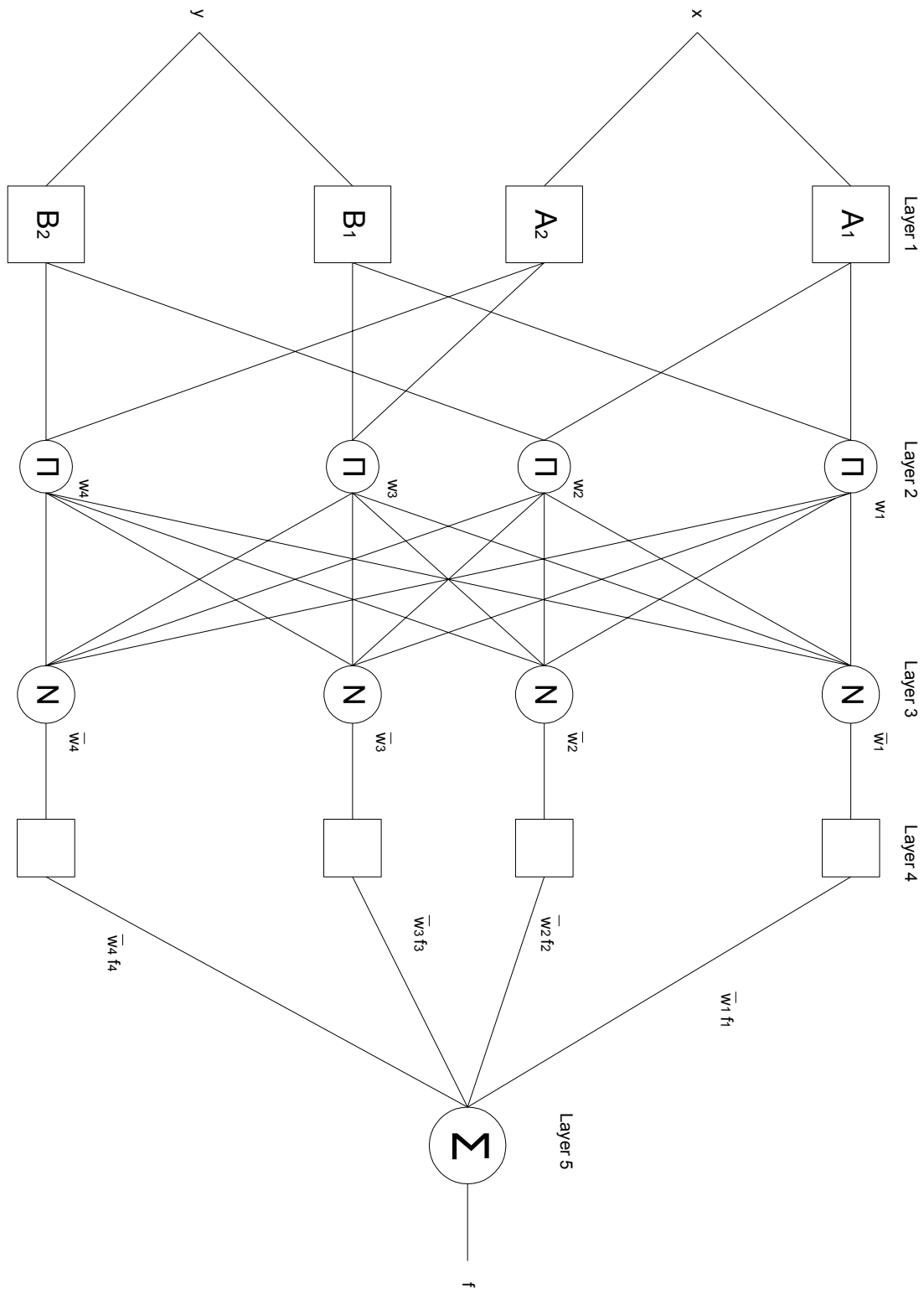
Αν (t) είναι small και (t-1) είναι small τότε (t+1) είναι very small

Αν (t) είναι small και (t-1) είναι large τότε (t+1) είναι small

Αν (t) είναι large και (t-1) είναι small τότε (t+1) είναι large

Αν (t) είναι large και (t-1) είναι large τότε (t+1) είναι very large

Λαμβάνοντας υπόψη τα στοιχεία που αναφέρθηκαν παραπάνω φαίνεται στη συνέχεια (Σχήμα 6) η αρχιτεκτονική του μοντέλου ANFIS. Ύστερα θα δούμε αναλυτικά τι συμβαίνει σε κάθε επίπεδο του μοντέλου και τέλος θα κάνουμε χρήση ενός παραδείγματος για την καλύτερη κατανόησή του. (Εδώ παρουσιάζουμε την έξοδο του i -οστού επιπέδου στο επίπεδο l σαν $O_{l,i}$.)



Σχήμα 6: Αρχιτεκτονική του ANFIS

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

$$O_{1,i} = \mu_{A_i}(x) \text{ για } i = 1,2 \text{ ή}$$

$$O_{1,i} = \mu_{B_{i-2}}(y) \text{ για } i = 3,4,$$

όπου x (ή y) – η είσοδος στον κόμβο i

A_i (ή B_{i-2})- 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 ($= A_1, A_2, B_1$ or B_2) και καθορίζει το βαθμό στον οποίο η είσοδος x (ή y) ικανοποιεί τον ποσοτικοποιητή A . Εδώ η συνάρτηση συμμετοχής του A είναι τύπου Gaussian2:

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

όπου a_i, c_i είναι το σύνολο 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:

$$O_{2,i} = w_i = \mu_{A_1}(x) * \mu_{B_1}(y), \text{ για } i = 1,2 \text{ και}$$

$$O_{2,i} = w_i = \mu_{A_2}(x) * \mu_{B_2}(y), \text{ για } i = 3,4.$$

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

$$O_{3,i} = \bar{w}_i = \frac{w_i}{w_1 + w_2 + w_3 + w_4}, \quad i=1,2,3,4.$$

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.

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

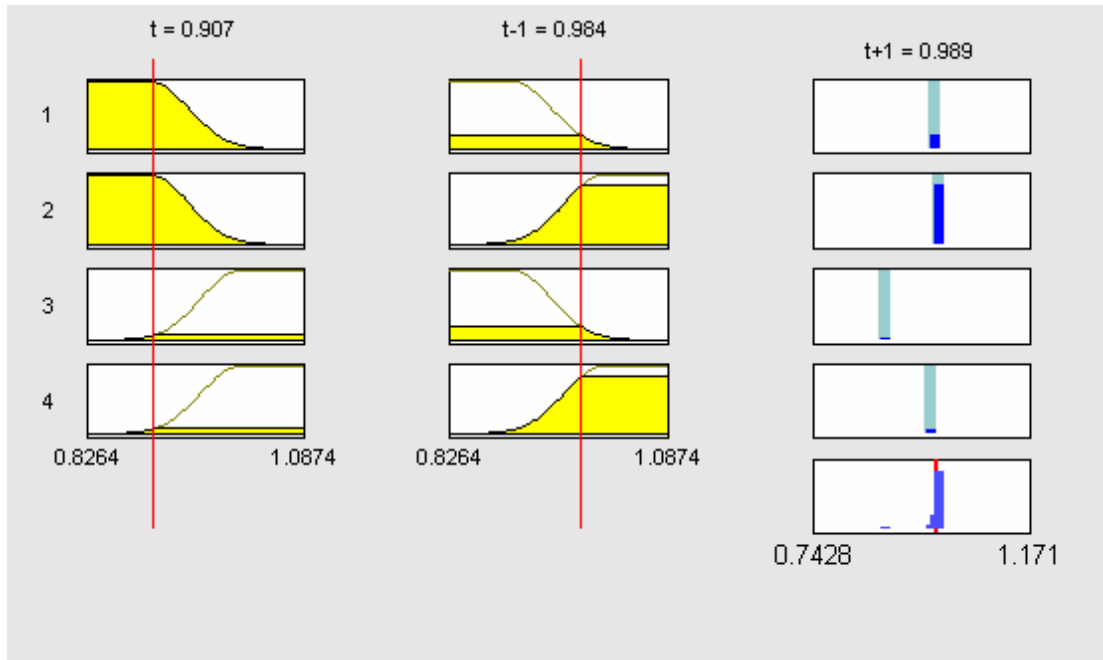
όπου: \bar{w}_i - η έξοδος του επιπέδου 3

$\{p_i, q_i, r_i\}$ - το σύνολο 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:

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

Αφού αναλύσαμε κάθε επίπεδο ξεχωριστά θα προχωρήσουμε τώρα στην παρουσίαση ενός παραδείγματος. Στο επόμενο σχήμα βλέπουμε πως το μοντέλο ANFIS βρίσκει το αποτέλεσμα που προκύπτει για δύο τυχαίες τιμές *t* και *t*-1:



Σχήμα 7: Οι υπολογισμοί του ANFIS

Όπως βλέπουμε στο σχήμα αν $t = 0.907$ και $t-1 = 0.984$ τότε $t+1 = 0.989$. Ας δούμε τώρα πως βγαίνει αυτό το αποτέλεσμα.

Επίπεδο 1 Εδώ χρησιμοποιούμε τη συνάρτηση συμμετοχής Gaussian2 για να υπολογίσουμε το $O_{1,i}$ για κάθε κόμβο. Το σύνολο παραμέτρων φαίνεται στον membership function editor και υπολογίζεται αυτόματα. Έτσι έχουμε:

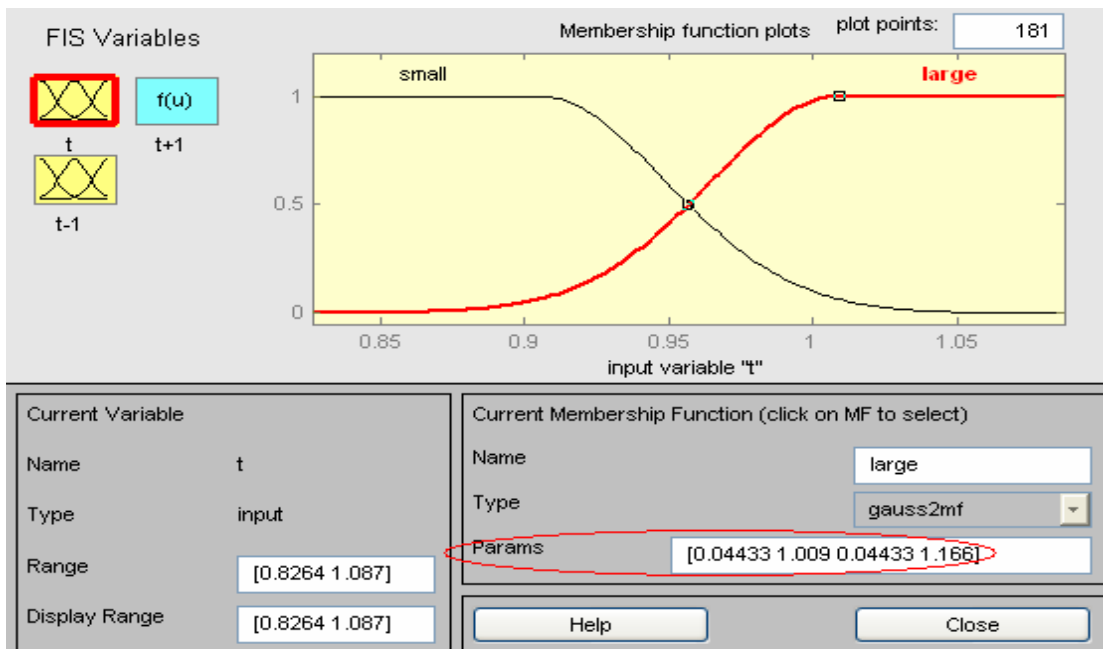
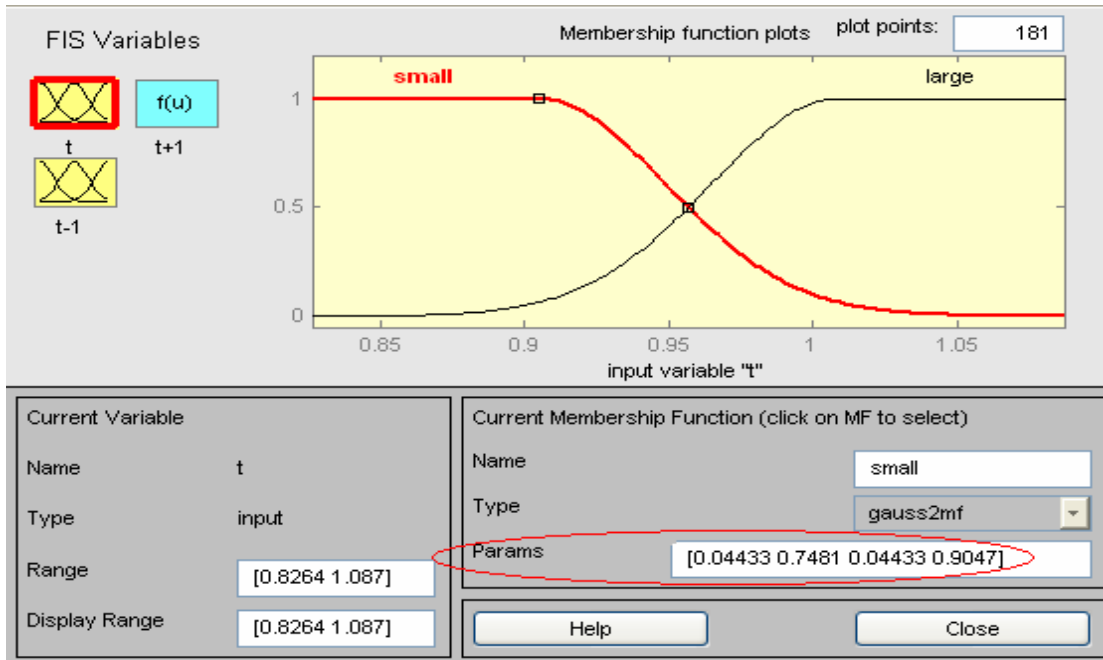
$$[a_1, c_1, a_2, c_2] = [0.04433, 0.7481, 0.04433, 0.9047]$$

αν η γλωσσική μεταβλητή είναι small και

$$[a_1, c_1, a_2, c_2] = [0.04433, 0.1009, 0.04433, 1.166]$$

αν η γλωσσική μεταβλητή είναι large

όπου a_1, c_1 είναι για την αριστερή καμπύλη της συνάρτησης Gaussian2 και a_2, c_2 είναι για την δεξιά καμπύλη. Επίσης, αφού και οι δύο εισοδοι είναι ίδιου τύπου, το σύνολο παραμέτρων είναι το ίδιο και για τις δύο εισόδους.



Σχήμα 8: Membership function editor

Από το σχήμα 7 βλέπουμε ότι τα A_1 και B_1 είναι **small** και στη **δεξιά** καμπύλη, ενώ τα A_2 και B_2 είναι **large** και στην **αριστερή** καμπύλη. Αυτό σημαίνει ότι για τα A_1 και B_1 θα χρησιμοποιήσουμε το σύνολο παραμέτρων $[0.04433, 0.9047]$ και για τα A_2 και B_2 το σύνολο παραμέτρων $[0.04433, 1.009]$.

Τώρα χρησιμοποιώντας όλα αυτά και για $x = 0.907$ και $y = 0.984$ παίρνουμε:

$$O_{1,1} = \mu_{A_1}(x) = 0.999$$

$$O_{1,2} = \mu_{A_2}(x) = 0.701$$

$$O_{1,3} = \mu_{B_1}(y) = 0.202$$

$$O_{1,4} = \mu_{B_2}(y) = 0.853$$

Επίπεδο 2 Εδώ κάνουμε τους υπολογισμούς χρησιμοποιώντας τις εξισώσεις που αναφέρθηκαν παραπάνω και έχουμε:

$$O_{2,1} = w_1 = \mu_{A_1}(x) * \mu_{B_1}(y) = 0.999 * 0.202 = 0.2016$$

$$O_{2,2} = w_2 = \mu_{A_1}(x) * \mu_{B_2}(y) = 0.999 * 0.853 = 0.8518$$

$$O_{2,3} = w_3 = \mu_{A_2}(x) * \mu_{B_1}(y) = 0.701 * 0.202 = 0.0143$$

$$O_{2,4} = w_4 = \mu_{A_2}(x) * \mu_{B_2}(y) = 0.701 * 0.853 = 0.0604$$

Επίπεδο 3 Όπως και στο επίπεδο 2 εδώ κάνουμε τους υπολογισμούς χρησιμοποιώντας την εξίσωση που αναφέρθηκε παραπάνω:

$$O_{3,1} = \bar{w}_1 = \frac{w_1}{w_1 + w_2 + w_3 + w_4} = \frac{0.2016}{0.2016 + 0.8518 + 0.0143 + 0.0604} = 0.1787$$

$$O_{3,2} = \bar{w}_2 = \frac{w_2}{w_1 + w_2 + w_3 + w_4} = \frac{0.8518}{0.2016 + 0.8518 + 0.0143 + 0.0604} = 0.755$$

$$O_{3,3} = \bar{w}_3 = \frac{w_3}{w_1 + w_2 + w_3 + w_4} = \frac{0.0143}{0.2016 + 0.8518 + 0.0143 + 0.0604} = 0.0127$$

$$O_{3,4} = \bar{w}_4 = \frac{w_4}{w_1 + w_2 + w_3 + w_4} = \frac{0.0604}{0.2016 + 0.8518 + 0.0143 + 0.0604} = 0.0536$$

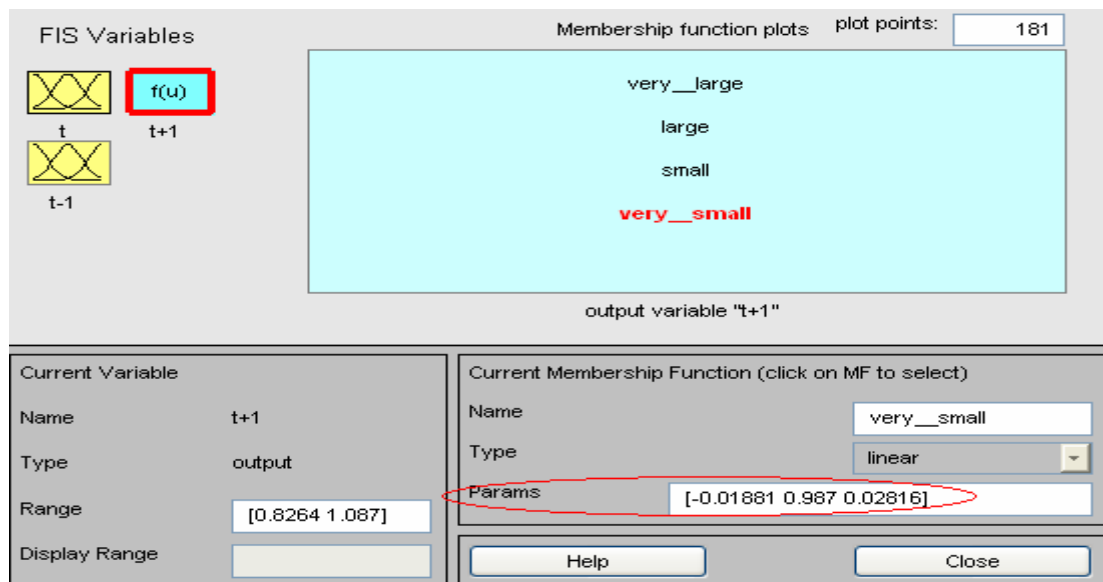
Επίπεδο 4 Για να μπορέσουμε να πραγματοποιήσουμε τους απαραίτητους υπολογισμούς σε αυτό το επίπεδο χρειαζόμαστε το σύνολο παραμέτρων για κάθε κόμβο. Αυτές οι παράμετροι φαίνονται στον membership function editor επιλέγοντας t+1 και είναι:

$[p_1, q_1, r_1] = [-0.01881, 0.987, 0.02816]$ αν η γλωσσική μεταβλητή είναι very small

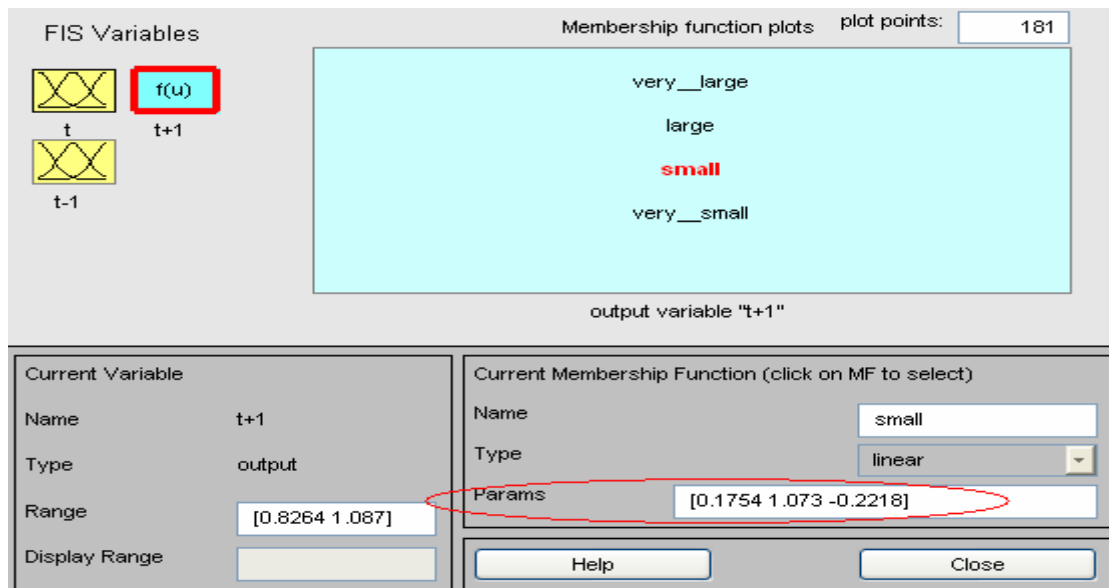
$[p_2, q_2, r_2] = [0.1754, 1.073, -0.2218]$ αν η γλωσσική μεταβλητή είναι small

$[p_3, q_3, r_3] = [0.9956, -0.2648, 0.2437]$ αν η γλωσσική μεταβλητή είναι large

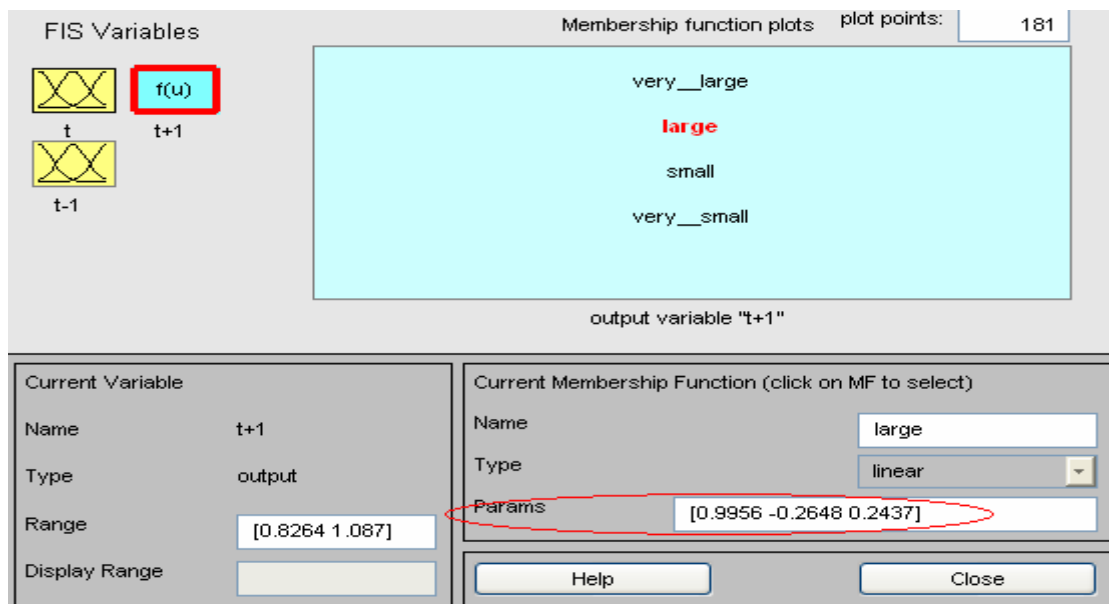
$[p_4, q_4, r_4] = [0.09453, 0.9042, 0.0006383]$ αν η γλωσσική μεταβλητή είναι very large



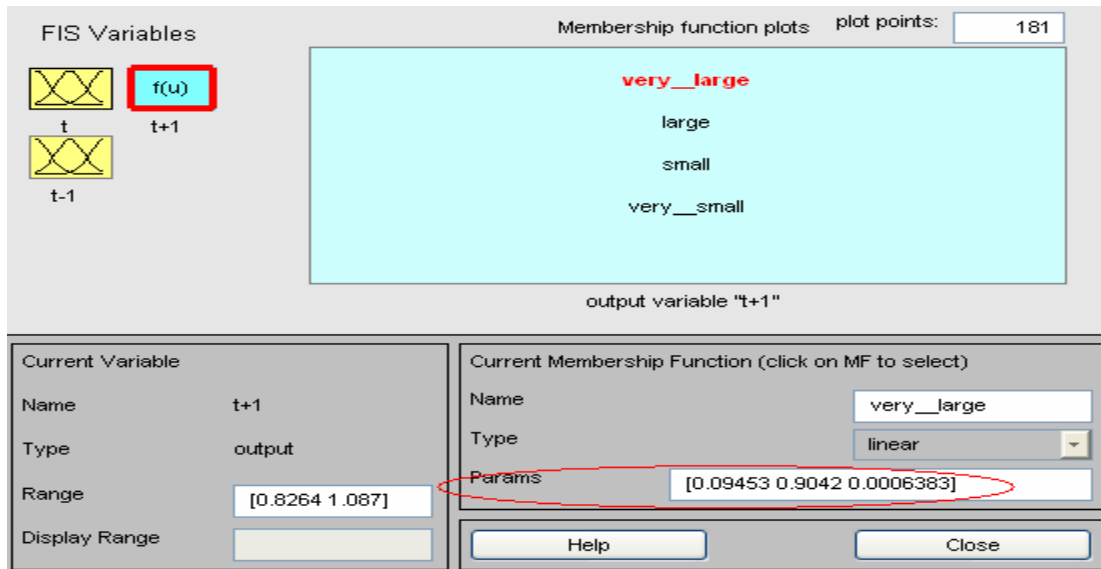
Σχήμα 9: Σύνολο παραμέτρων για very small



Σχήμα 10: Σύνολο παραμέτρων για small



Σχήμα 11: Σύνολο παραμέτρων για large



Σχήμα 12: Σύνολο παραμέτρων για very large

Χρησιμοποιώντας τα σύνολα παραμέτρων έχουμε:

$$O_{4,1} = \bar{w}_1(p_1 * x + q_1 * y + r_1) = 0.1787(-0.01881 * 0.907 + 0.987 * 0.984 + 0.02816) = 0.176$$

$$O_{4,2} = \bar{w}_2(p_2 * x + q_2 * y + r_2) = 0.755(0.1754 * 0.907 + 1.073 * 0.984 - 0.2218) = 0.75$$

$$O_{4,3} = \bar{w}_3(p_3 * x + q_3 * y + r_3) = 0.0127(0.9956 * 0.907 - 0.2648 * 0.984 + 0.2437) = 0.011$$

$$O_{4,4} = \bar{w}_4(p_4 * x + q_4 * y + r_4) = 0.0536(0.09453 * 0.907 + 0.9042 * 0.984 + 0.0006383) = 0.052$$

Επίπεδο 5 Στο τελευταίο αυτό επίπεδο αυτό βρίσκουμε την τιμή για t+1 χρησιμοποιώντας την εξίσωση που αναφέρθηκε στη θεωρία:

$$O_{5,1} = \sum_i \bar{w}_i * f_i = 0.176 + 0.75 + 0.011 + 0.052 = 0.989$$

ΚΕΦΑΛΑΙΟ 3: Μέθοδοι για προσδιορισμό συστήματος

3.1 Προσδιορισμός συστήματος: Μια εισαγωγή

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:

- Για να προβλέψει μια συμπεριφορά του συστήματος, όπως στη πρόβλεψη χρονοσειρών και στη πρόβλεψη καιρού.
- 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.
- Για να σχεδιάσει έναν ελεγκτή βασισμένο στο μοντέλο ενός συστήματος, όπως στον έλεγχο αεροσκαφών και σκαφών. Επίσης για να κάνουμε την προσομοίωση υπολογιστών του συστήματος υπό έλεγχο, χρειαζόμαστε ένα μοντέλο του συστήματος.

Ο προσδιορισμός συστημάτων περιλαμβάνει γενικά δύο βήματα:

Προσδιορισμός δομής Σε αυτό το βήμα, πρέπει να εφαρμόσουμε την *a priori* γνώση για το σύστημα στόχο για να καθορίσουμε μια κατηγορία μοντέλων μέσα στην οποία πρόκειται να διεξαχθεί η αναζήτηση του καταλληλότερου μοντέλου. Συνήθως αυτή η κατηγορία μοντέλων δείχνεται από μια παραμετρική συνάρτηση $y = f(\mathbf{u}; \boldsymbol{\theta})$, όπου y είναι η έξοδος του μοντέλου, \mathbf{u} είναι το διάνυσμα εισόδου και $\boldsymbol{\theta}$ 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 κάνουμε είναι να εφαρμόσουμε τεχνικές βελτιστοποίησης για να καθορίσουμε το διάνυσμα παραμέτρου $\theta = \hat{\theta}$ τέτοιο ώστε το προκύπτον μοντέλο $\hat{y} = f(\mathbf{u}; \hat{\theta})$ να μπορεί να περιγράψει το σύστημα κατάλληλα.

Εάν δεν έχουμε οποιαδήποτε *a priori* γνώση για το σύστημα στόχο, τότε ο προσδιορισμός δομής γίνεται ένα δύσκολο πρόβλημα και πρέπει να επιλέξουμε τη δομή με δοκιμή και σφάλμα. Ευτυχώς, ξέρουμε πάρα πολλά 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 κάνουμε προσδιορισμό δομής και παραμέτρου επανειλημμένα έως ότου να βρούμε ένα ικανοποιητικό πρότυπο ως εξής:

1. Καθορίζουμε και παραμετροποιούμε μια κατηγορία μαθηματικών μοντέλων που αντιπροσωπεύουν το σύστημα που προσδιορίζεται.
2. Εκτελούμε τον προσδιορισμό παραμέτρου για να επιλέξουμε τις παραμέτρους που ταιριάζουν καλύτερα στο σύνολο δεδομένων εκπαίδευσης.
3. Πραγματοποιούμε δοκιμές επικύρωσης για να δούμε 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, επικυρώσης ή ελέγχου.**)

4. Ολοκληρώνουμε τη διαδικασία μόλις τα αποτελέσματα της δοκιμής επικύρωσης είναι ικανοποιητικά. Διαφορετικά, μια άλλη κατηγορία μοντέλων επιλέγεται και τα βήματα 2 μέχρι 4 επαναλαμβάνονται.

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 περιγράψουμε τη μέθοδο ελαχίστων τετραγώνων και τη μέθοδο οπισθόδρομης βαθμωτής ελαχιστοποίησης και έπειτα θα δούμε πως 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 παραμετροποιημένη έκφραση

$$y = \theta_1 f_1(\mathbf{u}) + \theta_2 f_2(\mathbf{u}) + \dots + \theta_n f_n(\mathbf{u}), \quad (3.2.1)$$

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 βάλεις δεδομένα χρησιμοποιώντας ένα γραμμικό

μοντέλο ανφέρεται σαν γραμμική παλινδρόμηση. Κατά συνέπεια η εξίσωση (3.2.1) καλείται επίσης συνάρτηση παλινδρόμησης, και τα θ_i ονομάζονται συντελεστές παλινδρόμησης.

Για να προσδιορίσουμε τις άγνωστες παραμέτρους θ_i , συνήθως πρέπει να εκτελέσουμε πειράματα για να λάβουμε 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 (3.2.1) παράγει ένα σύνολο γραμμικών εξισώσεων m :

$$\begin{aligned} f_1(\mathbf{u}_1)\theta_1 + f_2(\mathbf{u}_1)\theta_2 + \dots + f_n(\mathbf{u}_1)\theta_n &= y_1 \\ f_1(\mathbf{u}_2)\theta_1 + f_2(\mathbf{u}_2)\theta_2 + \dots + f_n(\mathbf{u}_2)\theta_n &= y_2 \\ &\vdots \\ f_1(\mathbf{u}_m)\theta_1 + f_2(\mathbf{u}_m)\theta_2 + \dots + f_n(\mathbf{u}_m)\theta_n &= y_m \end{aligned} \quad (3.2.2)$$

Χρησιμοποιώντας τη μορφή πινάκων, μπορούμε να ξαναγράψουμε τις προηγούμενες εξισώσεις σε μια συνοπτική μορφή:

$$\mathbf{A}\boldsymbol{\theta} = \mathbf{y} \quad (3.2.3)$$

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

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

$\boldsymbol{\theta}$ είναι ένα $n \times 1$ διάνυσμα άγνωστων παραμέτρων:

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

και \mathbf{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 μας βασίζεται στους πίνακες \mathbf{A} και \mathbf{y} , μερικές φορές αναφερόμαστε στο $(\mathbf{a}_i^T; y_i)$ σαν το i -οστό ζευγάρι δεδομένων του συνόλου δεδομένων εκπαίδευσης.

Για να προσδιορίσουμε μεμονωμένα το άγνωστο διάνυσμα $\boldsymbol{\theta}$, είναι απαραίτητο να ισχύει $m \geq n$. Αν ο \mathbf{A} είναι τετραγωνικός ($m = n$) και αντιστρέψιμος, τότε μπορούμε να λύσουμε ως προς τον άγνωστο \mathbf{x} από την εξίσωση (3.2.3) και έχουμε

$$\boldsymbol{\theta} = \mathbf{A}^{-1} \mathbf{y} \quad . \quad (3.2.4)$$

Εντούτοις, ο m είναι συνήθως μεγαλύτερος από τον n , δείχνοντας έτσι ότι έχουμε 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 (3.2.3) 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:

$$\mathbf{A}\boldsymbol{\theta} + \mathbf{e} = \mathbf{y} \quad . \quad (3.2.5)$$

Τώρα, αντί της εύρεσης της ακριβούς λύσης στην εξίσωση (3.2.3), θέλουμε να ψάξουμε για το $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ που ελαχιστοποιεί το άθροισμα του τετραγωνικού σφάλματος και ορίζεται ως

$$E(\boldsymbol{\theta}) = \sum_{i=1}^m (y_i - \mathbf{a}_i^T \boldsymbol{\theta})^2 = \mathbf{e}^T \mathbf{e} = (\mathbf{y} - \mathbf{A}\boldsymbol{\theta})^T (\mathbf{y} - \mathbf{A}\boldsymbol{\theta}) \quad (3.2.6)$$

όπου $\mathbf{e} = \mathbf{y} - \mathbf{A}\boldsymbol{\theta}$ είναι το διάνυσμα λάθους που παράγεται από μια συγκεκριμένη επιλογή του $\boldsymbol{\theta}$. Σημειώστε 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 3.2.1 *Εκτιμητής ελαχίστων τετραγώνων*

Το τετραγωνικό σφάλμα στην εξίσωση (3.2.6) ελαχιστοποιείται όταν $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, ο οποίος καλείται εκτιμητής ελαχίστων τετραγώνων (LSE για συντομία) και ο οποίος ικανοποιεί την κανονική εξίσωση

$$\mathbf{A}^T \mathbf{A} \hat{\boldsymbol{\theta}} = \mathbf{A}^T \mathbf{y} \quad (3.2.7)$$

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

$$\hat{\boldsymbol{\theta}} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y} \quad . \quad (3.2.8)$$

Απόδειξη: Υπάρχουν διάφορες μέθοδοι διαθέσιμες στη βιβλιογραφία για την εύρεση του εκτιμητή ελαχίστων τετραγώνων για την εξίσωση (3.2.3). Μια απλή προσέγγιση είναι να τεθεί το παράγωγο του $E(\boldsymbol{\theta})$ ως προς $\boldsymbol{\theta}$ ίσο σε μηδέν. Σημειώνοντας ότι ο $\boldsymbol{\theta}^T \mathbf{A}^T \mathbf{y} = \mathbf{y}^T \mathbf{A} \boldsymbol{\theta}$ είναι βαθμωτός, μπορούμε να επεκταθούμε για $E(\boldsymbol{\theta})$:

$$E(\boldsymbol{\theta}) = (\mathbf{y}^T - \boldsymbol{\theta}^T \mathbf{A}^T)(\mathbf{y} - \mathbf{A}\boldsymbol{\theta}) = \boldsymbol{\theta}^T \mathbf{A}^T \mathbf{A} \boldsymbol{\theta} - 2\mathbf{y}^T \mathbf{A} \boldsymbol{\theta} + \mathbf{y}^T \mathbf{y} \quad (3.2.9)$$

Τότε η παράγωγος του $E(\boldsymbol{\theta})$ είναι

$$\frac{\partial E(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 2\mathbf{A}^T \mathbf{A} \boldsymbol{\theta} - 2\mathbf{A}^T \mathbf{y}. \quad (3.2.10)$$

Θέτοντας $\frac{\partial E(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 0$ στο $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, παίρνουμε την κανονική εξίσωση

$$\mathbf{A}^T \mathbf{A} \hat{\boldsymbol{\theta}} = \mathbf{A}^T \mathbf{y}. \quad (3.2.11)$$

Αν ο $\mathbf{A}^T \mathbf{A}$ είναι αντιστρέψιμος, τότε ο $\hat{\boldsymbol{\theta}}$ μπορεί να επιλυθεί μοναδικά:

$$\hat{\boldsymbol{\theta}} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y}. \quad (3.2.12)$$

Εφόσον το $E(\boldsymbol{\theta})$ είναι μια τετραγωνική συνάρτηση $\boldsymbol{\theta}$, το σφάλμα ελαχίστων τετραγώνων που επιτυγχάνεται για $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ μπορεί να βρεθεί να είναι

$$E(\hat{\boldsymbol{\theta}}) = (\mathbf{y} - \mathbf{A}\hat{\boldsymbol{\theta}})^T (\mathbf{y} - \mathbf{A}\hat{\boldsymbol{\theta}}) = \mathbf{y}^T \mathbf{y} - \mathbf{y}^T \mathbf{A} (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y}. \quad (3.2.13)$$

Όμως, αν ο $\mathbf{A}^T \mathbf{A}$ είναι μη αντιστρέψιμος, τότε ο LSE δεν είναι μοναδικός και πρέπει να εφαρμόσουμε την έννοια του γενικευμένου αντιστρόφου για να βρούμε τον $\hat{\boldsymbol{\theta}}$. Χωρίς απώλεια γενικότητας, θα υποθέσουμε ότι ο $\mathbf{A}^T \mathbf{A}$ είναι αντιστρέψιμος.

Η προηγούμενη παραγωγή είναι βασισμένη στην υπόθεση ότι κάθε στοιχείο του διανύσματος λάθους \mathbf{e} έχει το ίδιο βάρος προς το γενικό τετραγωνικό σφάλμα. Μια περαιτέρω γενίκευση είναι να αφήσουμε κάθε όρο σφάλματος να σταθμιστεί διαφορετικά. Συγκεκριμένα, ας πούμε ότι ο \mathbf{W} είναι ο επιθυμητός πίνακας βαρύτητας, ο οποίος είναι συμμετρικός και θετικά ορισμένος. Τότε το σταθμισμένο τετραγωνικό σφάλμα είναι

$$E_{\mathbf{W}}(\boldsymbol{\theta}) = (\mathbf{y} - \mathbf{A}\boldsymbol{\theta})^T \mathbf{W}(\mathbf{y} - \mathbf{A}\boldsymbol{\theta}). \quad (3.2.14)$$

Ελαχιστοποιώντας το $E_{\mathbf{W}}(\boldsymbol{\theta})$ ως προς $\boldsymbol{\theta}$ παράγουμε το σταθμισμένο εκτιμητή ελαχίστων τετραγώνων $\hat{\boldsymbol{\theta}}_{\mathbf{W}}$:

$$\hat{\boldsymbol{\theta}}_{\mathbf{W}} = (\mathbf{A}^T \mathbf{W} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{W} \mathbf{y} \quad (3.2.15)$$

Προφανώς, ο $\hat{\boldsymbol{\theta}}_{\mathbf{W}}$ μειώνεται σε $\hat{\boldsymbol{\theta}}$ όταν ο \mathbf{W} επιλέγεται σαν ένας μοναδιαίος πίνακας.

3.3 Οπισθοδρόμηση για πρωστοροδοτούμενα 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 πως να πάρουμε 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 πάρουμε την 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. Θα αναλύσουμε αυτόν τον κανόνα εκμάθησης στην συνέχεια αυτής της ενότητας.

Υποθέστε ότι δεδομένο πρωστοροφοδοτούμενο 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, υποθέτουμε 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, έχουμε την ακόλουθη γενική έκφραση για τη συνάρτηση των κόμβων $f_{l,i}$:

$$x_{l,i} = f_{l,i}(x_{l-1,1}, \dots, x_{l-1,N(l-1)}, \alpha, \beta, \gamma, \dots), \quad (3.3.1)$$

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

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

$$E_p = \sum_{k=1}^{N(l)} (d_k - x_{L,k})^2, \quad (3.3.2)$$

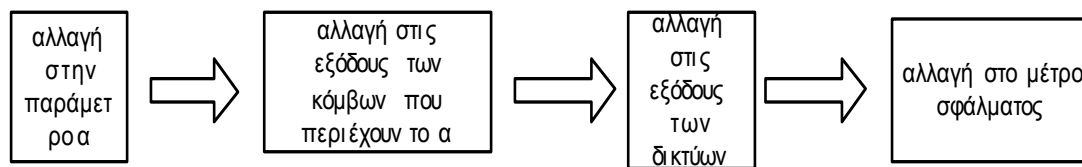
όπου d_k είναι το k -οστό συστατικό του p -οστού επιθυμητού διανύσματος εξόδου και $x_{L,k}$ είναι το k -οστό συστατικό του πραγματικού διανύσματος εξόδου που παράγεται με την παρουσίαση του p -οστού διανύσματος εισόδου στο δίκτυο. (Για σημειογραφική απλότητα, παραλείπουμε το δείκτη p και από το d_k 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 μας εδώ είναι να ελαχιστοποιήσουμε ένα συνολικό μέτρο σφάλματος, που ορίζεται ως

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

Θυμηθείτε ότι ο ορισμός του E_p στην εξίσωση (3.3.2) δεν είναι καθολικός αλλά είναι δυνατός και άλλος ορισμός για το E_p για συγκεκριμένες καταστάσεις ή εφαρμογές. Επομένως, θα αποφύγουμε μια ρητή έκφραση για το μέτρο σφάλματος E_p για να δώσουμε έμφαση στη γενικότητα. Επιπλέον, υποθέτουμε ότι το E_p εξαρτάται μόνο από τους κόμβους εξόδου.

Για να χρησιμοποιήσουμε τη βαθμωτή ελαχιστοποίηση ώστε να ελαχιστοποιήσουμε το μέτρο σφάλματος, πρέπει πρώτα να βρούμε το διάνυσμα κλίσης. Πριν υπολογίσουμε το διάνυσμα κλίσης, πρέπει να παρατηρήσουμε τις ακόλουθες αιτιώδεις σχέσεις:



όπου τα βέλη = 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 περάσουμε μια σειρά από πληροφορίες παραγώγων ξεκινώντας από το επίπεδο εξόδου και πηγαίνοντας ανάποδα από επίπεδο σε επίπεδο έως ότου φθάσουμε στο επίπεδο εισόδου.

Για να διευκολύνει τη συζήτηση, ορίζουμε το σήμα σφάλματος $\in_{l,i}$ 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,

$$\in_{l,i} = \frac{\partial^+ E_p}{\partial x_{l,i}} \quad (3.3.3)$$

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 βλέπουμε τη συνάρτηση να 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 διαδωθεί μέσω έμμεσων πορειών στο επίπεδο εξόδου και συνεπώς θα προκαλέσει μια αντίστοιχη αλλαγή στην τιμή του E_p . Επομένως, το $\in_{l,i}$ 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:

$$\in_{L,i} = \frac{\partial^+ E_p}{\partial x_{L,i}} = \frac{\partial E_p}{\partial x_{L,i}} \quad (3.3.4)$$

Αυτό ισούται με $\in_{L,i} = -2(d_i - x_{L,i})$ αν το E_p ορίζεται όπως στην εξίσωση (3.3.2). Για τον εσωτερικό κόμβο στην i -οστή θέση του επιπέδου l , το σήμα σφάλματος μπορεί να βρεθεί από τον κανόνα αλυσίδας:

$$\epsilon_{l,i} = \underbrace{\frac{\partial^+ E_p}{\partial x_{l,i}}}_{\text{error signal at layer } l} = \sum_{m=1}^{N(l+1)} \underbrace{\frac{\partial^+ E_p}{\partial x_{l+1,m}}}_{\text{error signal at layer } l+1} \frac{\partial f_{l+1,m}}{\partial x_{l,i}} = \sum_{m=1}^{N(l+1)} \epsilon_{l+1,m} \frac{\partial f_{l+1,m}}{\partial x_{l,i}}, \quad (3.3.5)$$

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 οποιουδήποτε l και i [και $1 \leq i \leq N(l)$], μπορούμε να βρούμε το

$$\epsilon_{l,i} = \frac{\partial^+ E_p}{\partial x_{l,i}} \text{ εφαρμόζοντας πρώτα την εξίσωση (3.3.4) μία φορά για να πάρουμε τα}$$

σήματα σφάλματος στο επίπεδο εξόδου, και έπειτα εφαρμόζοντας την εξίσωση (3.3.5) επαναληπτικά έως ότου φθάσουμε 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 εφαρμόσουμε τον κανόνα αλυσίδας ξανά για να βρούμε το διάνυσμα κλίσης. Εάν a είναι μια παράμετρος του i -οστού κόμβου στο επίπεδο l , έχουμε:

$$\frac{\partial^+ E_p}{\partial a} = \frac{\partial^+ E_p}{\partial x_{l,i}} \frac{\partial f_{l,i}}{\partial a} = \epsilon_{l,i} \frac{\partial f_{l,i}}{\partial a}. \quad (3.3.6)$$

Σημειώστε ότι αν επιτρέψουμε στην παράμετρο a να μοιράζεται μεταξύ διαφορετικών κόμβων, τότε η εξίσωση (3.3.6) πρέπει να αλλαχθεί σε μια πιο γενική μορφή:

$$\frac{\partial^+ E_p}{\partial a} = \sum_{x^* \in S} \frac{\partial^+ E_p}{\partial x^*} \frac{\partial f^*}{\partial a}, \quad (3.3.7)$$

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

$$\frac{\partial^+ E}{\partial a} = \sum_{p=1}^P \frac{\partial^+ E_p}{\partial a} . \quad (3.3.8)$$

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

$$\Delta a = -\eta \frac{\partial^+ E_p}{\partial a} , \quad (3.3.9)$$

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

$$\eta = \frac{\kappa}{\sqrt{\sum_{\alpha} \left(\frac{\partial E}{\partial \alpha} \right)^2}} , \quad (3.3.10)$$

όπου το κ είναι το μέγεθος βήματος, το μήκος δηλαδή κάθε μετάβασης κατά μήκος της κατεύθυνσης κλίσης στο διάστημα παραμέτρου. Συνήθως μπορούμε να αλλάξουμε το μέγεθος βήματος για να μεταβάλλουμε την ταχύτητα της σύγκλισης.

Όταν ένα πρωστοτροφοδοτούμενο δίκτυο n -κόμβων αναπαριστάται στην τοπολογική του διάταξη, μπορούμε να υπολογίσουμε 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, έχουμε τον ακόλουθο συνοπτικό τύπο για τον υπολογισμό του σήματος

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

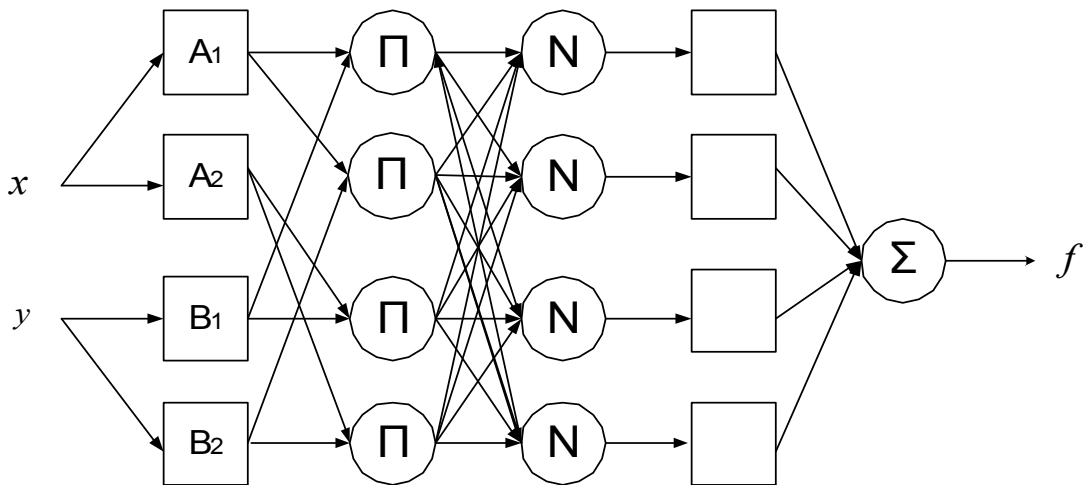
$$\frac{\partial^+ E_p}{\partial x_i} = \frac{\partial f_{n+1}}{\partial x_i} + \sum_{i < j \leq n} \epsilon_j \frac{\partial^+ E_p}{\partial x_j} \frac{\partial f_j}{\partial x_i} , \quad (3.3.11)$$

ή

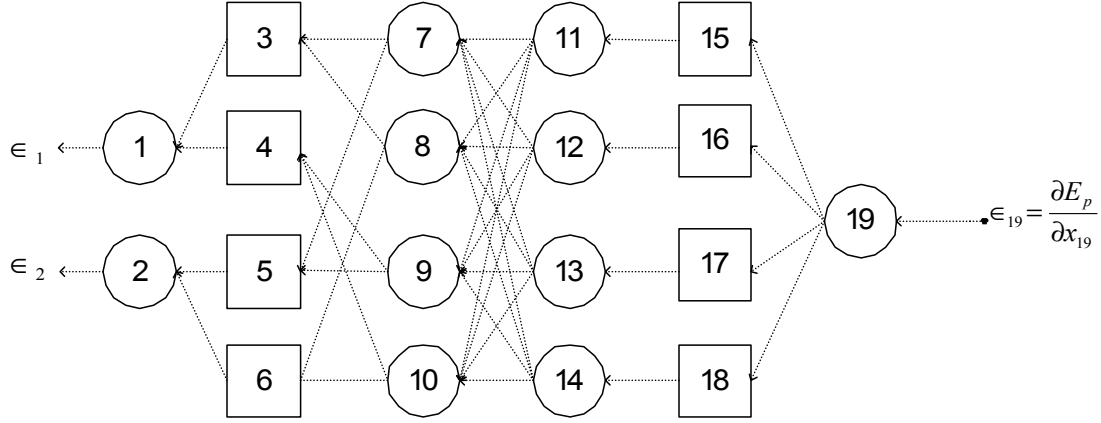
$$\epsilon_i = \frac{\partial f_{n+1}}{\partial x_i} + \sum_{i < j \leq n} \epsilon_j \frac{\partial f_j}{\partial x_i} , \quad (3.3.12)$$

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 βρούμε 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 1(a): Το προσαρμόσιμο δίκτυο



Σχήμα 1(b): Το δίκτυο διάδοσης σφάλματος

Τώρα θα υπολογίσουμε τα σήματα σφάλματος στους εσωτερικούς κόμβους. Χρησιμοποιούμε f_i και x_i για να δείξουμε τη συνάρτηση και την έξοδο του κόμβου i . Η έξοδος του κόμβου i είναι το σήμα σφάλματος αυτού του κόμβου στο πραγματικό προσαρμόσιμο δίκτυο. Σε σύμβολα, αν επιλέξουμε το τετραγωνικό μέτρο σφάλματος για E_p , τότε έχουμε 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 έχουμε:

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

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

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

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

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:

$$\begin{aligned}\epsilon_{14} &= \frac{\partial^+ E_p}{\partial x_{14}} = \frac{\partial^+ E_p}{\partial x_{18}} \frac{\partial f_{18}}{\partial x_{14}} = \epsilon_{18} \frac{\partial f_{18}}{\partial x_{14}}, \\ \epsilon_{13} &= \frac{\partial^+ E_p}{\partial x_{13}} = \frac{\partial^+ E_p}{\partial x_{17}} \frac{\partial f_{17}}{\partial x_{13}} = \epsilon_{17} \frac{\partial f_{17}}{\partial x_{13}}, \\ \epsilon_{12} &= \frac{\partial^+ E_p}{\partial x_{12}} = \frac{\partial^+ E_p}{\partial x_{16}} \frac{\partial f_{16}}{\partial x_{12}} = \epsilon_{16} \frac{\partial f_{16}}{\partial x_{12}}, \\ \epsilon_{11} &= \frac{\partial^+ E_p}{\partial x_{11}} = \frac{\partial^+ E_p}{\partial x_{15}} \frac{\partial f_{15}}{\partial x_{11}} = \epsilon_{15} \frac{\partial f_{15}}{\partial x_{11}}.\end{aligned}$$

Αντιθέτως, οι κόμβοι 7, 8, 9 και 10 εξαρτώνται από τέσσερις διαφορετικούς κόμβους όπως φαίνεται στο σχήμα 1. Έτσι, έχουμε:

$$\begin{aligned}\epsilon_{10} &= \frac{\partial^+ E_p}{\partial x_{10}} = \epsilon_{14} \frac{\partial f_{14}}{\partial x_{10}} + \epsilon_{13} \frac{\partial f_{13}}{\partial x_{10}} + \epsilon_{12} \frac{\partial f_{12}}{\partial x_{10}} + \epsilon_{11} \frac{\partial f_{11}}{\partial x_{10}}, \\ \epsilon_9 &= \frac{\partial^+ E_p}{\partial x_9} = \epsilon_{14} \frac{\partial f_{14}}{\partial x_9} + \epsilon_{13} \frac{\partial f_{13}}{\partial x_9} + \epsilon_{12} \frac{\partial f_{12}}{\partial x_9} + \epsilon_{11} \frac{\partial f_{11}}{\partial x_9}, \\ \epsilon_8 &= \frac{\partial^+ E_p}{\partial x_8} = \epsilon_{14} \frac{\partial f_{14}}{\partial x_8} + \epsilon_{13} \frac{\partial f_{13}}{\partial x_8} + \epsilon_{12} \frac{\partial f_{12}}{\partial x_8} + \epsilon_{11} \frac{\partial f_{11}}{\partial x_8}, \\ \epsilon_7 &= \frac{\partial^+ E_p}{\partial x_7} = \epsilon_{14} \frac{\partial f_{14}}{\partial x_7} + \epsilon_{13} \frac{\partial f_{13}}{\partial x_7} + \epsilon_{12} \frac{\partial f_{12}}{\partial x_7} + \epsilon_{11} \frac{\partial f_{11}}{\partial x_7}.\end{aligned}$$

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

$$\begin{aligned}
\epsilon_6 &= \frac{\partial^+ E_p}{\partial x_6} = \frac{\partial^+ E_p}{\partial x_{10}} \frac{\partial f_{10}}{\partial x_6} + \frac{\partial^+ E_p}{\partial x_8} \frac{\partial f_8}{\partial x_6} = \epsilon_{10} \frac{\partial f_{10}}{\partial x_6} + \epsilon_8 \frac{\partial f_8}{\partial x_6}, \\
\epsilon_5 &= \frac{\partial^+ E_p}{\partial x_5} = \frac{\partial^+ E_p}{\partial x_9} \frac{\partial f_9}{\partial x_5} + \frac{\partial^+ E_p}{\partial x_7} \frac{\partial f_7}{\partial x_5} = \epsilon_9 \frac{\partial f_9}{\partial x_5} + \epsilon_7 \frac{\partial f_7}{\partial x_5}, \\
\epsilon_4 &= \frac{\partial^+ E_p}{\partial x_4} = \frac{\partial^+ E_p}{\partial x_{10}} \frac{\partial f_{10}}{\partial x_4} + \frac{\partial^+ E_p}{\partial x_9} \frac{\partial f_9}{\partial x_4} = \epsilon_{10} \frac{\partial f_{10}}{\partial x_4} + \epsilon_9 \frac{\partial f_9}{\partial x_4}, \\
\epsilon_3 &= \frac{\partial^+ E_p}{\partial x_3} = \frac{\partial^+ E_p}{\partial x_8} \frac{\partial f_8}{\partial x_3} + \frac{\partial^+ E_p}{\partial x_7} \frac{\partial f_7}{\partial x_3} = \epsilon_8 \frac{\partial f_8}{\partial x_3} + \epsilon_7 \frac{\partial f_7}{\partial x_3}.
\end{aligned}$$

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

$$\begin{aligned}
\epsilon_2 &= \frac{\partial^+ E_p}{\partial x_2} = \frac{\partial^+ E_p}{\partial x_6} \frac{\partial f_6}{\partial x_2} + \frac{\partial^+ E_p}{\partial x_5} \frac{\partial f_5}{\partial x_2} = \epsilon_6 \frac{\partial f_6}{\partial x_2} + \epsilon_5 \frac{\partial f_5}{\partial x_2}, \\
\epsilon_1 &= \frac{\partial^+ E_p}{\partial x_1} = \frac{\partial^+ E_p}{\partial x_4} \frac{\partial f_4}{\partial x_1} + \frac{\partial^+ E_p}{\partial x_3} \frac{\partial f_3}{\partial x_1} = \epsilon_4 \frac{\partial f_4}{\partial x_1} + \epsilon_3 \frac{\partial f_3}{\partial x_1}
\end{aligned}$$

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

Αν και μπορούμε να εφαρμόσουμε οπισθοδρόμηση ή την βαθμωτή ελαχιστοποίηση για να προσδιορίσουμε τις παραμέτρους σε ένα προσαρμόσιμο δίκτυο, αυτή η απλή μέθοδος βελτιστοποίησης παίρνει συνήθως πολύ χρόνο προτού να συγκλίνει. Μπορούμε να παρατηρήσουμε, όμως, ότι η έξοδος ενός προσαρμόσιμου δικτύου είναι γραμμική σε μερικές από τις παραμέτρους του δικτύου. Έτσι μπορούμε να προσδιορίσουμε αυτές τις γραμμικές παραμέτρους με τη γραμμική μέθοδο ελαχίστων τετραγώνων που περιγράψαμε 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 δούμε πως αυτό λειτουργεί.

Το προσαρμόσιμο δίκτυό μας έχει μία έξοδο που αναπαριστάται ως

$$o = F(i, S), \quad (3.4.1)$$

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

$$S = S_1 \oplus S_2, \quad (3.4.2)$$

(όπου \oplus αντιπροσωπεύει άμεσο άθροισμα) τέτοια ώστε η $H \circ F$ είναι γραμμική στα στοιχεία του S_2 , τότε εφαρμόζοντας την H στην εξίσωση (3.4.1), έχουμε

$$H(o) = H \circ F(i, S), \quad (3.4.3)$$

που είναι γραμμική στα στοιχεία του S_2 . Τώρα με δεδομένα στοιχεία του S_1 , μπορούμε να βάλουμε P δεδομένα εκπαίδευσης στην εξίσωση (3.4.2) και να πάρουμε μια εξίσωση πινάκων:

$$\mathbf{A}\boldsymbol{\theta} = \mathbf{y} \quad (3.4.4)$$

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) $\boldsymbol{\theta}^*$:

$$\boldsymbol{\theta}^* = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y}, \quad (3.4.5)$$

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 μπορούμε να συνδυάσουμε τη βαθμωτή ελαχιστοποίηση και τον εκτιμητή ελαχίστων τετραγώνων για να ενημερώσουμε τις παραμέτρους 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, υπολογίζουμε τις εξόδους των κόμβων στο δίκτυο από επίπεδο σε επίπεδο μέχρι να βρούμε μια αντίστοιχη σειρά στους πίνακες \mathbf{A} και \mathbf{y} στην εξίσωση (3.4.4). Αυτή η διαδικασία επαναλαμβάνεται για όλα τα ζευγάρια δεδομένων εκπαίδευσης για να σχηματίσουμε τους ολοκληρωμένους \mathbf{A} και \mathbf{y} . Έπειτα οι παράμετροι στο S_2 προσδιορίζονται από τον ψευδοαντίστροφο τύπο στην εξίσωση (3.4.5). Αφού προσδιορίζονται οι παράμετροι στο S_2 , μπορούμε να υπολογίσουμε 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 (3.3.4) και (3.3.5)] 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 (3.3.9).

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 δείξαμε προηγουμένως, παρατηρούμε 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 έχουμε

S = σύνολο συνολικών παραμέτρων,

S_1 = σύνολο αρχικών παραμέτρων (μη γραμμικών),

S_2 = σύνολο επακόλουθων παραμέτρων (γραμμικών)

στην εξίσωση (3.4.2) ενώ $H(\cdot)$ και $F(\cdot, \cdot)$ είναι η μοναδιαία συνάρτηση και η συνάρτηση του ασαφούς συστήματος συμπερασμού (fuzzy inference system), αντίστοιχα στην εξίσωση (3.4.3). 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 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 1: Δύο περάσματα στην υβριδική διαδικασία εκμάθησης για το ANFIS

ΚΕΦΑΛΑΙΟ 4: Αποτελέσματα από τη χρήση του μοντέλου ANFIS

4.1 Εισαγωγή

Στο κεφάλαιο αυτό θα αναλύσουμε τα αποτελέσματα που παίρνουμε από τη χρήση του μοντέλου ANFIS. Αρχικά θα κάνουμε μια σύντομη περιγραφή των στοιχείων που χρησιμοποιήσαμε για την εκπαίδευση και τον έλεγχο του μοντέλου ANFIS. Στη συνέχεια θα παρουσιάσουμε τα αποτελέσματα για ορισμένες περιπτώσεις και θα τα συγκρίνουμε με τα αποτελέσματα από τη χρήση των μοντέλων AR και ARMA που αναφέραμε προηγουμένως. Τέλος, θα αναφέρουμε τα συμπεράσματα από τη χρήση του μοντέλου ANFIS.

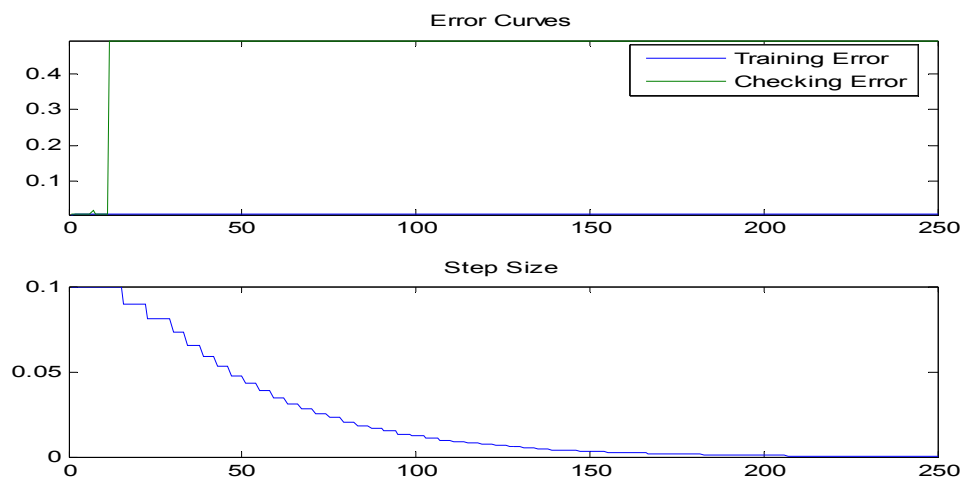
4.2 Περιγραφή του μοντέλου

Σε αυτή τη μελέτη θα χρησιμοποιηθεί ένα Προσαρμοστικό Νευροασαφές σύστημα συμπερασμού (Adaptive Neural Fuzzy Inference, ANFIS) μοντέλο για την πρόβλεψη της ισοτιμίας ευρώ-δολαρίου ένα βήμα μπροστά. Τα ασαφή συστήματα συμπερασμού (Fuzzy inference systems) που χρησιμοποιούν νευρωνικά δίκτυα προτάθηκαν ώστε να μπορέσουμε να αποφύγουμε 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 δύο εισόδους και μια έξοδο και η τιμή πρόβλεψης δίνεται από την ακόλουθη εξίσωση:

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

Τα δεδομένα που χρησιμοποιούνται σε αυτό το μοντέλο αφορούν τις καθημερινές τιμές της ισοτιμίας ευρώ/δολαρίου που παρουσιάζονται σαν χρονοσειρά. 1355 καθημερινές παρατηρήσεις χρησιμοποιούνται, από τις οποίες οι πρώτες 874 παρατηρήσεις χρησιμοποιούνται για να εκπαιδεύσουν το μοντέλο και οι 219 για να επαληθεύσουν το μοντέλο. Οι άλλες 242 παρατηρήσεις χρησιμοποιούνται προκειμένου να αξιολογηθεί η απόδοση του μοντέλου.

Όπως μπορούμε να δούμε στο ακόλουθο σχήμα το μοντέλο δίνει πολύ χαμηλό σφάλμα εκπαίδευσης και ελέγχου ενώ το μέγεθος βήματος τέθηκε στο 0.1.

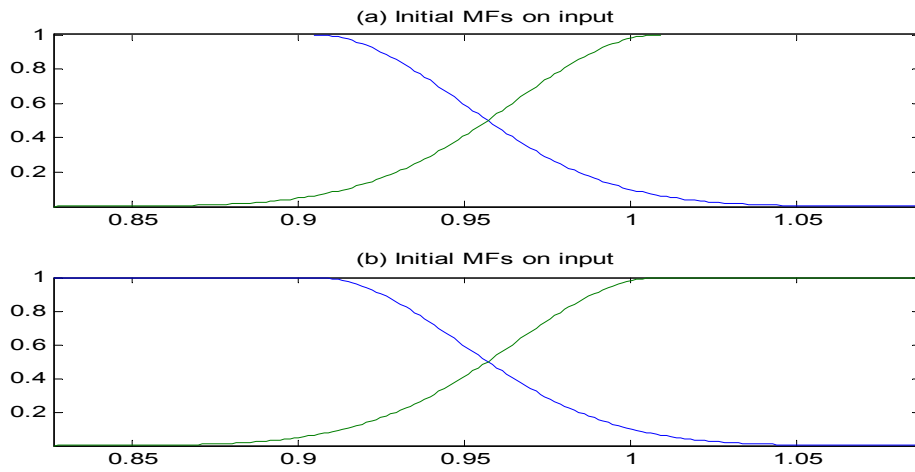


Σχήμα 1: Σφάλμα εκπαίδευσης-ελέγχου και μέγεθος βήματος

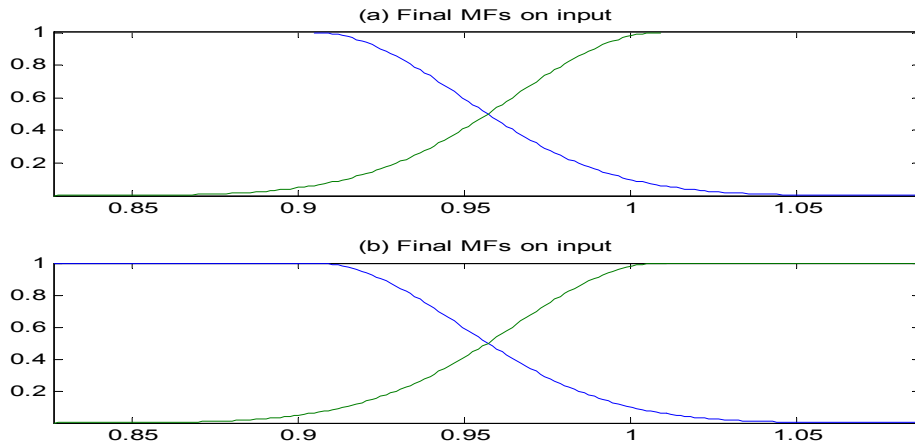
Χρησιμοποιήσαμε δύο συναρτήσεις συμμετοχής στο μοντέλο. Οι συναρτήσεις συμμετοχής που χρησιμοποιήθηκαν είναι τύπου gauss2mf. Ο τύπος μιας τέτοιας συνάρτησης για την εκπαίδευση του μοντέλου είναι:

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

Οι αρχικές και οι τελικές συναρτήσεις συμμετοχής παρουσιάζονται στα σχήματα 2 και 3.



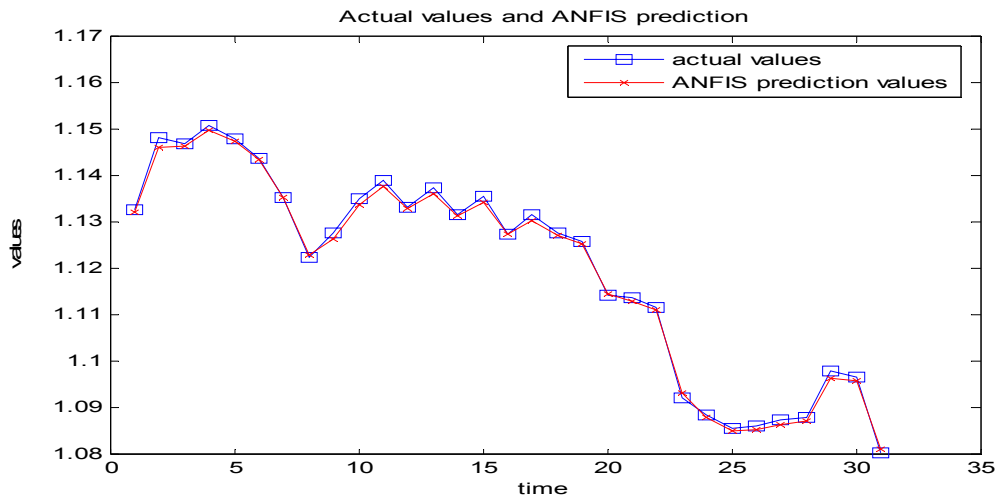
Σχήμα 2: Αρχικές συναρτήσεις συμμετοχής



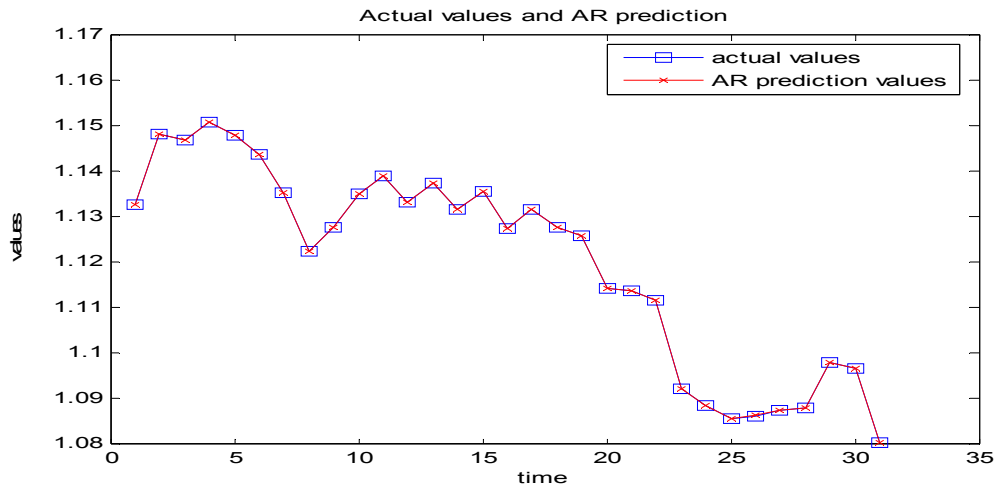
Σχήμα 3: Τελικές συναρτήσεις συμμετοχής

4.3 Αποτελέσματα

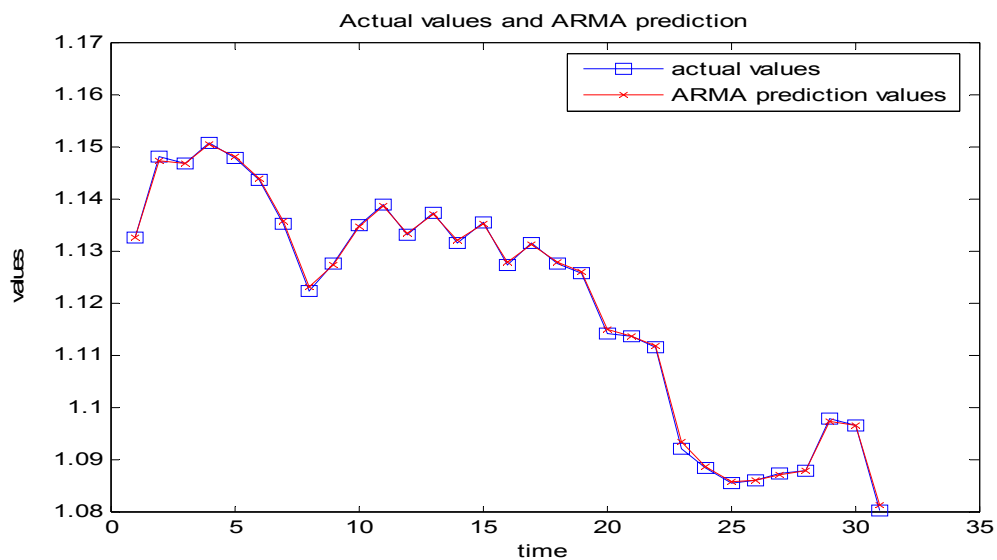
Στην παράγραφο αυτή θα παρουσιάσουμε τα αποτελέσματα από τη χρήση του μοντέλου μας και έπειτα θα τα συγκρίνουμε με τα αποτελέσματα από τη χρήση δύο συμβατικών στατιστικών μεθόδων πρόβλεψης, του αυτοπαλινδρόμενου μοντέλου (AR) και του αυτοπαλινδρόμενου κινούμενου μέσου όρου (ARMA). Για να έχουμε μια πιο ολοκληρωμένη εικόνα όσον αφορά την απόδοση του μοντέλου θα πάρουμε τρεις περιπτώσεις. Στην πρώτη περίπτωση θα πάρουμε τα δεδομένα που χρησιμοποιήσαμε για την εκπαίδευση και τον έλεγχο του μοντέλου. Στα σχήματα 4,5 και 6 παρουσιάζονται οι 30 τελευταίες τιμές.



Σχήμα 4: Πραγματικές τιμές ισοτιμίας και πρόβλεψη του ANFIS (εντός δείγματος)



Σχήμα 5: Πραγματικές τιμές ισοτιμίας και πρόβλεψη του AR (εντός δείγματος)



Σχήμα 6: Πραγματικές τιμές ισοτιμίας και πρόβλεψη του ARMA (εντός δείγματος)

Όπως βλέπουμε στο σχήμα 4 το μοντέλο ANFIS αποδίδει πολύ καλά και ακολουθεί επιτυχώς την κατεύθυνση της αλλαγής στη κίνηση συναλλαγματικής ισοτιμίας. Προκειμένου να συγκριθούν τα τρία μοντέλα θα χρησιμοποιήσουμε τα ακόλουθα τρία διαφορετικά μέτρα σφάλματος (Makridakis & Hibon, 1979).

Τετραγωνική ρίζα του μέσου τετραγωνικού σφάλματος

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

Μέσο απόλυτο σφάλμα

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

Μέσο απόλυτο ποσοστιαίο σφάλμα

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

όπου:

A_t - πραγματική τιμή

F_t - προβλεπόμενη τιμή

-

$$e_t = F_t - A_t$$

e_t -προβλεπόμενο σφάλμα

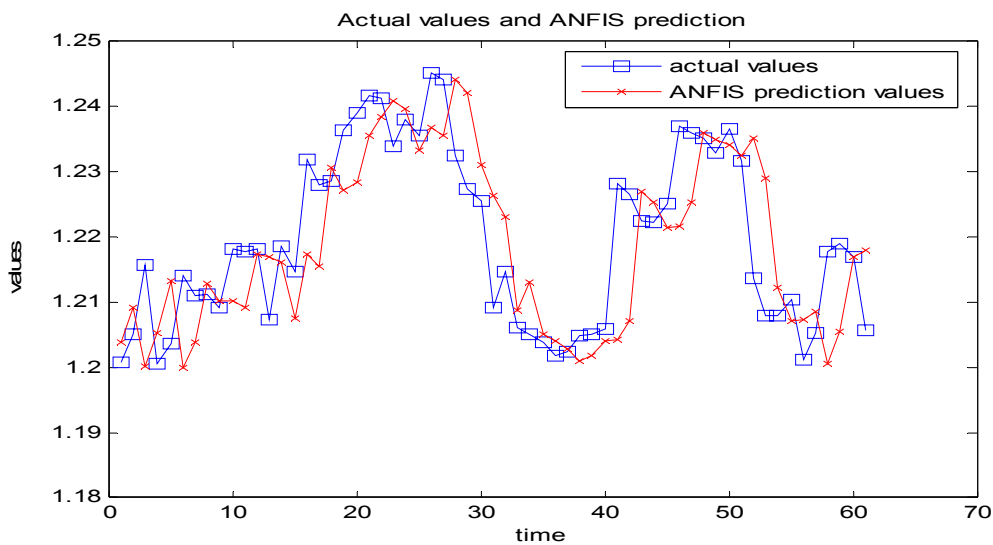
Τα αποτελέσματα παρουσιάζονται στον ακόλουθο πίνακα:

Πίνακας 1. Σφάλματα για προβλεπόμενα αποτελέσματα ένα βήμα μπροστά (εντός δείγματος)

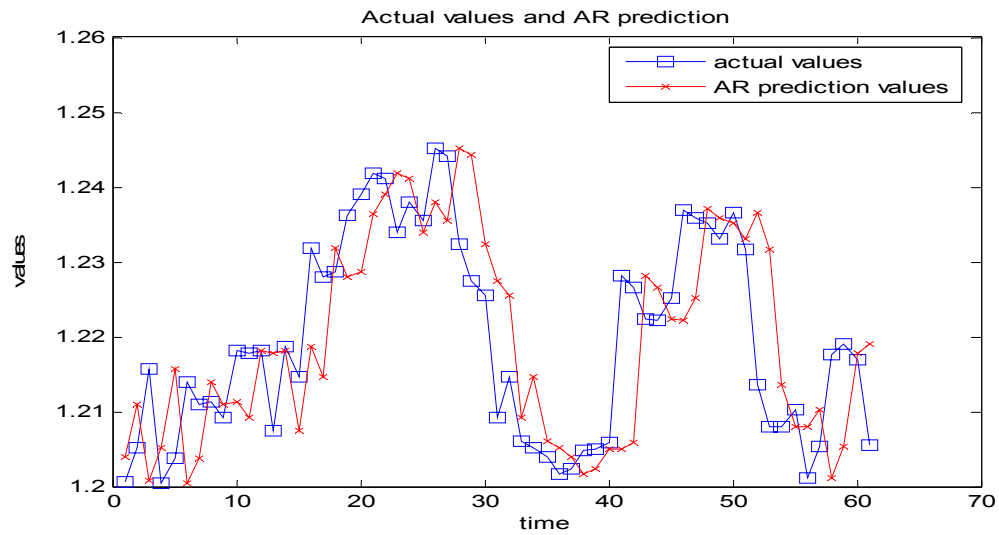
	ANFIS	AR	ARMA
RMSE	0.0006	0.0056	0.0054
MAE	0.0008	0.0080	0.0077
MAPE	0.0657	0.5937	0.5801

Από τα αποτελέσματα που παρουσιάζονται παραπάνω, μπορούμε να δούμε ότι το μοντέλο μας αποδίδει καλύτερα και έχει πολύ καλύτερα αποτελέσματα έναντι του αυτοπαλινδρόμενου μοντέλου (AR) και του αυτοπαλινδρόμενου κινούμενου μέσου (ARMA).

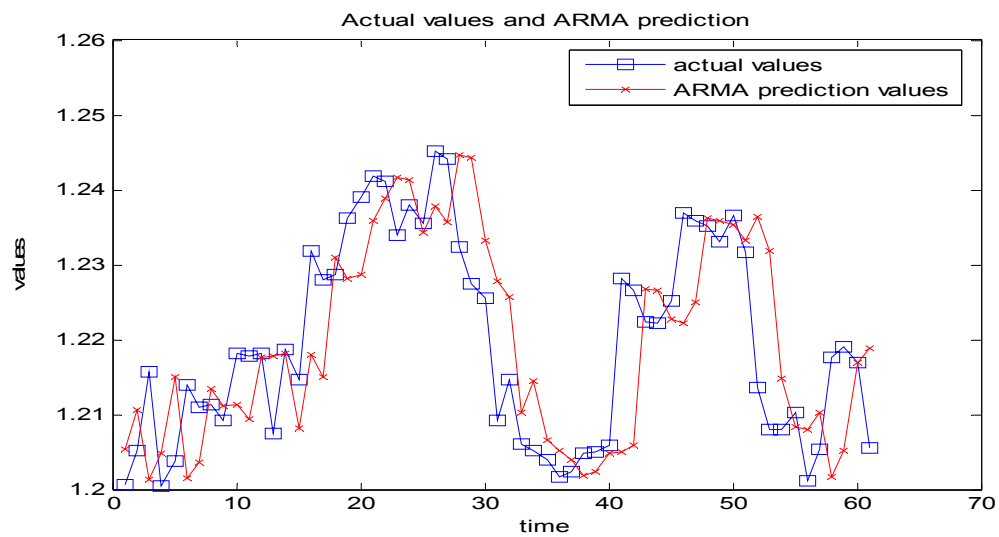
Τώρα θα δούμε πώς το μοντέλο ANFIS αποδίδει με δεδομένα διαφορετικά από εκείνα στα οποία έχει εκπαιδευθεί. Στα σχήματα 7,8 και 9 παρουσιάζουμε τις τελευταίες 60 τιμές των μοντέλων ANFIS, AR και ARMA.



Σχήμα 7: Πραγματικές τιμές ισοτιμίας και πρόβλεψη του ANFIS (εκτός δείγματος)



Σχήμα 8: Πραγματικές τιμές ισοτιμίας και πρόβλεψη του AR (εκτός δείγματος)



Σχήμα 9: Πραγματικές τιμές ισοτιμίας και πρόβλεψη του ARMA (εκτός δείγματος)

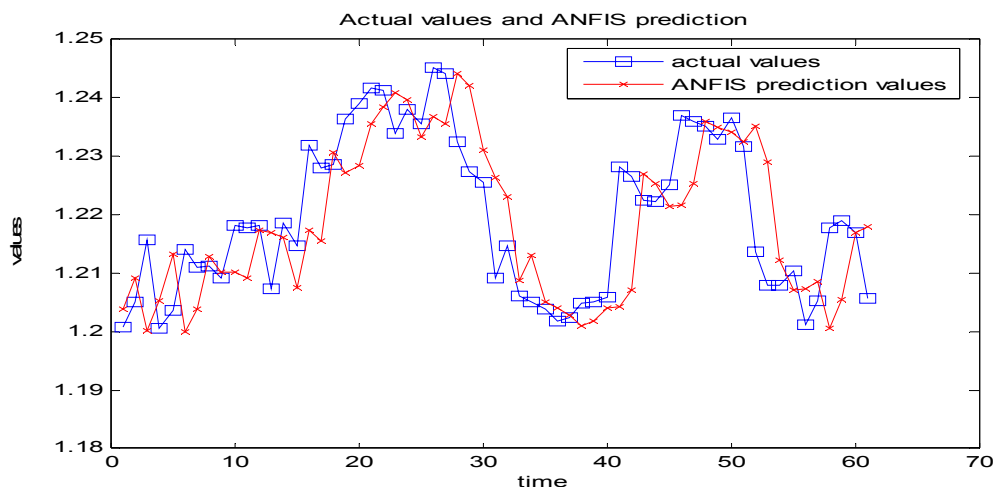
Τα αποτελέσματα παρουσιάζονται στον ακόλουθο πίνακα:

Πίνακας 2. Σφάλματα για προβλεπόμενα αποτελέσματα ένα βήμα μπροστά (εκτός δείγματος)

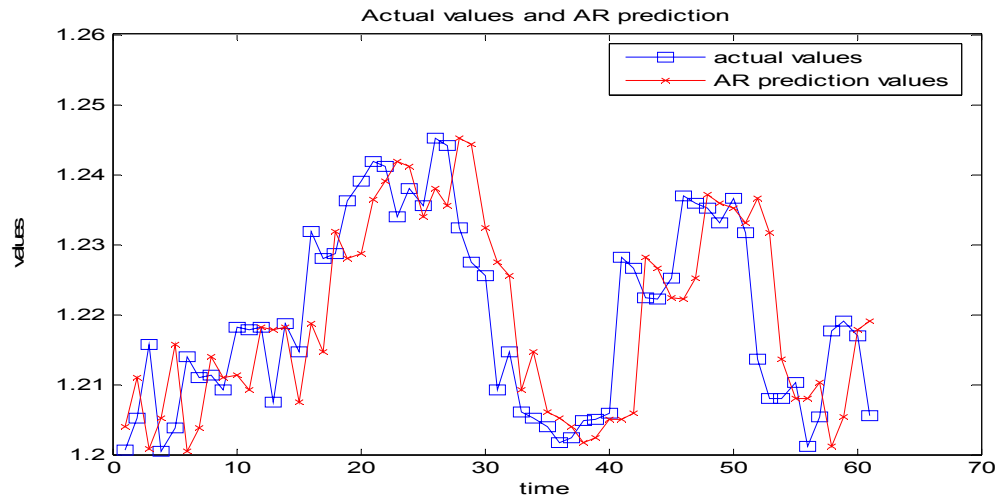
	ANFIS	AR	ARMA
RMSE	0.0087	0.0089	0.0089
MAE	0.0110	0.0114	0.0113
MAPE	0.7142	0.7383	0.7312

Όπως βλέπουμε από τα αποτελέσματα που παρουσιάζονται παραπάνω, το μοντέλο ANFIS αποδίδει ελαφρώς καλύτερα από το αυτοπαλινδρόμενο μοντέλο (AR) και το αυτοπαλινδρόμενο κινούμενου μέσου όρου (ARMA) όταν χρησιμοποιούμε δεδομένα διαφορετικά από τα δεδομένα εκπαίδευσης και ελέγχου.

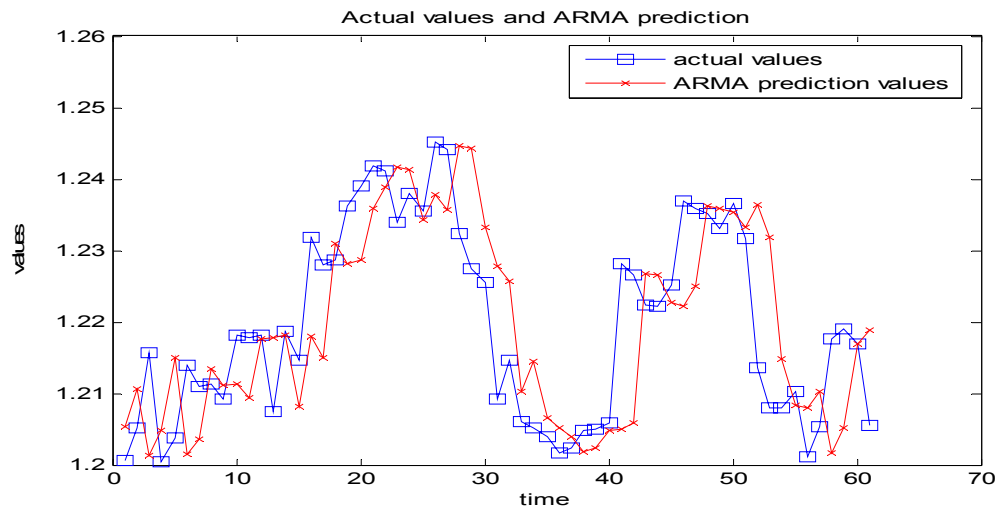
Τέλος, θα ελέγξουμε την απόδοση του μοντέλου ANFIS χρησιμοποιώντας δεδομένα εντός και εκτός δείγματος. Στα σχήματα 10,11 και 12 παρουσιάζουμε τις τελευταίες 60 τιμές των μοντέλων ANFIS, AR και ARMA.



Σχήμα 10: Πραγματικές τιμές ισοτιμίας και πρόβλεψη του ANFIS (εντός και εκτός δείγματος)



Σχήμα 11: Πραγματικές τιμές ισοτιμίας και πρόβλεψη του AR (εντός και εκτός δείγματος)



Σχήμα 12: Πραγματικές τιμές ισοτιμίας και πρόβλεψη του ARMA (εντός και εκτός δείγματος)

Τα αποτελέσματα παρουσιάζονται στον ακόλουθο πίνακα:

Πίνακας 3. Σφάλματα για προβλεπόμενα αποτελέσματα ένα βήμα μπροστά (εντός και εκτός δείγματος)

	ANFIS	AR	ARMA
RMSE	0.0072	0.0074	0.0074
MAE	0.0092	0.0095	0.0095
MAPE	0.7082	0.7329	0.7307

Τα αποτελέσματα δείχνουν ότι το μοντέλο ANFIS ξεπερνά και το αυτοπαλινδρόμικο μοντέλο (AR) και το αυτοπαλινδρόμικο κινούμενου μέσου όρου (ARMA) κατά τη χρησιμοποίηση δεδομένων εντός και εκτός δείγματος.

4.4 Συμπεράσματα

Στο κεφάλαιο αυτό δείξαμε ότι η δυναμική μεταξύ των συναλλαγματικών ισοτιμιών μπορεί να συλληφθεί από ένα σύστημα ANFIS και μπορεί να μας βοηθήσει για να προβλέψουμε τη βραχυπρόθεσμη τάση των ισοτιμιών συναλλάγματος. Τα αποτελέσματα είναι πολύ ενθαρρυντικά και έχουν δείξει ότι το μοντέλο ANFIS αποδίδει καλύτερα από τις άλλες συμβατικές μεθόδους πρόβλεψης. Αυτό, με τη σειρά του, σημαίνει ότι το μοντέλο ANFIS μπορεί επιτυχώς να χρησιμοποιηθεί στον οικονομικό στρατηγικό προγραμματισμό, δεδομένου ότι η εφαρμογή των τεχνικών πρόβλεψης στις αγορές συναλλάγματος έχει γίνει ένα βασικό μέρος στον οικονομικό στρατηγικό προγραμματισμό κατά τη διάρκεια των τελευταίων ετών.

ΒΙΒΛΙΟΓΡΑΦΙΑ

1. Box, G. E. P & Jenkins, G. M. (1970). *Time Series Analysis, Forecasting and Control*. San-Fransisco, CA: Holden-Day.
2. Brown, R. G. (1963). *Smoothing, Forecasting and Prediction*. Englewood Cliffs, NJ: Prentice Hall.
3. Chen, A. -S. & Leung, M. T. (2004). Regression neural network for error correction in foreign exchange forecasting and trading. *Computers & Operation Research*, 31, 1049-1068.
4. Chien, A. J. & Leung, M. T. (2003). A Bayesian vector error correction model for forecasting exchange rates. *Computers & Operation Research*, 30, 887-900.
5. El Shazly, M. R. & El Shazly E. Hassan (1997). Comparing the forecasting performance of neural networks and forward exchange rates. *Journal of Multinational Financial Management* 7, 345-356.
6. El Shazly, M. R. & El Shazly E. Hassan (1999). Forecasting currency prices using a genetically evolved neural network architecture. *International Review of Financial Analysis* 8, 67-82.
7. Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50, 987-1007.
8. Jang, J. S., (1993). ANFIS: Adaptive-Network-Based Fuzzy Inference Systems. *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 23(3), 665-685.

9. J. –S. R. Jang, et al. (1997). *Neuro-Fuzzy and Soft Computing-a computational approach to learning and machine intelligence*, Prentice Hall, USA.
10. M.T. Leung et al (2000). Forecasting exchange rates using general regression neural networks. *Computers & Operations Research* 27, 1093-1110
11. Li, L., Pang, W., Yu, W., Trout, M.D., (2004). Forecasting Short-Term Exchange Rates : A Recurrent Neural Network Approach. *Neural Networks in Business Forecasting. G. Peter Zhang Group Inc* 195-212.
12. Lisi, F., Schiavo, R.A. (1999). A comparison between neural networks and chaotic models for exchange rate prediction. *Computational Statistics & Data Analysis* 30, 87-102
13. Makridakis, S. & Hibon, M. (1979). Accuracy of forecasting: An empirical investigation. *Journal of the Royal Statistical Society, A*, 142, 97-145.
14. Qi, M., Wu, Y. (2003). Nonlinear prediction of exchange rates with monetary fundamentals. *Journal of Empirical Finance* 10, 623–640
15. Poddig, T., Rehkugler, H. (1996). A ‘world’ model of integrated financial markets using artificial neural networks. *Neurocomputing* 10, 251-273
16. Triggs, D. W. (1967). Exponential smoothing with adaptive response rate. *Operations Research Quarterly*, 18, 53-59.
17. Yao, J. & Tan, C.L. (2000). A case study on using neural networks to perform technical forecasting of forex. *Neurocomputing*, 34, 79-98.