

ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ

ΤΟΜΕΑΣ ΕΠΙΧΕΙΡΗΣΙΑΚΗΣ

ΕΡΕΥΝΑΣ



ΕΡΕΥΝΗΤΙΚΗ ΕΡΓΑΣΙΑ ΜΕ ΤΙΤΛΟ:
Πρόβλεψη των σημείων αναστροφής της τάσης
μετοχών με την χρήση του ANFIS και ενός
νευρωνικού δικτύου βασιζόμενο σε γενετικούς
αλγορίθμους

Επιμέλεια: Ηλίας Μπραϊμης

ΧΑΝΙΑ 2007

ΕΥΧΑΡΙΣΤΙΕΣ

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

Αρχικά, θα ήθελα να ευχαριστήσω τον κ. Σκιαδά Χρήστο, τον κ. Μυγδαλά Αθανάσιο και τον κ. Γρηγορούδη Ευάγγελο που μου έδωσαν την ευκαιρία να ασχοληθώ με ένα τόσο ενδιαφέρον θέμα.

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

Θα ήθελα ακόμα να ευχαριστήσω τους φίλους και συναδέλφους μου: Κατσαμάνη Γιώργο, Λεντζάκη Αντώνη, Μιχαηλίδη Νίκο και Στρατουδάκη Μάνο 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
Ιανουάριος 2007

ΠΕΡΙΕΧΟΜΕΝΑ

ΚΕΦΑΛΑΙΟ 1: Εισαγωγή.....	4
1.1 Σχετική βιβλιογραφία	4
1.2 Θεωρητικό υπόβαθρο του ANFIS.....	6
ΚΕΦΑΛΑΙΟ 2: Θεωρητική προσέγγιση των μοντέλων.....	14
2.1 Προσδιορισμός συστήματος.....	14
2.2 Εκτιμητής ελαχίστων τετραγώνων (Least-squares estimator).....	17
2.3 Οπισθοδρόμηση για πρωστοφοδοτούμενα δίκτυα	22
2.4 Υβριδικός αλγόριθμος εκμάθησης: Συνδυασμός βαθμωτής ελαχιστοποίησης και εκτιμητή ελαχίστων τετραγώνων	32
2.5 Νευρωνικά δίκτυα	36
2.5.1 Perceptrons. Αρχιτεκτονική και κανόνας εκμάθησης.....	36
2.5.2 Οπισθοδρομούμενα πολυεπίπεδα δίκτυα perceptron.....	40
2.5.3 Κανόνας εκμάθησης οπισθοδρόμησης	42
2.6 Γενετικοί Αλγόριθμοι.....	45
ΚΕΦΑΛΑΙΟ 3: Ανάλυση των μοντέλων	51
3.1 Εισαγωγή	51
3.2 Περιγραφή του μοντέλου ANFIS	51
3.3 Αρχιτεκτονική του ANFIS.....	57
3.4 Περιγραφή του νευρωνικού δικτύου	72
ΚΕΦΑΛΑΙΟ 4: Αποτελέσματα από τη χρήση των μοντέλων.....	74
4.1 Εισαγωγή	74
4.2 Εκπαίδευση του μοντέλου ANFIS.....	74
4.3 Αποτελέσματα του μοντέλου ANFIS.....	80
4.4 Αποτελέσματα του νευρωνικού δικτύου	83
ΚΕΦΑΛΑΙΟ 5: Συμπεράσματα.....	87
ΚΕΦΑΛΑΙΟ 6: Βιβλιογραφία.....	89

ΚΕΦΑΛΑΙΟ 1: Εισαγωγή

1.1 Σχετική βιβλιογραφία

Η πρόβλεψη των σημείων αλλαγής τάσης στο χρηματιστήριο είναι ένα δύσκολο και πολύπλοκο πρόβλημα. Η δυσκολία αυτή οφείλεται στην αβεβαιότητα και μεταβλητότητα των χρηματιστηριακών αγορών που εμποδίζουν τις προσπάθειες για μια καλή πρόβλεψη. Ο λόγος για αυτή την μεταβλητότητα είναι ότι τα χρηματιστήρια επηρεάζονται από πολλούς οικονομικούς, πολιτικούς ακόμα και ψυχολογικούς παράγοντες. Όμως εξαιτίας των αξιολογών κερδών που μπορούν να αποφέρουν σε κάποιον οι επενδύσεις σε αυτόν τον τομέα, πολλοί ερευνητές προσπάθησαν να λύσουν το πρόβλημα αυτό. Σαν αποτέλεσμα, ένας αριθμός μεθόδων προτάθηκαν με στόχο την πρόβλεψη των σημείων αλλαγής στο χρηματιστήριο.

Το 1998, οι Garcia-Ferrer και Queralto [16] 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., 1994). 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 [15] 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 [22] 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 [2] 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 στο άρθρο του 'On-line detection of turning points using non-parametric surveillance: The effect of the growth after the turn' [3] απέδειξε ότι η πιθανότητα να εντοπίσει ένα μοντέλο ανίχνευσης τα σημεία αλλαγής τάσης αυξάνει όταν αυξάνεται η άνοδος της αναμενόμενης τιμής μετά την αλλαγή.

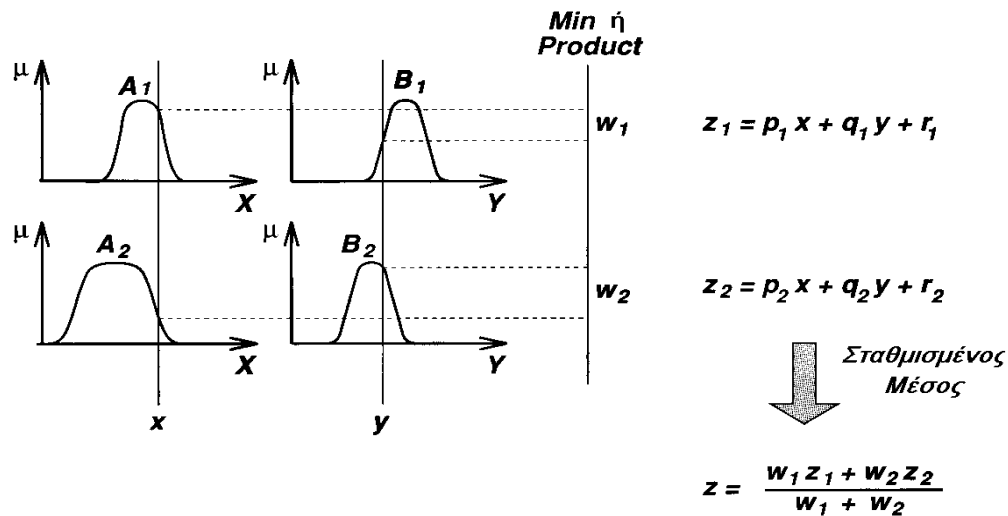
Στην παρούσα έρευνα χρησιμοποιείται το σύστημα ANFIS και ένα νευρωνικό δίκτυο βασισμένο σε γενετικούς αλγόριθμους για να προβλέψουν τα σημεία αλλαγής της κατεύθυνσης μιας μετοχής. Τόσο το νευρωνικό δίκτυο όσο και το μοντέλο ANFIS δίνουν αρκετά ικανοποιητικά αποτελέσματα.

1.2 Θεωρητικό υπόβαθρο του ANFIS

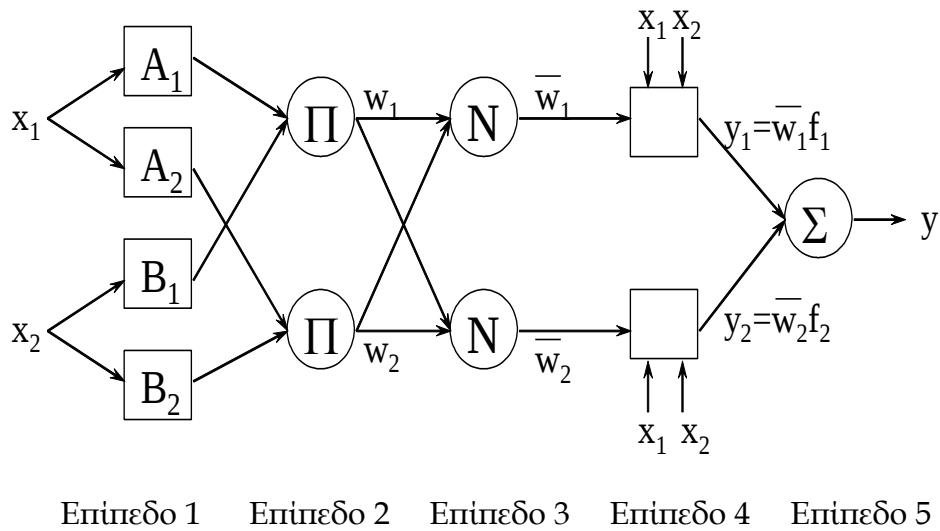
Ένα νευρο-ασαφές σύστημα ορίζεται ως ένας συνδυασμός τεχνητών νευρωνικών δικτύων (ANN) και Ασαφούς Συστήματος Συμπερασμού (Fuzzy Inference System, FIS) κατά τέτοιο τρόπο ώστε ο αλγόριθμος εκμάθησης νευρωνικών δικτύων χρησιμοποιείται για να καθορίσει τις παραμέτρους του FIS [1]. Το Προσαρμοστικό Νευροασαφές 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, 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 [20])



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 στο σχήμα 2. Οι συναρτήσεις κόμβων στο ίδιο επίπεδο είναι της ίδιας οικογένειας συναρτήσεων όπως περιγράψουμε 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 N . Ο i -ιστός 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:

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

Θεωρώντας ότι χρησιμοποιούνται όλες οι πιθανές παράμετροι, όπου ο αριθμός τους είναι συναρτήσει και του αριθμού των εισόδων και του αριθμού των συναρτήσεων συμμετοχής, τότε ο αριθμός όλων των κανόνων μπορεί να οριστεί ως:

$$Rule_n = \prod_{i=1}^{I_{n_n}} M \cdot f_i$$

και αν $premispara_n$ είναι ο αριθμός όλων των παραμέτρων που απαιτούνται για τη συνάρτηση συμμετοχής τότε ο αριθμός όλων των παραμέτρων ορίζεται ως:

$$para_n = premispara_n \sum_{i=1}^{I_{n_n}} I_{n_n} \cdot M \cdot f_i + Rule_n (I_{n_n} + 1)$$

Ο συγγραφέας A. Abraham [1], 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 [21]) 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 δείγματος εκπαίδευσης	Μέγεθος δείγματος 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].

ΚΕΦΑΛΑΙΟ 2: Θεωρητική προσέγγιση των μοντέλων

2.1 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:

- Για να προβλέψει μια συμπεριφορά του συστήματος, όπως στην πρόβλεψη χρονοσειρών και στην 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 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 $y = f(\mathbf{u}; \boldsymbol{\theta})$, 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 γίνει είναι να εφαρμοστούν τεχνικές βελτιστοποίησης για να καθοριστεί το διάνυσμα παραμέτρου $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ τέτοιο ώστε το προκύπτον μοντέλο $\hat{y} = f(\mathbf{u}; \hat{\boldsymbol{\theta}})$ 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:

1. Καθορίζεται και παραμετροποιείται μια κατηγορία μαθηματικών μοντέλων που αντιπροσωπεύουν το σύστημα που προσδιορίζεται.
2. Εκτελείται ο προσδιορισμός παραμέτρου για να επιλεχθούν οι παράμετροι που ταιριάζουν καλύτερα στο σύνολο δεδομένων εκπαίδευσης.
3. 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.**)
4. Ολοκληρώνεται η 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 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.

2.2 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 :

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

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

$$\mathbf{A}\boldsymbol{\theta} = \mathbf{y} \tag{2.2.3}$$

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

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

$\boldsymbol{\theta}$ 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 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 $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, ο οποίος καλείται εκτιμητής ελαχίστων τετραγώνων (LSE για συντομία) και ο οποίος ικανοποιεί την κανονική εξίσωση

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

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

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

Απόδειξη: Υπάρχουν διάφορες μέθοδοι διαθέσιμες στη βιβλιογραφία για την εύρεση του εκτιμητή ελαχίστων τετραγώνων για την εξίσωση (2.2.3). 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, μπορεί να γίνει επέκταση για $E(\boldsymbol{\theta})$:

$$E(\boldsymbol{\theta}) = (\mathbf{y}^T - \boldsymbol{\theta}^T \mathbf{A}^T)(\mathbf{y} - \mathbf{A}\boldsymbol{\theta}) = \boldsymbol{\theta}^T \mathbf{A}^T \mathbf{A} \boldsymbol{\theta} - 2\mathbf{y}^T \mathbf{A} \boldsymbol{\theta} + \mathbf{y}^T \mathbf{y} \quad (2.2.9)$$

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

$$\frac{\partial E(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 2\mathbf{A}^T \mathbf{A} \boldsymbol{\theta} - 2\mathbf{A}^T \mathbf{y}. \quad (2.2.10)$$

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

$$\mathbf{A}^T \mathbf{A} \hat{\boldsymbol{\theta}} = \mathbf{A}^T \mathbf{y}. \quad (2.2.11)$$

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

$$\hat{\boldsymbol{\theta}} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y}. \quad (2.2.12)$$

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

$$E(\hat{\boldsymbol{\theta}}) = (\mathbf{y} - \mathbf{A}\hat{\boldsymbol{\theta}})^T (\mathbf{y} - \mathbf{A}\hat{\boldsymbol{\theta}}) = \mathbf{y}^T \mathbf{y} - \mathbf{y}^T \mathbf{A} (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y}. \quad (2.2.13)$$

Όμως, αν ο $\mathbf{A}^T \mathbf{A}$ είναι μη αντιστρέψιμος, τότε ο LSE δεν είναι μοναδικός και πρέπει να εφαρμοστεί η έννοια του γενικευμένου αντιστρόφου για να βρεθεί ο $\hat{\boldsymbol{\theta}}$. Χωρίς απώλεια γενικότητας, θα υποθεθεί 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 κάθε όρος σφάλματος να σταθμιστεί διαφορετικά. Συγκεκριμένα, ας υποθεθεί ότι ο $\mathbf{W}$ είναι ο επιθυμητός πίνακας βαρύτητας, ο οποίος είναι συμμετρικός και θετικά ορισμένος. Τότε το σταθμισμένο τετραγωνικό σφάλμα είναι

$$E_{\mathbf{W}}(\boldsymbol{\theta}) = (\mathbf{y} - \mathbf{A}\boldsymbol{\theta})^T \mathbf{W}(\mathbf{y} - \mathbf{A}\boldsymbol{\theta}). \quad (2.2.14)$$

Ελαχιστοποιώντας το $E_{\mathbf{W}}(\boldsymbol{\theta})$ ως προς $\boldsymbol{\theta}$ παράγεται ο σταθμισμένος εκτιμητής ελαχίστων τετραγώνων $\hat{\boldsymbol{\theta}}_{\mathbf{W}}$:

$$\hat{\boldsymbol{\theta}}_{\mathbf{W}} = (\mathbf{A}^T \mathbf{W} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{W} \mathbf{y} \quad (2.2.15)$$

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

2.3 Οπισθοδρόμηση για πρωστοτροφοδοτούμενα δίκτυα

Η ενότητα αυτή παρουσιάζει ένα βασικό κανόνα εκμάθησης για προσαρμοσίμα 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 υποτεθεί 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 υποτεθεί 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 α, 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:

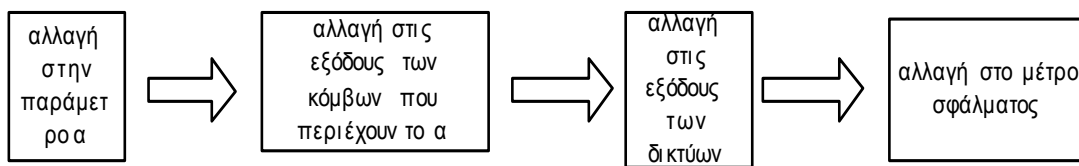
$$E_p = \sum_{k=1}^{N(l)} (d_k - x_{L,k})^2, \quad (2.3.2)$$

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 υποθεθεί 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 παραγωγού και της συνηθισμένης μερικής παραγωγού 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 $1 \leq i \leq N(l)$], μπορούν να

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

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

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

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

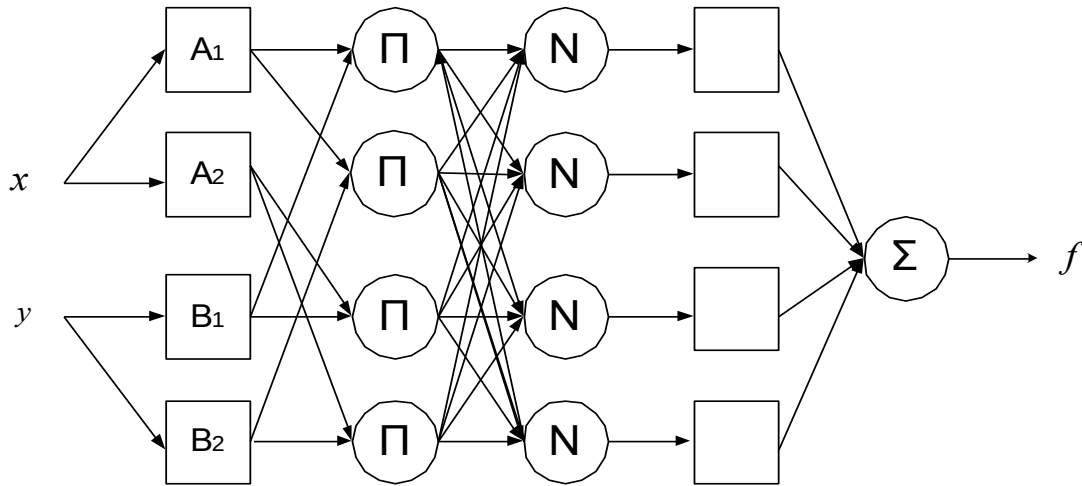
ή

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

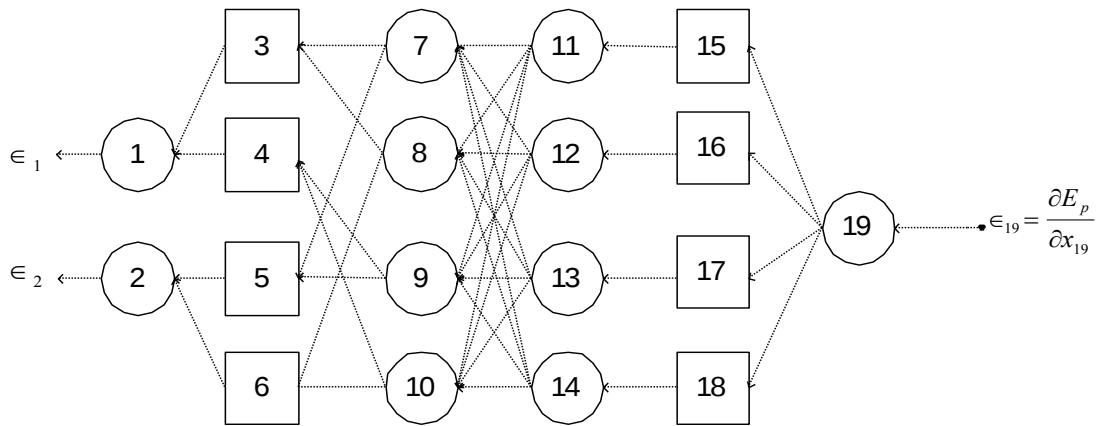
όπου ο πρώτος όρος δείχνει μια άμεση επίδραση του x_i στο E_p μέσω της άμεσης διαδρομής από τον κόμβο i στον κόμβο $n+1$ και κάθε όρος παραγώγου στο άθροισμα δείχνει την έμμεση επίδραση του x_i στο E_p . Μόλις βρεθεί το σήμα σφάλματος για κάθε κόμβο, τότε το διάνυσμα κλίσης για τις παραμέτρους παράγεται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

επίπεδο εξόδου ως εισοδοι 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:

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

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

$$\epsilon_2 = \frac{\partial^+ E_p}{\partial x_2} = \frac{\partial^+ E_p}{\partial x_6} \frac{\partial f_6}{\partial x_2} + \frac{\partial^+ E_p}{\partial x_5} \frac{\partial f_5}{\partial x_2} = \epsilon_6 \frac{\partial f_2}{\partial x_2} + \epsilon_5 \frac{\partial f_5}{\partial x_2},$$

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

2.4 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 πώς 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_1 , 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

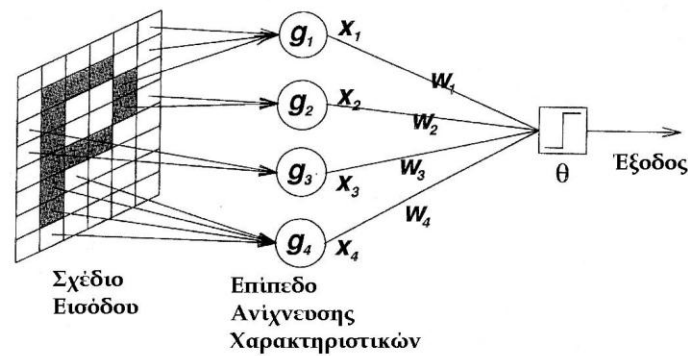
Πίνακας 2.1: Δύο περάσματα στην υβριδική διαδικασία εκμάθησης για το ANFIS

2.5 Νευρωνικά δίκτυα

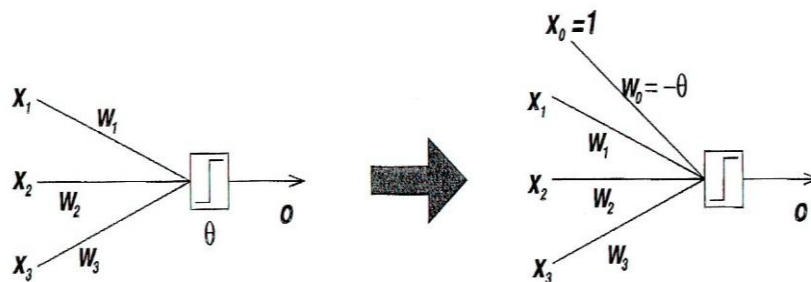
2.5.1 Perceptrons. Αρχιτεκτονική και κανόνες εκμάθησης

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 [24]. Αργότερα ο Rosenblatt [26] 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 :

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

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 $f(\cdot)$ είναι η 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

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 [26] 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 [25], στην οποία ανέλυσαν τα δίκτυα perceptron διεξοδικά και συμπέραναν ότι τα δίκτυα perceptron ενός επιπέδου μπορούν μόνο να χρησιμοποιηθούν για πάρα πολύ απλά προβλήματα.

2.5.2 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:

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

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

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

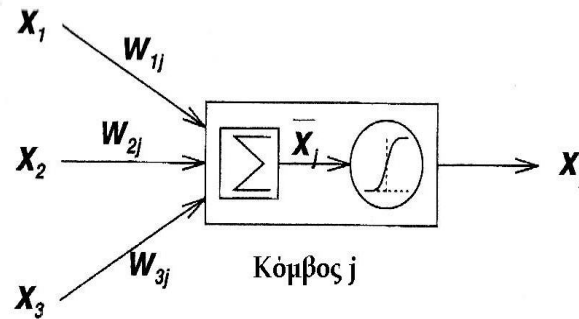
Η υπερβολική εφαπτομένη (hyperbolic tangent) και η λογιστική προσεγγίζουν τη σιγμοειδή (sigmoid) και τη βηματική συνάρτηση, αντίστοιχα. Μερικές φορές αυτές οι δύο συναρτήσεις ενεργοποίησης αναφέρονται ως squashing συναρτήσεις επειδή οι είσοδοι για αυτές τις συναρτήσεις «στριμώνονται» στο πεδίο $[0,1]$ 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.

2.5.3 Κανόνας εκμάθησης οπισθοδρόμησης

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



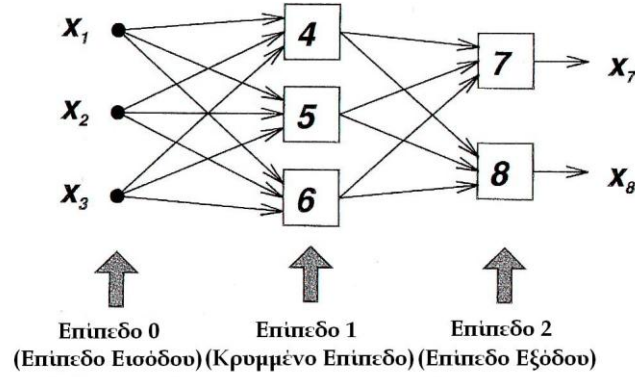
Σχήμα 2.4: Κόμβος j ενός οπισθοδρομούμενου πολυεπίπεδου δικτύου perceptron

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.

2.6 Γενετικοί Αλγόριθμοι

Οι γενετικοί αλγόριθμοι είναι μέθοδοι στοχαστικής βελτιστοποίησης χωρίς παραγώγους που βασίζονται στις αρχές της φυσικής επιλογής και των εξελικτικών διαδικασιών. Προτάθηκαν και ερευνήθηκαν για πρώτη φορά από τον John Holland στο πανεπιστήμιο του Μίσιγκαν το 1975 [19]. Σαν ένα εργαλείο βελτιστοποίησης, οι γενετικοί αλγόριθμοι χρησιμοποιούνται σε πολλές εφαρμογές. Οι λόγοι οι οποίοι έκαναν τους γενετικούς αλγόριθμους τόσο δημοφιλείς είναι ότι δεν εξαρτώνται από παραγώγους και ότι έχουν τα ακόλουθα χαρακτηριστικά:

- Οι γενετικοί αλγόριθμοι είναι διαδικασίες παράλληλης αναζήτησης που μπορούν να χρησιμοποιηθούν από μηχανές παράλληλης επεξεργασίας για να επιταχύνουν κατά πολύ τις λειτουργίες τους.
- Οι γενετικοί αλγόριθμοι μπορούν να εφαρμοστούν και σε συνεχή και σε διακριτά προβλήματα βελτιστοποίησης.
- Οι γενετικοί αλγόριθμοι είναι στοχαστικοί και επομένως πολύ δύσκολα μπορούν να παγιδευτούν σε τοπικά ελάχιστα, τα οποία αναπόφευκτα εμφανίζονται σε κάθε πρακτική εφαρμογή βελτιστοποίησης.
- Η ευελιξία των γενετικών αλγορίθμων διευκολύνει τόσο τη δομή όσο και τον προσδιορισμό των παραμέτρων σε πολύπλοκα μοντέλα όπως τα νευρωνικά δίκτυα και τα συστήματα ασαφούς συμπερασμού.

Οι γενετικοί αλγόριθμοι κωδικοποιούν κάθε σημείο σε ένα σύνολο παραμέτρων (ή λύσεων) σε μία αλυσίδα δυαδικών ψηφίων που ονομάζεται χρωμόσωμα, και κάθε σημείο συνδέεται με μία τιμή ταιριάσματος η οποία, για μεγιστοποίηση, είναι συνήθως ίση με την αντικειμενική συνάρτηση που αντιστοιχεί στο σημείο. Αντί για ένα απλό σημείο, οι γενετικοί αλγόριθμοι συνήθως κρατούν ένα σετ σημείων σαν πληθυσμό (ή δεξαμενή γονιδίων), ο οποίος στη συνέχεια εξελίσσεται συνεχώς προς μία καλύτερη συνολική τιμή ταιριάσματος. Σε κάθε γενιά, ο γενετικός αλγόριθμος κατασκευάζει ένα καινούργιο πληθυσμό χρησιμοποιώντας γενετικούς διαχειριστές (operators) όπως η διασταύρωση και η μετάλλαξη. Τα στοιχεία με μεγαλύτερες τιμές ταιριάσματος είναι πιο πιθανό να επιβιώσουν και να συμμετέχουν σε διαδικασίες διασταύρωσης. Μετά από έναν αριθμό γενεών, ο πληθυσμός περιέχει στοιχεία με καλύτερες τιμές ταιριάσματος. Αυτό είναι ανάλογο με τα μοντέλα του Δαρβίνου για την εξέλιξη με τυχαία μετάλλαξη και φυσική επιλογή. Οι γενετικοί αλγόριθμοι και οι παραλλαγές τους αναφέρονται μερικές φορές ως μέθοδοι βελτιστοποίησης βασισμένες σε πληθυσμούς που βελτιώνουν την απόδοση αναβαθμίζοντας ολόκληρους πληθυσμούς αντί για απλά στοιχεία.

Σημαντικά συστατικά των γενετικών αλγορίθμων είναι τα σχέδια κωδικοποίησης, η αξιολόγηση ταιριάσματος, η επιλογή γονέων, οι διαχειριστές διασταυρώσεων και οι διαχειριστές μεταλλάξεων. Αυτά εξηγούνται στη συνέχεια.

Σχέδια κωδικοποίησης Αυτά μετατρέπουν σημεία στο σύνολο λύσεων σε αλυσίδες δυαδικής απεικόνισης. Για παράδειγμα, ένα σημείο (11, 6, 9) σε ένα τρισδιάστατο σύνολο λύσεων αναπαριστάται σαν μια δυαδική αλυσίδα:

$$\underbrace{1011}_{11} \underbrace{0110}_{6} \underbrace{1001}_{9}$$

όπου κάθε τιμή συντεταγμένης κωδικοποιείται σαν ένα γονίδιο που αποτελείται από τέσσερα δυαδικά ψηφία χρησιμοποιώντας δυαδική κωδικοποίηση. Άλλα σχέδια κωδικοποίησης, όπως η κωδικοποίηση Gray, μπορούν επίσης να χρησιμοποιηθούν και όταν είναι απαραίτητο μπορούν να γίνουν διευθετήσεις για την κωδικοποίηση αρνητικών ή δεκαδικών. Τα σχέδια κωδικοποίησης παρέχουν ένα τρόπο μεταφοράς της γνώσης που ισχύει για κάθε πρόβλημα μέσα στο γενετικό αλγόριθμο και παίζουν ένα σημαντικό ρόλο στον καθορισμό της απόδοσης του γενετικού αλγορίθμου. Επιπλέον, γενετικοί διαχειριστές, όπως η διασταύρωση και η μετάλλαξη, μπορούν και πρέπει να σχεδιάζονται μαζί με το σχέδιο κωδικοποίησης για κάθε ξεχωριστή εφαρμογή.

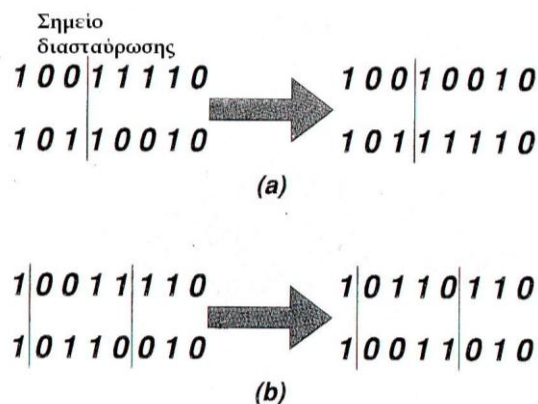
Αξιολόγηση ταιριάσματος Το πρώτο βήμα μετά τη δημιουργία μιας γενιάς είναι ο υπολογισμός της τιμής ταιριάσματος κάθε στοιχείου του πληθυσμού. Για προβλήματα μεγιστοποίησης, η τιμή ταιριάσματος f_i του i -οστού στοιχείου είναι συνήθως η αντικειμενική συνάρτηση που αντιστοιχεί σε αυτό το στοιχείο. Συνήθως χρειάζονται τιμές ταιριάσματος που είναι θετικές. Μια άλλη προσέγγιση είναι να χρησιμοποιηθεί κατάταξη των στοιχείων του πληθυσμού σαν τις τιμές ταιριάσματος τους. Το πλεονέκτημα αυτής της προσέγγισης είναι ότι η αντικειμενική συνάρτηση δεν χρειάζεται να είναι ακριβής εφόσον μπορεί να παρέχει σωστές πληροφορίες κατάταξης.

Επιλογή Μετά την αξιολόγηση πρέπει να δημιουργηθεί ένας νέος πληθυσμός από τη τρέχουσα γενιά. Η διαδικασία επιλογής καθορίζει ποιοι γονείς συμμετέχουν στην παραγωγή απογόνων για την επόμενη γενιά και είναι ανάλογη με την επιβίωση του πιο κατάλληλου στη φυσική επιλογή. Συνήθως τα στοιχεία επιλέγονται για διασταύρωση με μία πιθανότητα επιλογής που εξαρτάται από τις τιμές ταιριάσματος τους. Ο πιο συνηθισμένος τρόπος για να

γίνει η επιλογή είναι να τεθεί η πιθανότητα επιλογής ίση με $\frac{f_i}{\sum_{k=1}^n f_k}$, όπου n

είναι το μέγεθος του πληθυσμού. Ο σκοπός αυτής της μεθόδου επιλογής είναι να επιτρέψει σε στοιχεία με τιμές ταιριάσματος πάνω από το μέσο όρο να πολλαπλασιαστούν και να αντικαταστήσουν τα στοιχεία με τιμές ταιριάσματος κάτω από το μέσο όρο.

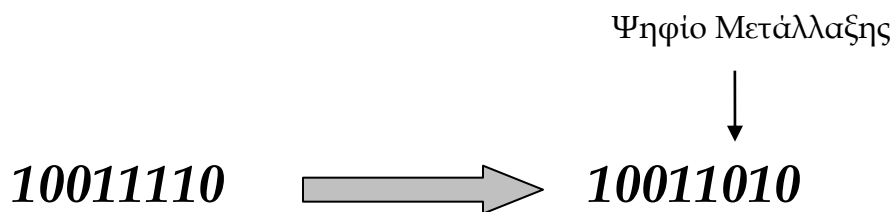
Διασταύρωση Η διασταύρωση χρησιμοποιείται για να παραχθούν νέα χρωμοσώματα με την ελπίδα να διατηρηθούν τα καλά χαρακτηριστικά από την προηγούμενη γενιά. Η διασταύρωση εφαρμόζεται συνήθως σε επιλεγμένα ζεύγη γονέων με πιθανότητα ίση με ένα δεδομένο ρυθμό διασταύρωσης. Η **διασταύρωση ενός σημείου** είναι ο πιο βασικός διαχειριστής διασταύρωσης, όπου επιλέγεται τυχαία ένα σημείο διασταύρωσης στο γενετικό κώδικα και τα χρωμοσώματα των γονέων αλλάζουν θέση σε αυτό το σημείο. **Στη διασταύρωση δύο σημείων**, επιλέγονται δύο σημεία διασταύρωσης και το μέρος της αλυσίδας του χρωμοσώματος μεταξύ αυτών των δύο σημείων αλλάζει για να παραχθούν δύο νέα γονίδια. Ομοίως ορίζεται η διασταύρωση n-σημείων.



Σχήμα 2.6: Διαχειριστές διασταύρωσης: α) διασταύρωση ενός σημείου
β) διασταύρωση δύο σημείων

Ο σκοπός της διασταύρωσης είναι όμοιος με αυτόν του ζευγαρώματος στη φυσική εξελικτική διαδικασία, στην οποία οι γονείς περνούν τμήματα των δικών τους χρωμοσωμάτων στα παιδιά τους. Επομένως, μερικά παιδιά μπορούν να ξεπεράσουν τους γονείς τους αν πάρουν «καλά» γονίδια ή γενετικά χαρακτηριστικά από τους γονείς τους.

Μετάλλαξη Η διασταύρωση εκμεταλλεύεται τρέχουσες δυνατότητες των γονιδίων, αλλά αν ο πληθυσμός δεν περιέχει όλες τις κωδικοποιημένες πληροφορίες που απαιτούνται για την λύση ενός συγκεκριμένου προβλήματος, τότε κανένας αριθμός συνδυασμών γονιδίων δεν μπορεί να παράξει επιθυμητό αποτέλεσμα. Για τον λόγο αυτό υπάρχει ένας διαχειριστής μετάλλαξης ικανός να παράγει από μόνος του νέα χρωμοσώματα. Ο πιο συνηθισμένος τρόπος για να γίνει μετάλλαξη είναι να αλλάξει ένα ψηφίο με πιθανότητα ίση με ένα δεδομένο πολύ μικρό ρυθμό μετάλλαξης. Ένας διαχειριστής μετάλλαξης μπορεί να εμποδίσει κάθε απλό ψηφίο απ'το να συγκλίνει σε μία τιμή μέσα σε ολόκληρο τον πληθυσμό και ακόμα μπορεί να εμποδίσει ολόκληρο τον πληθυσμό απ'το να συγκλίνει και να εγκλωβιστεί σε κάποιο τοπικό ελάχιστο. Ο ρυθμός μετάλλαξης επιλέγεται συνήθως μικρός ώστε να μη χάνονται τα καλά χρωμοσώματα που προκύπτουν από τη διασταύρωση. Αν ο ρυθμός μετάλλαξης είναι υψηλός (πάνω από 0.1), η απόδοση του γενετικού αλγορίθμου πλησιάζει αυτή μιας πρωτόγονης τυχαίας αναζήτησης.



Σχήμα 2.7: Διαχειριστής Μετάλλαξης

Στη συνέχεια περιγράφεται ένας απλός γενετικός αλγόριθμος για προβλήματα μεγιστοποίησης.

Βήμα 1: Αρχικοποίηση ενός πληθυσμού με τυχαία παραγόμενα στοιχεία και αξιολόγηση της τιμής ταιριάσματος κάθε στοιχείου.

Βήμα 2:

- α) Επιλογή δύο στοιχείων από τον πληθυσμό με πιθανότητες ανάλογες με τις τιμές ταιριάσματος τους.
- β) Εφαρμογή διασταύρωσης με πιθανότητα ίση με το ρυθμό διασταύρωσης.
- γ) Εφαρμογή μετάλλαξης με πιθανότητα ίση με το ρυθμό μετάλλαξης.
- δ) Επανάληψη των βημάτων α) έως δ) έως ότου έχουν παραχθούν αρκετά στοιχεία για να σχηματιστεί η επόμενη γενιά

Βήμα 3: Επανάληψη των βημάτων 1 και 2 έως ότου έχει εκπληρωθεί ένα κριτήριο διακοπής.

ΚΕΦΑΛΑΙΟ 3: Ανάλυση των μοντέλων

3.1 Εισαγωγή

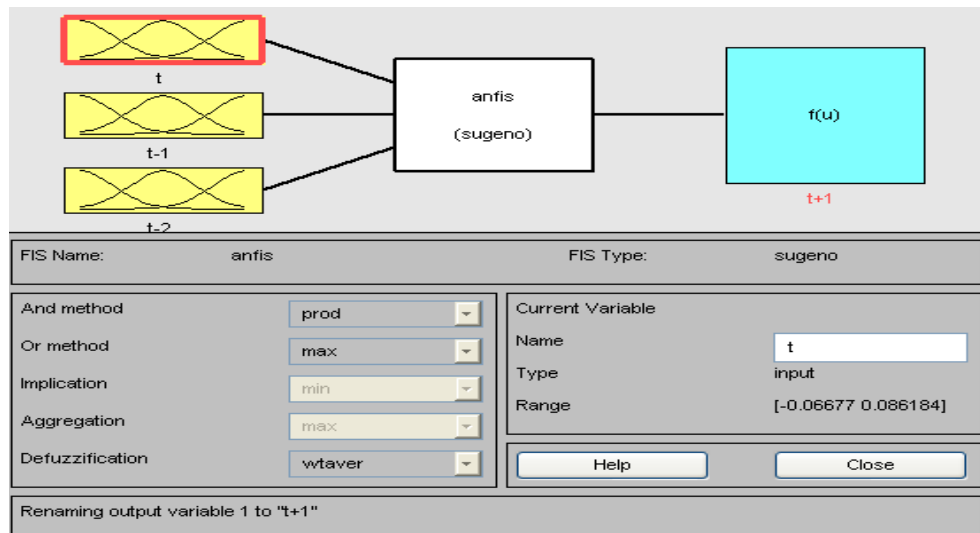
Στο κεφάλαιο αυτό θα γίνει ανάλυση των μοντέλων που χρησιμοποιήθηκαν στην έρευνά μας. Αρχικά θα γίνει μια περιγραφή του μοντέλου ANFIS και θα παρουσιαστεί η δομή του. Στη συνέχεια θα εξεταστεί κάθε επίπεδο του μοντέλου ξεχωριστά και θα παρουσιαστούν όλοι οι υπολογισμοί που γίνονται σε κάθε επίπεδο. Ακόμα, θα παρουσιαστεί 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. Τέλος, θα γίνει μια περιγραφή του νευρωνικού δικτύου που χρησιμοποιήθηκε.

3.2 Περιγραφή του μοντέλου ANFIS

Το μοντέλο που χρησιμοποιήθηκε αποτελείται από τρεις εισόδους και μία έξοδο. Το μοντέλο παίρνει σαν εισόδους συνεχόμενες τιμές της μετοχής της Εθνικής Τράπεζας και σαν έξοδο δίνει την πρόβλεψη της τιμής της μετοχής για την επόμενη ημέρα. Όλες οι τιμές της μετοχής έχουν μετασχηματιστεί στη μορφή του ρυθμού μεταβολής σύμφωνα με τη σχέση:

$$\frac{\text{τιμή}(k) - \text{τιμή}(k-1)}{\text{τιμή}(k-1)}$$

Η δομή του μοντέλου παρουσιάζεται στο ακόλουθο σχήμα:



Σχήμα 3.1: Δομή του μοντέλου

Από το σχήμα φαίνεται 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 συνάρτηση αρκεί να μπορεί να περιγράψει κατάλληλα την έξοδο του μοντέλου εντός της ασαφούς περιοχής που καθορίζεται από τα αίτια του κανόνα. Εδώ εξετάζεται ένα πρώτου βαθμού ασαφές μοντέλο τύπου Sugeno

το οποίο σημαίνει ότι το $f(x,y)$ είναι ένα πρώτου βαθμού πολυώνυμο. Στο μοντέλο ισχύουν οι παρακάτω κανόνες:

Αν (t) είναι small και (t-1) είναι small και (t-2) είναι small τότε (t+1) είναι 1

Αν (t) είναι small και (t-1) είναι small και (t-2) είναι large τότε (t+1) είναι 2

Αν (t) είναι small και (t-1) είναι large και (t-2) είναι small τότε (t+1) είναι 3

Αν (t) είναι small και (t-1) είναι large και (t-2) είναι large τότε (t+1) είναι 4

Αν (t) είναι large και (t-1) είναι small και (t-2) είναι small τότε (t+1) είναι 5

Αν (t) είναι large και (t-1) είναι small και (t-2) είναι large τότε (t+1) είναι 6

Αν (t) είναι large και (t-1) είναι large και (t-2) είναι small τότε (t+1) είναι 7

Αν (t) είναι large και (t-1) είναι large και (t-2) είναι large τότε (t+1) είναι 8

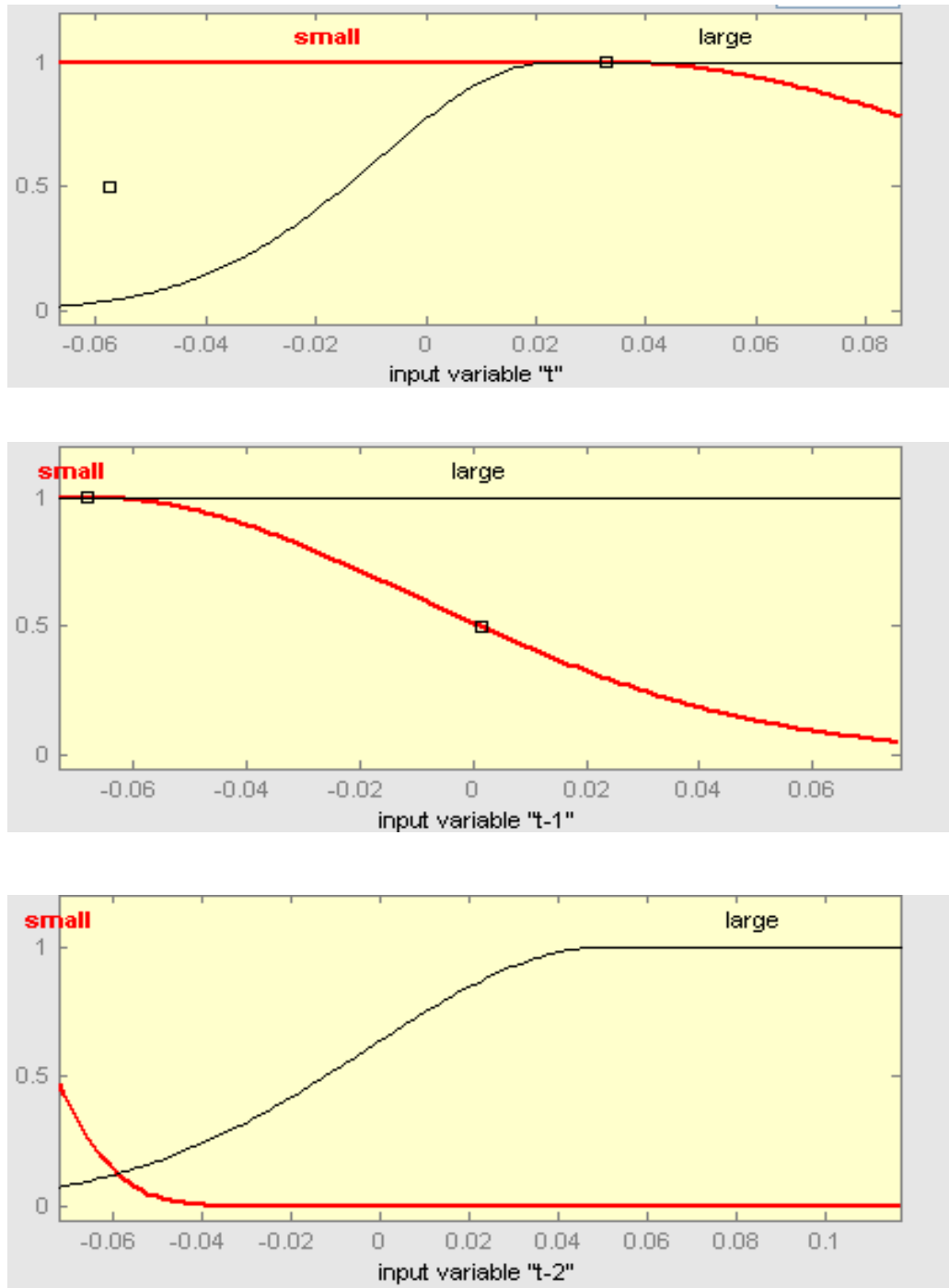
Για παράδειγμα, αν ο ρυθμός μεταβολής τη χρονική στιγμή (t) είναι small και ο ρυθμός μεταβολής τη χρονική στιγμή (t-1) είναι small και ο ρυθμός μεταβολής τη χρονική στιγμή (t-2) είναι small τότε ο ρυθμός μεταβολής τη χρονική στιγμή (t+1) είναι $O_{4,1} = \bar{w}_1(p_1 * x + q_1 * y + s_1 * z_1 + r_1)$.

Λεπτομερέστερη αναφορά στους κανόνες και πως αυτοί χρησιμοποιούνται από το μοντέλο για την πραγματοποίηση των υπολογισμών θα γίνει στην ενότητα 3.3.

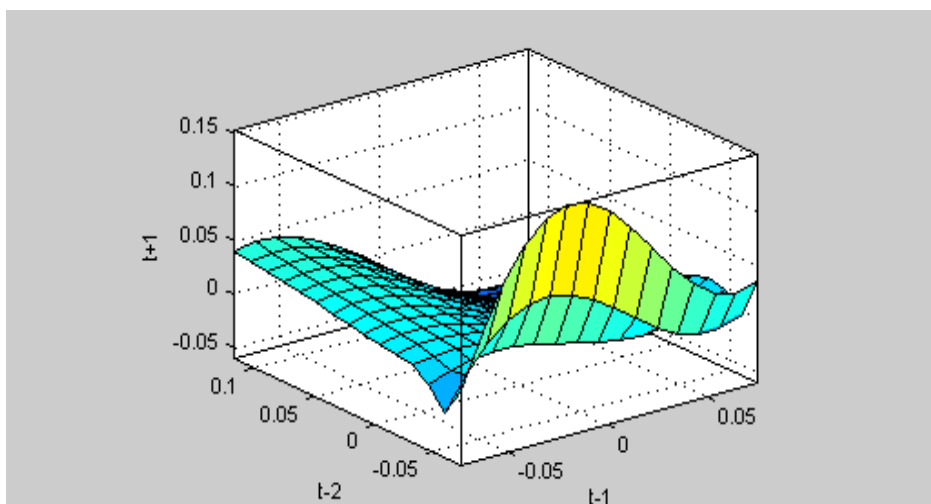
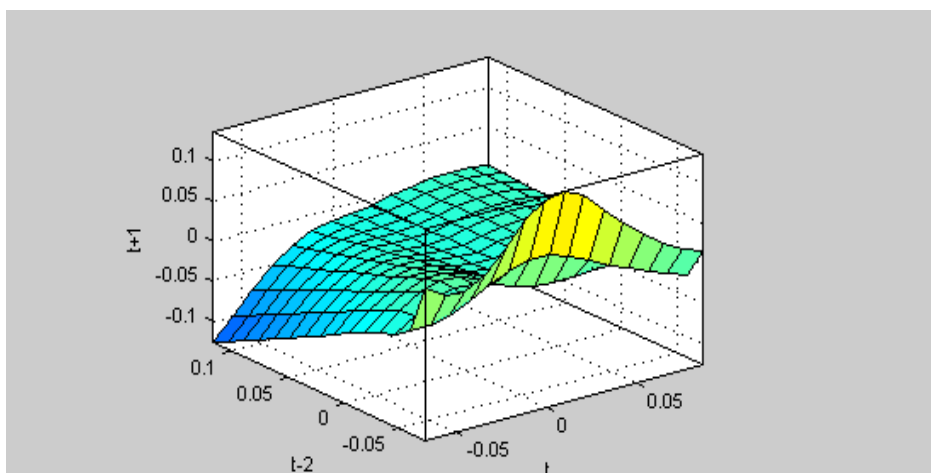
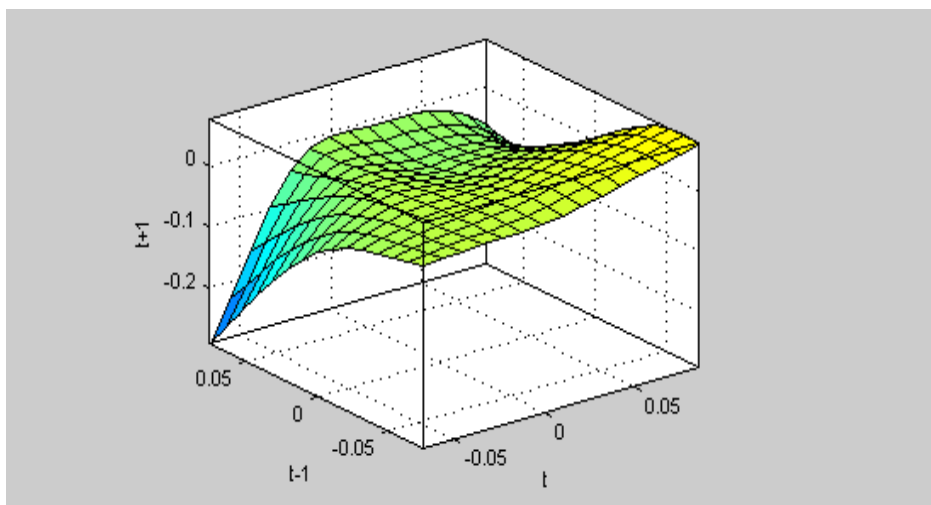
Οι συναρτήσεις συμμετοχής των εισόδων (t), (t-1) και (t-2) είναι οι ίδιες, αφού και οι τρεις συναρτήσεις είναι τύπου gauss2mf. Η εξίσωση των συναρτήσεων τύπου gauss2mf είναι:

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

όπου τα a_i, b_i, c_i είναι το σύνολο των παραμέτρων. Στο σχήμα 3.2 φαίνονται οι συναρτήσεις συμμετοχής των εισόδων (t), (t-1) και (t-2), ενώ στο σχήμα 3.3 είναι οι προκύπτουσες επιφάνειες εισόδου-εξόδου:

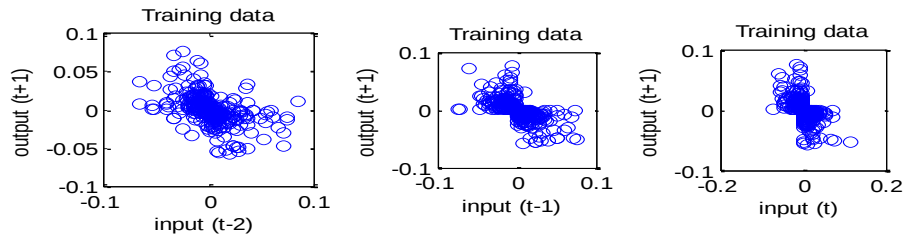


Σχήμα 3.2: Συναρτήσεις συμμετοχής

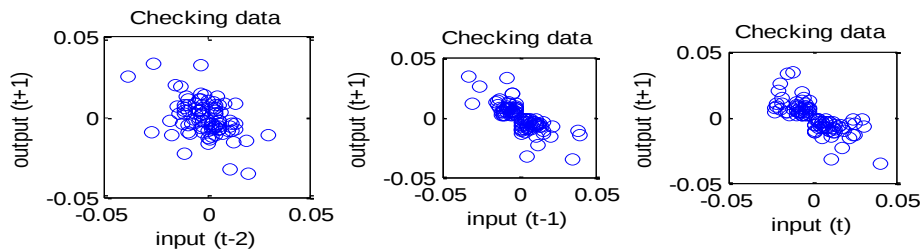


Σχήμα 3.3: Συνολικές επιφάνειες εισόδου-εξόδου

Στο σχήμα 3.3 παρουσιάζονται οι τρισδιάστατες επιφάνειες του ασαφούς μοντέλου. Όπως φαίνεται οι επιφάνειες είναι σχετικά ομαλές στα σημεία γύρω από το μηδέν, κάτι το οποίο είναι πολύ φυσικό αφού τα περισσότερα δεδομένα κινούνται γύρω από αυτό το σημείο. Αντιθέτως, στα ακραία σημεία παρατηρούνται αρκετές ανωμαλίες και αυτό οφείλεται στην έλλειψη επαρκούς αριθμού δεδομένων στις περιοχές αυτές. Το γεγονός αυτό εξηγείται καλύτερα με την εξέταση των γραφημάτων διασποράς στα επόμενα σχήματα:



Σχήμα 3.4: Γραφήματα διασποράς δεδομένων εκπαίδευσης



Σχήμα 3.5: Γραφήματα διασποράς δεδομένων ελέγχου

3.3 Αρχιτεκτονική του ANFIS

Σε αυτή την παράγραφο θα εξεταστεί η δομή του μοντέλου ANFIS. Συγκεκριμένα, θα παρουσιαστούν τα επίπεδα στα οποία χωρίζεται το μοντέλο και ποιοι υπολογισμοί γίνονται σε κάθε επίπεδο ούτως ώστε να προκύψει το τελικό αποτέλεσμα.

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

Αν (t) είναι small και $(t-1)$ είναι small και $(t-2)$ είναι small τότε $(t+1)$ είναι 1

Αν (t) είναι small και $(t-1)$ είναι small και $(t-2)$ είναι large τότε $(t+1)$ είναι 2

Αν (t) είναι small και $(t-1)$ είναι large και $(t-2)$ είναι small τότε $(t+1)$ είναι 3

Αν (t) είναι small και $(t-1)$ είναι large και $(t-2)$ είναι large τότε $(t+1)$ είναι 4

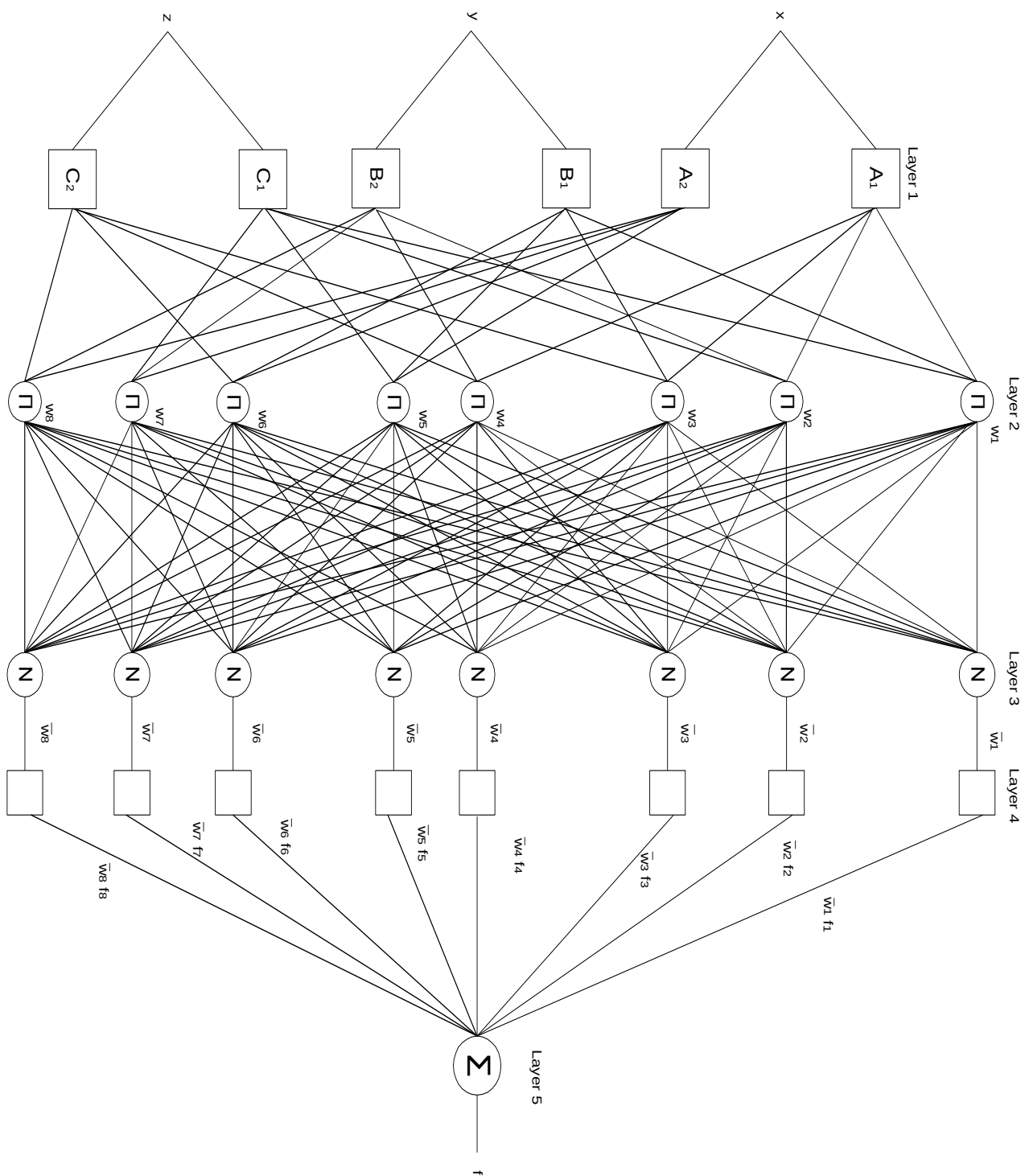
Αν (t) είναι large και $(t-1)$ είναι small και $(t-2)$ είναι small τότε $(t+1)$ είναι 5

Αν (t) είναι large και $(t-1)$ είναι small και $(t-2)$ είναι large τότε $(t+1)$ είναι 6

Αν (t) είναι large και $(t-1)$ είναι large και $(t-2)$ είναι small τότε $(t+1)$ είναι 7

Αν (t) είναι large και $(t-1)$ είναι large και $(t-2)$ είναι large τότε $(t+1)$ είναι 8

Λαμβάνοντας υπόψη τα στοιχεία που αναφέρθηκαν παραπάνω φαίνεται στη συνέχεια (Σχήμα 3.6) η αρχιτεκτονική του μοντέλου ANFIS. Ύστερα θα παρουσιαστεί αναλυτικά τι συμβαίνει σε κάθε επίπεδο του μοντέλου και τέλος θα γίνει χρήση ενός παραδείγματος για την καλύτερη κατανόησή του. (Η έξοδος του i -οστού επιπέδου στο επίπεδο l συμβολίζεται ως $O_{l,i}$.)



Σχήμα 3.6: Αρχιτεκτονική του ANFIS

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

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

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

$$O_{1,i} = \mu_{C_j}(y) \text{ για } i = j+4 \text{ και } j=1, 2,$$

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 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:

$$O_{2,i} = w_i = \mu_{A_1}(x) * \mu_{B_1}(y) * \mu_{C_1}(z), \text{ για } i = 1,2$$

$$O_{2,i} = w_i = \mu_{A_1}(x) * \mu_{B_{1-2}}(y) * \mu_{C_2}(z), \text{ για } i = 3,4$$

$$O_{2,i} = w_i = \mu_{A_2}(x) * \mu_{B_1}(y) * \mu_{C_j}(z), \text{ για } i = j+4, j=1,2$$

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

Επίπεδο 3: Κάθε κόμβος σε αυτό το επίπεδο είναι ένας σταθερός κόμβος N. 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.

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

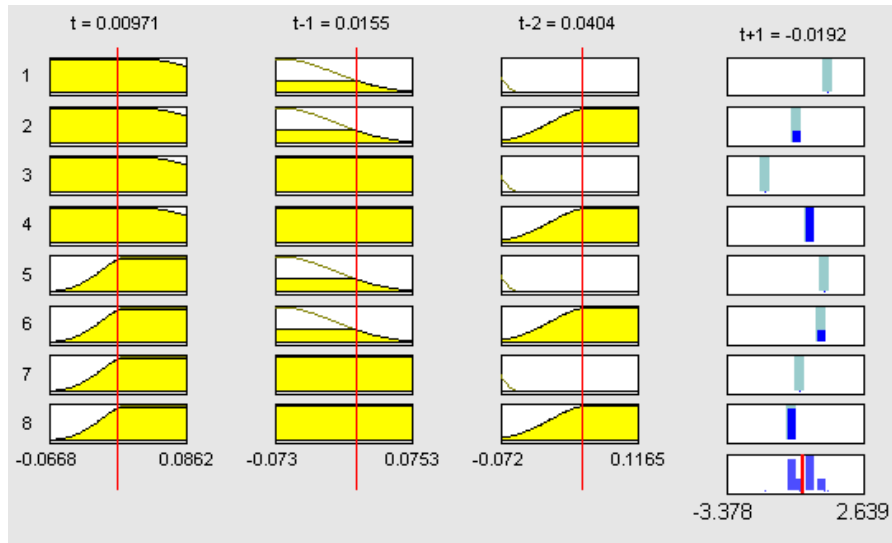
όπου: $\bar{w}_i$ - η έξοδος του επιπέδου 3

$\{p_i, q_i, s_i, r_i\}$ - 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 αναλύθηκε κάθε επίπεδο ξεχωριστά θα γίνει στη συνέχεια η παρουσίαση ενός παραδείγματος. Στο επόμενο σχήμα φαίνεται πως το μοντέλο ANFIS βρίσκει το αποτέλεσμα που προκύπτει για τρεις τυχαίες τιμές t, t-1 και t-2:



Σχήμα 3.7: Οι υπολογισμοί του ANFIS

Από το σχήμα φαίνεται ότι αν $t = 0.00971$, $t-1 = 0.0155$ και $t-2 = 0.0404$ τότε $t+1 = -0.0192$. Στη συνέχεια θα παρουσιαστεί πως προκύπτει αυτό το αποτέλεσμα.

Επίπεδο 1: Εδώ χρησιμοποιείται η συνάρτηση συμμετοχής Gaussian2 για να υπολογιστεί το $O_{1,i}$ για κάθε κόμβο. Το σύνολο παραμέτρων φαίνεται στον membership function editor και υπολογίζεται αυτόματα. Έτσι ισχύει:

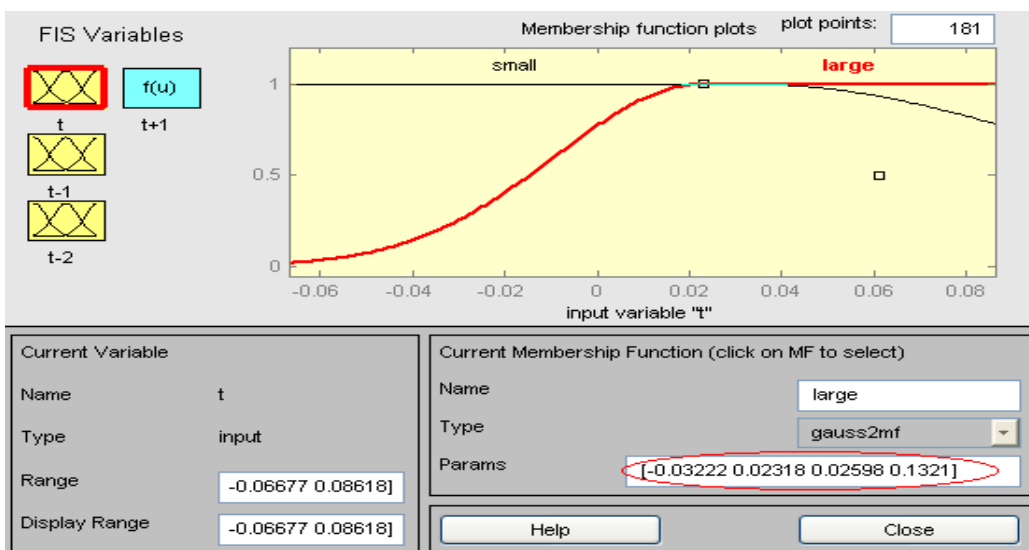
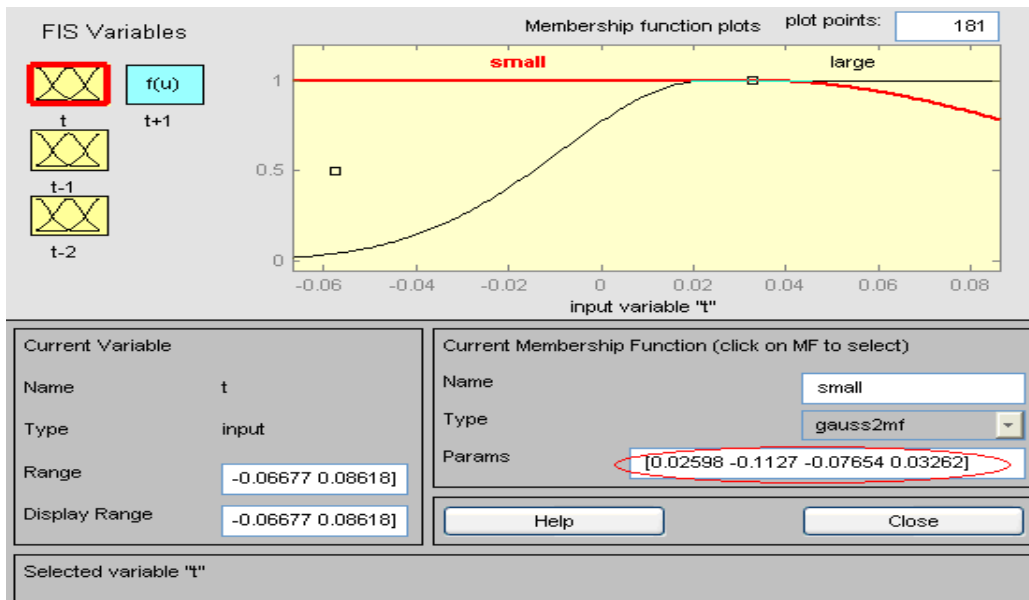
$$[a_1, c_1, a_2, c_2] = [0.02598, -0.1127, -0.07654, 0.03262]$$

αν η γλωσσική μεταβλητή είναι small και

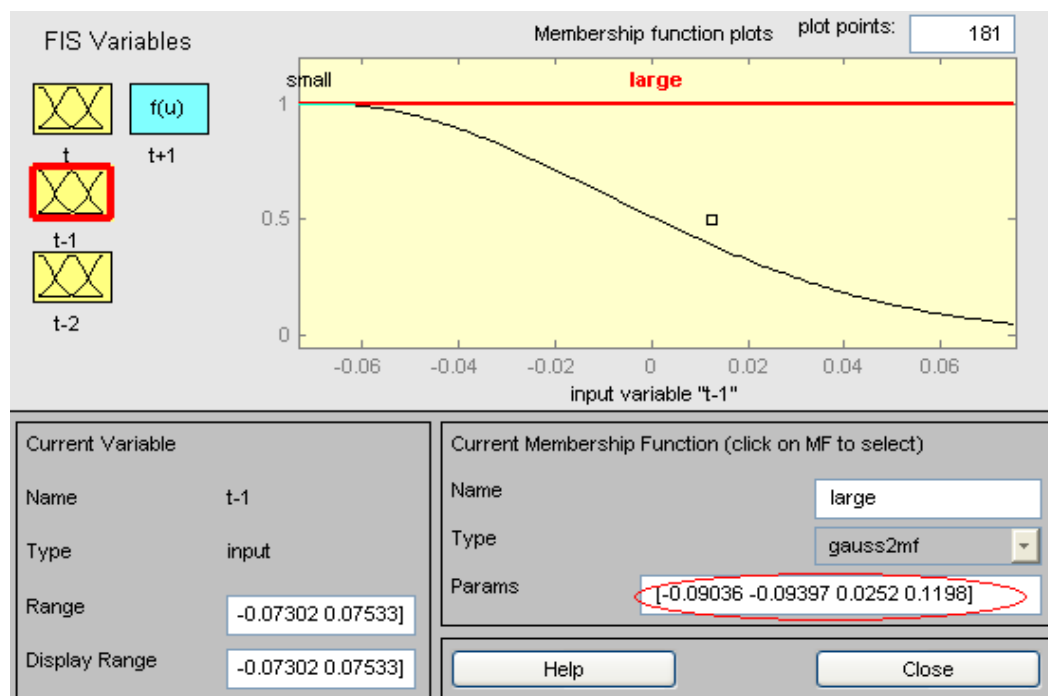
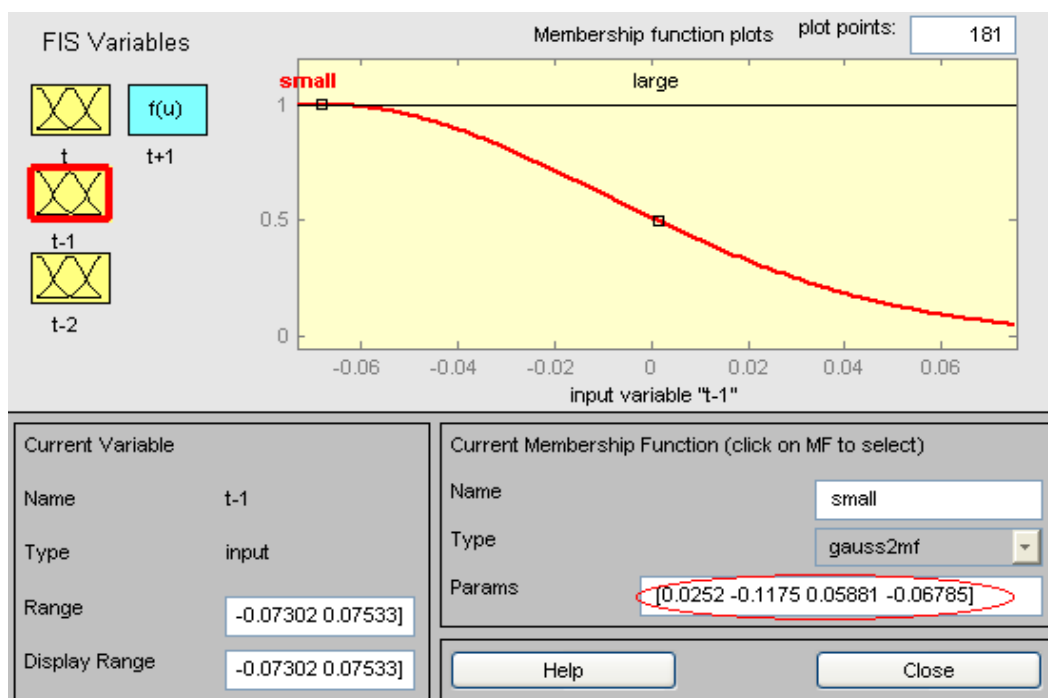
$$[a_1, c_1, a_2, c_2] = [-0.03222, 0.02318, 0.02598, 0.1321]$$

αν η γλωσσική μεταβλητή είναι large

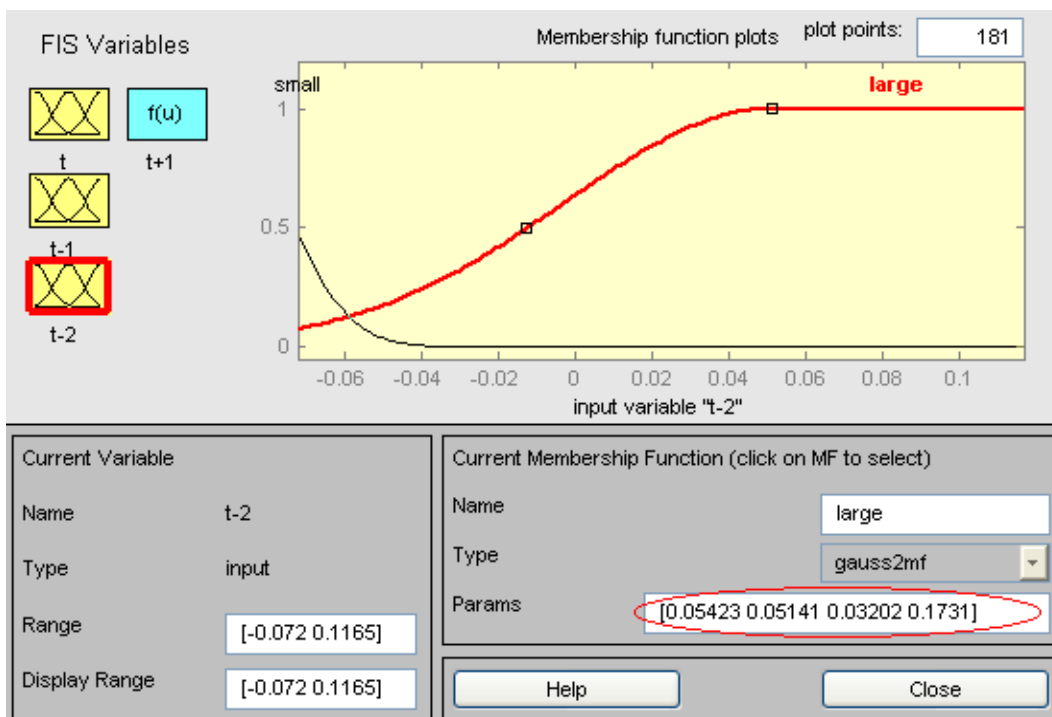
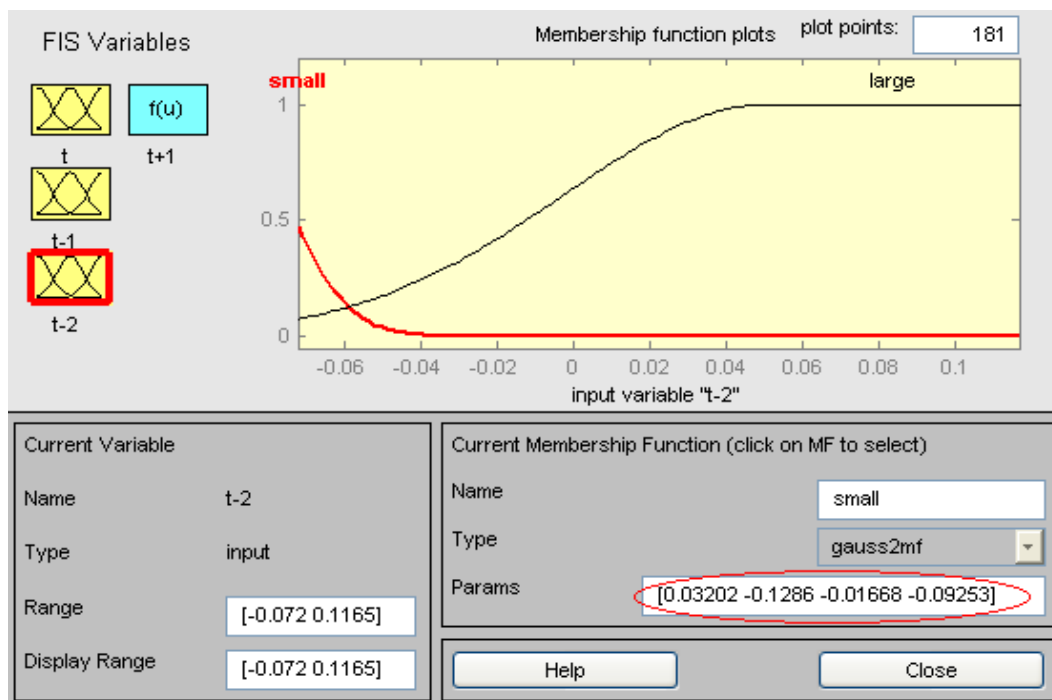
όπου a_1, c_1 είναι για την αριστερή καμπύλη της συνάρτησης Gaussian2 και a_2, c_2 είναι για την δεξιά καμπύλη. Το σύνολο αυτό των παραμέτρων ισχύει μόνο για την είσοδο t. Στο σχήμα 3.8 φαίνεται ο Membership function editor για την είσοδο t, ενώ στα σχήματα 3.9 και 3.10 φαίνεται ο Membership function editor για τις εισόδους t-1 και t-2 καθώς και το σύνολο παραμέτρων τους.



Σχήμα 3.8: Membership function editor για t



Σχήμα 3.9: Membership function editor για t-1



Σχήμα 3.10: Membership function editor για t-2

Από το σχήμα 3.7 φαίνεται ότι τα A_1 , B_1 και C_1 είναι **small** και άρα στη **δεξιά** καμπύλη, ενώ τα A_2 , B_2 και C_2 είναι **large** και άρα στην **αριστερή** καμπύλη. Αυτό σημαίνει ότι για τα A_1 , B_1 και C_1 θα χρησιμοποιηθεί το σύνολο παραμέτρων $[a_2, c_2]$ και για τα A_2 , B_2 και C_2 το σύνολο παραμέτρων $[a_1, c_1]$.

Τώρα χρησιμοποιώντας όλα αυτά και για $x = 0.00971$, $y = 0.0155$ και $z = 0.0404$ προκύπτει ότι:

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

$$O_{1,2} = \mu_{A_2}(x) = 0.916$$

$$O_{1,3} = \mu_{B_1}(y) = 0.366$$

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

$$O_{1,5} = \mu_{C_1}(z) = 1.61 * 10^{-14}$$

$$O_{1,6} = \mu_{C_2}(z) = 0.980$$

Επίπεδο 2: Εδώ γίνονται οι υπολογισμοί χρησιμοποιώντας τις εξισώσεις που αναφέρθηκαν παραπάνω και προκύπτουν τα ακόλουθα:

$$O_{2,1} = w_1 = \mu_{A_1}(x) * \mu_{B_1}(y) * \mu_{C_1}(z) = 1 * 0.366 * 1.61 * 10^{-14} = 5.92 * 10^{-15}$$

$$O_{2,2} = w_2 = \mu_{A_1}(x) * \mu_{B_1}(y) * \mu_{C_2}(z) = 1 * 0.366 * 0.980 = 0.359$$

$$O_{2,3} = w_3 = \mu_{A_1}(x) * \mu_{B_2}(y) * \mu_{C_1}(z) = 1 * 1 * 1.61 * 10^{-14} = 1.61 * 10^{-14}$$

$$O_{2,4} = w_4 = \mu_{A_1}(x) * \mu_{B_2}(y) * \mu_{C_2}(z) = 1 * 1 * 0.980 = 0.980$$

$$O_{2,5} = w_5 = \mu_{A_2}(x) * \mu_{B_1}(y) * \mu_{C_1}(z) = 0.916 * 0.366 * 1.61 * 10^{-14} = 5.42 * 10^{-15}$$

$$O_{2,6} = w_6 = \mu_{A_2}(x) * \mu_{B_1}(y) * \mu_{C_2}(z) = 0.916 * 0.366 * 0.980 = 0.329$$

$$O_{2,7} = w_7 = \mu_{A_2}(x) * \mu_{B_2}(y) * \mu_{C_1}(z) = 0.916 * 1 * 1.61 * 10^{-14} = 1.48 * 10^{-14}$$

$$O_{2,8} = w_8 = \mu_{A_2}(x) * \mu_{B_2}(y) * \mu_{C_2}(z) = 0.916 * 1 * 0.980 = 0.898$$

Επίπεδο 3: Όπως και στο επίπεδο 2 εδώ πραγματοποιούνται οι υπολογισμοί με χρήση της εξίσωσης που αναφέρθηκε παραπάνω:

$$\sum_{i=1}^8 w_i = 5.92 * 10^{-14} + 0.359 + 1.61 * 10^{-14} + 0.980 + 5.42 * 10^{-15} + 0.329 + 1.48 * 10^{-14} + 0.898 = 2.565$$

$$O_{3,1} = \bar{w}_1 = \frac{w_1}{\sum_{i=1}^8 w_i} = \frac{5.92 * 10^{-15}}{2.565} = 2.31 * 10^{-15}$$

$$O_{3,2} = \bar{w}_2 = \frac{w_2}{\sum_{i=1}^8 w_i} = \frac{0.359}{2.565} = 0.140$$

$$O_{3,3} = \bar{w}_3 = \frac{w_3}{\sum_{i=1}^8 w_i} = \frac{1.61 * 10^{-14}}{2.565} = 6.30 * 10^{-15}$$

$$O_{3,4} = \bar{w}_4 = \frac{w_4}{\sum_{i=1}^8 w_i} = \frac{0.980}{2.565} = 0.382$$

$$O_{3,5} = \bar{w}_5 = \frac{w_5}{\sum_{i=1}^8 w_i} = \frac{5.43 * 10^{-15}}{2.565} = 2.12 * 10^{-15}$$

$$O_{3,6} = \bar{w}_6 = \frac{w_6}{\sum_{i=1}^8 w_i} = \frac{0.329}{2.565} = 0.128$$

$$O_{3,7} = \bar{w}_7 = \frac{w_7}{\sum_{i=1}^8 w_i} = \frac{1.48 * 10^{-14}}{2.565} = 5.78 * 10^{-15}$$

$$O_{3,8} = \bar{w}_8 = \frac{w_8}{\sum_{i=1}^8 w_i} = \frac{0.898}{2.565} = 0.350$$

Επίπεδο 4: Για να μπορούν να πραγματοποιηθούν οι απαραίτητοι υπολογισμοί σε αυτό το επίπεδο χρειάζεται το σύνολο παραμέτρων για κάθε κόμβο. Αυτές οι παράμετροι φαίνονται στον membership function editor επιλέγοντας t+1 και είναι:

$$[p_1, q_1, s_1, r_1] = [-2.924, 8.521, -2.575, 1.115] \text{ αν η γλωσσική μεταβλητή είναι 1}$$

$$[p_2, q_2, s_2, r_2] = [-4.513, 1.906, -1.021, -0.2503] \text{ αν η γλωσσική μεταβλητή είναι 2}$$

$$[p_3, q_3, s_3, r_3] = [-4.063, 4.858, -6.491, -1.416] \text{ αν η γλωσσική μεταβλητή είναι 3}$$

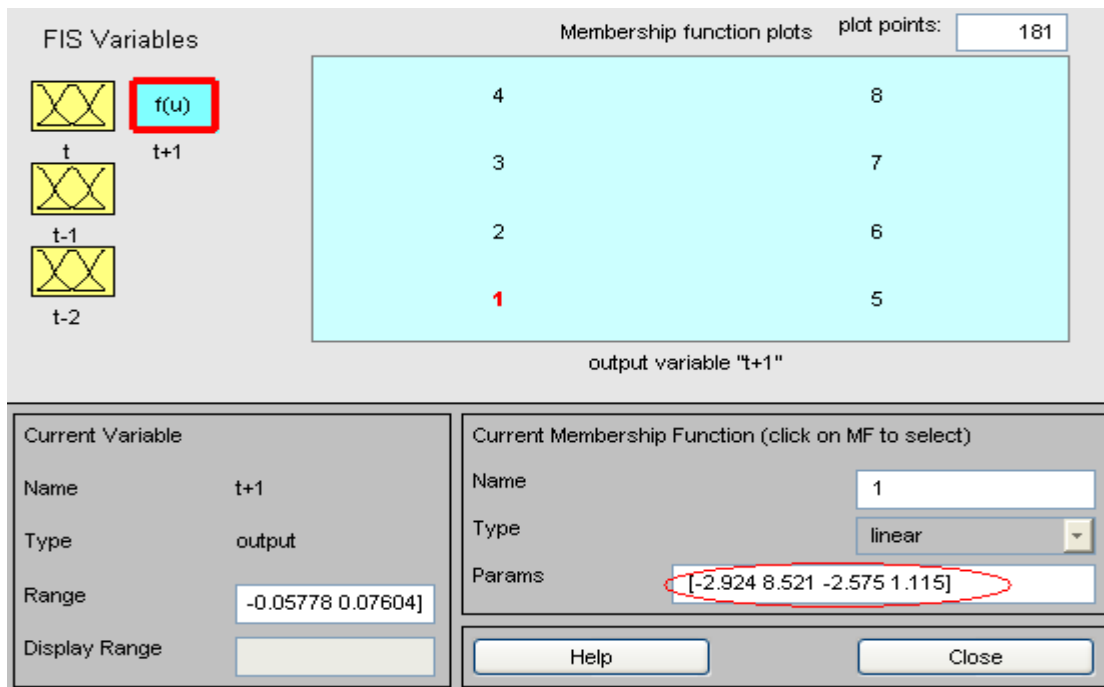
$$[p_4, q_4, s_4, r_4] = [5.592, -3.038, -0.6412, 0.2891] \text{ αν η γλωσσική μεταβλητή είναι 4}$$

$$[p_5, q_5, s_5, r_5] = [-1.483, 5.085, -2.135, 0.9413] \text{ αν η γλωσσική μεταβλητή είναι 5}$$

$$[p_6, q_6, s_6, r_6] = [7.281, -3.374, 3.104, 0.6468] \text{ αν η γλωσσική μεταβλητή είναι 6}$$

$$[p_7, q_7, s_7, r_7] = [-2.116, 3.266, -4.637, 0.008978] \text{ αν η γλωσσική μεταβλητή είναι 7}$$

$$[p_8, q_8, s_8, r_8] = [-5.505, 6.06, -0.485, -0.5383] \text{ αν η γλωσσική μεταβλητή είναι 8}$$



Σχήμα 3.11: Σύνολο παραμέτρων για γλωσσική μεταβλητή 1

FIS Variables

Membership function plots plot points: 181

 **f(u)**
 t t+1
 t
 t-1
 t-2

4	8
3	7
2	6
1	5

output variable "t+1"

Current Variable

Name t+1

Type output

Range -0.05778 0.07604

Display Range

Current Membership Function (click on MF to select)

Name 2

Type linear

Params [-4.513 1.906 -1.021 -0.2503]

Help Close

Σχήμα 3.12: Σύνολο παραμέτρων για γλωσσική μεταβλητή 2

FIS Variables

Membership function plots plot points: 181

 **f(u)**
 t t+1
 t
 t-1
 t-2

4	8
3	7
2	6
1	5

output variable "t+1"

Current Variable

Name t+1

Type output

Range -0.05778 0.07604

Display Range

Current Membership Function (click on MF to select)

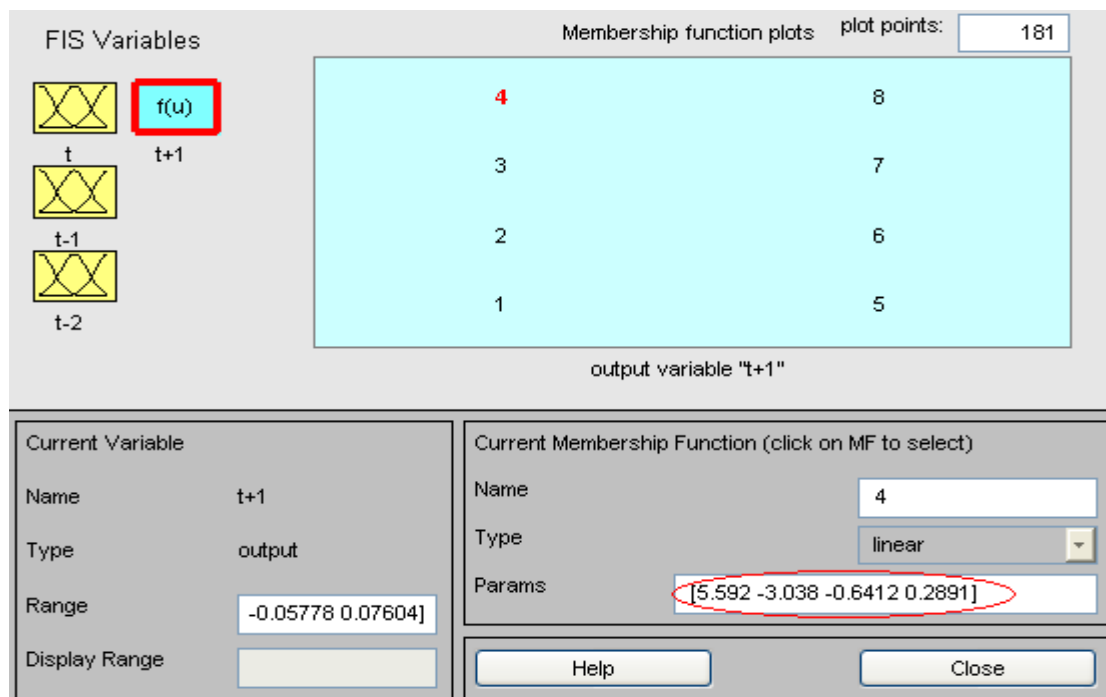
Name 3

Type linear

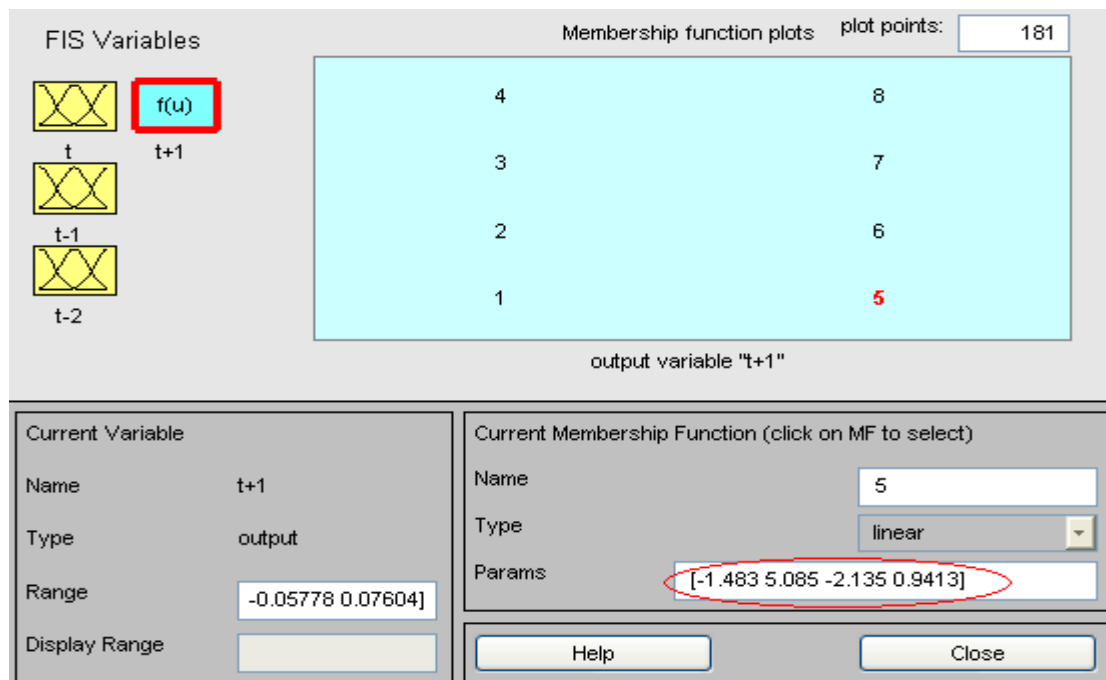
Params [-4.063 4.858 -6.491 -1.416]

Help Close

Σχήμα 3.13: Σύνολο παραμέτρων για γλωσσική μεταβλητή 3



Σχήμα 3.14: Σύνολο παραμέτρων για γλωσσική μεταβλητή 4



Σχήμα 3.15: Σύνολο παραμέτρων για γλωσσική μεταβλητή 5

FIS Variables

Membership function plots plot points: 181

4	8
3	7
2	6
1	5

output variable "t+1"

Current Variable

Name: t+1

Type: output

Range: -0.05778 0.07604

Display Range:

Current Membership Function (click on MF to select)

Name: 6

Type: linear

Params: [7.281 -3.374 3.104 0.6468]

Help Close

Σχήμα 3.16: Σύνολο παραμέτρων για γλωσσική μεταβλητή 6

FIS Variables

Membership function plots plot points: 181

4	8
3	7
2	6
1	5

output variable "t+1"

Current Variable

Name: t+1

Type: output

Range: -0.05778 0.07604

Display Range:

Current Membership Function (click on MF to select)

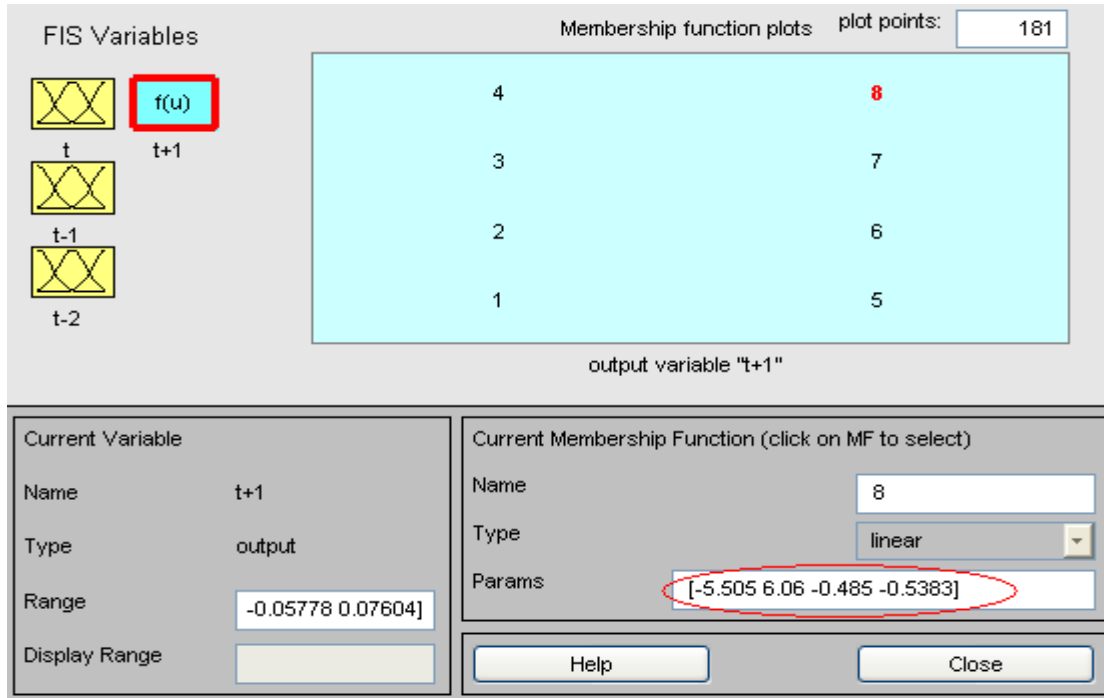
Name: 7

Type: linear

Params: [-2.116 3.266 -4.637 0.008978]

Help Close

Σχήμα 3.17: Σύνολο παραμέτρων για γλωσσική μεταβλητή 7



Σχήμα 3.18: Σύνολο παραμέτρων για γλωσσική μεταβλητή 8

Χρησιμοποιώντας τα σύνολα παραμέτρων προκύπτουν τα ακόλουθα:

$$O_{4,1} = \bar{w}_1 * (p_1 * x + q_1 * y + s_1 * z + r_1) =$$

$$= 2.31 * 10^{-15} * (-2.924 * 0.00971 + 8.521 * 0.0155 - 2.575 * 0.0404 + 1.115) = 2.57 * 10^{-15}$$

$$O_{4,2} = \bar{w}_2 * (p_2 * x + q_2 * y + s_2 * z + r_2) =$$

$$= 0.140 * (-4.513 * 0.00971 + 1.906 * 0.0155 - 1.021 * 0.0404 - 0.2503) = -0.043$$

$$O_{4,3} = \bar{w}_3 * (p_3 * x + q_3 * y + s_3 * z + r_3) =$$

$$= 6.30 * 10^{-15} * (-4.063 * 0.00971 + 4.858 * 0.0155 - 6.491 * 0.0404 - 1.416) = -1.04 * 10^{-14}$$

$$O_{4,4} = \bar{w}_4 * (p_4 * x + q_4 * y + s_4 * z + r_4) =$$

$$= 0.382 * (5.592 * 0.00971 - 3.038 * 0.0155 - 0.6412 * 0.0404 + 0.2891) = 0.103$$

$$O_{4,5} = \bar{w}_5 * (p_5 * x + q_5 * y + s_5 * z + r_5) =$$

$$= 2.12 * 10^{-15} * (-1.483 * 0.00971 + 5.085 * 0.0155 - 2.135 * 0.0404 + 0.9413) = 1.95 * 10^{-15}$$

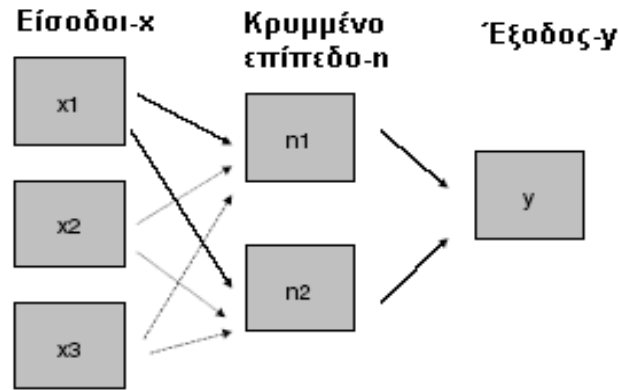
$$\begin{aligned}
O_{4,6} &= \bar{w}_6 * (p_6 * x + q_6 * y + s_6 * z + r_6) = \\
&= 0.128 * (7.281 * 0.00971 - 3.374 * 0.0155 + 3.104 * 0.0404 + 0.6468) = 0.101 \\
O_{4,7} &= \bar{w}_7 * (p_7 * x + q_7 * y + s_7 * z + r_7) = \\
&= 5.78 * 10^{-15} * (-2.116 * 0.00971 + 3.266 * 0.0155 - 4.637 * 0.0404 + 0.008978) = -8.56 * 10^{-16} \\
O_{4,8} &= \bar{w}_8 * (p_8 * x + q_8 * y + s_8 * z + r_8) = \\
&= 0.350 * (-5.505 * 0.00971 + 6.06 * 0.0155 - 0.485 * 0.0404 - 0.5383) = -0.181
\end{aligned}$$

Επίπεδο 5: Στο τελευταίο αυτό επίπεδο αυτό υπολογίζεται η τιμή για t+1 με χρήση της εξίσωσης που αναφέρθηκε στη θεωρία:

$$\begin{aligned}
O_{5,1} &= \sum_{i=1}^8 \bar{w}_i * f_i = \\
&= 2.57 * 10^{-15} - 0.043 - 1.04 * 10^{-14} + 0.103 + 1.95 * 10^{-15} + 0.101 - 8.56 * 10^{-16} - 0.181 = -0.0192
\end{aligned}$$

3.4 Περιγραφή του νευρωνικού δικτύου

Το νευρωνικό δίκτυο που χρησιμοποιήθηκε στην παρούσα ερευνητική εργασία χρησιμοποιεί γενετικούς αλγορίθμους για την εκπαίδευσή του. Είναι ένα προσωτροφοδοτούμενο νευρωνικό δίκτυο ενός επιπέδου με δύο νευρώνες. Η αρχιτεκτονική του νευρωνικού δικτύου εμφανίζεται στο ακόλουθο σχήμα.



Σχήμα 3.19: Αρχιτεκτονική του νευρωνικού δικτύου

Το μέγεθος του πληθυσμού που χρησιμοποιεί ο γενετικός αλγόριθμος είναι 50 ενώ ο αριθμός των γενεών είναι 20. Ο αριθμός των επαναλήψεων εκπαίδευσης που χρησιμοποιήθηκαν είναι 20.000. Η διαδικασία που χρησιμοποιήθηκε για την εκπαίδευση του νευρωνικού δικτύου είναι η εξής:

1. Έγινε αφαίρεση της τάσης ώστε να ομαλοποιηθούν τα δεδομένα εισόδου.
2. Με τη χρήση μιας squashing συνάρτησης έγινε συμπίεση των δεδομένων στο διάστημα $[0,1]$.
3. Επιλέχθηκαν τυχαία οι αρχικές τιμές για τις παραμέτρους του νευρωνικού δικτύου. Ο αριθμός των παραμέτρων εξαρτάται από τον αριθμό των επιπέδων και τον αριθμό των νευρώνων και υπολογίστηκε ότι είναι ίσος με 11.
4. Έγιναν επαναλήψεις του γενετικού αλγορίθμου ούτως ώστε να βρεθούν οι βέλτιστες τιμές για τις παραμέτρους του νευρωνικού δικτύου.
5. Έγινε αποσυμπίεση των δεδομένων για να βρεθούν οι πραγματικές τιμές των παραμέτρων.

ΚΕΦΑΛΑΙΟ 4: Αποτελέσματα από τη χρήση των μοντέλων

4.1 Εισαγωγή

Στο κεφάλαιο αυτό θα γίνει ανάλυση των αποτελεσμάτων που προκύπτουν από τη χρήση του μοντέλων. Αρχικά θα γίνει μια σύντομη περιγραφή των στοιχείων που χρησιμοποιήθηκαν για την εκπαίδευση και τον έλεγχο του μοντέλου ANFIS. Στη συνέχεια θα παρουσιαστούν τα αποτελέσματα του μοντέλου ANFIS και του νευρωνικού δικτύου και τέλος, θα γίνει αναφορά στα συμπεράσματα που προκύπτουν από τη χρήση του μοντέλων.

4.2 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 συμπερασματος 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 της μετοχής της Εθνικής Τράπεζας της Ελλάδας, που είναι μία από τις μεγαλύτερες Ελληνικές Τράπεζες και η μετοχή της διαπραγματεύεται στο Χρηματιστήριο Αξιών Αθηνών. Τα δεδομένα αφορούν μια περίοδο 19 ετών, από το 1987 έως 2005 και έχουν πάρει τη μορφή του ρυθμού μεταβολής σύμφωνα με τον ακόλουθο τύπο:

$$\frac{\text{τιμή}(k) - \text{τιμή}(k-1)}{\text{τιμή}(k-1)}$$

Τα ημερήσια δεδομένα εξετάστηκαν ένα προς ένα προκειμένου να εντοπιστούν τα σημεία αλλαγής κατεύθυνσης (turning point). Ανάλογα με την αλλαγή κατεύθυνσης της τιμής δημιουργήθηκαν δύο διαφορετικοί τύποι δεδομένων. Ο πρώτος τύπος δεδομένων αφορά σετ τεσσάρων τιμών τα οποία περιέχουν τρεις ανοδικές τιμές και μια καθοδική. Ο δεύτερος τύπος δεδομένων αφορά σετ τεσσάρων τιμών τα οποία περιέχουν τρεις καθοδικές τιμές και μια ανοδική. Παράδειγμα των δεδομένων παρουσιάζεται στον ακόλουθο πίνακα:

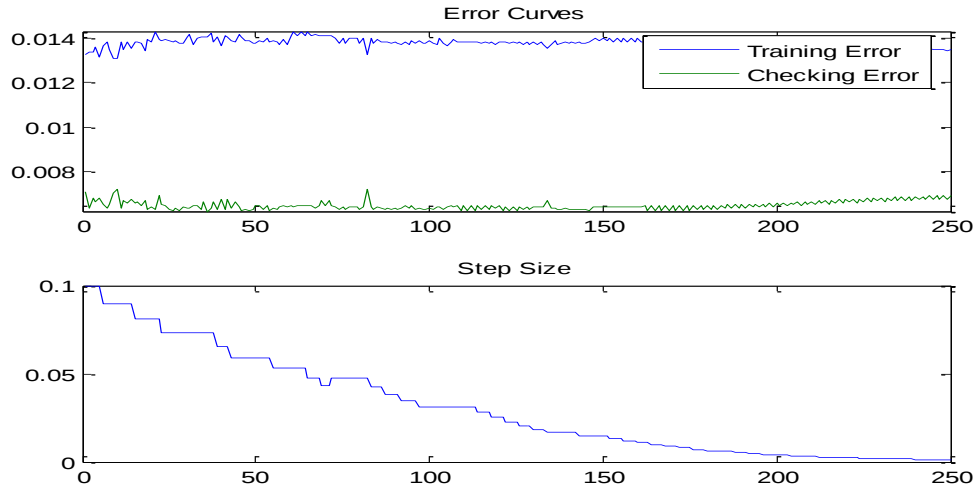
Πίνακας 4.1: Δείγμα των σετ δεδομένων

t-2	t-1	t	t+1
-0.007650	-0.018722	-0.004489	0.013529
0.024083	0.002240	0.031844	-0.001624
-0.002851	-0.017153	-0.002909	0.011085
0.005310	0.019366	0.006908	-0.009148
0.010462	0.003654	0.002427	-0.004843
0.004966	0.015442	0.013382	-0.025810
0.010914	0.007422	0.027461	-0.005215
-0.009881	-0.017964	-0.007453	0.003413
0.023194	0.001295	0.012290	-0.009585

Ο συνολικός αριθμός των σετ δεδομένων που χρησιμοποιήθηκαν είναι 488. Από αυτά, τα πρώτα 292 χρησιμοποιήθηκαν για την εκπαίδευση του μοντέλου ενώ τα επόμενα 98 χρησιμοποιήθηκαν για τον έλεγχο του μοντέλου. Τέλος, τα υπόλοιπα 98 χρησιμοποιήθηκαν για την αξιολόγηση του μοντέλου.

Προκειμένου να δειχθεί πόσο αποτελεσματικό είναι το μοντέλο με τη χρήση τυχαίων δεδομένων, πραγματοποιήθηκε αξιολόγηση του μοντέλου με τη χρήση κανονικών δεδομένων, δεδομένων δηλαδή τα οποία έχουν επιλεγεί τυχαία, στα οποία η μόνη επεξεργασία που πραγματοποιήθηκε είναι ο υπολογισμός του ημερήσιου ρυθμού μεταβολής.

Όπως φαίνεται στο ακόλουθο σχήμα, το μοντέλο δίνει πολύ χαμηλό σφάλμα εκπαίδευσης και ελέγχου ενώ το μέγεθος βήματος τέθηκε στο 0.1.

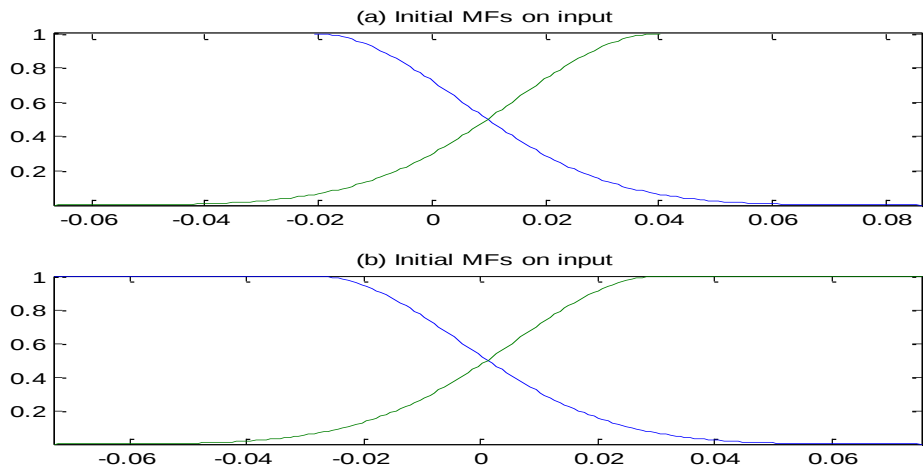


Σχήμα 4.1: Σφάλμα εκπαίδευσης-ελέγχου και μέγεθος βήματος

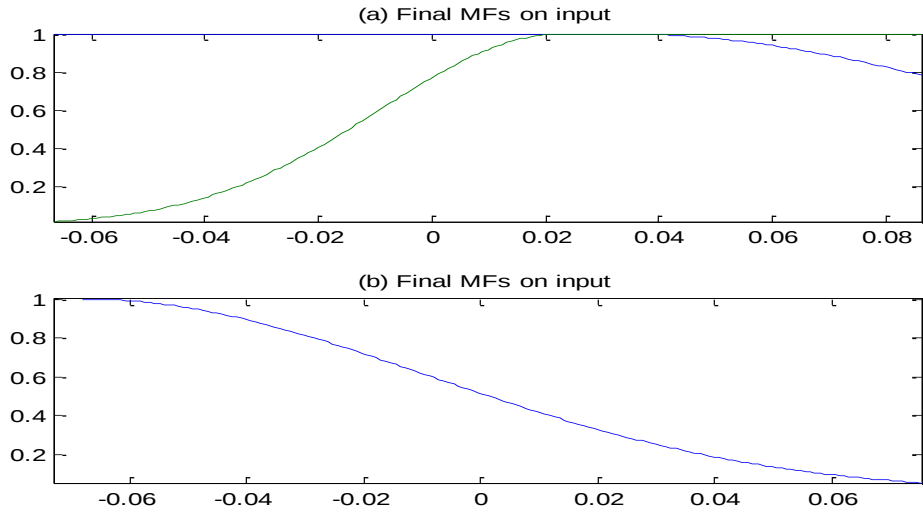
Στο μοντέλο χρησιμοποιήθηκαν δύο συναρτήσεις συμμετοχής. Οι συναρτήσεις συμμετοχής που χρησιμοποιήθηκαν είναι τύπου gauss2mf. Ο τύπος μιας τέτοιας συνάρτησης για την εκπαίδευση του μοντέλου είναι:

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

Οι αρχικές και οι τελικές συναρτήσεις συμμετοχής παρουσιάζονται στα σχήματα 4.2 και 4.3.

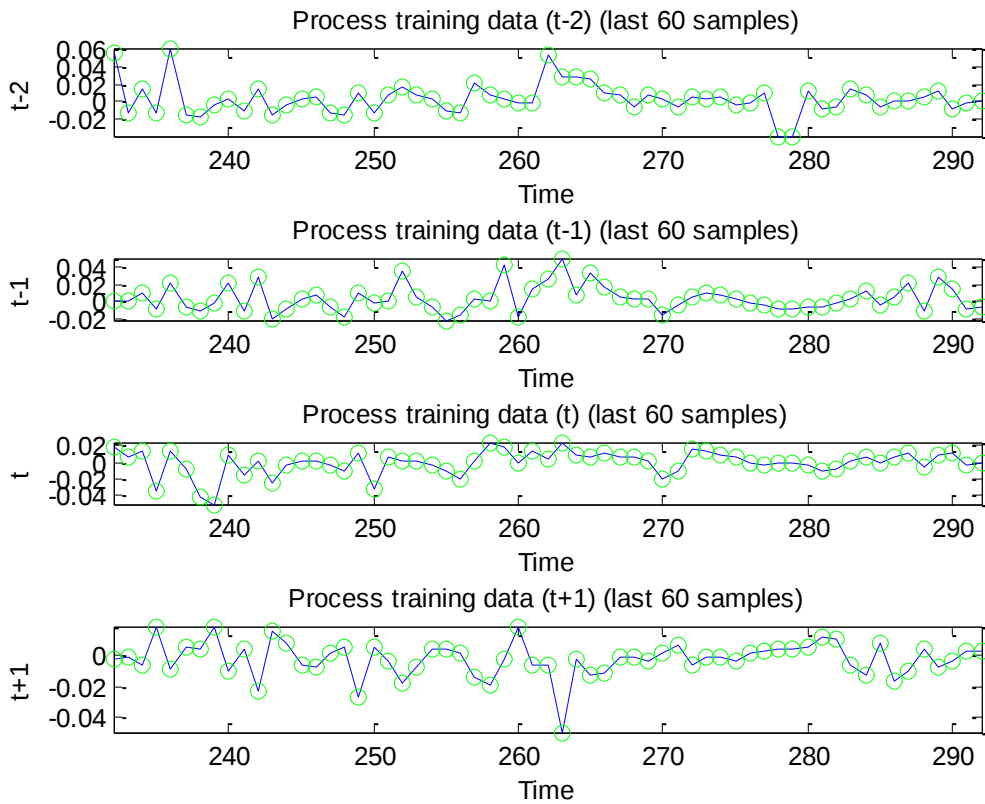


Σχήμα 4.2: Αρχικές συναρτήσεις συμμετοχής



Σχήμα 4.3: Τελικές συναρτήσεις συμμετοχής

Ακόμα, στο επόμενο σχήμα φαίνονται τα διαγράμματα των δεδομένων εκπαίδευσης.



Σχήμα 4.4: Διαγράμματα δεδομένων εκπαίδευσης

Τα ίδια δεδομένα χρησιμοποιήθηκαν και για την εκπαίδευση του νευρωνικού δικτύου. Από αυτά το 80% χρησιμοποιήθηκε για πρόβλεψη εντός δείγματος και το υπόλοιπο 20% για πρόβλεψη εκτός δείγματος. Στον παρακάτω πίνακα παρουσιάζονται συγκεντρωμένα τα χαρακτηριστικά του ANFIS.

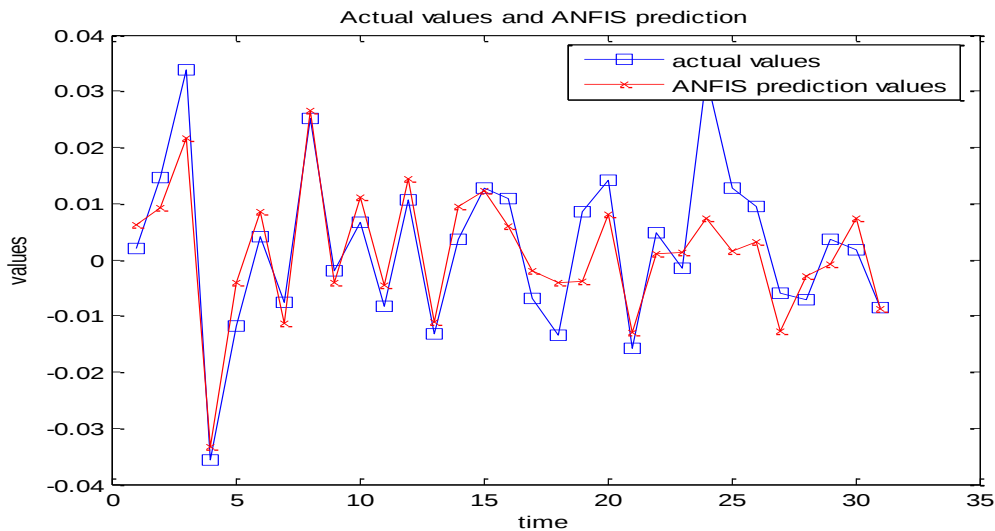
Πίνακας 4.2 : Χαρακτηριστικά του ANFIS

Χαρακτηριστικά του ANFIS	
Τύπος ασαφούς μηχανισμού συμπερασμού	Sugeno
Αριθμός εισόδων	3
Ονομασία 1ης εισόδου	t (input1)
Ονομασία 2ης εισόδου	$t - 1$ (input2)
Ονομασία 3ης εισόδου	$t - 2$ (input3)
Αριθμός εξόδων	1
Ονομασία εξόδου	$t + 1$ (output1)
Αριθμός συναρτήσεων συμμετοχής εισόδων	3 τύπου gaussian2 για κάθε είσοδο
Αριθμός συναρτήσεων εξόδου	8
Αριθμός κανόνων	8
And method	product
Or method	max
Imp. method	product
Agg. method	max
Defuzz. method	wtaver
Αριθμός κόμβων (nodes)	34
Συνολικός αριθμός παραμέτρων	56
Αριθμός γραμμικών παραμέτρων	32
Αριθμός μη γραμμικών παραμέτρων	24

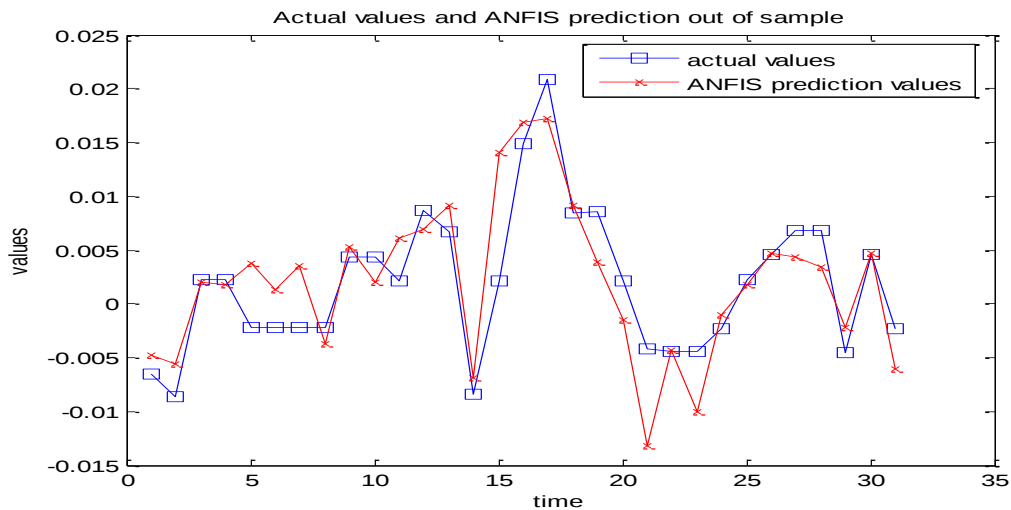
Στις επόμενες παραγράφους θα παρουσιαστούν τα αποτελέσματα από τη χρήση των δύο μοντέλων.

4.3 Αποτελέσματα του μοντέλου ANFIS

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).



Σχήμα 4.5: Πραγματικές τιμές και πρόβλεψη του ANFIS (εντός δείγματος)



Σχήμα 4.6: Πραγματικές τιμές και πρόβλεψη του ANFIS (εκτός δείγματος)

Από τα παραπάνω σχήματα φαίνεται ότι το μοντέλο ANFIS μπορεί να προβλέψει επιτυχώς τις αλλαγές τάσης καθώς και στις δύο περιπτώσεις η πρόβλεψη του μοντέλου ακολουθεί την κατεύθυνση της κίνησης των πραγματικών τιμών. Βέβαια, μεγαλύτερο ποσοστό επιτυχίας έχει το μοντέλο στα δεδομένα στα οποία έχει εκπαιδευτεί. Προκειμένου να συγκριθούν οι δύο περιπτώσεις θα χρησιμοποιηθεί το ποσοστό επιτυχίας του μοντέλου και τα ακόλουθα τρία διαφορετικά μέτρα σφάλματος (Makridakis & Hibon, 1979 [23]).

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

Ποσοστό επιτυχίας

$$Percentage \% = \frac{n}{N}$$

όπου:

n - αριθμός επιτυχημένων προβλέψεων

N - σύνολο δεδομένων που χρησιμοποιήθηκαν

Τα αποτελέσματα για εντός και εκτός δείγματος πρόβλεψη παρουσιάζονται στον ακόλουθο πίνακα:

Πίνακας 4.3: Ποσοστό επιτυχίας % και σφάλματα για εντός και εκτός δείγματος αξιολόγηση για το ANFIS

	εντός δείγματος αξιολόγηση	εκτός δείγματος αξιολόγηση
RMSE	0,0119	0,0073
MAE	0,0083	0,005
MAPE	164,18	98,593
Ποσοστό επιτυχίας %	93,316	92,857

Ο παραπάνω πίνακας δείχνει ότι το μοντέλο ANFIS μπορεί να προβλέπει τα σημεία αλλαγής τάσης με πολύ καλό ποσοστό επιτυχίας τόσο σε δείγματα στα οποία έχει εκπαιδευτεί το μοντέλο όσο και σε δείγματα στα οποία δεν έχει εκπαιδευτεί. Και στις δύο περιπτώσεις έχει γίνει προεπεξεργασία των δεδομένων όπως αναφέρθηκε στην προηγούμενη ενότητα.

Η προεπεξεργασία των δεδομένων είναι μια διαδικασία αρκετά χρονοβόρα. Το γεγονός αυτό καθιστά το μοντέλο δύσχρηστο αν είναι απαραίτητο να γίνεται κάθε φορά προεπεξεργασία των δεδομένων. Για το λόγο αυτό πρέπει να ελεγχθεί η προβλεπτική ικανότητα του μοντέλου χωρίς τη χρήση προεπεξεργασμένων δεδομένων. Στη συνέχεια λοιπόν θα εξεταστεί η ικανότητα του μοντέλου να προβλέπει τα σημεία αλλαγής τάσης χρησιμοποιώντας κανονικά δεδομένα, δεδομένα δηλαδή τα οποία είναι τυχαία και δεν έχουν υποστεί καμία προεπεξεργασία. Τα δεδομένα αυτά αποτελούνται από 100 ημερήσιες παρατηρήσεις της μετοχής της Εθνικής Τράπεζας της Ελλάδας.

Στον επόμενο πίνακα φαίνονται τα ποσοστά επιτυχίας του μοντέλου ANFIS με κανονικά δεδομένα για 3 διαφορετικές χρονικές περιόδους:

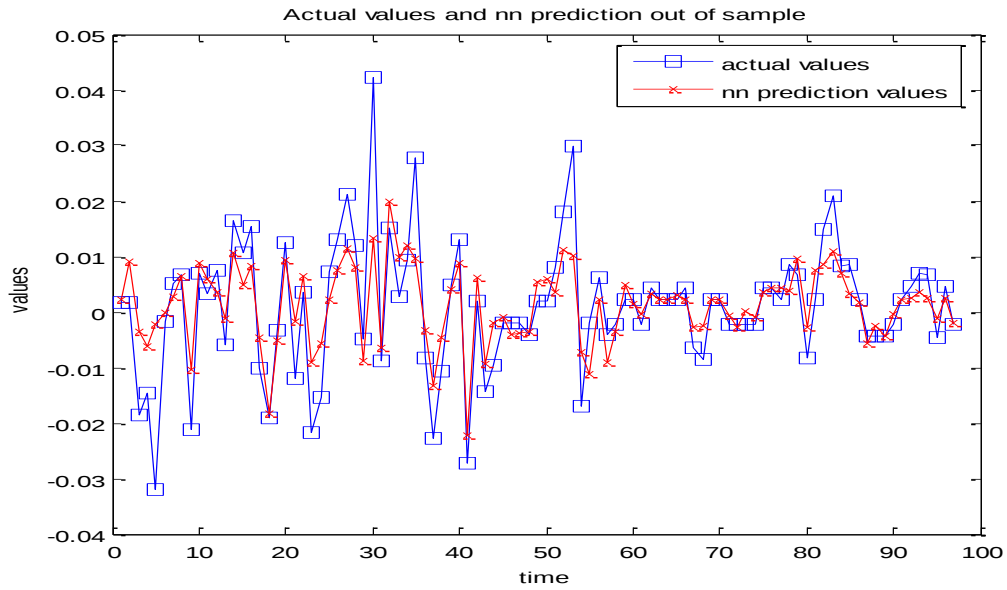
Πίνακας 4.4: Ποσοστό επιτυχίας % και RMSE για κανονικά δεδομένα

Περίοδος	2ο τρίμηνο 2005	3ο τρίμηνο 2004	4ο τρίμηνο 2004
Ποσοστό επιτυχίας %	64	46	53
RMSE	0,0186	0,0237	0,0293

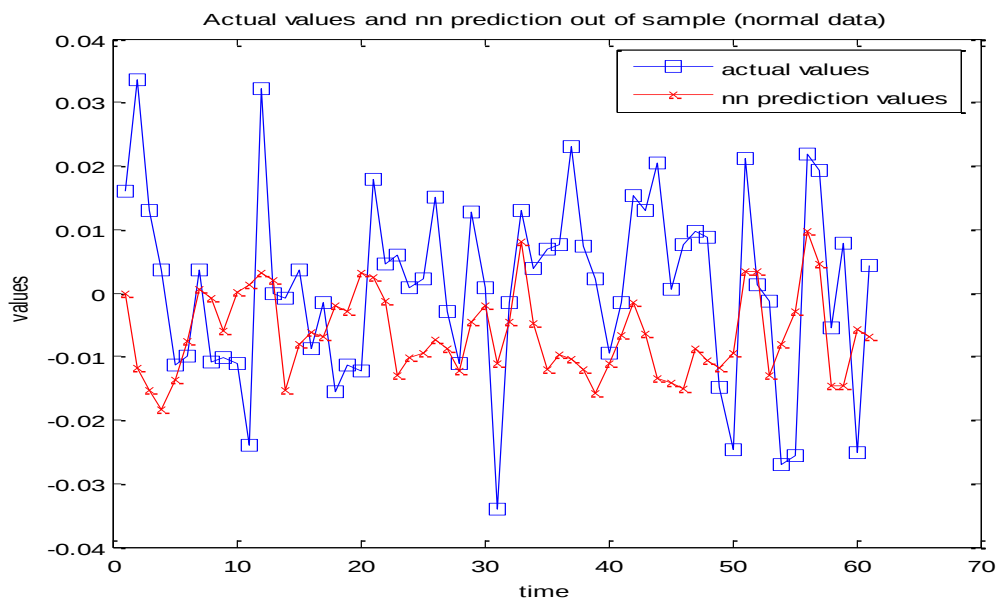
Η μέση ακρίβεια πρόβλεψης του μοντέλου ANFIS με τη χρήση κανονικών δεδομένων σύμφωνα με τον παραπάνω πίνακα είναι 54.33%, ποσοστό το οποίο είναι αρκετά καλό.

4.4 Αποτελέσματα του νευρωνικού δικτύου

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



Σχήμα 4.7: Πραγματικές τιμές και πρόβλεψη του νευρωνικού δικτύου (εκτός δείγματος)



Σχήμα 4.8: Πραγματικές τιμές και πρόβλεψη του νευρωνικού δικτύου (κανονικά δεδομένα)

Από τα παραπάνω σχήματα φαίνεται ότι το νευρωνικό δίκτυο προβλέπει τις αλλαγές τάσης με μεγαλύτερο ποσοστό επιτυχίας να έχει το μοντέλο στα δεδομένα τα οποία έχουν υποστεί προεπεξεργασία.

Τα αποτελέσματα για εντός και εκτός δείγματος πρόβλεψη παρουσιάζονται στον ακόλουθο πίνακα:

Πίνακας 4.5: Ποσοστό επιτυχίας % και σφάλματα για εντός και εκτός δείγματος αξιολόγηση του νευρωνικού δικτύου βασισόμενο σε γενετικούς αλγόριθμους

	εντός δείγματος αξιολόγηση	εκτός δείγματος αξιολόγηση
RMSE	0,0070	0,0136
MAE	0,0047	0,0087
MAPE	67,39	131,96
Ποσοστό επιτυχίας %	98,97	92,30

Ο παραπάνω πίνακας δείχνει ότι το νευρωνικό δίκτυο μπορεί να προβλέπει τα σημεία αλλαγής τάσης με πολύ καλό ποσοστό επιτυχίας τόσο σε δείγματα στα οποία έχει εκπαιδευτεί το μοντέλο όσο και σε δείγματα στα οποία δεν έχει εκπαιδευτεί. Μάλιστα, στα δείγματα στα οποία έχει εκπαιδευτεί το μοντέλο το νευρωνικό δίκτυο παρουσιάζει καλύτερο ποσοστό επιτυχίας από το μοντέλο ANFIS ενώ στα δείγματα στα οποία δεν έχει εκπαιδευτεί παρουσιάζει περίπου το ίδιο ποσοστό επιτυχίας.

Στον επόμενο πίνακα παρουσιάζονται τα ποσοστά επιτυχίας του νευρωνικού δικτύου με κανονικά δεδομένα για 6 διαφορετικές χρονικές περιόδους:

Πίνακας 4.6: Ποσοστό επιτυχίας % και RMSE για κανονικά δεδομένα

Περίοδος	2ο τρίμηνο 2005	1ο τρίμηνο 2005	4ο τρίμηνο 2004
Ποσοστό επιτυχίας %	49,18	54,09	53,82
RMSE	0,0169	0,0214	0,0148
Περίοδος	3ο τρίμηνο 2004	2ο τρίμηνο 2004	1ο τρίμηνο 2004
Ποσοστό επιτυχίας %	52,45	52,45	56,26
RMSE	0,0165	0,0249	0,0286

Από τον παραπάνω πίνακα φαίνεται ότι η μέση ακρίβεια πρόβλεψης του νευρωνικού δικτύου με τη χρήση κανονικών δεδομένων είναι 53.04%. Αυτό το αποτέλεσμα δείχνει ότι το μοντέλο ANFIS παρουσιάζει καλύτερο ποσοστό επιτυχίας από το νευρωνικό δίκτυο, όπως προκύπτει από την σύγκριση του Πίνακα 4.4 και του πίνακα 4.6.

ΚΕΦΑΛΑΙΟ 5: Συμπεράσματα

Στην έρευνα αυτή παρουσιάστηκε μια λύση του προβλήματος της πρόβλεψης των σημείων αλλαγής τάσης σε μια σειρά ημερήσιων χρονολογικών δεδομένων μιας μετοχής. Για την λύση του προβλήματος χρησιμοποιήθηκε το νευρο-ασαφές σύστημα ANFIS, και ένα πρόσθιας τροφοδότησης νευρωνικό δίκτυο το οποίο χρησιμοποιεί γενετικούς αλγόριθμους προκειμένου να υπολογιστεί το δίκτυο. Για την εκπαίδευση του ANFIS και του νευρωνικού δικτύου χρησιμοποιήθηκαν δείγματα τιμών τα οποία αποτελούντο από προσεκτικά επιλεγμένες τετράδες ιστορικών τιμών μιας μετοχής. Οι τρεις πρώτες τιμές αποτελούν τις τρεις εισόδους των μοντέλων και η τετάρτη τιμή την έξοδο των μοντέλων. Ένα μέρος των δεδομένων χρησιμοποιήθηκε για την εκπαίδευση και ένα άλλο μέρος για την αξιολόγηση των μοντέλων. Εκτός από την αξιολόγηση με βάση τα επιλεγμένα δεδομένα, ακολούθησε και αξιολόγηση των μοντέλων με κανονικά δεδομένα (χωρίς επιλογή συγκεκριμένων τετράδων). Αποδείχθηκε ότι τόσο το μοντέλο ANFIS όσο και το νευρωνικό δίκτυο μπορούν να χρησιμοποιηθούν για την πρόβλεψη των σημείων αλλαγών τάσης στο χρηματιστήριο. Τα αποτελέσματα είναι πολύ ενθαρρυντικά και έχουν δείξει ότι το μοντέλο ANFIS αποδίδει αρκετά καλύτερα από το νευρωνικό δίκτυο. Σε εκτός δείγματος επιλεγμένα δεδομένα το ποσοστό επιτυχής πρόβλεψης των σημείων αναστροφής της τάσης φτάνει το 92,86% για το ANFIS και το 92,30% για το νευρωνικό δίκτυο. Σε εκτός δείγματος κανονικών δεδομένων (μη επιλεγμένων) το μέσο ποσοστό επιτυχής πρόβλεψης των σημείων αναστροφής της τάσης φτάνει το 54,33% για το ANFIS και το 53,04% για το νευρωνικό δίκτυο. Αυτό σημαίνει ότι το μοντέλο ANFIS μπορεί επιτυχώς να χρησιμοποιηθεί στον προγραμματισμό των συναλλαγών, δεδομένου ότι η εφαρμογή των τεχνικών πρόβλεψης στις χρηματιστηριακές αγορές έχει

γίνει ένα βασικό εργαλείο των ατόμων που ασχολούνται με βραχυχρόνια πρόβλεψη τιμών κατά τη διάρκεια των τελευταίων ετών.

Το βασικό πρόβλημα που εμφανίζεται είναι η χρονοβόρα διαδικασία που απαιτείται για την προσεκτική επιλογή τετράδων δεδομένων. Πάντως, τα αποτελέσματα έδειξαν ότι το μοντέλο μπορεί να χρησιμοποιηθεί για την πρόβλεψη των σημείων αλλαγής τάσης με αρκετά καλό ποσοστό επιτυχίας ακόμα και με την χρήση κανονικών δεδομένων τα οποία δεν έχουν υποστεί καμία προεπεξεργασία.

Τέλος, έγινε μία προσπάθεια να χρησιμοποιηθεί το μοντέλο αυτό για τις μετοχές και άλλων εταιρειών χωρίς να γίνει εκπαίδευση του μοντέλου με καινούργια επιλεγμένα δεδομένα των συγκεκριμένων εταιρειών, όμως τα αποτελέσματα που προέκυψαν δεν ήταν τόσο καλά καθώς έδιναν ποσοστά επιτυχίας κάτω του 50%. Άρα πρέπει κάθε φορά να χρησιμοποιούνται επιλεγμένα δεδομένα της κάθε εταιρείας για την εκπαίδευση του μοντέλου ούτως ώστε να προκύπτουν περισσότερο αξιόλογα αποτελέσματα. Αυτό θα πρέπει να γίνει αντικείμενο μελλοντικής έρευνας.

Επίσης περαιτέρω έρευνα θα πρέπει να γίνει για να εξεταστεί η απόδοση των μοντέλων με περισσότερες εισόδους και άλλες τεχνικές μάθησης.

ΚΕΦΑΛΑΙΟ 6: Βιβλιογραφία

- [1] 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.
- [2] Andersson, E., (2002). Monitoring cyclical processes. A non-parametric approach. *Journal of Applied Statistics*, Vol. 29, No. 7, 973-990.
- [3] 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.
- [4] Atsalakis, G., Valavanis K., September (2006, a). A Neuro-Fuzzy controller for Stock market forecasting. Working paper August, Technical University of Crete.
- [5] Atsalakis, G., May (2006 b). Neuro-fuzzy and Technical Analysis for stock prediction. Working paper, Technical University of Crete.
- [6] 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.
- [7] Atsalakis, G., Ucenic, C., (2006, c). Electric load forecasting by Neuro-fuzzy approach. *WSEAS International Conference on Energy and Environmental Systems*, Chalkida Evia, Greece.
- [8] 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.
- [9] 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.

- [10] Atsalakis, G., Ucenic, C., Plokamakis G., (2005, b). Forecasting of Electricity Demand Using Neuro-fuzzy (ANFIS) Approach, *International Conference on NHIBE, Corfu, Greece*.
- [11] 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*.
- [12] 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*.
- [13] 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*.
- [14] Bengoechea, P., Camacho, M. Perez-Quiros, G. (2006). A useful tool for forecasting the Euro-area business cycle phases. *International Journal of Forecasting*.
- [15] Canova, F., Ciccarelli, M. (2004). Forecasting and turning point predictions in a Bayesian panel VAR model. *Journal of Econometrics*, 120, 327 – 359.
- [16] Garcia-Ferrer, A., Queralto, R. A. (1998). Can univariate models forecast turning points in seasonal economic time series? *International Journal of Forecasting*, 14, 433–446.
- [17] Harding, D., Pagan, A. (2006). Synchronization of cycles. *Journal of Econometrics*, 132, 59–79.
- [18] Harvey, A. C., Trimbur, T. M., Van Dijk, H. K.. Trends and cycles in economic time series: A Bayesian approach. *Journal of Econometrics*.
- [19] Holland, J. H. (1975). Adaptation in natural and artificial systems. *University of Michigan Press, Michigan*.
- [20] Jang, J.-S.R, C.-T. Sun, (1995). Neuro-fuzzy Modeling and Control, *Proceedings of the IEEE*, Vol. 83, No. 3, pp. 378-406.

- [21] 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.
- [22] 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.
- [23] Makridakis, S. & Hibon, M. (1979). Accuracy of forecasting: An empirical investigation. *Journal of the Royal Statistical Society, A*, 142, 97-145.
- [24] McCulloch, W. S. & Pitts, W. (1943). A logical calculus of the ideas imminent in neural nets. *Bulletin of Mathematical Biophysics*, 5, 115-137.
- [25] Minsky, M. & Papert, S. (1969). *Perceptrons*. MIT Press, Cambridge, M.A.
- [26] Rosenblatt, F. (1962). *Principles of neurodynamics: perceptrons and the theory of brain mechanisms*. Spartan, New York.
- [27] Van Nieuwerburgh, S., Veldkamp, L. (2006). Learning asymmetries in real business cycles. *Journal of Monetary Economics*, 53, 753–772.
- [28] 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*.
- [29] 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*.
- [30] 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*.

- [31] 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.*