



**Πολυτεχνείο Κρήτης
Τμήμα Μηχανικών Παραγωγής και Διοίκησης**

Διπλωματική Εργασία με θέμα:

**Πρόβλεψη κατανάλωσης ηλεκτρικής ενέργειας στην Κρήτη με το
νεύρο-ασαφές σύστημα (ANFIS), χρησιμοποιώντας χρονοσειρές
κατανάλωσης ηλεκτρικής ενέργειας.**

Δημήτριος Παπαϊωάννου

Επιβλέπων Καθηγητής: Μουστάκης Βασίλειος

**Εξεταστική Επιτροπή: Μουστάκης Βασίλειος
Μπιλάλης Νικόλαος
Κοντογιάννης Θωμάς**

Χανιά, Σεπτέμβριος 2004

Περιεχόμενα

– Εισαγωγή.....	3	– 1.5.1.1 Φθίνουσες Μέθοδοι (Descent Methods)	29
– Μέρος 1 ^ο	4	– 1.5.1.2 Μέθοδοι βασισμένες σε βαθμωτό διάνυσμα (Gradient-based Methods).....	30
– 1.1 Θεωρία Ασαφούς Λογικής (Fuzzy Logic Theory)	5	– 1.5.2 Απότομη Φθίνουσα Μέθοδος (The Method of Steepest Descent)....	32
– 1.2 ΘΕΜΕΛΙΑ ΤΗΣ ΑΣΑΦΟΥΣ ΛΟΓΙΚΗΣ.....	6	– 1.6 Προσδιορισμός μεγέθους βήματος (Step size).....	33
– 1.2.1 ΑΣΑΦΗ ΣΥΝΟΛΑ.....	6	– 1.6.1 Αρχικό Bracketing.....	33
– 1.2.2 Συναρτήσεις συμμετοχής (Membership Function).	7	– 1.6.2 Αναζήτηση ευθείας.....	35
– 1.2.2.1 Παραδείγματα Συναρτήσεων Συμμετοχής.....	9	– 1.6.3 Η μέθοδος του Newton.....	35
– 1.2.3 ΛΟΓΙΚΟΙ ΤΕΛΕΣΤΕΣ.....	10	– 1.6.4 Η μέθοδος της τέμνουσας..	35
– 1.2.3.1 Γενικά.....	10	– 1.6.5 Η μέθοδος Sectioning.....	36
– 1.2.3.2 Ειδικότερα.....	11	– 1.6.6 Πολωνυμική Παρεμβολή (Polynomial Interpolation).....	38
– 1.2.4 ΚΑΝΟΝΕΣ ΑΝ-ΤΟΤΕ (IF-THEN)	13	– 1.7 Προσαρμοστικά Δίκτυα (Adaptive Networks)	39
– 1.3 Ασαφή Συστήματα Εξαγωγής Συμπερασμάτων (Fuzzy Inference Systems)	15	– 1.7.1 Παράγωγοι Παραμετροποιήσιμων Συναρτήσεων συμμετοχής.....	40
– 1.3.1 FIS Τύπου Mamdani.....	16	– 1.7.2 Αρχιτεκτονική.....	40
– 1.3.2 FIS Τύπου Sugeno.....	17	– 1.7.2.1 Καταμερισμός παραμέτρων σε προσαρμοστικά δίκτυα.....	42
– 1.4 Μέθοδος Ελαχίστων Τετραγώνων Για Ταυτοποίηση Συστήματος.....	22	– 1.7.3 Παραδείγματα προσαρμοστικών δικτύων.....	44
– 1.4.1 Εκτιμητήρια Ελαχίστων Τετραγώνων	24	– 1.7.3.1 Προσαρμοστικό δίκτυο με έναν γραμμικό κόμβο (linear node) .	44
– 1.4.2 Θεώρημα Εκτιμήτριας Ελαχίστων Τετραγώνων	26	– 1.7.3.2 Δίκτυο perceptron.....	44
– 1.4.3 Γενίκευση τετραγωνικού σφάλματος με χρήση βαρών.....	26	– 1.7.3.3 Ένα πολυεπίπεδο perceptron.....	45
– 1.4.4 Πολλαπλές Έξοδοι.....	27	– 1.8 Υβριδικός κανόνας εκπαίδευσης: συνδυασμός απότομης φθίνουσας μεθόδου και σφάλματος ελαχίστων τετραγώνων.....	46
– 1.4.5 Εκτιμητήρια Ελαχίστων Τετραγώνων για Μη Γραμμικά Μοντέλα (Σύντομη αναφορά).....	28	– 1.8.1 Εκπαίδευση Off-line (μαζική εκπαίδευση σε παρτίδες) (Off-line learning) (Batch learning).....	46
– 1.5 Βελτιστοποίηση Με παραγώγου.....	29	– 1.8.2 Εκπαίδευση On-line (εκπαίδευση δείγμα προς δείγμα) (On-	

line learning) (Pattern–By–Pattern learning).....	48	– 1.10.4 Περιορισμοί του ANFIS....	60
– 1.9 Οπισθόδρομη διάδοση για δίκτυα χωρίς ανατροφοδότηση. (Backpropagation for feedforward networks)	49	– 1.10.5 Πρακτικά Ζητήματα.....	60
– 1.9.1 Γενικά.....	49	– 1.10.6 Πρόβλεψη Χρονοσειρών...61	
– 1.9.2 Ειδικότερα	49	– Μέρος 2ο	62
– 1.9.2.1 Προσδιορισμός του σήματος σφάλματος.....	51	– 2. Πρακτική Εφαρμογή.....	63
– 1.10 Προσαρμοστικό Νεύρο-Ασαφές Σύστημα Εξαγωγής Συμπερασμάτων (ΠΝΑΣΕΣ) (Adaptive Neuro-Fuzzy Inference Systems) (ANFIS).....	54	– 2.1 Γενική Περιγραφή του Προγράμματος Πρόβλεψης.....	63
– 1.10.1 Αρχιτεκτονική του ANFIS.	54	– 2.2 Παρουσίαση και Σύγκριση Αποτελεσμάτων.....	66
– 1.10.2 Υβριδικός αλγόριθμος εκπαίδευσης.....	57	– 2.2.1 Ειδικότερα:	66
– 1.10.3 Υπερπροσαρμογή (overfitting) δεδομένων.....	59	– 2.3 Πίνακες.....	67
		– 2.4 Γραφήματα.....	70
		– ΕΠΙΛΟΓΟΣ.....	80
		– Βιβλιογραφία:	81

Ένα Μεγάλο Ευχαριστώ

Ξεκινώντας θα ήθελα να ευχαριστήσω πάρα πολύ τον κ. Μουστάκη για την ευκαιρία που μου έδωσε να ασχοληθώ με ένα τόσο πρωτοποριακό και χρήσιμο θέμα και τον κ. Ατσαλάκη για την πολύτιμη βοήθεια του πάνω σε κάθε ζήτημα της εργασίας και για την επίλυση των πολλών αποριών μου πάνω στην ασαφή λογική.

Ακόμα, τον αδελφό μου Ανδρέα για την μεγάλη ψυχολογική υποστήριξη και την βοήθεια στην δακτυλογράφηση, αλλά και όλη την οικογένεια μου που θυσιάσαν χρόνια από την ζωή τους για να μου δώσουν την δυνατότητα να σπουδάσω και να αποκτήσω όλα αυτά τα εφόδια που θα μου χρειαστούν στην μετέπειτα ζωή μου.

Τέλος όλους τους φίλους που γνώρισα στα Χανιά και που μαζί περάσαμε όμορφες στιγμές «οργώνοντας» την Κρήτη και στηρίζοντας ο ένας τον άλλον σε κάθε εύκολη και δύσκολη στιγμή.

Εισαγωγή

Πάντοτε ψάχναμε μεθόδους με τις οποίες θα μπορούσαμε να βελτιστοποιήσουμε τους διαθέσιμους πόρους που έχουμε. Για το σκοπό αυτό αναζητούσαμε πληροφορίες στο παρελθόν από τις οποίες θα αποκτούσαμε την απαραίτητη γνώση έτσι ώστε στο παρόν και στο μέλλον να αποφύγουμε λάθη και σπατάλες.

Στον απαραίτητο για τον σύγχρονο πολιτισμό κλάδο της ηλεκτρικής ενέργειας είναι επιτακτική η ανάγκη της ορθής διαχείρισης των ενεργειακών πόρων για την ασφάλεια του δικτύου, την οικονομική ανάπτυξη της χώρας και της ενεργειακής εταιρείας, την ποιοτική παροχή υπηρεσιών στον τελικό καταναλωτή, καθώς όλα αυτά είναι μεταβλητές που χαρακτηρίζουν άμεσα το βιοτικό επίπεδο των πολιτών μιας χώρας.

Η ανάγκη αυτή καθίσταται υποχρέωση κάποιες φορές σε περιοχές όπως είναι η Κρήτη, όπου τα προβλήματα αδιάλειπτης τροφοδότησης με ηλεκτρισμό ταλαιπωρούν τους κατοίκους της για πολλές δεκαετίες χωρίς ακόμα να έχουν δοθεί επαρκείς λύσεις στο μείζον αυτό πρόβλημα. Είναι βασικό να μελετήσουν λοιπόν οι υπεύθυνοι της ενεργειακής εταιρείας τις χρονοσειρές κατανάλωσης ηλεκτρικής ενέργειας ούτως ώστε να έχουν την καλύτερη πρόβλεψη για την μελλοντική ζήτηση και κατά συνέπεια να κινηθούν αναλόγως στην αγορά και χρήση των πρώτων υλών ή στην ανάπτυξη νέων εγκαταστάσεων εάν παραστεί ανάγκη.

Λέγοντας χρονοσειρά εννοούμε ένα σύνολο παρατηρήσεων μιας μεταβλητής οι οποίες έχουν ληφθεί σε ίσα χρονικά διαστήματα. Τα δεδομένα της μπορούν να προέλθουν από μετρήσεις που έγιναν σε τακτά χρονικά διαστήματα. Η χρονική διάταξη των τιμών μιας χρονοσειράς είναι βασική γι' αυτό και απαιτούνται ειδικές μέθοδοι ανάλυσης τους. [2] ***

Η ανάλυση γίνεται συνήθως με στατιστικές μεθόδους. Κάποιες πολύ γνωστές και μελετημένες από αυτές είναι η αυτοπαλίνδρομη μέθοδος (AR) ή η μέθοδος του αυτοπαλίνδρουμου κινητού μέσου (ARMA). Στην παρούσα όμως εργασία θα χρησιμοποιήσουμε μια εναλλακτική μέθοδο, της οποίας τα αποτελέσματα θα συγκρίνουμε με τις δυο παραπάνω στατιστικές μεθόδους. Η μέθοδος αυτή είναι η πρόβλεψη με τη βοήθεια της ασαφούς λογικής. Το μοντέλο που θα χρησιμοποιήσουμε είναι το ANFIS (Adaptive Neuro-Fuzzy Inference System), που στα ελληνικά μεταφράζεται «Προσαρμοστικό Νεύρο-Ασαφές Σύστημα Εξαγωγής Συμπερασμάτων».

Γεγονός είναι ότι η ασαφής λογική απαιτεί μεγαλύτερη επεξεργαστική ισχύ από τους υπολογιστές σε σχέση με τις στατιστικές μεθόδους. Όμως, το κέρδος τους είναι ότι μπορούν να δώσουν τελικά καλύτερα αποτελέσματα όπως θα δειχθεί και στη συνέχεια.

Στο πρώτο μέρος αναπτύσσεται το θεωρητικό υπόβαθρο που είναι για την κατανόηση και την χρήση ενός ANFIS. Στο δεύτερο μέρος παρουσιάζεται η πρακτική εφαρμογή που είχε στην παρούσα εργασία, καθώς και τα αποτελέσματα που προέκυψαν.

*** Σε ολόκληρη την εργασία οι παραπομπές στην βιβλιογραφία σημειώνονται με τον αντίστοιχο αριθμό μέσα σε αγκύλες []. Έτσι ο αναγνώστης μπορεί εύκολα, στην τελευταία σελίδα να αναζητήσει τις αντίστοιχες πληροφορίες (συγγραφέα, τίτλο, έκδοτη).

ΜΕΡΟΣ 1^ο

1.1 ΘΕΩΡΙΑ ΑΣΑΦΟΥΣ ΛΟΓΙΚΗΣ (FUZZY LOGIC THEORY)

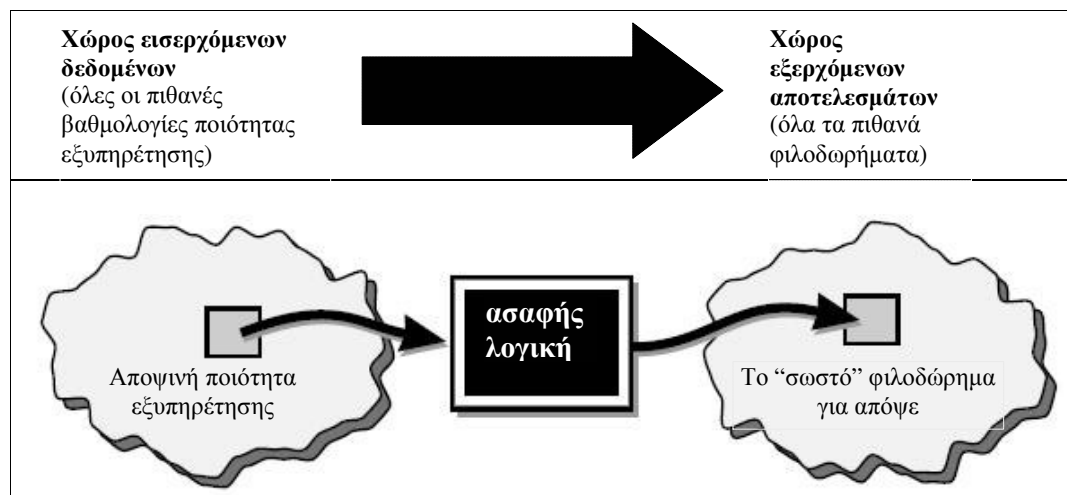
Η Ασαφής Λογική (Α.Λ.) είναι μια μεθοδολογία που ασχολείται με τη σχετική σημασία της ακρίβειας. [3]



Σκίτσο 1. Ακρίβεια και σημαντικότητα στη ζωή. [3]

Χαρακτηριστικά το βλέπουμε και στο Σκίτσο 1. Ή ακόμα πιο απλά «η κωδικοποίηση της κοινής αίσθησης». Ένας βολικός τρόπος δηλαδή, να χαρτογραφήσουμε τον χώρο των δεδομένων που εισάγουμε, στον χώρο των αποτελεσμάτων που εξάγουμε. Ο κύριος μηχανισμός για να κάνουμε αυτό είναι μια λίστα δηλώσεων αν-τότε (if-then), που ονομάζονται κανόνες. Όλοι οι κανόνες υπολογίζονται παράλληλα, και η σειρά δεν έχει καμία σημασία.

Για παράδειγμα στο πρόβλημα, πόσο φιλοδώρημα να αφήσουμε σε ένα μαγαζί, ένας χάρτης εισόδου εξόδου φαίνεται στο παρακάτω σχήμα 1:



Σχήμα 1. Χάρτης εισόδου εξόδου. [3]

1.2 ΘΕΜΕΛΙΑ ΤΗΣ ΑΣΑΦΟΥΣ ΛΟΓΙΚΗΣ

1.2.1 ΑΣΑΦΗ ΣΥΝΟΛΑ

Η ασαφής λογική ξεκινάει με τη σύλληψη του ασαφούς συνόλου, το οποίο δεν έχει ξεκάθαρα και σαφώς καθορισμένα όρια και μπορεί να περιέχει στοιχεία με μερικό βαθμό συγγένειας. [3]

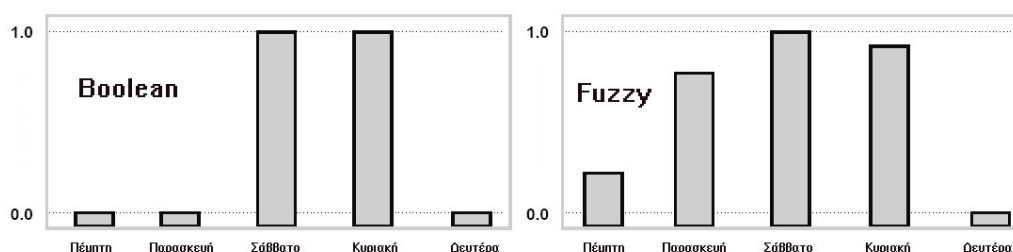
Θα λέγαμε ότι εισερχόμαστε στον κόσμο όπου η αυστηρή λογική ναι-όχι σταματάει να είναι χρήσιμη. Και το παρακάτω σχόλιο χαρακτηρίζει τα θεμέλια της ασαφούς λογικής:

Στην ασαφή λογική, η αλήθεια κάθε υπόθεσης γίνεται υπόθεση ενός βαθμού.

Κάθε υπόθεση μπορεί να είναι ασαφής. Το εργαλείο που δίνει η Α.Λ. είναι η ικανότητα να απαντάμε σε μια «ναι-όχι» ερώτηση με «περίπου ναι – ή – περίπου όχι»

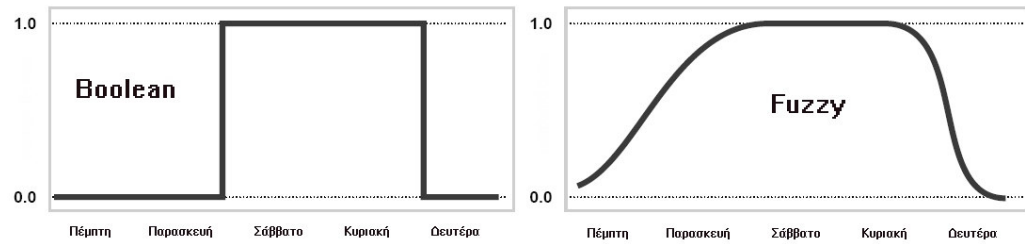
Η σκέψη στην ασαφή λογική είναι απλώς ένα πρόβλημα γενίκευσης της γνωστής Ναι-Όχι λογικής Boolean. Αν δώσουμε την τιμή 1 στο «αληθές» και την τιμή 0 στο «ψευδές», τότε λέμε ότι η Α.Λ. επιτρέπει επίσης ενδιάμεσες τιμές, όπως 0,2 ή 0,486 ή 0,79.

Για παράδειγμα στο σχήμα 2 φαίνονται οι αληθείς «τιμές» των ημερών που θεωρούμε ότι ανήκουν στο σαββατοκύριακο, θεωρώντας ότι ο χρόνος παίρνει διακεκριμένες τιμές, πρώτα με Boolean Ναι-Όχι, και δίπλα απαντήσεις Α.Λ. με ενδιάμεσες τιμές.



Σχήμα 2. Ημερες που θεωρούμε ότι ανήκουν στο σαββατοκύριακο

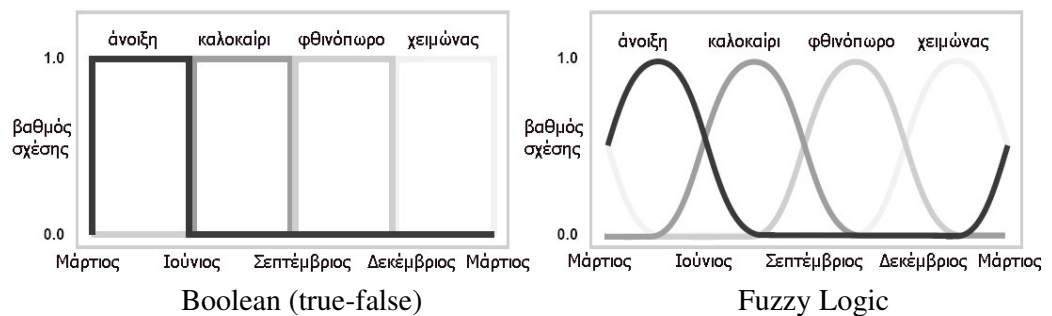
Το ίδιο πάλι παράδειγμα στο σχήμα 3, θεωρώντας όμως ότι ο χρόνος παίρνει συνεχείς τιμές, πρώτα με Boolean Ναι-Όχι, και δίπλα απαντήσεις Α.Λ. με ενδιάμεσες τιμές. Τώρα καθορίζουμε και το βαθμό στον οποίο κάθε στιγμή ανήκει στο σαββατοκύριακο, αντί για ολόκληρη τη μέρα κάθε φορά.



Σχήμα 3. Στιγμές που θεωρούμε ότι ανήκουν στο σαββατοκύριακο

Η καμπύλη που καθορίζει τις μέρες του σαββατοκύριακου κάθε στιγμής στο χρόνο, είναι μια συνάρτηση που αντιστοιχίζει τον χώρο των δεδομένων που εισάγουμε (χρόνος της εβδομάδας), στον χώρο των αποτελεσμάτων που εξάγουμε (χρόνος που ανήκει στο σαββατοκύριακο). Πιο συγκεκριμένα είναι γνωστή σαν συνάρτηση συμμετοχής (membership function).

Ένα άλλο παράδειγμα όμοιο με πριν αλλά για πολλαπλά γεγονότα φαίνεται στο σχήμα 4 όπου αναζητούμε ποια εποχή του χρόνου είναι τότε.

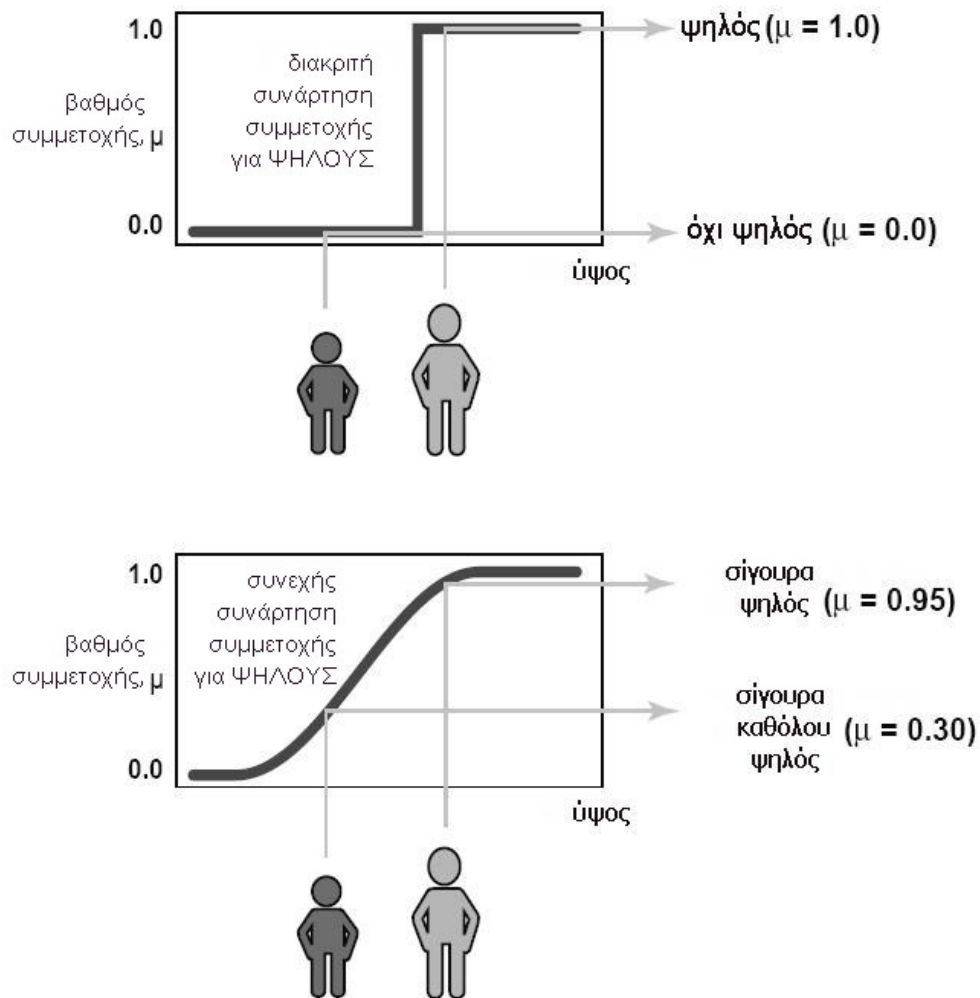


Σχήμα 4. Ποια εποχή του χρόνου είναι τότε.

1.2.2 ΣΥΝΑΡΤΗΣΕΙΣ ΣΥΜΜΕΤΟΧΗΣ (MEMBERSHIP FUNCTION).

Μια συνάρτηση συμμετοχής είναι μια καμπύλη, που προσδιορίζει πως κάθε σημείο στον χώρο εισόδου αντιστοιχίζεται σε μια τιμή συμμετοχής (ή βαθμό συμμετοχής) μεταξύ 0 και 1. [3]

Στον άξονα εξόδου (κάθετος y) αντιστοιχίζονται οι αριθμοί των βαθμών σχέσεων, μεταξύ 0 και 1. Η καμπύλη που είναι γνωστή και σαν συνάρτηση συμμετοχής και συχνά συμβολίζεται με το ελληνικό γράμμα «μ», προσδιορίζει τη μετάβαση από το «λίγο τόσο» στο «πολύ τόσο». Π.χ. αν αναφερόμαστε στο πόσο ψηλός είναι κάποιος, τότε προσδιορίζει τη μετάβαση από το «όχι τόσο ψηλός» στο «ψηλός». Καθώς όλοι οι άνθρωποι είναι ψηλοί σε κάποιον βαθμό, κάποιοι είναι σημαντικά λιγότερο ψηλοί από κάποιους άλλους.



Σχήμα 5. Διακριτές και συνεχείς συναρτήσεις συμμετοχής, «για ψηλούς». [3]

Η μόνη, ουσιαστικά, προϋπόθεση που πρέπει να ικανοποιεί μια συνάρτηση συμμετοχής είναι ότι πρέπει να διακυμαίνεται μεταξύ 0 και 1. Η συνάρτηση αυτή καθ' αυτή μπορεί να είναι μια αυθαίρετη καμπύλη της οποίας το σχήμα μπορούμε να προσαρμόσουμε όπως εμείς θέλουμε, σύμφωνα με τις ανάγκες μας για απλότητα, βολικότητα και επάρκεια.

Ένα κλασικό σύνολο θα μπορούσε να εκφραστεί ως εξής

$$A = \{x \mid x > 6\}$$

Ένα ασαφές σύνολο είναι μια προέκταση ενός κλασικού. Αν X είναι το σύμπαν ομιλίας και τα στοιχεία του δηλώνονται με x , τότε ένα ασαφές σύνολο A στο X , προσδιορίζεται σαν ένα σύνολο αντιστοιχισμένων ζευγών.

$$A = \{x, \mu_A(x) \mid x \in X\}$$

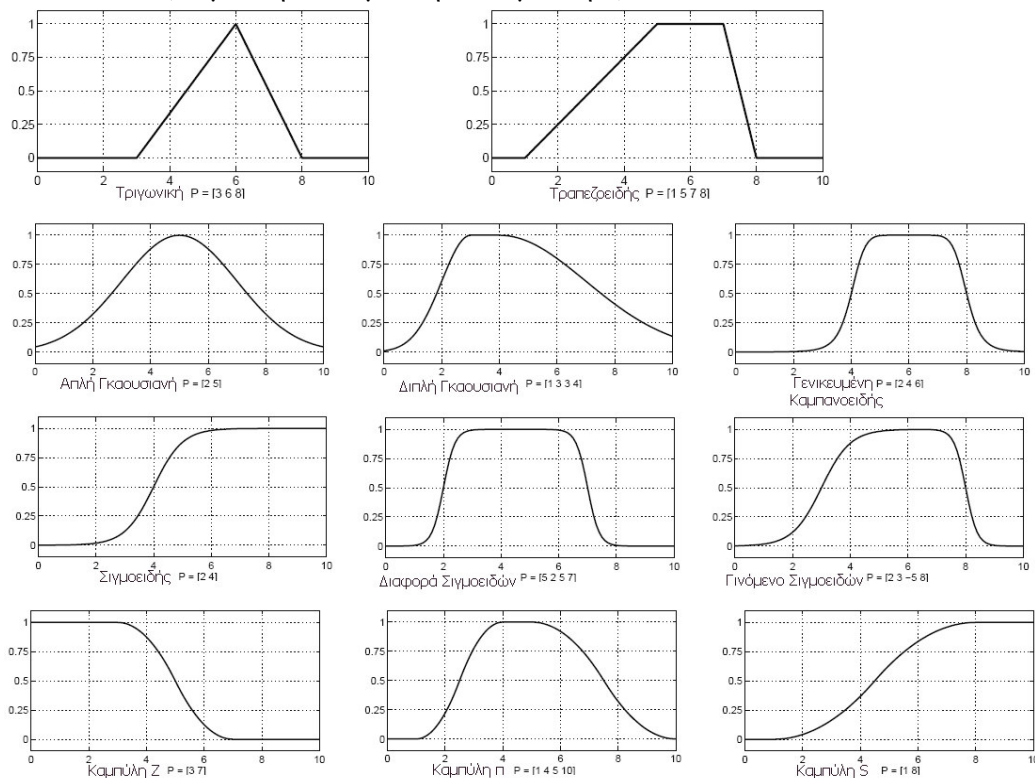
όπου $\mu_A(x)$ είναι η συνάρτηση συμμετοχής του x στο A .

Η συνάρτηση συμμετοχής αντιστοιχίζει κάθε στοιχείο του X σε μια τιμή σχέσης μεταξύ 0 και 1. Συνήθως ο χώρος των δεδομένων X που εισάγουμε, αναφέρεται και σαν **σύμπαν ομιλίας** ή απλά **σύμπαν** και μπορεί να αποτελείται από διακριτά (ταξινομημένα ή μη) αντικείμενα ή από συνεχές διάστημα.

1.2.2.1 ΠΑΡΑΔΕΙΓΜΑΤΑ ΣΥΝΑΡΤΗΣΕΩΝ ΣΥΜΜΕΤΟΧΗΣ

Κάποιες βασικές γνωστές συναρτήσεις συμμετοχής μπορεί:

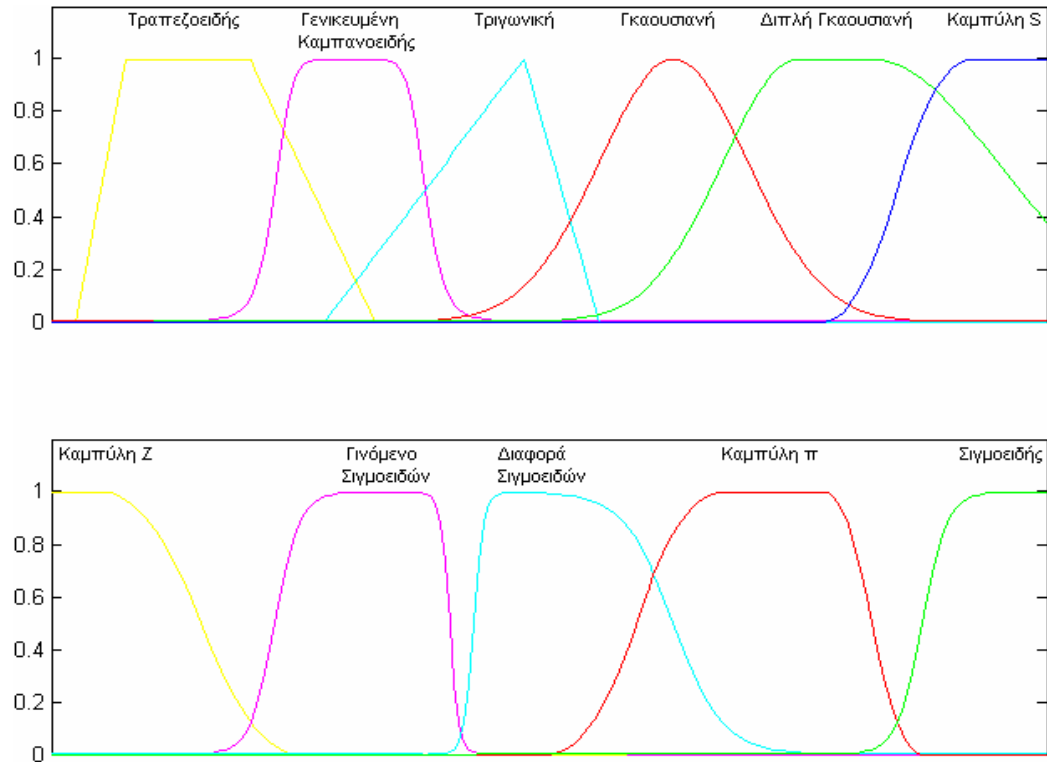
- να είναι **γραμμικές** (Τριγωνική, Τραπεζοειδής),
- να ακολουθούν την κατανομή **Gauss**
(Γκαουσσισιανή, Διπλή Γκαουσσισιανή, Γενικευμένη Καμπανοειδής),
- να είναι **σιγμοειδείς** καμπύλες
(Σιγμοειδής, Διαφορά Σιγμοειδών, Γινόμενο Σιγμοειδών),
- να είναι **τετραγωνικές** ή **κυβικές πολυωνυμικές** καμπύλες
(Καμπύλη Z, Καμπύλη π, Καμπύλη S).



Σχήμα 6. Παραδείγματα συναρτήσεων συμμετοχής. [3]

Στο σχήμα 6 βλέπουμε τις συναρτήσεις συμμετοχής (Σ.Σ.) κάθε μια ξεχωριστά, και τις παραμέτρους που χρειάζεται η κάθε μια, με το $P = [\dots]$ (τα νούμερα είναι τυχαία),

ενώ στο σχήμα 7 τις βλέπουμε συγκριτικά όλες μαζί.



Σχήμα 7. 2^ο παράδειγμα συναρτήσεων συμμετοχής. [3]

1.2.3 ΛΟΓΙΚΟΙ ΤΕΛΕΣΤΕΣ

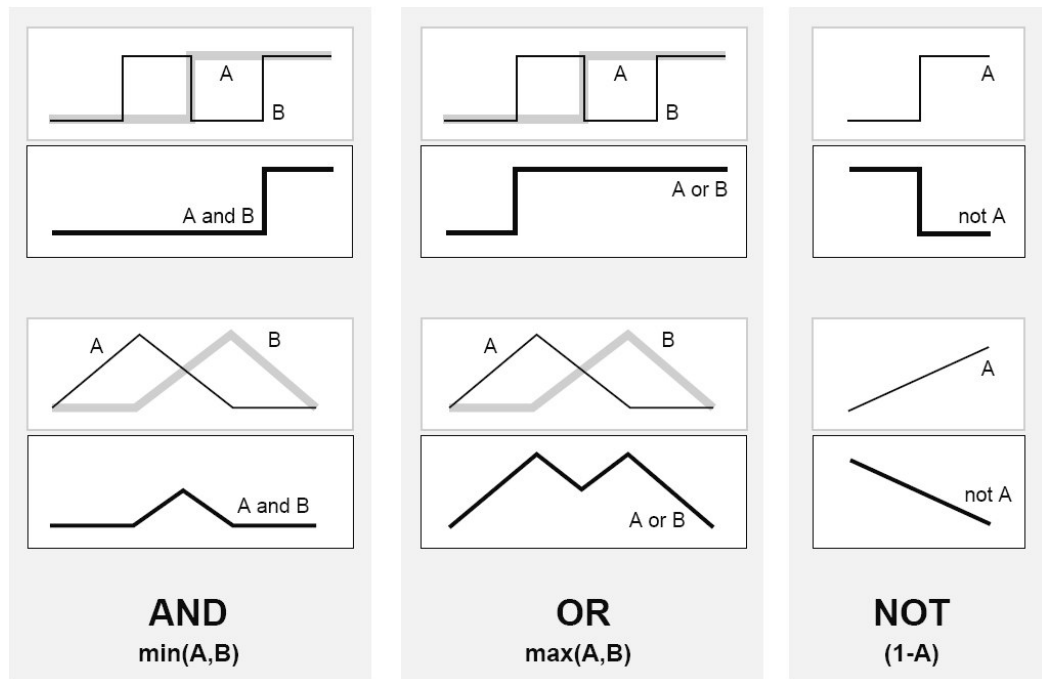
1.2.3.1 ΓΕΝΙΚΑ

Στην Α.Λ. μπορούμε να χρησιμοποιήσουμε τους τελεστές ΚΑΙ (AND), Ή (OR), ΟΧΙ (NOT), που είναι γενικοί μηχανισμοί πραγματοποίησης πράξεων. [3]

Το ΚΑΙ (AND) υλοποιείται μέσα από την πράξη \min ή την πράξη product (prod) (γινόμενο).

Το Ή (OR) υλοποιείται μέσα από την πράξη \max ή μέσα από το αλγεβρικό άθροισμα που ισούται με “ $a+b-ab$ ” (στα αγγλικά λέγεται επίσης και probabilistic OR, πιθανοτικό ή).

Με την βοήθεια των τελεστών μπορούμε να κατασκευάσουμε νέες καμπύλες από ήδη υπάρχουσες, όπως φαίνεται και από το σχήμα 8 καλύτερα.



Σχήμα 8. Κατασκευή νέων καμπυλών από ήδη υπάρχουσες, με την βοήθεια των τελεστών. [3]

1.2.3.2 ΕΙΔΙΚΟΤΕΡΑ

Η **τομή** δυο ασαφών συνόλων A και B προσδιορίζεται γενικώς από μια δυαδική αντιστοίχιση T , η οποία συνδυάζει δυο συναρτήσεις συμμετοχής ως εξής:

$$\mu_{A \ll B}(x) = T(\mu_A(x), \mu_B(x))$$

Οι τελεστές ασαφούς τομής αναφέρονται συνήθως ως τελεστές T -norm (Τριγωνικές νόρμες) και τηρούν τις εξής απαιτήσεις: [1]

$$\text{όρια: } T(0, 0) = 0, T(a, 1) = T(1, a) = a$$

$$\text{μονοτονία: } T(a, b) \leq T(c, d) \text{ if } a \leq c \text{ and } b \leq d$$

$$\text{αντιμεταθετικότητα: } T(a, b) = T(b, a)$$

$$\text{συσχέτιση: } T(a, T(b, c)) = T(T(a, b), c)$$

- Η πρώτη απαίτηση επιβάλλει την γενίκευση στα σαφή σύνολα.
- Η δεύτερη δείχνει, ότι μια μείωση στους βαθμούς (τιμές) συμμετοχής στο A ή στο B δεν μπορεί να αυξήσει τον βαθμό συμμετοχής στο $A \cap B$
- Η τρίτη δείχνει ότι ο τελεστής είναι αδιάφορος για την σειρά με την οποία θα εκτελεστούν οι πράξεις των ασαφών συνόλων.

- Η τέταρτη μας επιτρέπει να πάρουμε τομές μεταξύ οποιονδήποτε συνόλων σε οποιαδήποτε σειρά ομάδων ζευγών.

Τέσσερις από τους πιο γνωστούς τελεστές T-norm είναι:

- Minimum: $T_{min}(a,b) = \min(a,b) = a \wedge b$
 - Αλγεβρικό γινόμενο(algebraic product): $T_{ap}(a,b) = ab$
 - Καθορισμένο γινόμενο(bounded product): $T_{bp}(a,b) = 0 \vee (a + b - 1)$
 - Δραστικό γινόμενο (drastic product): $T_{dp}(a,b) = \begin{cases} a, \text{ αν } b = 1 \\ b, \text{ αν } a = 1 \\ 0, \text{ αν } a,b < 1 \end{cases}$
όπου $0 \leq a, b \leq 1$
-

Ο τελεστής ασαφούς **ένωσης** προσδιορίζεται γενικώς από μια δυαδική αντιστοίχιση S:

$$\mu_{A \cup B}(x) = S(\mu_A(x), \mu_B(x))$$

και αυτοί οι τελεστές συνήθως αναφέρονται σαν T-conorm (ή S-norm) και τηρούν τις εξής απαιτήσεις:

$$\text{όρια: } S(1, 1) = 1, S(0, a) = S(a, 0) = a$$

$$\text{μονοτονία: } S(a, b) \leq S(c, d) \text{ if } a \leq c \text{ and } b \leq d$$

$$\text{αντιμεταθετικότητα: } S(a, b) = S(b, a)$$

$$\text{συσχέτιση: } S(a, S(b, c)) = S(S(a, b), c)$$

οι επεξηγήσεις είναι ίδιες με παραπάνω στις T-norm

Τέσσερις από τους πιο γνωστούς τελεστές T-conorm είναι:

- Maximum: $S_{max}(a,b) = \max(a,b) = a \vee b$
- Αλγεβρικό άθροισμα(algebraic sum): $S_{as}(a,b) = a + b - ab$
- Καθορισμένο άθροισμα (bounded sum): $S_{bs}(a,b) = 1 \wedge (a + b)$
- Δραστικό άθροισμα (drastic sum): $S_{ds}(a,b) = \begin{cases} a, \text{ αν } b = 0 \\ b, \text{ αν } a = 0 \\ 1 \text{ αν } a,b > 0 \end{cases}$
όπου $0 \leq a, b \leq 1$

1.2.4 ΚΑΝΟΝΕΣ ΑΝ-ΤΟΤΕ (IF-THEN)

Ένας απλός κανόνας αν-τότε έχει την μορφή:

if x is A then y is B

όπου A και B είναι οι γλωσσικές προτάσεις που στα ασαφή σύνολα καθορίζονται στο πεδίο τιμών των X και Y αντιστοίχως.

Το πρώτο μέρος της πρότασης, if x is A, ονομάζεται **προϋπόθεση**, και το δεύτερο μέρος της πρότασης, then y is B, ονομάζεται **συμπέρασμα**. [3]

Η μετάφραση των κανόνων αν-τότε έχει διακριτά μέρη:

1^{ov} : υπολογίζουμε τις προϋποθέσεις, κάνοντας ασαφή τα δεδομένα της εισόδου, δίνοντας τους, τους βαθμούς σχέσης μεταξύ 0 και 1

2^{ov} : εφαρμόζουμε τους όποιους αναγκαίους ασαφείς τελεστές

3^{ov} : εφαρμόζουμε το αποτέλεσμα του 2^{ov} βήματος στα συμπεράσματα (γνωστό και σαν **επαγωγή** ή **implication**). Τα συμπεράσματα ενός ασαφούς κανόνα αναθέτουν ένα ολόκληρο ασαφές σύνολο στα δεδομένα της εξόδου. Το ασαφές σύνολο αντιπροσωπεύεται από μια συνάρτηση συμμετοχής που επιλέγεται για να καταδείξει τα ποιοτικά χαρακτηριστικά των προϋποθέσεων. Αν οι προϋποθέσεις είναι εν μέρη αληθείς (βαθμός <1) τότε το ασαφές σύνολο της εξόδου περικόπτεται σύμφωνα με την επαγωγική μέθοδο.

Η προϋπόθεση ενός κανόνα μπορεί να έχει πολλά κομμάτια

if sky is gray and wind is strong and barometer is falling, then ...

τα οποία όμως υπολογίζονται ταυτόχρονα σε ένα αποτέλεσμα με την βοήθεια των λογικών τελεστών.

Το συμπέρασμα ενός κανόνα μπορεί επίσης να έχει πολλά κομμάτια

If ... then hot water valve is open and cold water valve is shut

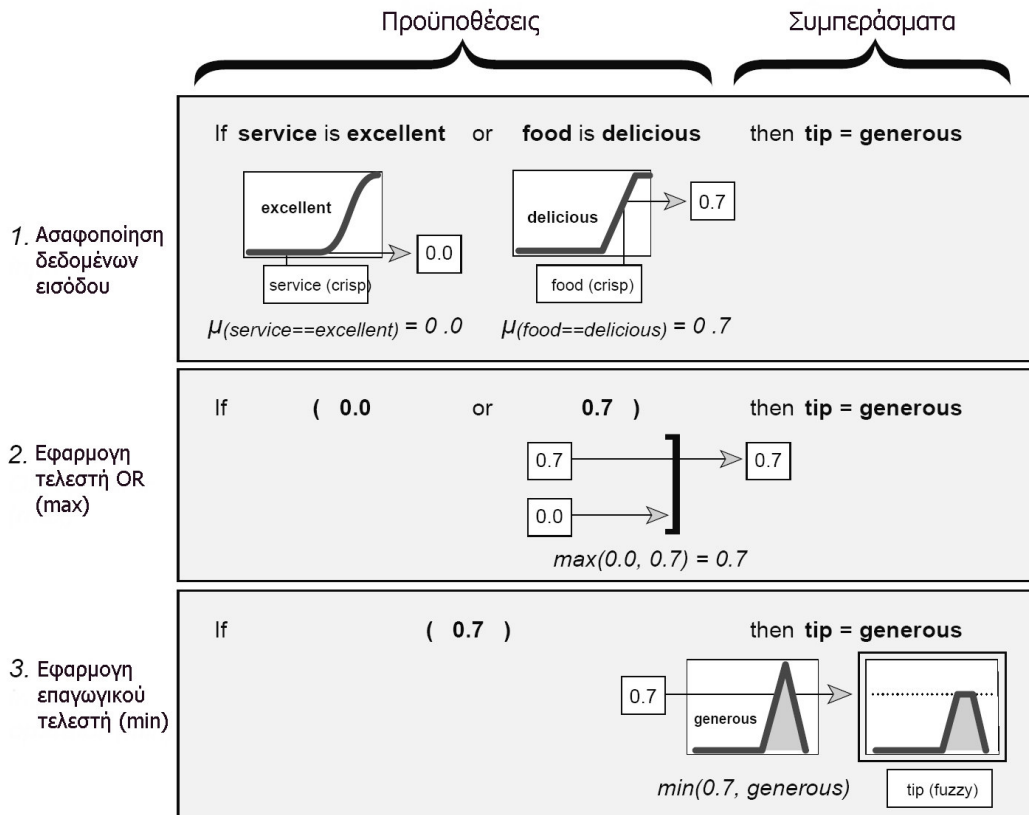
τα οποία επηρεάζονται ισοβαρώς από το αποτέλεσμα των προϋποθέσεων. Τα συμπεράσματα επηρεάζονται από τις προϋποθέσεις ως εξής:

Τα συμπεράσματα καθορίζουν ένα ασαφές σύνολο για την έξοδο.

Η επαγωγική συνάρτηση, τότε τροποποιεί το ασαφές σύνολο στο βαθμό που καθορίστηκε από τις προϋποθέσεις.

Ο συνηθέστερος τρόπος να τροποποιήσουμε την έξοδο του ασαφούς συνόλου είναι η περικοπή του συνόλου χρησιμοποιώντας την πράξη \min (ελάχιστο) ή η κλιμακοποίηση του χρησιμοποιώντας την πράξη product (γινόμενο).

Αυτά μπορούμε να τα παρατηρήσουμε ευκολότερα στο σχήμα 9, όπου στο τέλος γίνεται χρήση περικοπής. (Τα $=$ δηλώνουν το is)



Σχήμα 9. Οι κανόνες if-then στην ασαφή λογική. [3]

Γενικά ένας κανόνας από μόνος του δεν μπορεί να κάνει πολλά. Είναι απαραίτητοι δυο ή παραπάνω κανόνες. Η έξοδος κάθε κανόνα είναι ένα ασαφές σύνολο. Όλα αυτά τα σύνολα μετά ενώνονται σε ένα ενιαίο σύνολο. Τελικά το ενιαίο σύνολο αποασαφοποιείται για να μας δώσει έναν τελικό αριθμό.

1.3 ΑΣΑΦΗ ΣΥΣΤΗΜΑΤΑ ΕΞΑΓΩΓΗΣ ΣΥΜΠΕΡΑΣΜΑΤΩΝ (FUZZY INFERENCE SYSTEMS)

Το σημαντικότερο εργαλείο μοντελοποίησης το οποίο βασίζεται στην θεωρία ασαφών συνόλων είναι το FIS [1]. Τα FIS είναι η διαδικασία διατύπωσης και αντιστοίχισης κάποιων εισαγόμενων δεδομένων, σε κάποια εξαγόμενα δεδομένα κάνοντας χρήση ασαφούς λογικής. Η αντιστοίχιση των δεδομένων παρέχει μια βάση από την οποία θα εξάγουμε τις αποφάσεις ή θα διακρίνουμε υποδείξεις. [3]

Δυο γνωστοί τύποι FIS είναι:

→ το Mamdani που πρωτοδιατυπώθηκε το 1975 από τον Ebrahim Mamdani

→ και το Sugeno ή Takagi-Sugeno-Kang που παρουσιάστηκε το 1985.

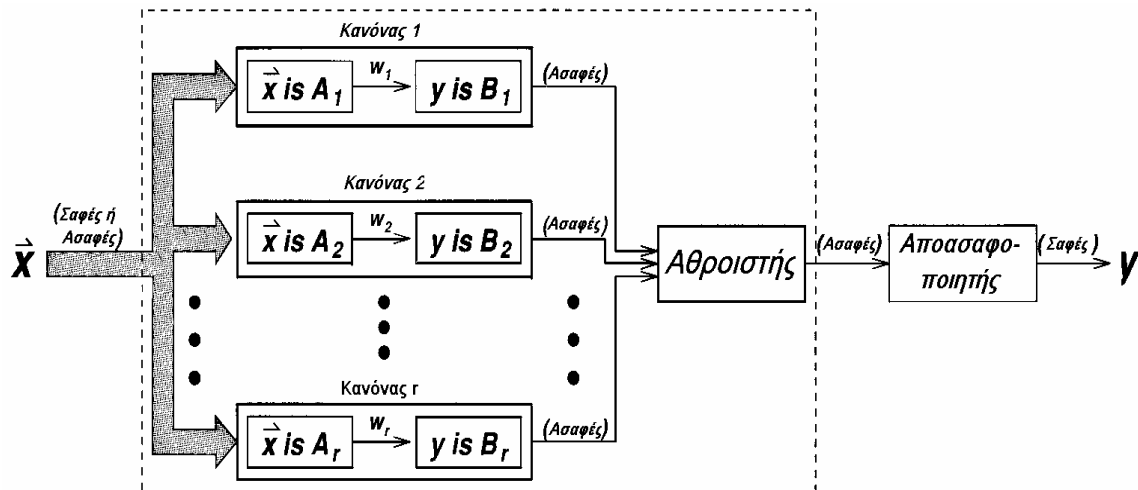
Εφαρμογές των FIS έχουμε τομείς όπως ο αυτόματος έλεγχος, ταξινόμηση δεδομένων, ανάλυση αποφάσεων, έμπειρα συστήματα και όραση υπολογιστών.

Η βασική δομή ενός FIS αποτελείται από τρία μέρη:

- μια βάση κανόνων, που περιέχει μια ομάδα ασαφών κανόνων
- μια βάση δεδομένων (ή λεξικό), που καθορίζει τις συναρτήσεις συμμετοχής που χρησιμοποιούνται στους ασαφείς κανόνες
- και τον συλλογιστικό μηχανισμό (Reasoning mechanism), που εκτελεί την εξαγωγή συμπερασμάτων με βάση τους κανόνες και τα δεδομένα, ώστε η έξοδος ή το συμπέρασμα να είναι αποδεκτό.

Ένα βασικό FIS μπορεί να πάρει είτε ασαφείς εισόδους, είτε σαφείς (συγκεκριμένες) με την μορφή ασαφών singletons. Η έξοδος όμως που παράγει είναι πάντα ασαφή σύνολα. Μερικές φορές είναι αναγκαίο να έχουμε σαφή έξοδο, ειδικά όταν το FIS χρησιμοποιείται ως ελεγκτής. Τότε χρειαζόμαστε μια μέθοδο αποασαφοποίησης, για να εξάγουμε ένα σαφές αποτέλεσμα που αντιπροσωπεύει με τον καλύτερο τρόπο το ασαφές σύνολο.

Με σαφείς εισόδους και εξόδους, ένα FIS υλοποιεί μια μη γραμμική αντιστοίχιση, από τον χώρο εισόδου σ' αυτόν της εξόδου, με έναν αριθμό κανόνων if-then, καθένας εκ των οποίων περιγράφει μια τοπική συμπεριφορά της αντιστοίχισης. Ειδικότερα, το μέρος των προϋποθέσεων ενός κανόνα προσδιορίζει μια ασαφή περιοχή στο χώρο εισόδου, ενώ αυτός των συμπερασμάτων προσδιορίζει την έξοδο σε μια ασαφή περιοχή.

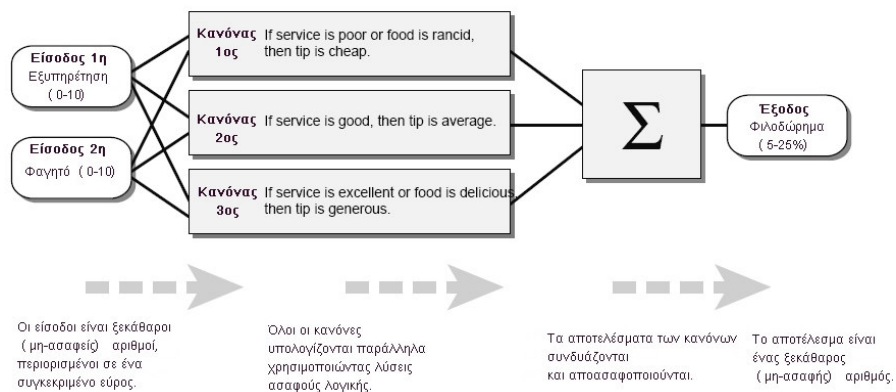


Σχήμα 10. Διάγραμμα ενός ασαφούς συστήματος εξαγωγής συμπερασμάτων. [1]

1.3.1 FIS ΤΥΠΟΥ MAMDANI

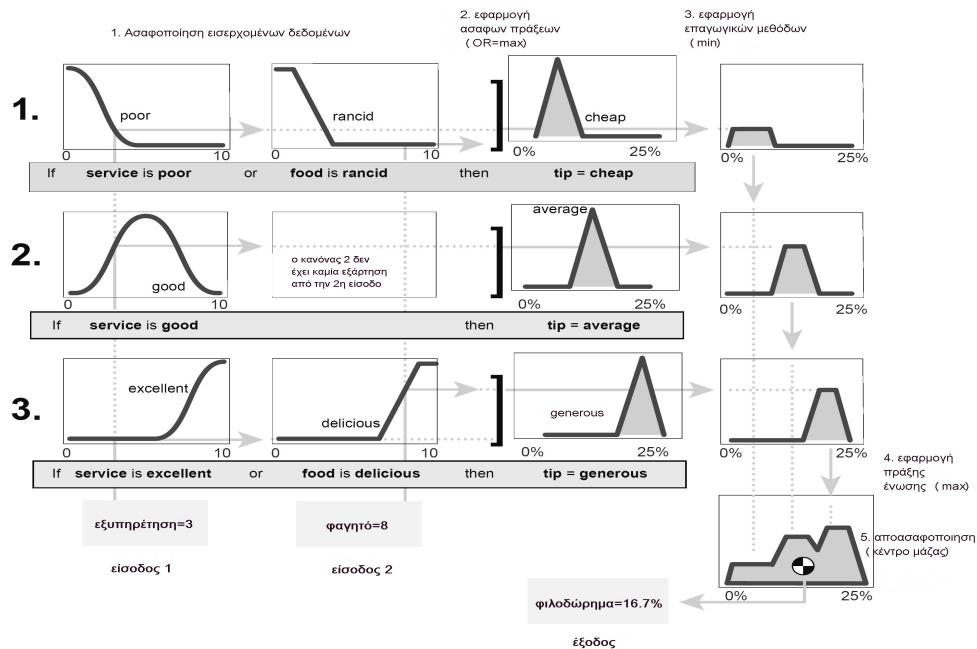
Στο Mamdani, πρέπει οι συναρτήσεις συμμετοχής να είναι ασαφή σύνολα. Μετά την διαδικασία της ενώσεως, υπάρχει ένα ασαφές σύνολο για κάθε μεταβλητή εξόδου που χρειάζεται αποασαφοποίηση.

Στο σχήμα 11 μπορούμε να δούμε τη λογική σχεδίαση ενός Mamdani στο ίδιο πρόβλημα του φιλοδώρηματος σε ένα εστιατόριο. [3]



Σχήμα 11. Σύστημα 2 εισόδων, 3 κανόνων και 1 έξοδος. [3]

Ενώ στο σχήμα 12 φαίνεται σε πλήρη ανάπτυξη ένα διάγραμμα FIS Mamdani.



Σχήμα 12. FIS Mamdani, για το ποσοστό φιλοδωρήματος.[3]

1.3.2 FIS ΤΥΠΟΥ SUGENO

Το ασαφές μοντέλο Sugeno δημιουργήθηκε σε μια προσπάθεια να αναπτυχθεί μια συστηματική προσέγγιση στην γέννηση ασαφών κανόνων από ένα δεδομένο σύνολο δεδομένων εισόδου – εξόδου. [1]

Σε πολλές περιπτώσεις είναι περισσότερο αποτελεσματική η χρήση μεμονωμένων σημείων-αιχμών σαν συναρτήσεις συμμετοχής εξόδου, από ένα συνεχές κατανομημένο ασαφές σύνολο. Μια τέτοια αιχμή είναι γνωστή με τον αγγλικό όρο συνάρτηση συμμετοχής εξόδου **singleton** και θα μπορούσε να χαρακτηριστεί σαν προαποασαφοποιημένο ασαφές σύνολο. Η συνάρτηση συμμετοχής ενός singleton είναι $\mu_A(x) = 1$.

Αυτές εμπλουτίζουν την επάρκεια της διαδικασίας αποασαφοποίησης διότι απλοποιούν ιδιαιτέρως τους απαιτούμενους υπολογισμούς σε σχέση με την πιο γενική μέθοδο Mamdani, στην οποία βρίσκουμε το κέντρο μάζας μιας δισδιάστατης συνάρτησης. Αντί λοιπόν να βρίσκουμε το κέντρο μάζας, χρησιμοποιούμε τον σταθμικό μέσο μερικών σημείων. Τα συστήματα Sugeno υποστηρίζουν αυτόν τον τύπο του μοντέλου.

Η μέθοδος Sugeno είναι παρόμοια με τη Mamdani από πολλές απόψεις. Τα πρώτα δύο μέρη της διαδικασίας, η ασαφοποίηση των δεδομένων εισόδου και η εφαρμογή των ασαφών τελεστών, είναι ακριβώς ίδια. Η κύρια διαφορά του Sugeno από το Mamdani είναι ότι στο Sugeno η συναρτήσεις συμμετοχής εξόδου είναι είτε γραμμικές είτε μια σταθερά. [3]

Ένας κανόνας σε ένα ασαφές μοντέλο Sugeno έχει τη μορφή:

$$\text{If } x = A \text{ and } y = B, \text{ then } z = f(x, y), [1]$$

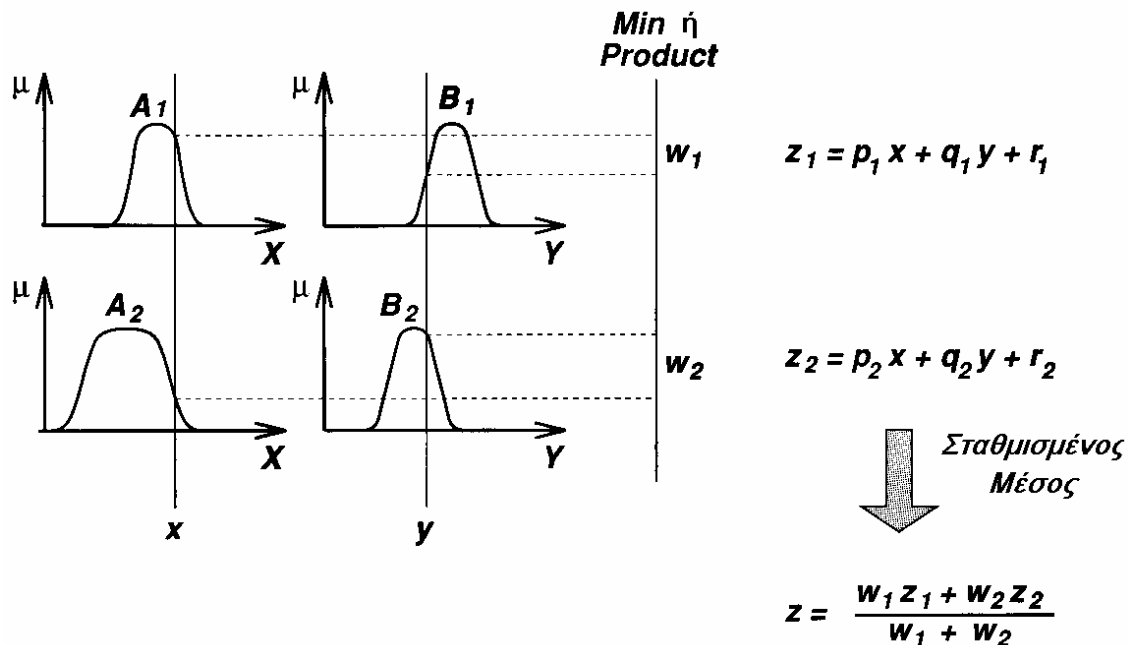
όπου A, B είναι ασαφή σύνολα των προϋποθέσεων και το $z = f(x, y)$ είναι μια σαφής συνάρτηση στον χώρο των συμπερασμάτων. Συνήθως το $f(x, y)$ είναι ένα πολυώνυμο των μεταβλητών εισόδου x, y , αλλά μπορεί να είναι μια οποιαδήποτε συνάρτηση, αρκεί να μπορεί να περιγράψει καταλλήλως την έξοδο του μοντέλου, μέσα στην ασαφή περιοχή που ορίστηκε από τις προϋποθέσεις των κανόνων. Όταν το $f(x, y)$ είναι πολυώνυμο πρώτου βαθμού, το FIS ονομάζεται ασαφές μοντέλο Sugeno πρώτου βαθμού [1]. Πχ:

$$\text{If Input 1} = x \text{ and Input 2} = y, \text{ then Output is } z = ax + by + c. [3]$$

Για ένα Sugeno μηδενικού βαθμού το επίπεδο εξόδου z είναι μια σταθερά ($a = b = 0$) [1] και μπορεί να θεωρηθεί και σαν μια ειδική περίπτωση FIS Mamdani, στο οποίο του κάθε κανόνας το συμπέρασμα, προσδιορίζεται από ένα ασαφές singleton (ή προαποασαφοποιημένο συμπέρασμα).

Η έξοδος ενός Sugeno μηδενικού βαθμού είναι μια ομαλή συνάρτηση των μεταβλητών εισόδου του, εφόσον οι γειτονικές ΣΣ στις προϋποθέσεις, έχουν αρκετή επικάλυψη.

Το σχήμα 13 δείχνει την διαδικασία της ασαφούς λογικής για ένα Sugeno πρώτου βαθμού.



Σχήμα 13. Το πρώτου βαθμού, ασαφές μοντέλο Sugeno. [1]

Το επίπεδο εξόδου z_i κάθε κανόνα είναι σταθμισμένο με βάρη, ή αλλιώς με μια **δύναμη πυροδότησης** (βαθμός ενεργοποίησης) w_i για κάθε κανόνα. [3]

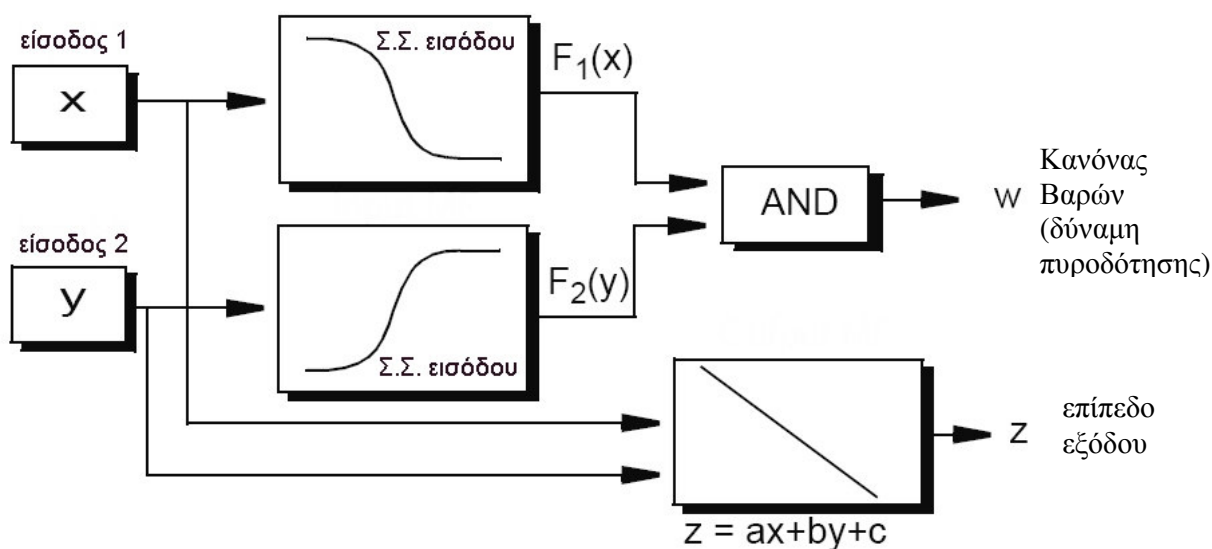
Για παράδειγμα για τον τελεστή ΚΑΙ, με 1^η είσοδο = x και 2^η είσοδο = y , η δύναμη πυροδότησης είναι :

$$w_i = \text{πράξη ΚΑΙ } (F_1(x), F_2(x))$$

όπου $F_{1,2}(\cdot)$ είναι οι συναρτήσεις συμμετοχής για τις εισόδους 1 και 2. Η τελική έξοδος του συστήματος είναι ο σταθμισμένος μέσος όλων των εξόδων και δίνεται από τον τύπο:

$$\text{Τελική Έξοδος} = \frac{\sum_{i=1}^N w_i z_i}{\sum_{i=1}^N w_i}$$

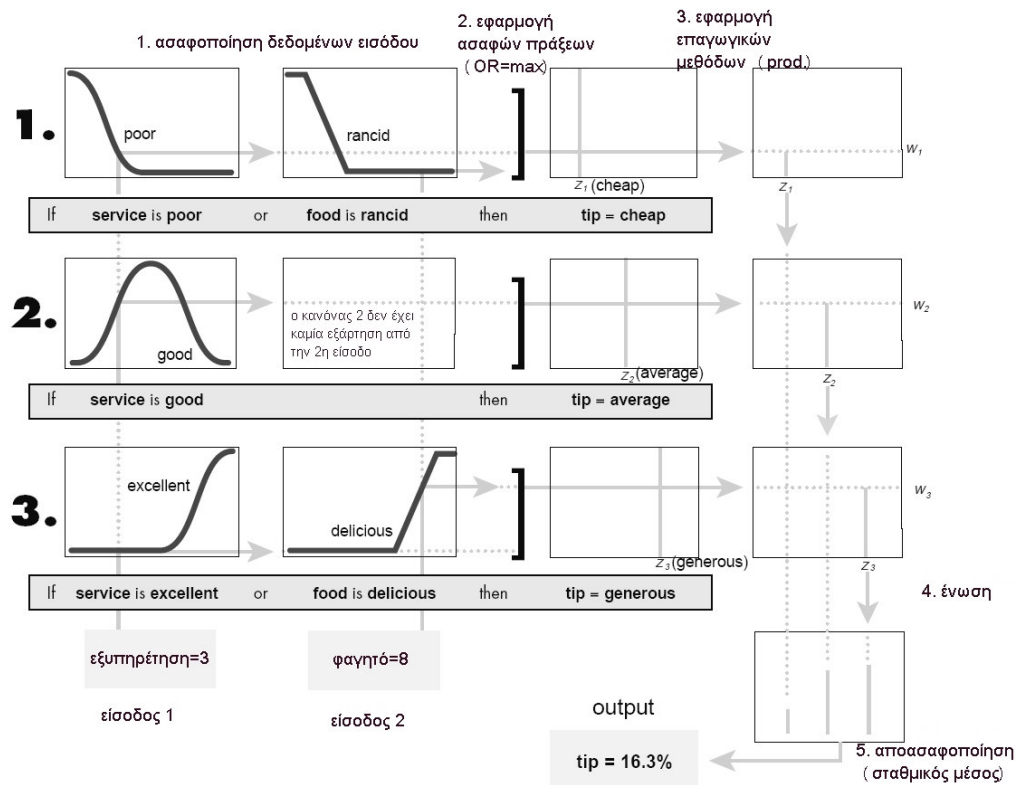
Ακόμα ένα σχήμα για την λειτουργία ενός Sugeno φαίνεται στο σχήμα 14



Σχήμα 14. Λειτουργία ενός Sugeno. [3]

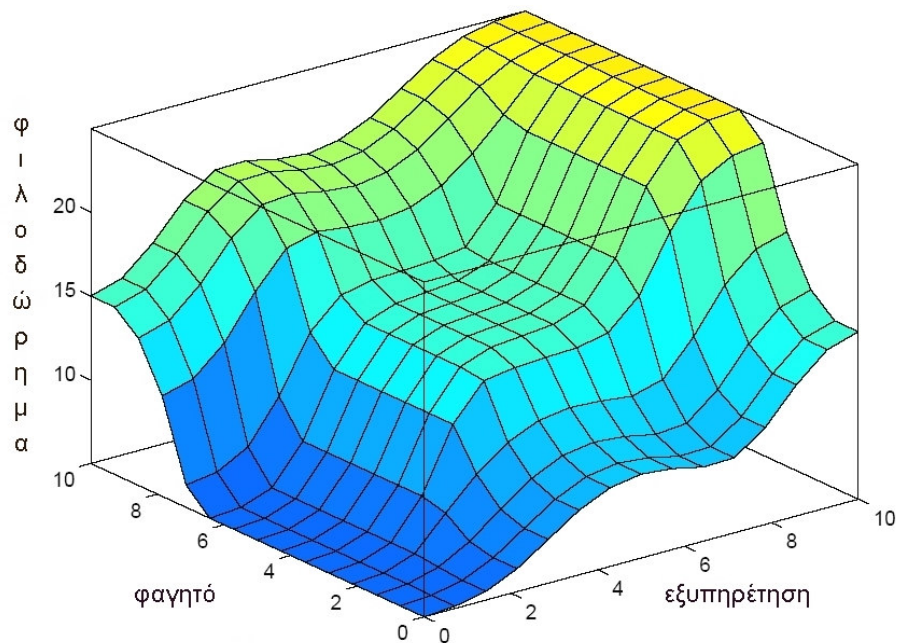
Ενώ στο σχήμα 15 φαίνεται σε πλήρη ανάπτυξη ένα διάγραμμα FIS Sugeno (για το πρόβλημα με το ποσοστό του φιλοδωρήματος). Για το δεδομένο πρόβλημα η χρήση συναρτήσεων singleton είναι απολύτως επαρκής.

Ένας εύκολος τρόπος να κατανοήσουμε τα πρώτου βαθμού Sugeno είναι να σκεφτούμε ότι κάθε κανόνας προσδιορίζει την θέση ενός «κινουμένου singleton». Οι αιχμές δηλαδή των singleton της εξόδου μπορούν να κινηθούν με γραμμικό τρόπο στον χώρο εξόδου, αναλόγως με το ποιες είναι οι εισόδους. Το σύστημα έτσι γίνεται αρκετά μικρό και επαρκές.



Σχήμα 15. FIS Sugeno, για το ποσοστό φιλοδωρήματος.[3]

Η τρισδιάστατη μορφή της τελικής επιφάνειας που αποτελεί την λύση του παραπάνω προβλήματος για όλες τις εισόδους και εξόδους (0-10) φαίνεται στο σχήμα 16



Σχήμα 16. Η επιφάνεια της λύση του προβλήματος για το φιλοδώρημα. [3]

Μπορούμε να κάνουμε και μοντέλα υψηλότερου βαθμού Sugeno, όμως τείνουν να παρουσιάζουν μεγάλη πολυπλοκότητα με μικρό προφανές κέρδος.

Τα συστήματα Sugeno λόγω του μικρότερου μεγέθους τους και του ότι είναι πιο επαρκή υπολογιστικά, σε σχέση με τα Mamdani, χρησιμοποιούνται σε προσαρμοστικές τεχνικές κατασκευής ασαφών μοντέλων.

Συνοπτικά η σύγκριση Sugeno – Mamdani στον πίνακα 1.1:

Πλεονεκτήματα της μεθόδου Sugeno
Επαρκέστερα υπολογιστικά
Καλή λειτουργία για γραμμικές τεχνικές
Καλή λειτουργία για βελτιστοποίηση και προσαρμοστικές τεχνικές
Εγγυημένη συνέχεια της τελικής επιφάνειας που αποτελεί την λύση
Καλά προσαρμοσμένο στην μαθηματική ανάλυση
Πλεονεκτήματα της μεθόδου Mamdani
Είναι διαισθητικά
Έχουν πολύ μεγάλη αποδοχή
Ευπροσάρμοστο στον ανθρώπινο τρόπο εισαγωγής των δεδομένα εισόδου

πίνακας 1.1. Τα συν των Sugeno – Mamdani.

1.4 ΜΕΘΟΔΟΣ ΕΛΑΧΙΣΤΩΝ ΤΕΤΡΑΓΩΝΩΝ ΓΙΑ ΤΑΥΤΟΠΟΙΗΣΗ ΣΥΣΤΗΜΑΤΟΣ

Το πρόβλημα του προσδιορισμού, ενός μαθηματικού μοντέλου για ένα άγνωστο σύστημα (ή σύστημα στόχος), με την βοήθεια μόνο που μας προσφέρουν οι παρατηρήσεις των δεδομένων εισόδου – εξόδου, ονομάζεται Ταυτοποίηση συστήματος. Αυτή μας βοηθάει:

- Να προβλέψουμε τη συμπεριφορά ενός συστήματος, όπως πρόβλεψη χρονοσειρών και μετεωρολογική πρόβλεψη.
- Να εξηγήσουμε τις αλληλεπιδράσεις και τις σχέσεις μεταξύ εισόδων και εξόδων ενός συστήματος.
- Να σχεδιάσουμε έναν ελεγκτή βασισμένο στο μοντέλο του ενός συστήματος, όπως για παράδειγμα έναν αυτόματο πιλότο.

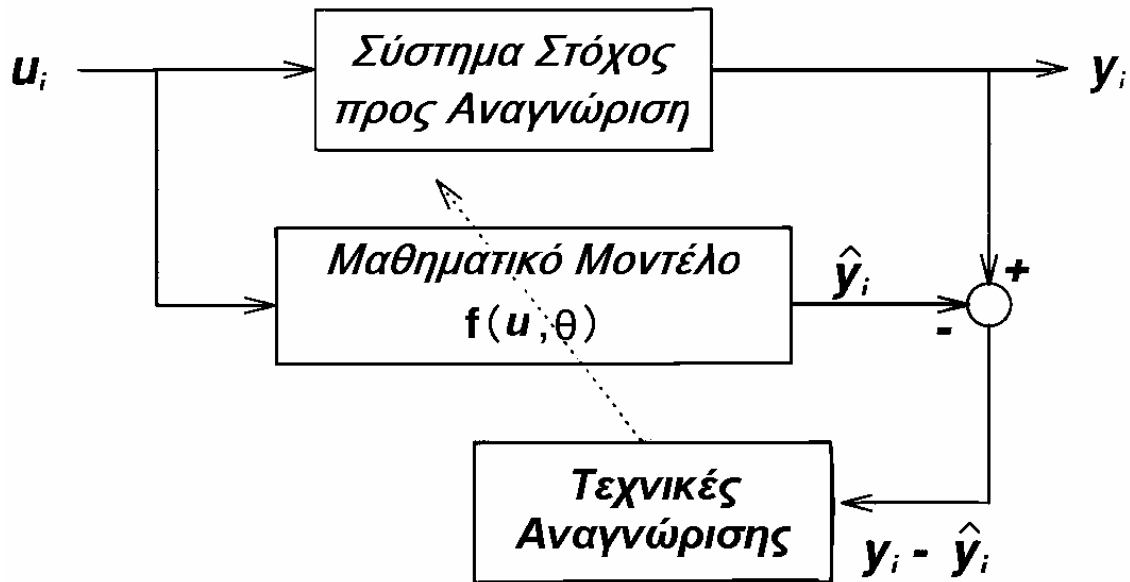
Στην Ταυτοποίηση συστήματος διακρίνουμε δυο βήματα:

- 1) **Ταυτοποίηση δομής.** Σ' αυτό το στάδιο, εφαρμόζουμε υπάρχουσες γνώσεις σχετικές με το σύστημα στόχος, για να αποφασίσουμε ποια κατηγορία μοντέλων μας ταιριάζει ώστε να ψάξουμε σ' αυτήν για το καταλληλότερο που θα εφαρμοστεί. Συνήθως αυτή η κατηγορία μοντέλων δηλώνεται από μια παραμετρική συνάρτηση $y = f(\mathbf{u}, \boldsymbol{\theta})$, όπου το y είναι η έξοδος του μοντέλου, \mathbf{u} το διάνυσμα εισόδου και $\boldsymbol{\theta}$ το διάνυσμα των παραμέτρων. Ο προσδιορισμός της συνάρτησης f εξαρτάται από το εκάστοτε πρόβλημα, την εμπειρία και την διαίσθηση του σχεδιαστή και τους νόμους που διέπουν το σύστημα στόχος.
- 2) **Ταυτοποίηση παραμέτρων.** Σ' αυτό το στάδιο, η δομή του μοντέλου είναι γνωστή και το μόνο που χρειαζόμαστε είναι να εφαρμόσουμε τεχνικές βελτιστοποίησης για να βρούμε το διάνυσμα των παραμέτρων $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ τέτοιο ώστε το προκύπτον μοντέλο $\hat{y} = f(\mathbf{u}, \hat{\boldsymbol{\theta}})$ να μπορεί να περιγράψει το σύστημα καταλλήλως.

Αν δεν έχουμε ήδη κάποια γνώση σχετική με το σύστημα στόχος, τότε η Ταυτοποίηση δομής είναι ένα δύσκολο πρόβλημα και πρέπει να επιλέξουμε την δομή του με την μέθοδο της δόκιμης και αποτυχίας.

Στο σχήμα 17 βλέπουμε ένα διάγραμμα ταυτοποίησης παραμέτρων, όπου μια είσοδος u_i εφαρμόζεται και στο σύστημα στόχος και στο μοντέλο, και η διαφορά μεταξύ της εξόδου y_i του συστήματος στόχος και της εξόδου \hat{y} του μοντέλου χρησιμοποιούνται καταλλήλως για την ενημέρωση του διανύσματος των παραμέτρων $\boldsymbol{\theta}$ για να μειωθεί η διαφορά. Το σύνολο των δεδομένων που δημιουργήθηκε από m επιθυμητά ζευγάρια εισόδου – εξόδου (u_i, y_i) , $i = 1, \dots, m$, καλείται **σύνολο δεδομένων εκπαίδευσης** ή **δειγματικό**

σύνολο δεδομένων. Στην πιο γενική περίπτωση τα u_i , y_i αντιπροσωπεύουν τα επιθυμητά διανύσματα εισόδου – εξόδου αντιστοίχως.



Σχήμα 17. Δομικό διάγραμμα για την ταυτοποίηση παραμέτρων. [1]

Γενικά η ταυτοποίηση συστήματος δεν είναι μια διαδικασία ενός περάσματος. Χρειάζεται να γίνει ταυτοποίηση δομής και παραμέτρων, επανειλημμένως, έως ότου βρεθεί ένα ικανοποιητικό μοντέλο, με βάση τα κάτωθι βήματα:

- 1) προσδιορισμός και παραμετροποίηση μιας κατηγορίας μαθηματικών μοντέλων που αντιπροσωπεύουν το σύστημα που πρόκειται να αναγνωρισθεί
- 2) Πραγματοποίηση ταυτοποίησης παραμέτρων, για να επιλεγούν αυτές που ταιριάζουν καλύτερα στο σύνολο δεδομένων εκπαίδευσης
- 3) Πραγματοποίηση ελέγχων επιβεβαίωσης για να δούμε αν το μοντέλο που αναγνωρίστηκε ανταποκρίνεται σωστά σε ένα άγνωστο σύνολο δεδομένων. Αυτό το σύνολο δεδομένων αποσπάται από το σύνολο δεδομένων εκπαίδευσης και ονομάζεται **δεδομένα ελέγχου**.
- 4) Τερματισμός της διαδικασίας μόλις τα αποτελέσματα του ελέγχου επιβεβαίωσης είναι ικανοποιητικά. Αλλιώς μια άλλη κατηγορία μοντέλων επιλέγεται και τα επαναλαμβάνονται τα βήματα 2 έως 4.

Για γραμμικά (αλλά και για μη γραμμικά) μοντέλα και στατικά (ή χωρίς μνήμη) συστήματα η μέθοδος των ελαχίστων τετραγώνων (Μ.Ε.Τ.) στην Ταυτοποίηση συστήματος είναι ένα μαθηματικό εργαλείο αρκετά ισχυρό και αρκετά μελετημένο, το οποίο έχει χρησιμοποιηθεί σε τομείς όπως ο προσαρμοστικός έλεγχος, η επεξεργασία σήματος και η στατιστική. Σε

προβλήματα με νεύρο-ασαφή μοντέλα, οι γραμμικές MET μας παρέχουν το βασικότερο και σημαντικότερο μαθηματικό θεμέλιο για την επίλυση τους.

Λέγοντας γραμμικά μοντέλα εννοούμε αυτά που είναι γραμμικά στις παραμέτρους τους. Γι' αυτό και ένα γραμμικό μοντέλο μπορεί να μην είναι γραμμικό στις εισόδους του. Οι MET μας παρέχουν μαθηματικές διαδικασίες με τις οποίες ένα γραμμικό μοντέλο μπορεί να επιτύχει την καλύτερη προσαρμογή σε πειραματικά δεδομένα με την έννοια του καλύτερου μέσου τετραγωνικού σφάλματος.

Λέγοντας στατικά συστήματα εννοούμε ότι η έξοδος του συστήματος στόχος εξαρτάται από τις παρούσες εισόδους μόνο και όχι από την ιστορία που έχουν οι είσοδοι.

1.4.1 ΕΚΤΙΜΗΤΡΙΑ ΕΛΑΧΙΣΤΩΝ ΤΕΤΡΑΓΩΝΩΝ

Στο γενικό πρόβλημα των ελαχίστων τετραγώνων, η έξοδος ενός γραμμικού μοντέλου y δίνεται από μια γραμμικώς παραμετρική σχέση

$$y = \theta_1 f_1(\mathbf{u}) + \theta_2 f_2(\mathbf{u}) + \dots + \theta_n f_n(\mathbf{u}) \quad (1)$$

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 άγνωστοι προς Εκτίμηση παράμετροι. Στην στατιστική, η διαδικασία της προσαρμογής δεδομένων, χρησιμοποιώντας ένα γραμμικό μοντέλο αναφέρεται και ως **γραμμική παλινδρόμηση**. Γι' αυτό και η εξίσωση (1) ονομάζεται και **συνάρτηση παλινδρόμησης** και τα θ_i **συντελεστές παλινδρόμησης**.

Για να αναγνωρίσουμε τις άγνωστες παραμέτρους θ_i , συνήθως πρέπει να πραγματοποιήσουν κάποιες δόκιμες για να αποκτήσουμε ένα σύνολο δεδομένων εκπαίδευσης κατασκευασμένο από ζεύγη δεδομένων

$\{(\mathbf{u}_i, y_i), i = 1, \dots, m\}$ που αντιπροσωπεύουν τα επιθυμητά ζεύγη δεδομένων εισόδου – εξόδου του προς μοντελοποίηση συστήματος στόχος. Αντικαθιστώντας κάθε ζεύγος μέσα στην εξίσωση (1), προκύπτει ένα σύνολο m γραμμικών εξισώσεων:

$$\left\{ \begin{array}{l} f_1(\mathbf{u}_1) \theta_1 + f_2(\mathbf{u}_1) \theta_2 + \dots + f_n(\mathbf{u}_1) \theta_n = y_1, \\ f_1(\mathbf{u}_2) \theta_1 + f_2(\mathbf{u}_2) \theta_2 + \dots + f_n(\mathbf{u}_2) \theta_n = y_2, \\ \dots \\ f_1(\mathbf{u}_m) \theta_1 + f_2(\mathbf{u}_m) \theta_2 + \dots + f_n(\mathbf{u}_m) \theta_n = y_m, \end{array} \right.$$

Σε μορφή πινάκων ξαναγράφουμε τις παραπάνω εξισώσεις σε μια πιο σύντομη μορφή:

$$\mathbf{A} \boldsymbol{\theta} = \mathbf{y}, \quad (2)$$

όπου \mathbf{A} είναι ένας πίνακας $m \times n$:

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

όπου $\boldsymbol{\theta}$ είναι ένα διάνυσμα $n \times 1$ των αγνώστων παραμέτρων:

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

και \mathbf{y} είναι ένα $m \times 1$ διάνυσμα εξόδου:

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

Η i -οστή σειρά του επαυξημένου πίνακα δεδομένων $[\mathbf{A}:\mathbf{y}]$, που συμβολίζεται με $[\mathbf{a}_i^T : y_i]$, σχετίζεται με το i -οστό ζεύγος δεδομένων εισόδου – εξόδου

(\mathbf{u}_i, y_i) μέσα από τη σχέση:

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

Για να αναγνωρίσουμε ένα μοναδικό $\boldsymbol{\theta}$, είναι απαραίτητο να είναι $m \geq n$. Αν ο \mathbf{A} είναι τετραγωνικός και με ορίζουσα διάφορη του μηδενός, τότε μπορούμε να βρούμε το $\boldsymbol{\theta}$ επιλύοντας την εξίσωση

$$\boldsymbol{\theta} = \mathbf{A}^{-1} \mathbf{y}.$$

Συνήθως το m είναι όντως μεγαλύτερο του n , υποδηλώνοντας ότι έχουμε περισσότερα ζεύγη δεδομένων απ' ότι παραμέτρους για προσαρμογή. Σ' αυτήν την περίπτωση η ακριβής λύση που να ικανοποιεί και τις m εξισώσεις δεν είναι πάντα δυνατή, μιας και τα δεδομένα μπορεί να εμπεριέχουν θόρυβο, ή το μοντέλο να μην είναι κατάλληλο για να περιγράψει το σύστημα στόχος. Γι' αυτό και η εξίσωση (2) τροποποιείται έτσι ώστε να ενσωματώνει το **διάνυσμα σφάλματος \mathbf{e}** για να λογαριάσουμε τον τυχαίο θόρυβο ή το σφάλμα του μοντέλου, ως ακολούθως:

$$\mathbf{A} \boldsymbol{\theta} + \mathbf{e} = \mathbf{y}.$$

Έτσι αντί να βρίσκουμε την ακριβή λύση της (2), ψάχνουμε να βρούμε ένα

$\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ που να ελαχιστοποιεί το άθροισμα του τετραγωνικού σφάλματος, που προσδιορίζεται από την εξίσωση

$$E(\boldsymbol{\theta}) = \sum_{i=1}^m (y_i - \mathbf{a}_i^T \boldsymbol{\theta})^2 = \mathbf{e}^T \mathbf{e} = (\mathbf{y} - \mathbf{A} \boldsymbol{\theta})^T (\mathbf{y} - \mathbf{A} \boldsymbol{\theta}), \quad (3)$$

όπου $\mathbf{e} = \mathbf{y} - \mathbf{A} \boldsymbol{\theta}$ είναι το διάνυσμα σφάλματος που προκύπτει από μια συγκεκριμένη επιλογή ενός $\boldsymbol{\theta}$. Το $E(\boldsymbol{\theta})$ είναι πολυώνυμο δευτέρου βαθμού και έχει ένα μοναδικό ελάχιστο στο $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$.

1.4.2 ΘΕΩΡΗΜΑ ΕΚΤΙΜΗΤΡΙΑΣ ΕΛΑΧΙΣΤΩΝ ΤΕΤΡΑΓΩΝΩΝ

Το τετραγωνικό σφάλμα στην εξίσωση (3) ελαχιστοποιείται όταν $\theta = \hat{\theta}$, το οποίο καλείται Εκτιμήτρια ελαχίστων Τετραγώνων (least squares estimator ή LSE για συντομία) και ικανοποιεί την κανονική εξίσωση

$$\mathbf{A}^T \mathbf{A} \hat{\theta} = \mathbf{A}^T \mathbf{y}.$$

Αν το $\mathbf{A}^T \mathbf{A}$ έχει ορίζουσα διάφορη του μηδενός, το $\hat{\theta}$ είναι μοναδικό και δίνεται από την εξίσωση

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

1.4.3 ΓΕΝΙΚΕΥΣΗ ΤΕΤΡΑΓΩΝΙΚΟΥ ΣΦΑΛΜΑΤΟΣ ΜΕ ΧΡΗΣΗ ΒΑΡΩΝ

Στα προηγούμενα θεωρούσαμε ότι κάθε στοιχείο του διανύσματος σφάλματος \mathbf{e} είχε το ίδιο βάρος στο τελικό τετραγωνικό σφάλμα. Μια μεγαλύτερη γενίκευση πετυχαίνουμε αν σε κάθε σφάλμα αποδώσουμε και ένα διαφορετικό βάρος. Τότε το σταθμικό τετραγωνικό σφάλμα θα είναι:

$$E_w(\theta) = (\mathbf{y} - \mathbf{A} \theta)^T \mathbf{W} (\mathbf{y} - \mathbf{A} \theta), \quad (4)$$

όπου \mathbf{W} είναι ο πίνακας των βαρών και είναι συμμετρικός και θετικά ορισμένος.

Η ελαχιστοποίηση του $E_w(\theta)$ ως προς το θ μας δίνει την **σταθμική Εκτιμήτρια ελαχίστων τετραγώνων $\hat{\theta}_w$** :

$$\hat{\theta}_w = (\mathbf{A}^T \mathbf{W} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{W} \mathbf{y}.$$

Το $\hat{\theta}_w$ μετατρέπεται σε $\hat{\theta}$ όταν το \mathbf{W} επιλέγεται ως μοναδιαίος πίνακας.

Στο σύστημα $\mathbf{A} \theta = \mathbf{y}$, όσο αυξάνουμε τους βαθμούς ελευθερίας, δηλαδή χρησιμοποιούμε στον πίνακα \mathbf{A} πολυώνυμα μεγαλύτερου βαθμού, μπορούμε να πάρουμε όλο και μικρότερο τετραγωνικό σφάλμα. Η καλύτερη προσαρμογή όμως στο σύνολο δεδομένων εκπαίδευσης, ενός πολυωνύμου μεγαλύτερου βαθμού δεν σημαίνει ότι θα αντιπροσωπεύει και καλύτερα τα πραγματικά χαρακτηριστικά του αναζητούμενου συστήματος.

Πρέπει λοιπόν να κάνουμε την βέλτιστη επιλογή σφάλματος, λαμβάνοντας υπόψιν την υπερπροσαρμογή και προσπαθώντας να την αποφύγουμε για πάρουμε το μοντέλο που θα έχει την καλύτερη γενίκευση, για να ανταποκρίνεται ικανοποιητικά στα πραγματικά άγνωστα δεδομένα ή στα δεδομένα ελέγχου.

Ένας τρόπος για να δούμε αν επιλέξαμε το πολυώνυμο κατάλληλου βαθμού είναι να εφαρμόσουμε ένα άλλο σύνολο δεδομένων εισόδου – εξόδου

που το ονομάζουμε **σύνολο δεδομένων δόκιμης** ή **σύνολο δεδομένων ελέγχου αξιοπιστίας**. Αυτό το σύνολο δεν πρέπει να έχει χρησιμοποιηθεί κατά την εκπαίδευση για την κατασκευή του πολυωνύμου των ελαχίστων τετραγώνων.

1.4.4 ΠΟΛΛΑΠΛΕΣ ΈΞΟΔΟΙ

Αν το σύστημα μας έχει αντί για μια, q εξόδους που εκφράζονται ως $\mathbf{y} = [y_1, \dots, y_q]^T$, με $q > 1$ τότε έχουμε ένα σύνολο γραμμικών εξισώσεων σε μορφή πινάκων:

$$\mathbf{A} \boldsymbol{\Theta} + \mathbf{E} = \mathbf{Y}$$

όπου \mathbf{A} είναι ένας πίνακας $m \times n$, όπως και στην προηγούμενη περίπτωση

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

όπου $\boldsymbol{\Theta}$ είναι ένας πίνακας $n \times q$ των αγνώστων παραμέτρων:

$$\boldsymbol{\Theta} = \begin{bmatrix} \theta_{11} & \cdots & \theta_{1q} \\ \vdots & \vdots & \vdots \\ \theta_{n1} & \cdots & \theta_{nq} \end{bmatrix}$$

και \mathbf{Y} είναι ένας $m \times q$ πίνακας εξόδου, με y_{ij} να υποδηλώνει την j -οστή τιμή εξόδου στο i -οστό ζεύγος δεδομένων:

$$\mathbf{Y} = \begin{bmatrix} y_{11} & \cdots & y_{1q} \\ \vdots & \vdots & \vdots \\ y_{m1} & \cdots & y_{mq} \end{bmatrix}.$$

Το αντίστοιχο τετραγωνικό σφάλμα που ελαχιστοποιούμε είναι:

$$E(\boldsymbol{\Theta}) = \sum_{i=1}^m \sum_{j=1}^q e_{ij}^2 = \sum_{j=1}^q \sum_{i=1}^m e_{ij}^2$$

Το $\sum_{i=1}^m e_{ij}^2$ είναι το τετραγωνικό μήκος της j -οστής στήλης του πίνακα \mathbf{E} , που εξαρτάται από την j -οστή στήλη του πίνακα $\boldsymbol{\Theta}$ μόνο. Έτσι:

$$\min_{\Theta} \sum_{j=1}^q \left(\sum_{i=1}^m e_{ij} \right) = \sum_{j=1}^q \left(\min_{\theta_j} \sum_{i=1}^m e_{ij} \right)$$

όπου θ_j είναι η j -οστή στήλη του πίνακα Θ . Δηλαδή το $\theta_j = \hat{\theta}_j$ που ελαχιστοποιεί το $\sum_{i=1}^m e_{ij}^2$ είναι η Εκτιμήτρια ελαχίστων τετραγώνων του:

$$\mathbf{A} \theta_j + \mathbf{e}_j = \mathbf{y}_j,$$

όπου \mathbf{e}_j και \mathbf{y}_j είναι οι j -αστές στήλες των πινάκων \mathbf{E} και \mathbf{Y} αντιστοίχως. Το αποτέλεσμα είναι:

$$\hat{\theta}_j = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y}_j,$$

και

$$\hat{\Theta} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{Y}.$$

1.4.5 ΕΚΤΙΜΗΤΡΙΑ ΕΛΑΧΙΣΤΩΝ ΤΕΤΡΑΓΩΝΩΝ ΓΙΑ ΜΗ ΓΡΑΜΜΙΚΑ ΜΟΝΤΕΛΑ (ΣΥΝΤΟΜΗ ΑΝΑΦΟΡΑ)

Αυτή η εκτιμήτρια χρησιμοποιείται για δεδομένα που σχετίζονται σε μοντέλα με μη γραμμικές παραμέτρους. Τα μη γραμμικά μοντέλα (δηλαδή αυτά των οποίων οι προς εκτίμηση παράμετροι είναι μη γραμμικές) μπορούν να χωριστούν σε δυο τύπους, οι οποίοι αναφέρονται ως εγγενώς γραμμικά (intrinsically linear) και εγγενώς μη γραμμικά. Με κατάλληλους μετασχηματισμούς των μεταβλητών εισόδου-εξόδου και των προσαρμοζόμενων παραμέτρων, ένα εγγενώς γραμμικό μοντέλο μπορεί να εκφραστεί με τον συνήθη τρόπο από την εξίσωση (1). Έτσι μπορούμε να εφαρμόσουμε την συνήθη μέθοδο των ελαχίστων τετραγώνων για να προσεγγίσουμε αποτελεσματικά τις βέλτιστες παραμέτρους. (Εξ' αιτίας της χρήσης των μετασχηματισμών, η λύση δεν είναι ακριβώς βέλτιστη στην ελαχιστοποίηση του μέτρου του τετραγωνικού σφάλματος).

Αν ένα μη γραμμικό μοντέλο δεν μπορεί να εκφραστεί σε γραμμική μορφή μετά από μετασχηματισμούς, τότε είναι εγγενώς μη γραμμικό. Για μη γραμμικά μοντέλα αυτού του τύπου, πρέπει να εφαρμόσουμε μη γραμμικές μεθόδους ελαχίστων τετραγώνων.

Μπορούμε ακόμα να προχωρήσουμε σε επαναλαμβανόμενους μετασχηματισμούς για να μετατρέψουμε ένα πολύπλοκο μη γραμμικό μοντέλο, σε ένα γραμμικό. Εφαρμόζοντας την εκτιμήτρια των ελαχίστων τετραγώνων για γραμμικές παραμέτρους και μεθόδους επαναλαμβανόμενων βελτιστοποιήσεων για τις μη γραμμικές οδηγούμαστε στην αποκαλούμενη **υβριδική εκπαίδευση**, στην οποία θα αναφερθούμε παρακάτω.

1.5 ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΜΕ ΠΑΡΑΓΩΓΟΥΣ

1.5.1.1 ΦΘΙΝΟΥΣΕΣ ΜΕΘΟΔΟΙ (DESCENT METHODS)

Η φθίνουσες μέθοδοι μας επιτρέπουν να ελαχιστοποιήσουμε μια αντικειμενική συνάρτηση E , που παίρνει αληθείς τιμές, και ορίζεται (προσδιορίζεται) σε έναν χώρο εισόδου $\theta = [\theta_1, \theta_2, \dots, \theta_n]^T$ n -διαστάσεων. Η εύρεση ενός (πιθανού τοπικού) ελαχίστου σημείου $\theta = \theta^*$ που ελαχιστοποιεί την $E(\theta)$ είναι ο βασικός στόχος.

Γενικά μια αντικειμενική συνάρτηση E μπορεί να έχει μια μη γραμμική μορφή ως προς μια προσαρμοζόμενη παράμετρο θ . Εξαιτίας της πολυπλοκότητας της E , συχνά καταφεύγουμε σε έναν επαναληπτικό αλγόριθμο για να μελετήσουμε επαρκώς τον χώρο εισόδου. Σε επαναληπτικές φθίνουσες μεθόδους, το επόμενο σημείο θ_{next} βρίσκεται σε επίπεδο χαμηλότερα από το παρών σημείο θ_{now} σε ένα **διάνυσμα κατεύθυνσης \mathbf{d}** (**direction vector \mathbf{d}**)

$$\theta_{\text{next}} = \theta_{\text{now}} + \eta \mathbf{d}$$

όπου η είναι κάποιο θετικό **μέγεθος βήματος** που ρυθμίζει την έκταση που θα προχωρήσουμε στην συγκεκριμένη κατεύθυνση. Ο όρος μέγεθος βήματος η είναι αυτό που στα νευρο-ασαφή ονομάζεται ρυθμός εκπαίδευσης. Εναλλακτικά χρησιμοποιούμε τον παρακάτω τύπο:

$$\theta_{k+1} = \theta_k + \eta_k \mathbf{d}_k \quad (k = 1, 2, 3, \dots)$$

όπου το k δηλώνει τον αριθμό της παρούσας επανάληψης, και τα θ_{now} και θ_{next} αντιπροσωπεύουν διαδοχικά στοιχεία σε μια δημιουργούμενη ακολουθία από υποψήφιες λύσεις $\{\theta_k\}$. Το θ_k θέλουμε να συγκλίνει σε ένα τοπικό ελάχιστο θ^* .

Οι επαναληπτικές φθίνουσες μέθοδοι υπολογίζουν το k -οστό βήμα $\eta_k \mathbf{d}_k$ με δυο διαδικασίες:

- πρώτα καθορίζουμε την κατεύθυνση του \mathbf{d}
- και μετά υπολογίζουμε το μέγεθος βήματος η .

Το επόμενο σημείο θ_{next} πρέπει να ικανοποιεί την επόμενη ανισότητα:

$$E(\theta_{\text{next}}) = E(\theta_{\text{now}} + \eta \mathbf{d}) < E(\theta_{\text{now}}) \quad (5)$$

Οι βασικές διαφορές μεταξύ διαφόρων φθινουσών αλγορίθμων βρίσκεται στην πρώτη διαδικασία για τον καθορισμό της επιτυχούς κατεύθυνσης. Όταν αποφασίσουμε, όλοι οι αλγόριθμοι συνεχίζουν για να βρουν μια (τοπική) ελάχιστη απόσταση από ένα σημείο της ευθείας που προκύπτει από το παρών σημείο θ_{now} και την κατεύθυνση του \mathbf{d} . Δηλαδή για την δεύτερη διαδικασία, το βέλτιστο μέγεθος βήματος μπορεί να βρεθεί με την ελαχιστοποίηση της απόστασης από την ευθεία:

$$\eta^* = \arg \min_{\eta > 0} \phi(\eta), \quad (6)$$

όπου

$$\phi(\eta) = E(\theta_{now} + \eta \mathbf{d}).$$

Η αναζήτηση του η^* γίνεται με μεθόδους αναζήτησης ευθείας (ή αναζήτηση μιας διάστασης).

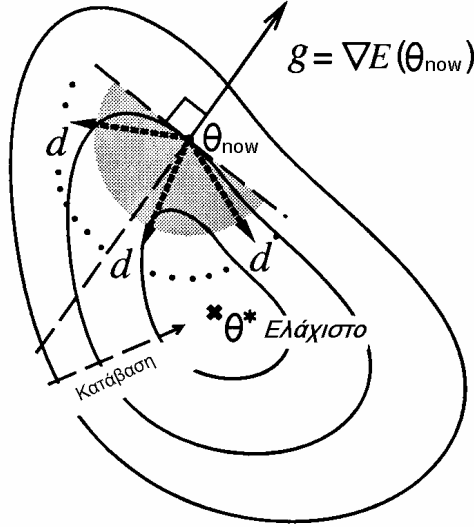
1.5.1.2 ΜΕΘΟΔΟΙ ΒΑΣΙΣΜΕΝΕΣ ΣΕ ΒΑΘΜΩΤΟ ΔΙΑΝΥΣΜΑ (GRADIENT-BASED METHODS)

Φθίνουσες μέθοδοι βασισμένες σε βαθμωτό διάνυσμα ονομάζονται αυτές όπου η άμεση κατηφορική κατεύθυνση \mathbf{d} βρίσκεται από την βάση ενός **βαθμωτου διανύσματος** (\mathbf{g}), μιας αντικειμενικής συνάρτησης E .

Το βαθμωτό διάνυσμα μιας διαφορίσιμης συνάρτησης $E: R^n \rightarrow R$ στο θ είναι το διάνυσμα των πρώτων παραγώγων της E , και συμβολίζεται με \mathbf{g} :

$$\mathbf{g}(\theta) (= \nabla E(\theta)) \stackrel{\text{def}}{=} \left[\frac{\partial E(\theta)}{\partial \theta_1}, \frac{\partial E(\theta)}{\partial \theta_2}, \dots, \frac{\partial E(\theta)}{\partial \theta_n} \right]^T.$$

Για ευκολία χρησιμοποιούμε το \mathbf{g} μαζί με το θ κλεισμένο σε παρενθέσεις, $\mathbf{g}(\theta)$.



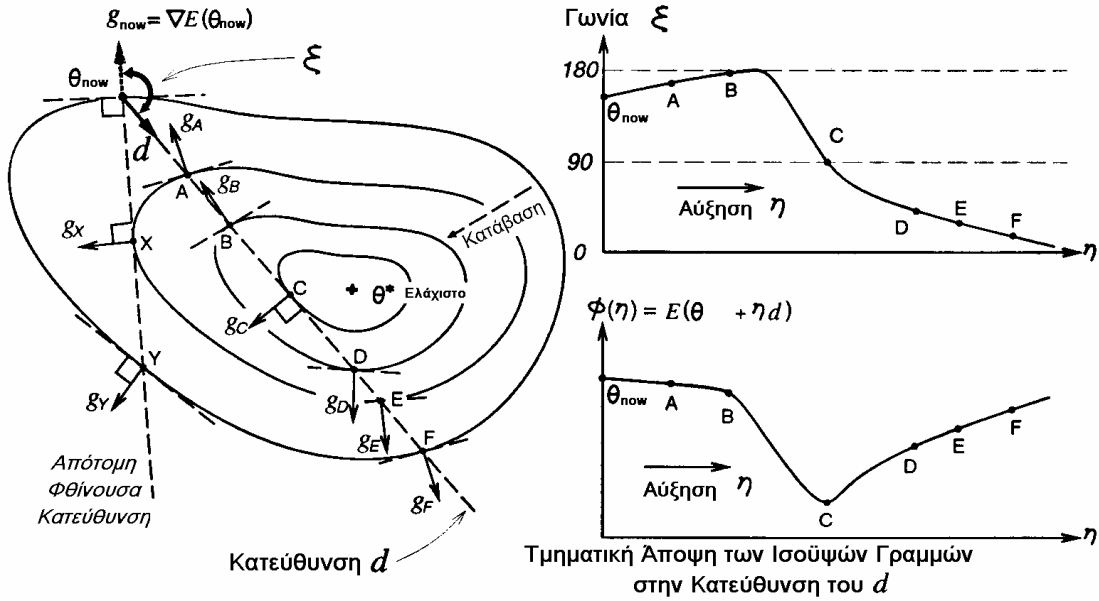
Σχήμα 18. Εφικτές φθίνουσες κατευθύνσεις. Κατευθύνσεις με σημείο εκκίνησης το θ_{now} στην σκιαγραμμισμένη περιοχή είναι πιθανά υποψήφια διανύσματα. Όταν $\mathbf{d} = -\mathbf{g}$, το \mathbf{d} είναι η απότομη φθίνουσα κατεύθυνση σε ένα τοπικό σημείο θ_{now} . [1]

Γενικά, δεδομένου ενός βαθμωτου ανύσματος, οι καθοδικές κινήσεις υπακούουν στην ακόλουθη προϋπόθεση για εφικτές Φθίνουσες κατευθύνσεις:

$$\phi'(0) = \left. \frac{dE(\theta_{now} + \eta \mathbf{d})}{d\eta} \right|_{\eta=0} = \mathbf{g}^T \mathbf{d} = \|\mathbf{g}\| \|\mathbf{d}\| \cos(\xi(\theta_{now})) < 0, \quad (7)$$

όπου το ξ δηλώνει την γωνία μεταξύ \mathbf{g} και $\boldsymbol{\theta}$, και $\xi(\boldsymbol{\theta}_{\text{now}})$ δηλώνει την γωνία μεταξύ \mathbf{g}_{now} και \mathbf{d} στο παρόν σημείο $\boldsymbol{\theta}_{\text{now}}$, όπως παρουσιάζεται και στο σχήμα 18 και 19. Αυτό μπορεί να επιβεβαιωθεί και από το ανάπτυγμα της E με σειρές Taylor.

$$E(\boldsymbol{\theta}_{\text{now}} + \eta \mathbf{d}) = E(\boldsymbol{\theta}_{\text{now}}) + \eta \mathbf{g}^T \mathbf{d} + O(\eta^2). \quad (8)$$



Σχήμα 19. Η γωνία ξ μεταξύ των βαθμωτών κατευθύνσεων \mathbf{g} και μιας φθίνουσας κατεύθυνση \mathbf{d} , η οποία ευρίσκεται με έναν συγκεκριμένο αλγόριθμο στο παρόν σημείο $\boldsymbol{\theta}_{\text{now}}$. N είναι το σύνολο όλων των πιθανών επομένων σημείων, $N \supset \{A, B, C, D, E, F, X, Y\}$. Στην μονόδρομη κατηφορική κατεύθυνση \mathbf{d} , το επόμενο σημείο $\boldsymbol{\theta}_{\text{next}}$ μπορεί να είναι ένα από τα έξι σημεία A, B, C, D, E ή F – ή να είναι γύρω απ αυτά, αναλόγως των μεγεθών των βημάτων. Με σύγκριση, στην απότομη φθίνουσα μέθοδο, το $\boldsymbol{\theta}_{\text{next}}$ μπορεί να είναι είτε το X είτε το Y , ή κοντά σε αυτά. [1]

Ο δεύτερος όρος στη δεξιά πλευρά της εξίσωσης θα επικρατήσει του τρίτου όρου και άλλων όρων υψηλότερου βαθμού του η όταν $\eta \rightarrow 0$. Με τόσο μικρά θετικά η , η ανισότητα (5) έχει βάση προφανώς όταν $\mathbf{g}^T \mathbf{d} < 0$. Η σκιασμένη επιφάνεια στο σχήμα 18 υποδηλώνει όλες τις εφικτές Φθίνουσες κατευθύνσεις που ικανοποιούν την προϋπόθεση (7). Εμφανώς οι ανυσματικές κατευθύνσεις είναι πάντα κάθετες στο περίγραμμα των καμπυλών.

Μια κατηγορία φθίνουσών μεθόδων βασισμένων σε βαθμωτό άνυσμα έχει την ακόλουθη θεμελιώδη μορφή, κατά την οποία εφικτές Φθίνουσες κατευθύνσεις μπορούν να προσδιοριστούν παρεκτρέποντας τα βαθμωτά ανύσματα μέσω πολλαπλασιασμού με \mathbf{G} (π.χ. αποκλίνοντα βαθμωτά ανύσματα):

$$\boldsymbol{\theta}_{\text{next}} = \boldsymbol{\theta}_{\text{now}} - \eta \mathbf{G} \mathbf{g}, \quad (9)$$

με κάποιο θετικό μέγεθος βήματος η και κάποιον θετικά ορισμένο πίνακα \mathbf{G} . Προφανώς, όταν $\mathbf{d} = -\mathbf{G}\mathbf{g}$, η προϋπόθεση (7) για την φθίνουσα κατεύθυνση έχει βάση μιας και $\mathbf{g}^T \mathbf{d} = -\mathbf{g}^T \mathbf{G} \mathbf{g} < 0$.

Στην ιδανική περίπτωση, επιθυμούμε να βρούμε μια τιμή της θ_{next} που ικανοποιεί το:

$$\mathbf{g}(\theta_{\text{next}}) = \left. \frac{\partial E(\theta)}{\partial \theta} \right|_{\theta = \theta_{\text{next}}} = 0. \quad (10)$$

Αυτή η εξίσωση είναι απαραίτητη προϋπόθεση, διότι το βαθμωτό \mathbf{g} είναι μηδέν σε κάθε **στάσιμο (stationary) σημείο**, δηλαδή σε μέγιστο, ελάχιστο ή σαγματικό σημείο.

Στην πράξη όμως είναι δύσκολο να λύσουμε την εξίσωση (10) αναλυτικά. Για να ελαχιστοποιήσουμε την αντικειμενική συνάρτηση, οι Φθίνουσες διαδικασίες επαναλαμβάνονται τυπικά, μέχρι ένα από τα ακόλουθα κριτήρια τερματισμού ικανοποιηθεί:

- 1) οι τιμές της αντικειμενικής συνάρτησης είναι ικανοποιητικά μικρές
- 2) το μήκος του βαθμωτού διανύσματος \mathbf{g} είναι μικρότερο από μια συγκεκριμένη τιμή
- 3) ο καθορισμένος χρόνος υπολογισμού έχει ξεπεραστεί.

1.5.2 ΑΠΟΤΟΜΗ ΦΘΙΝΟΥΣΑ ΜΕΘΟΔΟΣ (THE METHOD OF STEEPEST DESCENT)

Η **απότομη φθίνουσα μέθοδος**, γνωστή και σαν **μέθοδος βαθμωτου διανύσματος**, είναι μια από τις παλαιότερες τεχνικές για ελαχιστοποίηση μιας δεδομένης συνάρτησης ορισμένη σε έναν πολυδιάστατο χώρο εισόδου. Αυτή η μέθοδος σχηματίζει την βάση για πολλές άμεσες μεθόδους (direct methods) που χρησιμοποιούνται στην βελτιστοποίηση προβλημάτων με περιορισμούς ή χωρίς (constrained – unconstrained). Επιπλέον, πάρα την αργή της σύγκλιση, η μέθοδος είναι η περισσότερο χρησιμοποιούμενη μη γραμμική τεχνική βελτιστοποίησης εξαιτίας της απλότητας της.

Όταν $\mathbf{G} = \mathbf{I}$ (ο μοναδιαίος πίνακας), η εξίσωση (9) γίνεται ο τύπος της απότομης φθίνουσας μεθόδου:

$$\theta_{\text{next}} = \theta_{\text{now}} - \eta \mathbf{g}.$$

Στις εξισώσεις (7) και (8), αν $\cos \zeta = -1$ (που σημαίνει ότι το \mathbf{d} σημαδεύει προς την ίδια κατεύθυνση με το αρνητικό βαθμωτό $-\mathbf{g}$) η αντικειμενική συνάρτηση E μπορεί να μειωθεί τοπικά κατά το μεγαλύτερο βαθμό από το παρών σημείο θ_{now} . Αυτό σημαίνει ότι το αρνητικό βαθμωτό $-\mathbf{g}$ σημαδεύει στην πιο **απότομη τοπική κατηφορική** κατεύθυνση. Από γενικότερη θεώρηση, το να πηγαίνουμε κατά την κατεύθυνση του αρνητικού βαθμωτου μπορεί και να μην είναι ο συντομότερος δρόμος για να φτάσουμε το ελάχιστο σημείο θ^* (σχήμα 18,19).

Αν στην απότομη φθίνουσα (καταβατική, κατηφορική) μέθοδο εμπλέκεται και ελαχιστοποίηση ευθείας (line minimization) στην εξίσωση (6), που σημαίνει ότι αν το ελάχιστο σημείο η^* σε μια κατεύθυνση του \mathbf{d} , βρίσκεται σε κάθε επανάληψη, τότε θα έχουμε:

$$\begin{aligned}\phi'(\eta) &= \frac{dE(\boldsymbol{\theta}_{\text{now}} - \eta \mathbf{g}_{\text{now}})}{d\eta} \\ &= \nabla^T E(\boldsymbol{\theta}_{\text{now}} - \eta \mathbf{g}_{\text{now}}) \mathbf{g}_{\text{now}} \\ &= \mathbf{g}_{\text{next}}^T \mathbf{g}_{\text{now}} \\ &= 0,\end{aligned}$$

όπου \mathbf{g}_{next} είναι το βαθμωτό διάνυσμα στο επόμενο σημείο. Η προηγούμενη εξίσωση δείχνει ότι το επόμενο βαθμωτό διάνυσμα \mathbf{g}_{next} είναι πάντα ορθογώνιο στο παρών βαθμωτό διάνυσμα \mathbf{g}_{now} . Στο σχήμα 19 απεικονίζεται αυτή η κατάσταση στο σημείο X , όπου $\mathbf{g}_{\text{next}} = \mathbf{g}_X$.

1.6 ΠΡΟΣΔΙΟΡΙΣΜΟΣ ΜΕΓΕΘΟΥΣ ΒΗΜΑΤΟΣ (STEP SIZE)

Ο τύπος μιας κατηγορίας μεθόδων βασισμένων σε βαθμωτό άνυσμα δίνεται από την εξίσωση (9):

$$\boldsymbol{\theta}_{\text{next}} = \boldsymbol{\theta}_{\text{now}} + \eta \mathbf{d} = \boldsymbol{\theta}_{\text{now}} - \eta \mathbf{G} \mathbf{g}.$$

Αυτός ο τύπος συνεπάγεται τον αποτελεσματικό προσδιορισμό του μεγέθους βήματος η . Η επάρκεια του υπολογισμού του μεγέθους βήματος, επηρεάζει ολόκληρη την διαδικασία ελαχιστοποίησης. Για μια γενική συνάρτηση E , η αναλυτική επίλυση της εξίσωσης (6) όπως όταν

$$\phi'(\eta) = 0, \quad \text{όπου} \quad \phi(\eta) = E(\boldsymbol{\theta}_{\text{now}} + \eta \mathbf{d})$$

είναι συχνά αδύνατη. Που σημαίνει ότι η μιας μεταβλητής συνάρτηση $\phi(\eta)$ θα πρέπει να ελαχιστοποιηθεί πάνω στην ευθεία που προσδιορίστηκε από το παρών σημείο $\boldsymbol{\theta}_{\text{now}}$ και την κατεύθυνση του \mathbf{d} . Αυτό επιτυγχάνεται με τις μεθόδους αναζήτησης ευθείας (ή μονοδιάστατης αναζήτησης).

1.6.1 ΑΡΧΙΚΟ BRACKETING

Οι μέθοδοι αναζήτησης ευθείας, βασικά υποθέτουν ότι η περιοχή αναζήτησης, ή το προσδιορισθέν διάστημα, περιέχει ένα μόνο σχετικό ελάχιστο. Δηλαδή η εξίσωση E έχει μόνο ένα βέλτιστο στο κλειστό διάστημα. Ο προσδιορισμός του αρχικού διαστήματος στο οποίο το σχετικό ελάχιστο πρέπει να βρίσκεται, έχει σπουδαία σημασία. Για να αρχίσουμε την αναζήτηση ευθείας, κάποιοι αλγόριθμοι πρέπει να έχουν ένα υποτιθέμενο ελάχιστο μέσα στο διάστημα της

εκκίνησης για το αρχικό Bracketing. Αυτού του είδους η διαδικασία μπορεί πρόχειρα να κατηγοριοποιηθεί σε δυο σχήματα:

- 1) στο πρώτο, από τον υπολογισμό της συνάρτησης, για την εύρεση τριών σημείων που ικανοποιούν το:

$$E(\theta_{k-1}) > E(\theta_k) < E(\theta_{k+1}), \quad \theta_{k-1} < \theta_k < \theta_{k+1}.$$

- 2) Στο δεύτερο, παίρνουμε τις πρώτες παραγώγους για την εύρεση δυο σημείων που ικανοποιούν την:

$$E'(\theta_k) < 0, \quad E'(\theta_{k+1}) > 0, \quad \theta_k < \theta_{k+1}.$$

Για παράδειγμα για το πρώτο σχήμα, ο κοινός αλγόριθμος θα μπορούσε να περιγράψει ως εξής:

Αλγόριθμος 1: διαδικασία αρχικού Bracketing, για την αναζήτηση τριών σημείων θ_1 , θ_2 και θ_3 .

- 1) Δεδομένου ενός αρχικού σημείου θ_0 και $h \in R$, ορίζουμε το θ_1 να ισούται με $\theta_0 + h$. Υπολογισμός $E(\theta_1)$.

$$\text{AN } E(\theta_0) \geq E(\theta_1), \quad i \leftarrow 1,$$

(π.χ. όταν κατεβαίνουμε) ΠΗΓΑΙΝΕ στο (2)

ΑΛΛΙΩΣ

$$h \leftarrow -h, \text{ (π.χ. θέτουμε οπισθόδρομη κατεύθυνση)}$$

(π.χ. όταν ανεβαίνουμε)

$$E(\theta_{-1}) \leftarrow E(\theta_1),$$

$$\theta_1 \leftarrow \theta_0 + h,$$

$$i \leftarrow 0,$$

ΠΗΓΑΙΝΕ στο (3).

- 2) ΘΕΣΕ το επόμενο σημείο με: $h \leftarrow 2h$, $\theta_{i+1} \leftarrow \theta_i + h$,

- 3) ΥΠΟΛΟΓΙΣΕ το $E(\theta_{i+1})$

$$\text{AN } E(\theta_i) \geq E(\theta_{i+1}) \quad i \leftarrow i + 1$$

(π.χ. όταν ακόμα κατεβαίνουμε)

ΠΗΓΑΙΝΕ στο (2)

ΑΛΛΙΩΣ

ΔΙΑΤΑΞΕ τα θ_{i-1} , θ_i και θ_{i+1} σε φθίνουσα σειρά. Τότε έχουμε τα τρία σημεία $(\theta_1, \theta_2, \theta_3)$.

ΤΕΛΟΣ

1.6.2 ΑΝΑΖΗΤΗΣΗ ΕΥΘΕΙΑΣ

Η διαδικασία προσδιορισμού του η^* που ελαχιστοποιεί την μονοδιάστατη συνάρτηση $\phi(\eta)$ επιτυγχάνεται με την αναζήτηση πάνω στην γραμμή για το ελάχιστο. Η μέθοδος της αναζήτησης ευθείας (ή μονοδιάστατης αναζήτησης) είναι σημαντική διότι *τα προβλήματα μεγαλύτερων διαστάσεων σε τελική ανάλυση λύνονται με επαναλαμβανόμενες αναζητήσεις ευθείας*. Επίσης οι αλγόριθμοι αναζήτησης ευθείας, συνήθως περιλαμβάνουν δυο συνιστώσες: το sectioning (ή bracketing), και την πολυωνυμική παρεμβολή (polynomial interpolation).

1.6.3 Η ΜΕΘΟΔΟΣ ΤΟΥ NEWTON

Όταν για επόμενο σημείο θ_{now} επιλέγεται το ελάχιστο σημείο $\hat{\theta}$ όπως ορίζεται από την παρακάτω εξίσωση, έχουμε την κλασική μέθοδο του Newton:

$$\hat{\theta} = \theta_{\text{now}} - \mathbf{H}^{-1} \mathbf{g}. \quad (11)$$

Το βήμα $-\mathbf{H}^{-1} \mathbf{g}$ αποκαλείται **βήμα Newton** και η κατεύθυνση του αποκαλείται **κατεύθυνση Newton**. Όπου \mathbf{H} είναι ο πίνακας του Hess (Hessian matrix), και αποτελείται από τις *δεύτερες μερικές παραγώγους της $E(\theta)$* . Για να ισχύει η (11) θα πρέπει να υπάρχει ο αντίστροφος του \mathbf{H} .

Όταν οι $\phi(\eta_k)$, $\phi'(\eta_k)$, $\phi''(\eta_k)$, είναι διαθέσιμες, η κλασική μέθοδος του Newton από την εξίσωση (11), μπορεί να εφαρμοστεί για την επίλυση της εξίσωσης $\phi'(\eta_k) = 0$:

$$\eta_{k+1} = \eta_k - \frac{\phi'(\eta_k)}{\phi''(\eta_k)}. \quad (12)$$

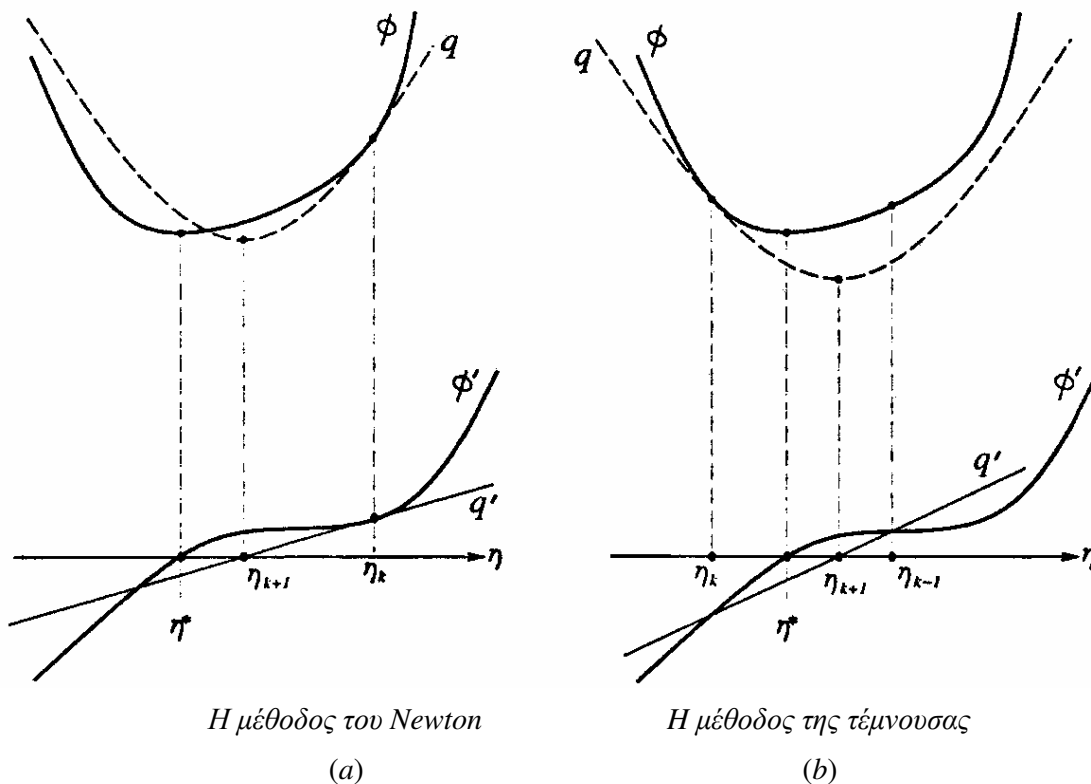
Το σχήμα 20α δείχνει ότι ο προηγούμενος τύπος προσδιορίζει το επόμενο βήμα η_{k+1} .

1.6.4 Η ΜΕΘΟΔΟΣ ΤΗΣ ΤΕΜΝΟΥΣΑΣ

Αν θέλουμε να χρησιμοποιήσουμε και το η_k και το η_{k-1} για να προσεγγίσουμε την δεύτερη παράγωγο στην εξίσωση (12) και αν οι πρώτες παράγωγοι είναι διαθέσιμες, τότε έχουμε ένα εκτιμώμενο η_{k+1} :

$$\eta_{k+1} = \eta_k - \frac{\phi'(\eta_k)}{\frac{\phi'(\eta_k) - \phi'(\eta_{k-1})}{\eta_k - \eta_{k-1}}}.$$

Αυτή είναι η αποκαλούμενη **μέθοδος της εσφαλμένης θέσης** ή η **μέθοδος της τέμνουσας**, όπως απεικονίζεται και στο σχήμα 20β.



Σχήμα 20. Η μέθοδος του Newton (αριστερά) και η μέθοδος της τέμνουσας (δεξιά) για το προσδιορισμό του μεγέθους βήματος. Η q είναι μια προσεγγιστική τετραγωνική συνάρτηση. [1]

1.6.5 Η ΜΕΘΟΔΟΣ SECTIONING

Ένας αλγόριθμος Sectioning ξεκινάει με ένα διάστημα $[a_1, b_1]$ στο οποίο πρέπει να βρίσκεται το ελάχιστο η^* , και έπειτα μειώνει το μήκος του διαστήματος σε κάθε επανάληψη με τον υπολογισμό του των τιμών του ϕ σε έναν συγκεκριμένο αριθμό σημείων. Τα δυο άκρα a_1 και b_1 μπορούν να βρεθούν με το αρχικό bracketing.

Η **μέθοδος της διχοτόμησης** είναι μια από τις πιο απλές μεθόδους sectioning για την επίλυση του $\phi'(\eta^*) = 0$, αν οι πρώτες παράγωγοι είναι διαθέσιμες.

Αλγόριθμος της μεθόδου της διχοτόμησης:

- 1) Δεδομένης μιας μικρής τιμής $\varepsilon \in R$ και ενός αρχικού διαστήματος με δυο άκρα a_1 και a_2 τέτοια ώστε $a_1 < a_2$ και $\phi'(a_1)\phi'(a_2) < 0$:

$$\eta_{\text{left}} \leftarrow a_1$$

$$\eta_{\text{right}} \leftarrow a_2$$

- 2) Υπολογισμός του μέσου του διαστήματος η_{mid} : $\eta_{\text{mid}} \leftarrow \frac{(\eta_{\text{right}} + \eta_{\text{left}})}{2}$.

$$\text{Αν } \phi'(\eta_{\text{right}})\phi'(\eta_{\text{mid}}) < 0, \quad \eta_{\text{left}} \leftarrow \eta_{\text{mid}}$$

$$\text{Αλλιώς,} \quad \eta_{\text{right}} \leftarrow \eta_{\text{mid}}$$

- 3) Έλεγχος Αν $|\eta_{\text{left}} - \eta_{\text{right}}| < \varepsilon$. Αν ισχύει, τερματίζεται ο αλγόριθμος. Αλλιώς ΠΗΓΑΙΝΕ στο (2).

Η μέθοδος της διχοτόμησης αντικαθιστά το δεξιό ή το αριστερό άκρο με το μέσον του διαστήματος, βάση των υπολογισμών της συνάρτησης στο μέσον. Το μήκος του διαστήματος bracketing ημιδιπλασιάζεται σε κάθε επανάληψη.

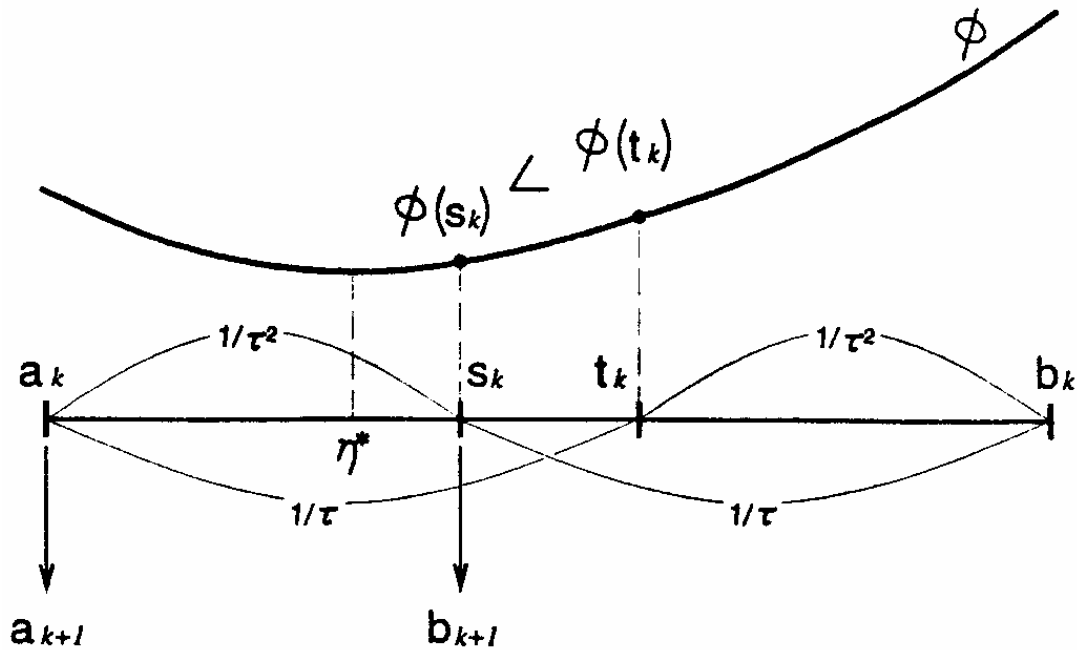
Η αποτελεσματική μείωση του μήκους του διαστήματος επιτυγχάνεται με την μέθοδο της **αναζήτησης της χρυσής τομής** που βασίζεται μόνο σε υπολογισμούς της συνάρτησης. Η χρυσή τομή απαιτεί το ϕ να μην είναι συνεχής συνάρτηση ούτε και διαφορίσιμη.

Δεδομένου ενός αρχικού διαστήματος $[a_1, b_1]$ ($\exists \eta^*$), τα σημεία της επόμενης δόκιμης (s_k, t_k) μέσα στο διάστημα, προσδιορίζονται με τον **λόγο τ της χρυσής τομής**:

$$s_k = b_k - \frac{1}{\tau}(b_k - a_k) = b_k + \frac{\tau-1}{\tau}(b_k - a_k),$$

$$t_k = a_k + \frac{1}{\tau}(b_k - a_k),$$

όπου $\tau = \frac{1+\sqrt{5}}{2} \approx 1.618$. Αυτή η διαδικασία εγγυάται ότι $a_k < s_k < t_k < b_k$, όπως φαίνεται και στο σχήμα 21.



Σχήμα 21. Η αναζήτηση της χρυσής τομής για τον προσδιορισμό του μήκους του βήματος.[1]

Ο αλγόριθμος γεννάει μια ακολουθία των δυο άκρων, a_k και b_k , σύμφωνα με το:

$$\begin{array}{lll} \text{ΑΝ } \phi(s_k) > \phi(t_k), & a_{k+1} = s_k, & b_{k+1} = b_k, \\ \text{ΑΛΛΙΩΣ,} & a_{k+1} = a_k, & b_{k+1} = t_k, \end{array}$$

Το ελάχιστο σημείο η^* βρίσκεται σε ένα διάστημα μόλις $\frac{1}{\tau}$ (≈ 0.618 , περίπου δυο-τρίτα) φορές το μήκος του προηγούμενου διαστήματος. Μετά την k -οστή επανάληψη, το μήκος του bracketing διαστήματος συρρικνώνεται στο $(b_1 - a_1) \left(\frac{1}{\tau}\right)^{k-1}$.

1.6.6 ΠΟΛΥΩΝΥΜΙΚΗ ΠΑΡΕΜΒΟΛΗ (POLYNOMIAL INTERPOLATION)

Οι μέθοδοι της πολυωνυμικής παρεμβολής βασίζονται σε διαδικασίες προσαρμογής καμπυλών, οι οποίες είναι αποτελεσματικές όταν η αντικειμενική συνάρτηση έχει ένα συγκεκριμένο βαθμό ομαλότητας.

Μια μέθοδος **τετραγωνικής παρεμβολής** κατασκευάζει μια ομαλή τετραγωνική καμπύλη q η οποία περνάει από τρία υπολογισμένα σημεία,

$(\eta_1, \phi_1), (\eta_2, \phi_2), (\eta_3, \phi_3)$:

$$q(\eta) = \sum_{i=1}^3 \phi_i \frac{\prod_{j \neq i} (\eta - \eta_j)}{\prod_{j \neq i} (\eta_i - \eta_j)},$$

όπου $\phi_i \equiv \phi(\eta_i)$, $i = 1, 2, 3$.

Η τετραγωνική συνάρτηση έχει ένα μοναδικό ελάχιστο σημείο, το οποίο μπορεί εύκολα να προσδιοριστεί, λύνοντας την $q'(\eta) = 0$. Έτσι το επόμενο σημείο δοκιμής η_{next} δίνεται από την:

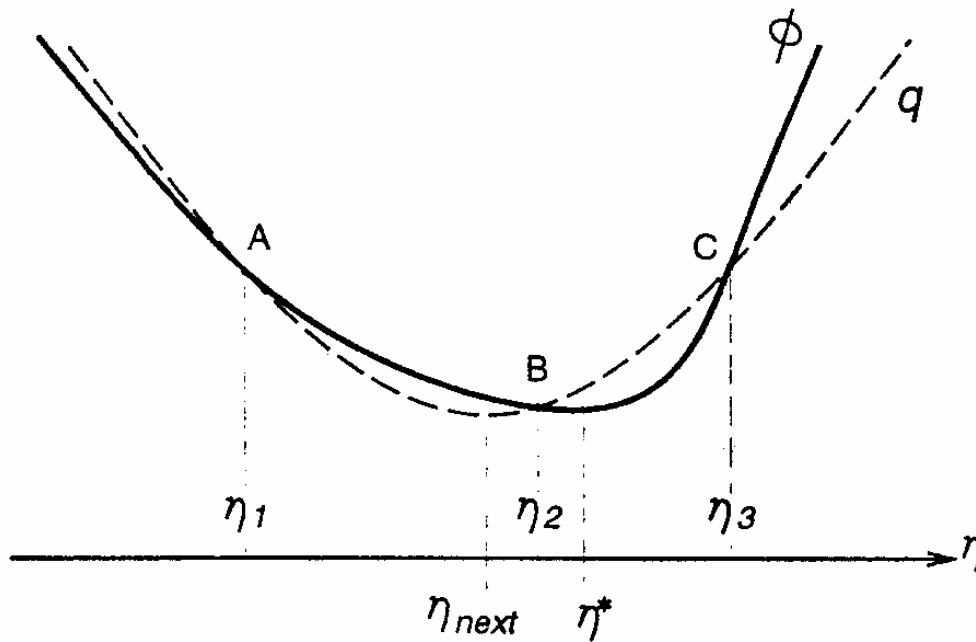
$$\eta_{\text{next}} = \frac{1}{2} \frac{(\eta_2^2 - \eta_3^2)\phi_1 + (\eta_3^2 - \eta_1^2)\phi_2 + (\eta_1^2 - \eta_2^2)\phi_3}{(\eta_2 - \eta_3)\phi_1 + (\eta_3 - \eta_1)\phi_2 + (\eta_1 - \eta_2)\phi_3},$$

και φαίνεται και στο σχήμα 22.

Όταν οι τέσσερις τιμές $\phi(\eta_1)$, $\phi'(\eta_1)$, $\phi(\eta_2)$, $\phi'(\eta_2)$ είναι διαθέσιμες, μια μέθοδος κυβικής παρεμβολής, μπορεί να κατασκευάσει μια κυβική εξίσωση, όπως επίσης προσδιορίσει και το επόμενο σημείο η_{next} ως το σχετικό ελάχιστο σημείο της κυβικής εξίσωσης:

$$\eta_{\text{next}} = \eta_2 - (\eta_2 - \eta_1) \left\{ \frac{\phi'(\eta_2) - \beta + \gamma}{\phi'(\eta_2) - \phi'(\eta_1) + 2\gamma} \right\}$$

όπου $\beta = \phi'(\eta_1) + \phi'(\eta_2) - 3 \frac{\phi(\eta_1) - \phi(\eta_2)}{\eta_1 - \eta_2}$ και $\gamma = \sqrt{\beta^2 - \phi'(\eta_1)\phi'(\eta_2)}$.



Σχήμα 22. Η τετραγωνική παρεμβολή[1]

1.7 ΠΡΟΣΑΡΜΟΣΤΙΚΑ ΔΙΚΤΥΑ (ADAPTIVE NETWORKS)

Ένα προσαρμοστικό δίκτυο είναι αυτό που στη δομή του περιέχει έναν αριθμό κόμβων συνδεδεμένων μέσω κατευθυντικών συνδέσεων. Κάθε κόμβος αναπαριστά μια μονάδα επεξεργασίας, και οι σύνδεσμοι μεταξύ κόμβων προσδιορίζουν την αιτιολογική σχέση (causal relationship) μεταξύ των συνδεδεμένων κόμβων. Προσαρμοστικός είναι ο κόμβος όλος ή μέρος του, που σημαίνει ότι οι Έξοδοι αυτών των κόμβων εξαρτώνται από προσαρμοζόμενες-τροποποιήσιμες παραμέτρους που ανήκουν σ' αυτούς τους κόμβους. Οι κανόνες εκπαίδευσης προσδιορίζουν πως αυτές οι παράμετροι θα πρέπει να αλλάζουν ώστε να ελαχιστοποιείται το προκαθορισμένο μέτρο σφάλματος (error measure), το οποίο είναι μια μαθηματική έκφραση που μετράει την ασυμφωνία μεταξύ της πραγματικής εξόδου του δικτύου και της επιθυμητής εξόδου.

Τα προσαρμοστικά δίκτυα χρησιμοποιούνται στην ταυτοποίηση συστήματος. Εμείς με τη σειρά μας θα πρέπει να βρούμε μια κατάλληλη αρχιτεκτονική για το δίκτυο και να θέσουμε ένα σύνολο παραμέτρων που μπορούν να μοντελοποιήσουν καλύτερα ένα σύστημα στόχος, το οποίο περιγράφεται από ένα σύνολο ζευγών δεδομένων εισόδου-εξόδου.

Ο βασικός κανόνας ενός προσαρμοστικού δικτύου είναι η απότομη φθίνουσα μέθοδος, στην οποία το βαθμωτό διάλυμα προκύπτει από επιτυχείς επικλήσεις του κανόνα της αλυσίδας. Αυτή η μέθοδος χρησιμοποιείται και για την εύρεση του βαθμωτού σε ένα νευρωνικό δίκτυο πολλών επιπέδων. Αυτή η

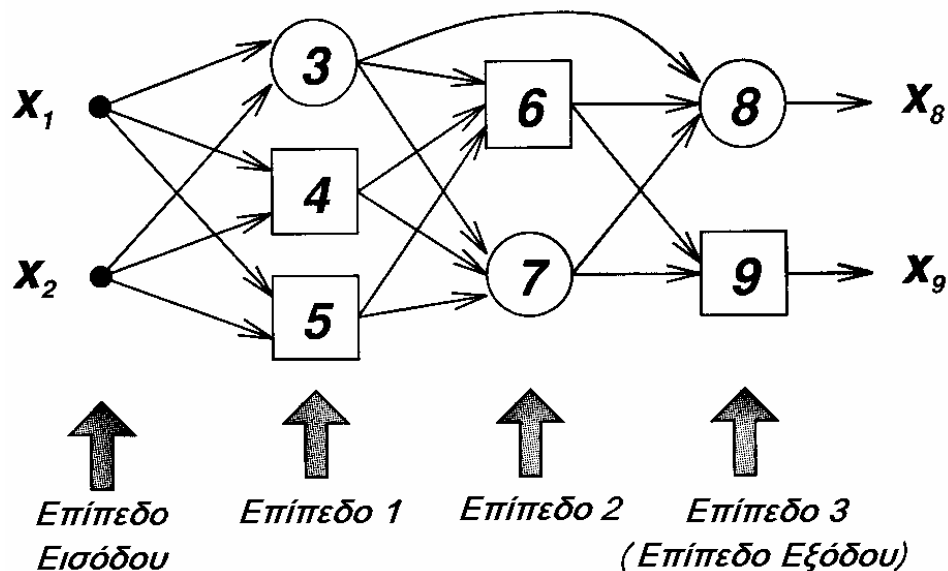
μέθοδος ονομάζεται κανόνας εκπαίδευσης οπισθοδρόμης διάδοσης (backpropagation learning rule).

1.7.1 ΠΑΡΑΓΩΓΟΙ ΠΑΡΑΜΕΤΡΟΠΟΙΗΣΙΜΩΝ ΣΥΝΑΡΤΗΣΕΩΝ ΣΥΜΜΕΤΟΧΗΣ.

[1] Για να κάνουμε ένα ασαφές σύστημα, προσαρμοστικό, θα πρέπει να ξέρουμε τις παραγώγους μιας ΣΣ ως προς το όρισμα εισόδου και τις παραμέτρους. Αυτές οι πληροφορίες από τις παραγώγους παίζουν σημαντικό ρόλο στην μάθηση ή την προσαρμογή ενός ασαφούς συστήματος.

1.7.2 ΑΡΧΙΤΕΚΤΟΝΙΚΗ ΠΡΟΣΑΡΜΟΣΤΙΚΩΝ ΔΙΚΤΥΩΝ

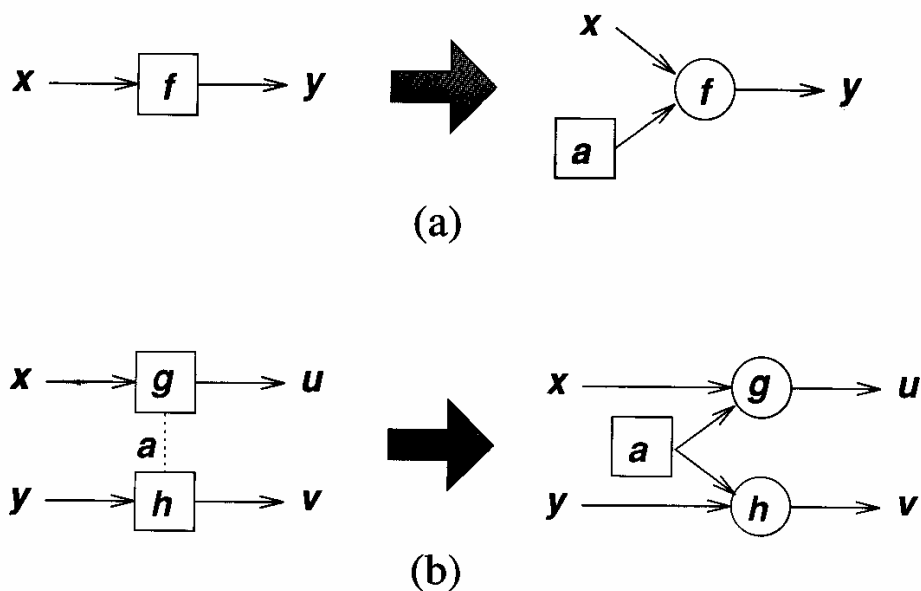
Όπως υποδηλώνει και το όνομα, ένα **προσαρμοστικό δίκτυο** (σχήμα 23) είναι μια δομή δικτύου που η συνολική του συμπεριφορά εισόδου–εξόδου διαμορφώνεται από ένα σύνολο τροποποιήσιμων παραμέτρων. Συγκεκριμένα η σύνθεση ενός προσαρμοστικού δικτύου πραγματοποιείται από ένα σύνολο κόμβων συνδεδεμένων με κατευθυνθείς συνδέσμους, όπου κάθε κόμβος εκτελεί μια συγκεκριμένη **λειτουργία κόμβου** στα εισερχόμενα σήματα, για να παράξει μια μονήρη **έξοδο κόμβου** και κάθε σύνδεσμος καθορίζει την κατεύθυνση ροής σήματος από τον έναν κόμβο στον άλλο. Συνήθως μια συνάρτηση κόμβου είναι μια παραμετρική συνάρτηση με τροποποιήσιμες παραμέτρους. Αλλάζοντας τις τελευταίες, μπορούμε να αλλάξουμε την λειτουργία κόμβου όπως και την συνολική συμπεριφορά του προσαρμοστικού δικτύου.



Σχήμα 23. Ένα εμπροσθόδρομο προσαρμοστικό δίκτυο σε αναπαράσταση με επίπεδα[1]

Οι σύνδεσμοι σε ένα προσαρμοστικό δίκτυο χρησιμοποιούνται απλώς για να προσδιορίσουν την κατεύθυνση διάδοσης των εξόδων κόμβου. Γενικά δεν υπάρχουν βάρη ή παράμετροι σχετιζόμενοι με συνδέσμους. Το σχήμα 23 είναι ένα αντιπροσωπευτικό προσαρμοστικό δίκτυο με δυο εισόδους και με δυο εξόδους.

Οι παράμετροι ενός προσαρμοστικού δίκτυο είναι κατανεμημένοι μέσα στους κόμβους, ώστε κάθε κόμβος να έχει ένα τοπικό σύνολο παραμέτρων. Η ένωση αυτών των τοπικών παραμέτρων είναι το ολικό σύνολο παραμέτρων του δικτύου. Αν το σύνολο παραμέτρων ενός κόμβου δεν είναι κενό, τότε η λειτουργία του κόμβου εξαρτάται από τις τιμές των παραμέτρων. Στα σχήματα, αναπαριστούμε τον **προσαρμοστικό κόμβο** αυτού του είδους με ένα τετράγωνο. Από την άλλη αν ο κόμβος έχει κενό σύνολο παραμέτρων, τότε η λειτουργία του είναι προκαθορισμένη. Ακόμα αναπαριστούμε τον τύπο **προκαθορισμένου κόμβου** με έναν κύκλο. Κάθε προσαρμοστικός κόμβος μπορεί να διασπαστεί σε ένα προκαθορισμένο κόμβο συν ένα ή περισσότερους κόμβους παραμέτρων.

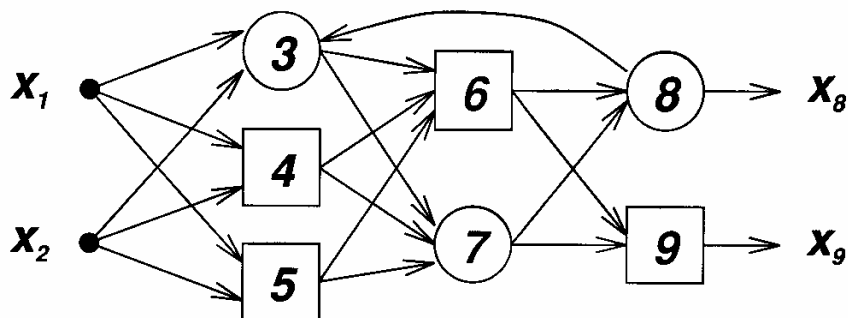


Σχήμα 24. Ανάλυση των προσαρμοστικών κόμβων: (a) ένας μονός κόμβος
(b) ένα πρόβλημα έχον κοινές παραμέτρους.[1]

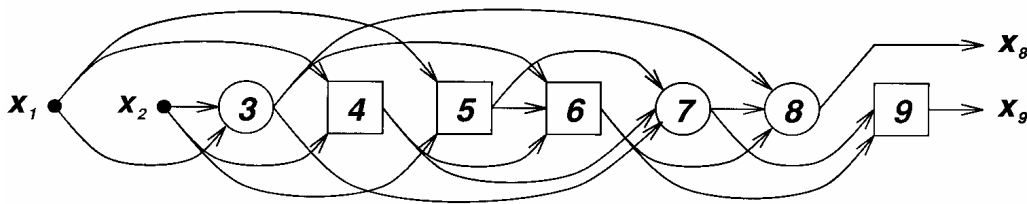
1.7.2.1 ΚΑΤΑΜΕΡΙΣΜΟΣ ΠΑΡΑΜΕΤΡΩΝ ΣΕ ΠΡΟΣΑΡΜΟΣΤΙΚΑ ΔΙΚΤΥΑ

Το σχήμα 24 δείχνει ένα προσαρμοστικό δίκτυο με έναν μόνο κόμβο, το οποίο μπορεί να αναπαρασταθεί σαν $y = f(x, a)$, όπου x, y είναι η είσοδος και έξοδος αντιστοίχως και a είναι η παράμετρος του κόμβου. Μια ισοδύναμη απεικόνιση είναι να μετακινήσουμε την παράμετρο έξω από τον κόμβο και να την τοποθετήσουμε σε έναν **κόμβο παράμετρο**, όπως φαίνεται στο σχήμα 24α. Αυτός ο κόμβος παράμετρος είναι μια ειδική περίπτωση ενός προσαρμοστικού κόμβου στο οποίο δεν υπάρχουν είσοδοι και η έξοδος είναι η ίδια η παράμετρος. Ο κόμβος παράμετρος είναι χρήσιμος στην επίλυση συγκεκριμένων αντιπροσωπευτικών προβλημάτων, όπως το πρόβλημα καταμερισμού παραμέτρων στο σχήμα 24α όπου δυο προσαρμοστικοί κόμβοι $u = g(x, a)$ και $v = h(y, a)$ μοιράζονται την ίδια παράμετρο a , όπως δηλώνεται και από την διάστικτη γραμμή που ενώνει αυτούς τους δυο κόμβους. Βγάζοντας έξω την παράμετρο και βάζοντας την μέσα σε έναν κόμβο παράμετρο, μπορούμε να ενσωματώσουμε τις απαιτήσεις του καταμερισμού παραμέτρων μέσα στην αρχιτεκτονική σχεδίαση του δικτύου. Αυτό απλοποιεί την αναπαράσταση του δικτύου όπως και την εφαρμογή του σε λογισμικό

Τα προσαρμοστικά δίκτυα είναι γενικώς κατηγοριοποιημένα σε δυο κατηγορίες με βάση τον τύπο των διασυνδέσεων που έχουν: εμπροσθόδρομα δίκτυα (feedforward) και επαναληπτικά. Το προσαρμοστικό δίκτυο που φαίνεται στο σχήμα 23 είναι εμπροσθόδρομο, μιας και η έξοδος κάθε κόμβου διαδίδεται από την πλευρά της εισόδου (αριστερά) προς την πλευρά της εξόδου (δεξιά) πάντα. Αν υπάρχει σύνδεσμος ανάδρασης που σχηματίζει ένα κυκλικό μονοπάτι στο δίκτυο, τότε το δίκτυο είναι επαναληπτικό. Το σχήμα 25 είναι ένα παράδειγμα. Στα γραφήματα, ένα εμπροσθόδρομο δίκτυο αναπαριστάται από ένα ακυκλικό κατευθυνόμενο γράφημα που δεν περιέχει κατευθυνόμενους κύκλους, ενώ ένα επαναληπτικό δίκτυο περιέχει πάντα τουλάχιστον έναν κατευθυνόμενο κύκλο.



Σχήμα 25. Ένα επαναλαμβανόμενο προσαρμοστικό δίκτυο[1]



Σχήμα 26. Ένα εμπροσθόδρομο προσαρμοστικό δίκτυο σε αναπαράσταση τοπολογικής διάταξης.[1]

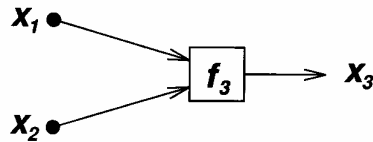
Στην **αναπαράσταση με επίπεδα** του εμπροσθόδρομου προσαρμοστικού δικτύου στο σχήμα 23, δεν υπάρχουν σύνδεσμοι μεταξύ κόμβων στο ίδιο επίπεδο, και οι εξοδοί των κόμβων σε ένα συγκεκριμένο επίπεδο είναι πάντα συνδεδεμένοι με κόμβους σε διαδοχικά επίπεδα. Αυτή η αναπαράσταση είναι συνήθως προτιμητέα εξ' αιτίας της ευκολίας διαμόρφωσης της, εξ' αιτίας του ότι κόμβοι στο ίδιο επίπεδο έχουν την ίδια λειτουργία ή γεννούν το ίδιο επίπεδο αφαίρεσης (level of abstraction), σχετικά με τα διανύσματα εισόδου.

Μια άλλη αναπαράσταση εμπροσθόδρομου δικτύου είναι η **αναπαράσταση τοπολογικής ταξινόμησης** (topological ordering representation), η οποία ετικετοποιεί τους κόμβους σε μια διατεταγμένη ακολουθία 1, 2, 3, ..., τέτοια που να μην υπάρχουν σύνδεσμοι από τον κόμβο i στον κόμβο j , οποτεδήποτε $i \geq j$. Το σχήμα 26 είναι η αναπαράσταση τοπολογικής ταξινόμησης (topological ordering representation) του δικτύου στο σχήμα 23. Αυτή η αναπαράσταση είναι λιγότερο διαμορφώσιμη απ' ό,τι η αναπαράσταση με επίπεδα, όμως διευκολύνει τον σχηματισμό κανόνων εκπαίδευσης. Η αναπαράσταση τοπολογικής ταξινόμησης (topological ordering representation) είναι στην πραγματικότητα μια ειδική περίπτωση της αναπαράστασης με επίπεδα, με έναν κόμβο ανά επίπεδο.

Ένα εμπροσθόδρομο προσαρμοστικό δίκτυο είναι στην πραγματικότητα μια στατική αντιστοίχιση μεταξύ των χώρων εισόδου και εξόδου. Αυτή η αντιστοίχιση μπορεί να είναι είτε μια απλή γραμμική σχέση ή μια υψηλού βαθμού μη γραμμική, εξαρτώμενη από την δομή του δικτύου (διάταξη κόμβων και συνδέσεων, κοκ.) και την λειτουργία κάθε κόμβου. Για να κατασκευάσουμε ένα προσαρμοστικό δίκτυο χρησιμοποιούμε ένα σύνολο δεδομένων προς εκπαίδευση και κάποιες διαδικασίες όπως οι κανόνες εκπαίδευσης ή οι προσαρμοστικοί αλγόριθμοι για την τροποποίηση των παραμέτρων για να βελτιώσουμε την απόδοση του δικτύου. Συνήθως η απόδοση ενός δικτύου μετριέται ως η ασυμφωνία (διαφορά) μεταξύ της επιθυμητής εξόδου και της πραγματικής κάτω από τις ίδιες συνθήκες. Αυτή η ασυμφωνία (διαφορά) ονομάζεται μέτρηση σφάλματος και μπορούμε να ορίσουμε διάφορες μορφές για διάφορες εφαρμογές. Γενικά, ένας κανόνας εκπαίδευσης εξάγεται από την εφαρμογή μιας συγκεκριμένης τεχνικής βελτιστοποίησης για μια δεδομένη μέτρηση σφάλματος. [1]

1.7.3 ΠΑΡΑΔΕΙΓΜΑΤΑ ΠΡΟΣΑΡΜΟΣΤΙΚΩΝ ΔΙΚΤΥΩΝ

1.7.3.1 ΠΡΟΣΑΡΜΟΣΤΙΚΟ ΔΙΚΤΥΟ ΜΕ ΕΝΑΝ ΓΡΑΜΜΙΚΟ ΚΟΜΒΟ (LINEAR NODE)



Σχήμα 27. Ένα γραμμικό προσαρμοστικό δίκτυο με έναν κόμβο.[1]

Στο σχήμα 27 είναι προσαρμοστικό δίκτυο με έναν κόμβο. Ορίζεται με:

$$x_3 = f_3(x_1, x_2; a_1, a_2, a_3) = a_1x_1 + a_2x_2 + a_3,$$

όπου x_1, x_2 είναι οι είσοδοι και a_1, a_2, a_3 είναι οι τροποποιήσιμες παράμετροι. Η συνάρτηση προσδιορίζει ένα επίπεδο σε έναν χώρο $x_1 - x_2 - x_3$, και θέτοντας τις κατάλληλες τιμές για τις παραμέτρους, μπορούμε να τοποθετήσουμε αυτό το επίπεδο αυθαίρετα όπως εμείς θέλουμε. Χρησιμοποιώντας το τετραγωνικό σφάλμα ως το σφάλμα μέτρησης για το δίκτυο, μπορούμε να αναγνωρίσουμε τις βέλτιστες παραμέτρους μέσω της μεθόδου εκτίμησης των γραμμικών ελαχίστων τετραγώνων.

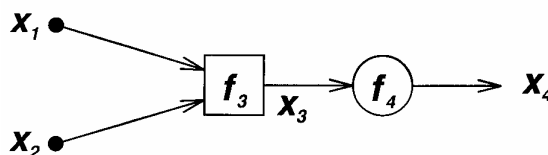
1.7.3.2 ΔΙΚΤΥΟ PERCEPTRON

Εάν προσθέσουμε και άλλον έναν κόμβο ώστε να επιτρέψουμε στην έξοδο του δικτύου στο σχήμα 27 να παίρνει μόνο δυο τιμές 0 και 1, τότε παίρνουμε το μη γραμμικό δίκτυο που φαίνεται σχήμα 28. Συγκεκριμένα, οι έξοδοι των κόμβων εκφράζονται ως εξής:

$$x_3 = f_3(x_1, x_2; a_1, a_2, a_3) = a_1x_1 + a_2x_2 + a_3,$$

$$\text{και } x_4 = f_4(x_3) = \begin{cases} 1 & \text{if } x_3 \geq 0 \\ 0 & \text{if } x_3 < 0 \end{cases},$$

όπου f_3 είναι μια γραμμική παραμετρική συνάρτηση και f_4 είναι μια βηματική συνάρτηση που αντιστοιχεί το x_3 είτε στο 0 είτε στο 1. Την ολική (overall) συνάρτηση του δικτύου μπορούμε να την δούμε σαν έναν **γραμμικό ταξινομητή**: Ο πρώτος κόμβος σχηματίζει ένα όριο απόφασης (decision boundary) σαν μια ευθεία γραμμή στον χώρο των $x_1 - x_2$, και ο δεύτερος κόμβος δείχνει σε ποιο ημιεπίπεδο ανήκει το διάνυσμα εισόδου (x_1, x_2) . Προφανώς μπορούμε να σχηματίσουμε ένα ισοδύναμο ημιεπίπεδο με ένα μόνο κόμβο του οποίου η συνάρτηση είναι σύνθεση των f_3 και f_4 . Ο κόμβος που προκύπτει είναι το δομικό υλικό του κλασσικού perceptron.



Σχήμα 28. Ένα μη γραμμικό προσαρμοστικό δίκτυο με έναν κόμβο.[1]

Μιας και η βηματική συνάρτηση είναι ασυνεχής σε ένα σημείο και επίπεδη σε όλα τα άλλα τα σημεία, δεν είναι κατάλληλη για εκπαιδευτικές διαδικασίες βασισμένες σε παραγώγους. Ένας τρόπος να αποφύγουμε αυτή τη δυσκολία είναι να χρησιμοποιήσουμε την σιγμοειδή συνάρτηση σαν μια συνθλιπτική (squashing) συνάρτηση που παίρνει τιμές μεταξύ 0 και 1:

$$x_4 = f_4(x_3) = \frac{1}{1 + e^{-x_3}}.$$

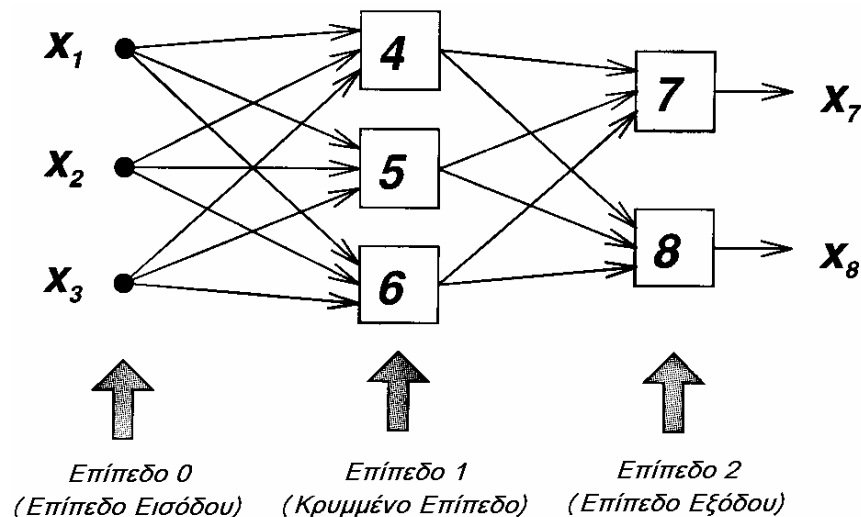
Αυτή είναι μια συνεχής και διαφορίσιμη προσέγγιση της βηματικής συνάρτησης. Η σύνθεση της f_3 και της διαφορίσιμης f_4 είναι το δομικό υλικό του πολυεπίπεδου perceptron.

1.7.3.3 ΈΝΑ ΠΟΛΥΕΠΙΠΕΔΟ PERCEPTRON

Στο σχήμα 29 φαίνεται μια χαρακτηριστική αρχιτεκτονική για ένα πολυεπίπεδο perceptron με τρεις εισόδους, δυο εξόδους και τρεις κρυμμένους κόμβους που δεν συνδέονται απευθείας ούτε σε είσοδο, ούτε σε έξοδο. Κάθε κόμβος σε ένα δίκτυο αυτού του τύπου έχει την ίδια συνάρτηση κόμβου, που είναι η σύνθεση μιας γραμμικής f_3 και μιας σιγμοειδούς f_4 όπως παραπάνω στο perceptron. Για παράδειγμα η συνάρτηση κόμβου στον κόμβο 7 στο σχήμα 29 είναι:

$$x_7 = \frac{1}{1 + \exp[-(w_{4,7}x_4 + w_{5,7}x_5 + w_{6,7}x_6 + t_7)]},$$

όπου x_4 , x_5 και x_6 είναι οι εξοδοί από τους κόμβους 4,5 και 6, αντιστοίχως, και το σύνολο παραμέτρων του κόμβου 7 δηλώνεται με το $\{w_{4,7}, w_{5,7}, w_{6,7}, t_7\}$. Συνήθως βλέπουμε το $w_{i,j}$ σαν το βάρος που σχετίζεται με τον σύνδεσμο που ενώνει τον i -οστό και τον j -οστό κόμβο, και το t_j σαν το κατώφλι (threshold) που συνδέεται με τον κόμβο j . Παρ' όλ' αυτά, αυτή η συσχέτιση βάρους-συνδέσμου είναι έγκυρη μόνο σ' αυτόν το τύπο δικτύου. Γενικά, ένας σύνδεσμος μας δείχνει μόνο την κατεύθυνση της ροής σήματος μεταξύ συνδεδεμένων κόμβων. [1]



Σχήμα 29. Ένα νευρωνικό δίκτυο 3-3-2.[1]

1.8 ΥΒΡΙΔΙΚΟΣ ΚΑΝΟΝΑΣ ΕΚΠΑΙΔΕΥΣΗΣ: ΣΥΝΔΥΑΣΜΟΣ ΑΠΟΤΟΜΗΣ ΦΘΙΝΟΥΣΑΣ ΜΕΘΟΔΟΥ ΚΑΙ ΣΦΑΛΜΑΤΟΣ ΕΛΑΧΙΣΤΩΝ ΤΕΤΡΑΓΩΝΩΝ

Παρ' όλο που μπορούμε να εφαρμόσουμε οπισθόδρομη διάδοση ή απότομη φθίνουσα μέθοδο για να αναγνωρίσουμε τις παραμέτρους σε ένα προσαρμοστικό δίκτυο, αυτή η απλή μέθοδος βελτιστοποίησης συνήθως συγκλίνει με αργό ρυθμό. Μπορούμε να παρατηρήσουμε παρ' όλ' αυτά, ότι η έξοδος ενός προσαρμοστικού δικτύου (υποθέτοντας ότι είναι μόνο μια) είναι γραμμική σε κάποιες από τις παραμέτρους του δικτύου. Γι' αυτό μπορούμε να αναγνωρίσουμε αυτές τις γραμμικές παραμέτρους με την γραμμική μέθοδο των ελαχίστων τετραγώνων. Αυτή η προσέγγιση οδηγεί στον υβριδικό κανόνα εκπαίδευσης ο οποίος συνδυάζει την απότομη φθίνουσα μέθοδο και την εκτιμήτρια ελαχίστων τετραγώνων για γρήγορη ταυτοποίηση των παραμέτρων. [1]

1.8.1 ΕΚΠΑΙΔΕΥΣΗ OFF-LINE (ΜΑΖΙΚΗ ΕΚΠΑΙΔΕΥΣΗ ΣΕ ΠΑΡΤΙΔΕΣ) (OFF-LINE LEARNING) (BATCH LEARNING)

Για ευκολία, υποθέτουμε ότι το υπό θεώρηση προσαρμοστικό δίκτυο έχει μόνο μια έξοδο που αναπαριστάται από την:

$$o = F(\mathbf{i}, S), (13)$$

όπου \mathbf{i} είναι το διάνυσμα των μεταβλητών εισόδου, S είναι το σύνολο των παραμέτρων και F είναι η ολική συνάρτηση εφαρμοσμένη από το προσαρμοστικό δίκτυο. Αν υπάρχει μια συνάρτηση H τέτοια ώστε η σύνθετη συνάρτηση $H \circ F$ είναι γραμμική σε κάποια από τα στοιχεία του S , τότε αυτά τα στοιχεία μπορούν να βρεθούν από την μέθοδο των ελαχίστων τετραγώνων. Ακόμα πιο τυπικά, αν το σύνολο των παραμέτρων S , μπορεί να διααιρεθεί σε δυο σύνολα

$$S = S_1 \oplus S_2, (14)$$

(όπου το \oplus αντιπροσωπεύει την άμεση άθροιση (direct sum)) τέτοια ώστε η $H \circ F$ είναι γραμμική στα στοιχεία του S_2 , τότε εφαρμόζοντας την H στην εξίσωση (13) έχουμε

$$H(o) = H \circ F(B\mathbf{i}, S), (15)$$

η οποία είναι γραμμική στα στοιχεία του S_2 . Δοθέντων των τιμών των στοιχείων του S_1 , μπορούμε να βάλουμε P δεδομένα εκπαίδευσης στην εξίσωση (15) και να πάρουμε μια εξίσωση πινάκων:

$$\mathbf{A} \boldsymbol{\theta} = \mathbf{y} (16)$$

Όπου $\boldsymbol{\theta}$ είναι ένα άγνωστο διάνυσμα του οποίου τα στοιχεία είναι παράμετροι στο S_2 . Αυτό είναι ένα σύννηθες γραμμικό πρόβλημα ελαχίστων τετραγώνων, και η καλύτερη λύση για το $\boldsymbol{\theta}$, η οποία ελαχιστοποιεί το $\|\mathbf{A} \boldsymbol{\theta} - \mathbf{y}\|^2$, είναι η εκτιμήτρια ελαχίστων τετραγώνων $\boldsymbol{\theta}^*$:

$$\theta^* = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y}, (17)$$

όπου είναι ο ανάστροφος του \mathbf{A} και $(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T$ είναι ο ψευδοαντίστροφος (pseudoinverse) του \mathbf{A} αν ο $\mathbf{A}^T \mathbf{A}$ έχει ορίζουσα διάφορη του μηδενός.

Θα μπορούσαμε να εφαρμόσουμε και τον τύπο της επαναληπτικής εκτιμήτριας των ελαχίστων τετραγώνων για να υπολογίσουμε το θ^* . Συγκεκριμένα, το i -οστό διάνυσμα γραμμή του πίνακα \mathbf{A} που προσδιορίζεται από την εξίσωση (16) θα είναι το \mathbf{a}_i^T και το i -οστό στοιχείο του \mathbf{y} θα είναι το y_i^T . Τότε το θ μπορεί να υπολογιστεί ως ακολούθως:

$$\left. \begin{aligned} \theta_{i+1} &= \theta_i + \mathbf{P}_{i+1} \mathbf{a}_{i+1} (y_{i+1}^T - \mathbf{a}_{i+1}^T \theta_i) \\ \mathbf{P}_{i+1} &= \mathbf{P}_i - \frac{\mathbf{P}_i \mathbf{a}_{i+1} \mathbf{a}_{i+1}^T \mathbf{P}_i}{1 + \mathbf{a}_{i+1}^T \mathbf{P}_i \mathbf{a}_{i+1}}, \quad i = 0, 1, \dots, P-1 \end{aligned} \right\} (18)$$

όπου η εκτιμήτρια ελαχίστων τετραγώνων θ^* είναι ίση με θ_P . Οι αρχικές συνθήκες που είναι αναγκαίες για την (18) είναι $\theta_0 = 0$ και $\mathbf{P}_0 = \gamma \mathbf{I}$, όπου γ είναι ένας θετικός μεγάλος αριθμός και \mathbf{I} είναι ο μοναδιαίος πίνακας διαστάσεων $M \times M$. Όταν έχουμε να κάνουμε με προσαρμοστικά δίκτυα με πολλαπλές εξόδους [το o στην (13) είναι ένα διάνυσμα στήλη], η εξίσωση (18) μπορεί και πάλι να εφαρμοστεί, με τη διαφορά ότι το y_i^T είναι η i -οστή γραμμή του πίνακα \mathbf{y} .

Έτσι μπορούμε να εφαρμόσουμε την απότομη φθίνουσα μέθοδο και την εκτιμήτρια ελαχίστων τετραγώνων, για να ανανεώσουμε τις παραμέτρους σε ένα προσαρμοστικό δίκτυο. Στην υβριδική εκπαίδευση που εφαρμόζεται σε κατάσταση μαζικής εκπαίδευσης σε παρτίδες, κάθε πέρασμα (epoch) αναλύεται σε ένα **εμπροσθόδρομο πέρασμα** και σε ένα **οπισθόδρομο πέρασμα**. Στο εμπροσθόδρομο πέρασμα, μετά την εισαγωγή του διανύσματος εισόδου, υπολογίζουμε τις εξόδους των κόμβων στο δίκτυο, επίπεδο προς επίπεδο, έως ότου βρεθεί μια αντιστοιχούσα σειρά στους πίνακες \mathbf{A} και \mathbf{y} από την εξίσωση (16). Αυτή η διαδικασία επαναλαμβάνεται για όλα τα ζεύγη των δεδομένων εκπαίδευσης για να σχηματιστούν τελείως οι \mathbf{A} και \mathbf{y} . Τότε οι παράμετροι στο S_2 ταυτοποιούνται είτε με τον τύπο του ψευδοαντιστρώφου στην εξίσωση (17) είτε με τους επαναληπτικούς τύπους των ελαχίστων τετραγώνων στην εξίσωση (18). αφού ταυτοποιηθούν οι παράμετροι στη S_2 , μπορούμε να υπολογίσουμε το μέτρο σφάλματος για κάθε ζεύγος δεδομένων εκπαίδευσης. Στο οπισθόδρομο πέρασμα, τα σήματα σφάλματος (η παράγωγος του μέτρου σφάλματος ως προς κάθε έξοδο κόμβου, βλέπε εξισώσεις (20) και (21)) διαδίδονται από το τέλος της εξόδου προς το τέλος της εισόδου. Το βαθμωτό διάνυσμα συμπληρώνεται με κάθε εγγραφή δεδομένου εκπαίδευσης. Στο τέλος του οπισθόδρομου περάσματος για όλα τα δεδομένα εκπαίδευσης, οι παράμετροι στο S_1 ενημερώνονται από την απότομη φθίνουσα μέθοδο με την εξίσωση (23).

Για δεδομένες προκαθορισμένες τιμές στο S_1 , οι παράμετροι στο S_2 που βρίσκονται, είναι εγγυημένες να είναι το γενικό βέλτιστο σημείο στον χώρο των παραμέτρων του S_2 εξαιτίας της επιλογής του μέτρου του τετραγωνικού σφάλματος. Ο υβριδικός κανόνας εκπαίδευσης μπορεί να μειώσει το μέγεθος του χώρου αναζήτησης που ελέγχεται από την «γνήσια» απότομη φθίνουσα μέθοδο, και κατά συνέπεια θα μειώσει ουσιαστικά και τον

χρόνο που απαιτείται για να φθάσουμε σε σύγκλιση. Παρ' όλα αυτά, κάποιες φορές οι παράμετροι του S_2 είναι κρυμμένες και χρειαζόμαστε μια συγκεκριμένη μέθοδο μετασχηματισμού για να τις ανακτήσουμε. [1]

1.8.2 ΕΚΠΑΙΔΕΥΣΗ ON-LINE (ΕΚΠΑΙΔΕΥΣΗ ΔΕΙΓΜΑ ΠΡΟΣ ΔΕΙΓΜΑ) (ON-LINE LEARNING) (PATTERN-BY-PATTERN LEARNING)

Αν οι παράμετροι ενημερώνονται μετά από την χρησιμοποίηση κάθε στοιχείου των δεδομένων, έχουμε την μέθοδο της On-line εκπαίδευσης ή αλλιώς της εκπαίδευσης δείγμα προς δείγμα. Αυτή η «στρατηγική» εκπαίδευσης είναι ουσιώδης στην On-line ταυτοποίηση παραμέτρων για συστήματα με μεταβαλλόμενα χαρακτηριστικά. Για να τροποποιήσουμε τον κανόνα της μαζικής εκπαίδευσης σε παρτίδες για να αποκτήσουμε μια On-line εκδοχή, θα πρέπει να βασιστεί η απότομη φθίνουσα μέθοδος στο E_p (από την εξίσωση (22)) αντί του E . Στην πραγματικότητα αυτή δεν είναι μια αληθινή βαθμωτή διαδικασία αναζήτησης για την ελαχιστοποίηση του E , θα προσεγγίσει το ένα όμως, αν ο ρυθμός εκπαίδευσης είναι μικρός.

Για την μέθοδο των επαναληπτικών ελαχίστων τετραγώνων για να εξηγήσουμε τα χαρακτηριστικά των εισερχόμενων δεδομένων που εξαρτώνται από το χρόνο, οι συνέπειες των ζευγών των παλαιών δεδομένων πρέπει να εξασθενούν καθώς τα νέα ζεύγη δεδομένων γίνονται διαθέσιμα. Μια απλή μέθοδος είναι να διατυπώσουμε το μέτρο του τετραγωνικού σφάλματος σε μια σταθμική εκδοχή με βάρη που δίνει υψηλότερο παράγοντα βάρους στα πιο πρόσφατα ζεύγη δεδομένων. Αυτό ισοδυναμεί με την προσθήκη ενός **παράγοντα λησμονιάς λ (forgetting factor λ)** στην αρχική επαναληπτική μέθοδο:

$$\left. \begin{aligned} \theta_{i+1} &= \theta_i + \mathbf{P}_{i+1} \mathbf{a}_{i+1} (y_{i+1}^T - \mathbf{a}_{i+1}^T \theta_i) \\ \mathbf{P}_{i+1} &= \frac{1}{\lambda} \left[\mathbf{P}_i - \frac{\mathbf{P}_i \mathbf{a}_{i+1} \mathbf{a}_{i+1}^T \mathbf{P}_i}{\lambda + \mathbf{a}_{i+1}^T \mathbf{P}_i \mathbf{a}_{i+1}} \right] \end{aligned} \right\},$$

όπου η συνήθης τιμή του λ στην πράξη είναι μεταξύ 0,9 και 1. Όσο μικρότερο είναι το λ , το συντομότερο εξασθενούν τα παλαιά δεδομένα. Ένα μικρό λ μερικές φορές προκαλεί αριθμητική αστάθεια, γι' αυτό και θα πρέπει να αποφεύγεται. [1]

1.9 ΟΠΙΣΘΟΔΡΟΜΗ ΔΙΑΔΟΣΗ ΓΙΑ ΔΙΚΤΥΑ ΧΩΡΙΣ ΑΝΑΤΡΟΦΟΔΟΤΗΣΗ.

(BACKPROPAGATION FOR FEEDFORWARD NETWORKS)

1.9.1 ΓΕΝΙΚΑ

Αυτή η τεχνική είναι χρήσιμη μόνο για δίκτυα χωρίς ανατροφοδότηση ή απλούστερα για όσα δεν έχουν συνδέσεις που προκαλούν επαναλήψεις. Ο όρος είναι η συντομογραφία του "backwards propagation of errors" (διάδοση του σφάλματος προς τα πίσω).

Η ουσία της τεχνικής είναι η ακόλουθη:

1. εισάγουμε ένα δείγμα προς εκπαίδευση στο νευρωνικό δίκτυο
2. συγκρίνουμε την έξοδο του νευρωνικού δικτύου με την απαιτούμενη εξόδο του δείγματος και υπολογίζουμε το σφάλμα σε κάθε νευρώνα.
3. για κάθε νευρώνα, υπολογίζουμε απ' το λάθος, την πραγματική έξοδο, και έναν παράγοντα κλιμάκωσης, για το πόσο χαμηλότερα ή ψηλότερα θα έπρεπε να είναι. Αυτό είναι το τοπικό σφάλμα.
4. χρησιμοποιώντας τα βάρη των νευρώνων στις εισερχόμενες συνδέσεις, αναθέτουμε μια «μομφή» για το τοπικό σφάλμα, σε νευρώνες του προηγούμενου επιπέδου (*level*).
5. επαναλαμβάνουμε τα παραπάνω βήματα στους νευρώνες στα προηγούμενα επίπεδα, χρησιμοποιώντας την «μομφή» του καθενός σαν το σφάλμα τους.

Όπως υπονοεί και το όνομα του αλγορίθμου, τα σφάλματα (συνεπώς και η μάθηση) διαδίδονται προς τα πίσω, από τους κόμβους εξόδο προς τους εσωτερικούς κόμβους. Η backpropagation χρησιμοποιείται για να υπολογίσουμε βαθμωτό άνυσμα του σφάλματος ενός δικτύου, λαμβάνοντας υπόψη τα μεταβαλλόμενα βάρη του δικτύου. Το βαθμωτό άνυσμα τότε χρησιμοποιείται σε έναν απλό στοχαστικό αλγόριθμο φθίνοντος βαθμωτού διανύσματος (*stochastic gradient descent algorithm*) για να βρει βάρη που ελαχιστοποιούν το σφάλμα. Συχνά ο όρος backpropagation χρησιμοποιείται με γενικότερο τρόπο, αναφερόμενος σε μια ολόκληρη διαδικασία περιγράφοντας και τον υπολογισμό του βαθμωτού και την χρήση του στον αλγόριθμο του στοχαστικού φθίνοντος βαθμωτού διανύσματος. Με το backpropagation συνήθως έχουμε γρήγορη σύγκλιση με ικανοποιητικά τοπικά ελάχιστα στα είδη των δικτύων για τα οποία είναι κατάλληλος. [4]

1.9.2 ΕΙΔΙΚΟΤΕΡΑ

Το κεντρικό κομμάτι αυτού του κανόνα εκμάθησης των νευρωνικών δικτύων αφορά το πώς με τρόπο περιοδικά επαναλαμβανόμενο θα αποκτήσει,

όπως είπαμε, ένα βαθμωτό διάνυσμα στο οποίο κάθε στοιχείο προσδιορίζεται ως η παράγωγος μιας μέτρησης σφάλματος για κάποια παράμετρο. Αυτό γίνεται με διαδικασίες του κανόνα της αλυσίδας (chain rule) μια βασική μέθοδος διαφορίσης σύνθετων συναρτήσεων. Η διαδικασία της εύρεσης ενός βαθμωτού διανύσματος σε μια δικτυακή δομή αναφέρεται γενικώς ως backpropagation εξαιτίας του ότι το βαθμωτό διάνυσμα υπολογίζεται στη κατεύθυνση που είναι αντίθετη προς αυτήν της εξόδου κάθε κόμβου. Όταν ένα βαθμωτό διάνυσμα βρεθεί, έχουμε διαθέσιμες για την ανανέωση / ενημέρωση των αναζητούμενων παραμέτρων, κάποιες μεθόδους βελτιστοποίησης με βάση κάποιες παραγωγίσεις καθώς και τεχνικές παλινδρόμησης. Συγκεκριμένα εάν χρησιμοποιήσουμε το βαθμωτό διάνυσμα σε μια απλή απότομη φθίνουσα (καταβατική, κατηφορική) μέθοδο (simple steepest descent method), τότε αυτή ονομάζεται κανόνας εκπαίδευσης backpropagation.

Ας υποθέσουμε ότι έχουμε ένα προσαρμοστικό δίκτυο με L επίπεδα και το επίπεδο l ($l = 0, 1, \dots, L$; $l = 0$ είναι το επίπεδο εισόδου) έχει $N(l)$ κόμβους. Τότε η έξοδος και η συνάρτηση του κόμβου i [$i = 1, \dots, N(l)$] στο επίπεδο l μπορεί να αναπαρασταθεί σαν $x_{l,i}$ και $f_{l,i}$ αντιστοίχως, όπως φαίνεται και στο σχήμα 30α. Υποθέτουμε ακόμα ότι δεν υπάρχουν συνδέσεις που να ενώνουν μη συνεχόμενα επίπεδα. Μιας και η έξοδος κάθε κόμβου εξαρτάται από τα εισερχόμενα σήματα και το σύνολο των παραμέτρων του κόμβου, η γενική έκφραση για την συνάρτηση $f_{l,i}$ του κόμβου είναι:

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

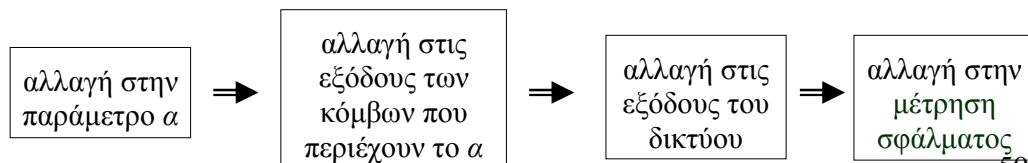
όπου α, β, γ , κλπ. είναι οι παράμετροι αυτού του κόμβου.

Υποθέτοντας ότι το συγκεκριμένο σύνολο δεδομένων εκπαίδευσης έχει P καταχωρήσεις, μπορούμε να προσδιορίσουμε μια μέτρηση **σφάλματος για την p -οστή** ($1 \leq p \leq P$) **καταχώρηση**, ως το σύνολο των τετραγωνικών σφαλμάτων:

$$E_p = \sum_{k=1}^{N(L)} (d_k - x_{L,k})^2, \quad (19)$$

όπου d_k είναι η k -οστή τιμή του p -οστού επιθυμητού διανύσματος εξόδου και $x_{L,k}$ είναι η k -οστή τιμή του πραγματικού διανύσματος εξόδου ευρισκόμενο, εισάγοντας το p -οστό διάνυσμα εισόδου στο δίκτυο (για λόγους απλότητας παραλήφθηκε ο δείκτης p για το d_k και το $x_{L,k}$). Προφανώς όταν το E_p είναι ίσο με μηδέν, το δίκτυο είναι ικανό να αναπαραγάγει ακριβώς το επιθυμητό διάνυσμα εξόδου στο p -οστό ζεύγος δεδομένων εκπαίδευσης. Γι' αυτό ο σκοπός μας είναι να ελαχιστοποιήσουμε το ολικό σφάλμα μέτρησης, που προσδιορίζεται από το $E_p = \sum_{p=1}^P E_p$ (το E_p εδώ εξαρτάται από τους κόμβους εξόδου μόνο).

Κοιτώντας τις ακόλουθες σχέσεις παρατηρούμε τα εξής:



μια μικρή αλλαγή στη παράμετρο a θα επηρεάσει την έξοδο του κόμβου που το περιέχει. Αυτό με τη σειρά του θα επηρεάσει την έξοδο του τελικού επιπέδου και κατ' επέκτασιν και τη μέτρηση σφάλματος. Γι' αυτό η βασική ιδέα στο υπολογισμό του βαθμωτού διανύσματος είναι να περάσουμε μια φόρμα πληροφοριών από παραγώγους, αρχίζοντας από το επίπεδο εξόδου και πηγαίνοντας στο ανάποδα, επίπεδο προς επίπεδο, μέχρι να φτάσουμε στο επίπεδο εισόδου.

1.9.2.1 ΠΡΟΣΔΙΟΡΙΣΜΟΣ ΤΟΥ ΣΗΜΑΤΟΣ ΣΦΑΛΜΑΤΟΣ.

Σφάλμα σήματος $\varepsilon_{l,i}$ είναι η παράγωγος της μέτρησης σφάλματος E_p ως προς την έξοδο του κόμβου i στο επίπεδο l , λαμβάνοντας υπόψιν τις άμεσες και έμμεσες συνδέσεις:

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

αυτός ο τύπος ονομάζεται πεπλεγμένη παράγωγος (ordered derivative) του Werbos. Η διαφορά μεταξύ πεπλεγμένης και κανονικής μερικής παραγώγου είναι η εξής: Για μια 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 μηδέν, μιας και το E_p δεν εξαρτάται από το $x_{l,i}$ άμεσα. Παρ' όλ' αυτά, είναι φανερό ότι το E_p εξαρτάται από το $x_{l,i}$ έμμεσα, καθώς μια αλλαγή στο $x_{l,i}$ θα εξαπλωθεί μέσω εμμέσων διαδρομών στο επίπεδο εξόδου και γι' αυτό έχουμε και αλλαγή στην τιμή του E_p . Γι' αυτό το $\varepsilon_{l,i}$ μπορεί να χαρακτηριστεί και σαν η αναλογία αυτών των δυο απειροελάχιστων αλλαγών.

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

$$\varepsilon_{L,i} = \frac{\partial^+ E_p}{\partial x_{L,i}} = \frac{\partial E_p}{\partial x_{L,i}} \quad (20)$$

που ισούται με $\varepsilon_{L,i} = -2 (d_i - x_{L,i})$ αν το E_p υπολογίζεται όπως στην εξίσωση (19). Για έναν εσωτερικό κόμβο στην i -οστή θέση του επιπέδου l , το σήμα σφάλματος μπορεί να υπολογιστεί από τον κανόνα της αλυσίδας (Leibniz):

$$\varepsilon_{l,i} = \underbrace{\frac{\partial^+ E_p}{\partial x_{l,i}}}_{\text{σήμα σφάλματος στο επίπεδο } l} = \sum_{m=1}^{N(l+1)} \underbrace{\frac{\partial^+ E_p}{\partial x_{l+1,m}}}_{\text{σήμα σφάλματος στο επίπεδο } l+1} \frac{\partial f_{l+1,m}}{\partial x_{l,i}} = \sum_{m=1}^{N(l+1)} \varepsilon_{l+1,m} \frac{\partial f_{l+1,m}}{\partial x_{l,i}} \quad (21)$$

όπου $0 \leq l \leq L - 1$. Δηλαδή το σήμα σφάλματος ενός εσωτερικού κόμβου στο επίπεδο l , μπορεί να εκφραστεί σαν γραμμικός συνδυασμός του σήματος σφάλματος των κόμβων στο επίπεδο $l + 1$. Γι' αυτό για κάθε l και i

$[0 \leq l \leq L \text{ και } 1 \leq i \leq N(l)]$, μπορούμε να βρούμε το $\varepsilon_{l,i} = \frac{\partial^+ E_p}{\partial x_{l,i}}$, εφαρμόζοντας πρώτα την εξίσωση (20) για να βρούμε τα σήματα σφάλματος του επιπέδου εξόδου και έπειτα εφαρμόζοντας την (21). Επαναλαμβάνοντας συνεχώς αυτά τα βήματα μπορούμε να φτάσουμε στο επιθυμητό επίπεδο l .

Η συνεπαγόμενη διαδικασία ονομάζεται οπισθοδρομή διάδοση (backpropagation), διότι τα σήματα σφάλματος αποκτώνται διαδοχικά από το επίπεδο εξόδου πίσω προς το επίπεδο εισόδου.

Ως βαθμωτό διάνυσμα ορίζεται η παράγωγος της μέτρησης σφάλματος ως προς κάθε παράμετρο, γι' αυτό πρέπει να εφαρμόσουμε τον κανόνα της αλυσίδας (Leibniz) ξανά, για να βρούμε το βαθμωτό διάνυσμα το οποίο είναι και προαπαιτούμενο για να χρησιμοποιήσουμε την απότομη φθίνουσα μέθοδο (steepest descent). Αν α είναι μια παράμετρος του i -οστού κόμβου στο επίπεδο l , θα έχουμε:

$$\frac{\partial^+ E_p}{\partial \alpha} = \frac{\partial^+ E_p}{\partial x_{l,i}} \frac{\partial f_{l,i}}{\partial a} = \varepsilon_{l,i} \frac{\partial f_{l,i}}{\partial a} \quad (22)$$

Αν θέλουμε το α να είναι κοινό σε διαφορετικούς κόμβους, θα πρέπει η εξίσωση (22) πάρει μια πιο γενική μορφή:

$$\frac{\partial^+ E_p}{\partial \alpha} = \sum_{x^* \in S} \frac{\partial^+ E_p}{\partial x^*} \frac{\partial f^*}{\partial a}$$

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

Η παράγωγος της συνολικής μέτρησης σφάλματος E ως προς το α είναι:

$$\frac{\partial^+ E}{\partial \alpha} = \sum_{p=1}^P \frac{\partial^+ E_p}{\partial a}$$

Ο γενικός τύπος ενημέρωσης για την νέα παράμετρο α είναι:

$$\Delta \alpha = -\eta \frac{\partial^+ E}{\partial \alpha} \quad (23)$$

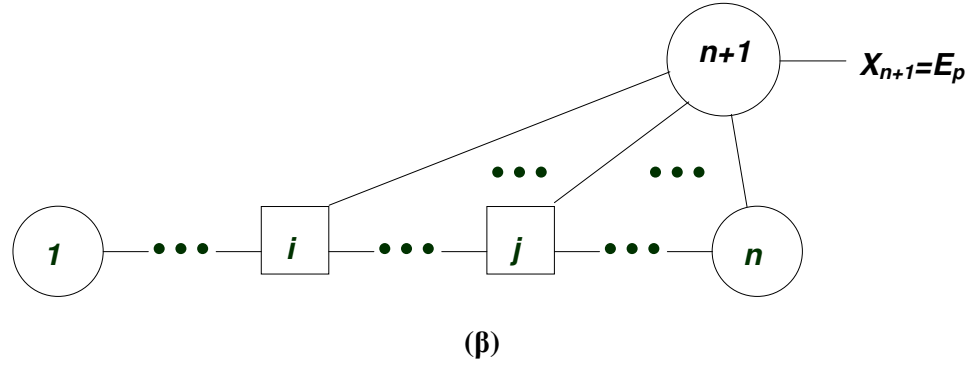
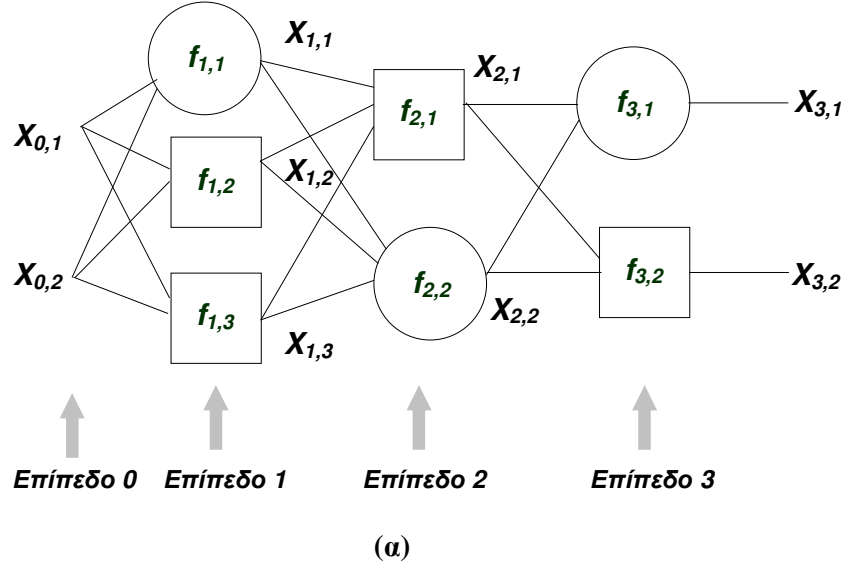
στο οποίο το η είναι ο ρυθμός εκμάθησης (learning rate), ο οποίος μπορεί να αναλυθεί ως:

$$\eta = \frac{\kappa}{\sqrt{\sum_a \left(\frac{\partial E}{\partial a} \right)^2}}$$

όπου κ είναι το μέγεθος βήματος (step size), το μήκος κάθε μετάβασης κατά μήκος της βαθμωτής κατεύθυνσης στον χώρο των παραμέτρων. Αλλάζοντας το μέγεθος βήματος, αλλάζουμε ταχύτητα σύγκλισης.

Όταν ένα δίκτυο χωρίς ανατροφοδότηση n -κόμβων αναπαριστάται στην τοπολογική του σειρά μπορούμε να ορίσουμε την μέτρηση σφάλματος E_p ως την έξοδο ενός επιπρόσθετου κόμβου με δείκτη $n + 1$, του οποίου

κόμβου η συνάρτηση f_{n+1} μπορεί να οριστεί από τις εξόδους κάθε κόμβου με μικρότερο δείκτη. σχήμα 30



Σχήμα 30. (α) αναπαράσταση επιπέδων, (β) αναπαράσταση τοπολογικής σειράς [1]

(Γι' αυτό, το E_p μπορεί να εξαρτάται απ' ευθείας από οποιουδήποτε κόμβους.) Εφαρμόζοντας τον κανόνα της αλυσίδας ξανά, παίρνουμε τον παρακάτω τύπο για να υπολογίσουμε το σήμα σφάλματος $\varepsilon_i = \frac{\partial E_p}{\partial x_i}$:

$$\frac{\partial^+ E_p}{\partial x_i} = \frac{\partial f_{n+1}}{\partial x_i} + \sum_{i < j \leq n} \frac{\partial^+ E_p}{\partial x_j} \frac{\partial f_j}{\partial x_i}$$

$$\text{ή } \varepsilon_i = \frac{\partial f_{n+1}}{\partial x_i} + \sum_{i < j \leq n} \varepsilon_j \frac{\partial f_j}{\partial x_i},$$

όπου η πρώτη έκφραση δείχνει την άμεση επίδραση του x_i στο E_p μέσω της άμεσης σύνδεσης από τον κόμβο i στον κόμβο $n+1$, και κάθε παραγόμενος όρος στην άθροιση συμβολίζει την έμμεση επίδραση του x_i στο E_p . Αφού βρούμε το σήμα σφάλματος για κάθε κόμβο, το βαθμωτό διάνυσμα για τις παραμέτρους εξάγεται όπως πριν. [1]

1.10 ΠΡΟΣΑΡΜΟΣΤΙΚΟ ΝΕΥΡΟ-ΑΣΑΦΕΣ ΣΥΣΤΗΜΑ ΕΞΑΓΩΓΗΣ ΣΥΜΠΕΡΑΣΜΑΤΩΝ (ΠΝΑΣΕΣ)

(ADAPTIVE NEURO-FUZZY INFERENCE SYSTEMS) (ANFIS)

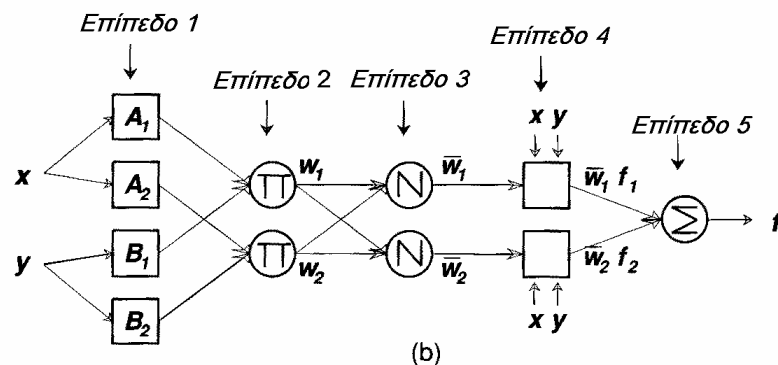
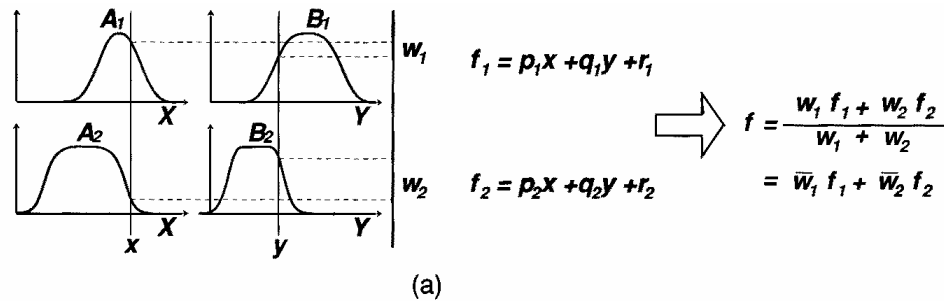
Από θέμα λειτουργικότητας δεν υπάρχουν περιορισμοί στο ποιες συναρτήσεις μπορούμε να χρησιμοποιήσουμε στους κόμβους ενός προσαρμοστικού δικτύου, εκτός από την προϋπόθεση της διαφορισιμότητας με ξεχωριστά βήματα (piecewise differentiability). Στην αρχιτεκτονική του σχεδίαση ο μόνο περιορισμός είναι ότι θα πρέπει να είναι εμπροσθόδρομου τύπου δίκτυο (σε κάποιες περιπτώσεις μετά το ανάπτυγμα του), αν δεν θέλουμε να χρησιμοποιήσουμε τα μοντέλα με την πιο πολύπλοκη και ασύγχρονη λειτουργία. Εξαιτίας αυτών των ελαχίστων περιορισμών τα προσαρμοστικά δίκτυα μπορούν να χρησιμοποιηθούν απευθείας σε μια ευρεία γκάμα εφαρμογών μοντελοποίησης, λήψης αποφάσεων, επεξεργασίας σήματος, και ελέγχου. [1]

1.10.1 ΑΡΧΙΤΕΚΤΟΝΙΚΗ ΤΟΥ ANFIS

Για απλότητα, υποθέτουμε ότι το υπό ανάλυση FIS έχει δυο εισόδους x και y και μια έξοδο z . Για ένα κοινό ασαφές μοντέλο Sugeno πρώτου βαθμού, ένα τυπικό σύνολο κανόνων με δυο ασαφείς κανόνες if-then είναι το παρακάτω:

Κανόνας 1^{ος}: if x is A_1 and y is B_1 , then $f_1 = p_1 x + q_1 y + r_1$,

Κανόνας 2^{ος}: if x is A_2 and y is B_2 , then $f_2 = p_2 x + q_2 y + r_2$.



Σχήμα 31. (α) ένα ασαφές μοντέλο Sugeno πρώτου βαθμού δυο εισόδων με δυο κανόνες (β) το ισοδύναμο ANFIS. [1]

Το σχήμα 31α απεικονίζει τον συλλογιστικό μηχανισμό (reasoning) γι' αυτό το μοντέλο Sugeno. Η αντίστοιχη ισοδύναμη αρχιτεκτονική του ANFIS φαίνεται στο σχήμα 31β, όπου κόμβοι του ιδίου επιπέδου έχουν παρόμοιες συναρτήσεις όπως θα δειχθεί και στη συνέχεια. (Η έξοδος του i -οστού κόμβου στο επίπεδο l δηλώνεται με $O_{l,i}$.)

Επίπεδο 1: κάθε κόμβος i σε αυτό το επίπεδο είναι ένας προσαρμοστικός κόμβος με συνάρτηση κόμβου:

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

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

όπου x (ή y) είναι η είσοδος στον κόμβο i και A_i (ή B_{i-2}) είναι η γλωσσική περιγραφική λέξη (όπως μικρός «small» ή μεγάλος «large») που σχετίζεται με αυτόν τον κόμβο. Δηλαδή $O_{1,i}$ είναι ο βαθμός συμμετοχής (membership grade) ενός ασαφούς συνόλου A ($= A_1, A_2, B_1$ ή B_2) και προσδιορίζει το βαθμό στον οποίο η δεδομένη είσοδος x (ή y) ικανοποιεί τον ποσοτικοποιητή (quantifier) A . Εδώ η συνάρτηση συμμετοχής για το A μπορεί να είναι οποιαδήποτε κατάλληλη παραμετρική συνάρτηση συμμετοχής (όπως πχ. η γενικευμένη καμπανοειδής: $\mu_A(x) = \frac{1}{1 + \left| \frac{x - c_i}{a_i} \right|^{2b}}$, όπου $\{a_i, b_i, c_i\}$ είναι το σύνολο

των παραμέτρων). Με την αλλαγή των τιμών αυτών των παραμέτρων τροποποιείται και η συνάρτηση αναλόγως, γι' αυτό και έχουμε διάφορες μορφές συναρτήσεων συμμετοχής για το ασαφές σύνολο A . Οι παράμετροι σ' αυτό το επίπεδο αναφέρονται ως **παράμετροι προϋποθέσεων** (premise parameters).

Επίπεδο 2: Κάθε κόμβος σ' αυτό το επίπεδο είναι ένας σταθερός κόμβος συμβολιζόμενος με Π , του οποίου η έξοδος είναι το γινόμενο όλων των εισερχόμενων σημάτων:

$$O_{2,i} = w_i = \mu_{A_i}(x) \mu_{B_i}(y), i = 1, 2.$$

Κάθε έξοδος κόμβου αντιπροσωπεύει την δύναμη πυροδότησης ενός κανόνα. Γενικά κάθε τελεστής T-norm που εκτελεί ασαφή ΚΑΙ (AND) μπορεί να χρησιμοποιηθεί ως συνάρτηση κόμβου σ' αυτό το επίπεδο.

Επίπεδο 3: Κάθε κόμβος σ' αυτό το επίπεδο είναι ένας σταθερός κόμβος (fixed node) συμβολιζόμενος με N . Ο i -οστός κόμβος υπολογίζει τον λόγο της δύναμης πυροδότησης του i -οστού κανόνα προς το σύνολο όλων των δυνάμεων πυροδότησης των κανόνων:

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

Για ευκολία, οι έξοδοι σ' αυτό το επίπεδο ονομάζονται κανονικοποιημένες **δυνάμεις πυροδότησης**.

Επίπεδο 4: Κάθε κόμβος i σ' αυτό το επίπεδο είναι ένας προσαρμοστικός κόμβος με συνάρτηση κόμβου:

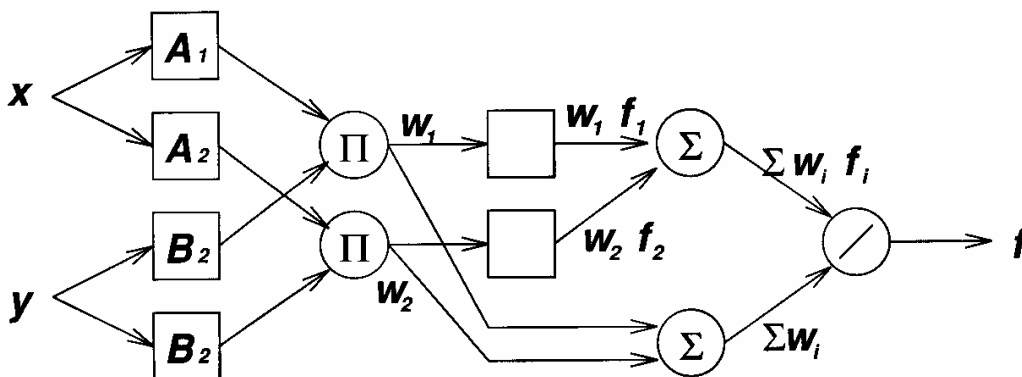
$$O_{4,i} = \bar{w}_i f_i = \bar{w}_i (p_i x + q_i y + r_i),$$

όπου \bar{w}_i είναι μια κανονικοποιημένη δύναμη πυροδότησης από το επίπεδο 3 και $\{p_i, q_i, r_i\}$ είναι το σύνολο των παραμέτρων αυτού του κόμβου. Οι παράμετροι σ' αυτό το επίπεδο αναφέρονται ως **παράμετροι συμπερασμάτων** (consequent parameters).

Επίπεδο 5: Ο μοναδικός κόμβος σ' αυτό το επίπεδο είναι ένας σταθερός κόμβος συμβολιζόμενος με Σ , ο οποίος υπολογίζει την ολική έξοδο ως το άθροισμα όλων των εισερχόμενων σημάτων:

$$\text{ολική έξοδος} = O_{5,1} = \sum_i \bar{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i}$$

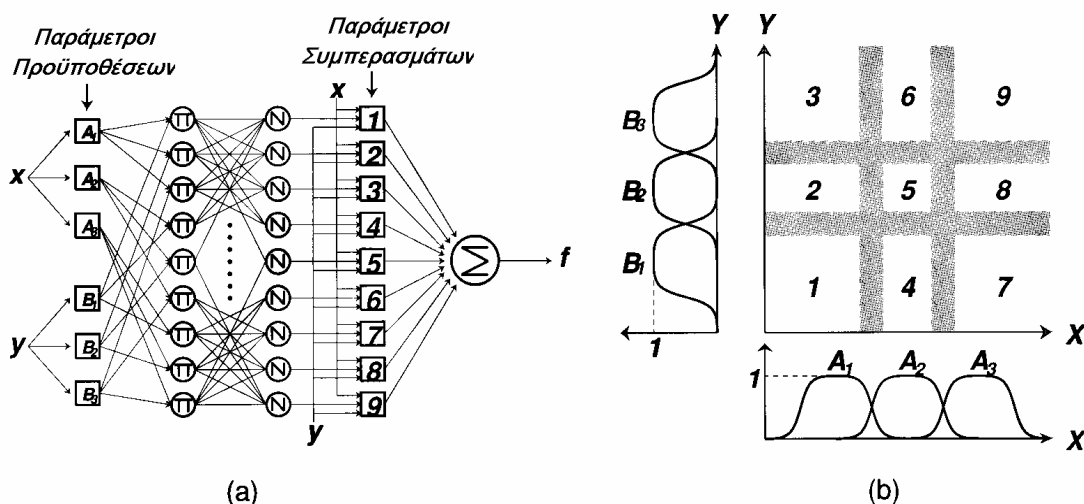
Αυτό το προσαρμοστικό δίκτυο είναι λειτουργικώς ισοδύναμο με το ασαφές μοντέλο Sugeno. Η δομή αυτό του προσαρμοστικού δικτύου δεν είναι μοναδική. Μπορούμε να συνδυάσουμε τα επίπεδα 3 και 4 για να αποκτήσουμε ένα ισοδύναμο δίκτυο με τέσσερα μόνο επίπεδα. Με το ίδιο δείγμα μπορούμε να πραγματοποιήσουμε την κανονικοποίηση βαρών στο τελευταίο επίπεδο. Το σχήμα 32 απεικονίζει ένα ANFIS αυτού του τύπου. Στην ακραία περίπτωση, μπορούμε να συρρικνώσουμε ακόμα και όλο το δίκτυο, σε έναν μόνο προσαρμοστικό κόμβο με το ίδιο σύνολο παραμέτρων. Προφανώς, η ανάθεση συναρτήσεων κόμβων και η σύνθεση του δικτύου είναι αυθαίρετες, εφόσον κάθε κόμβος και κάθε επίπεδο πραγματοποιούν λειτουργίες που είναι σημαντικές και έχουν την δυνατότητα να αποτελούνται από επιμέρους τμήματα.



Σχήμα 32. Η αρχιτεκτονική του ANFIS για το ασαφές μοντέλο Sugeno, όπου η κανονικοποίηση βαρών πραγματοποιείται στο τελευταίο επίπεδο. [1]

Εκτός από το Sugeno, μπορούμε να κατασκευάσουμε ANFIS και με το μοντέλο Mamdani. Προκύπτει όμως ένα περισσότερο περίπλοκο ANFIS σε σχέση με το Sugeno, που δεν σημαίνει ότι έχει και καλύτερες δυνατότητες εκπαίδευσης και υπολογισμού. Το Sugeno χρησιμοποιείται περισσότερο μιας και διακρίνεται για την διαφάνεια και την αποτελεσματικότητά του.

Στο σχήμα 33α βλέπουμε μια αρχιτεκτονική ANFIS που είναι ισοδύναμη με ένα ασαφές μοντέλο Sugeno πρώτου βαθμού δυο εισόδων και εννέα κανόνων, όπου σε κάθε είσοδο θεωρούμε ότι αντιστοιχούν τρεις συναρτήσεις συμμετοχής. Το σχήμα 33β απεικονίζει πως ο δυο διαστάσεων χώρος εισόδου είναι χωρισμένος σε εννέα υπέρθετες (overlapping) ασαφείς περιοχές, εκ των οποίων, κάθε μια ελέγχεται από ένα ασαφή κανόνα if-then. Δηλαδή το μέρος των προϋποθέσεων ενός κανόνα προσδιορίζει μια ασαφή περιοχή, ενώ το μέρος των συμπερασμάτων προσδιορίζει την έξοδο μέσα στην περιοχή.



Σχήμα 33. (α) Η αρχιτεκτονική του ANFIS για το ασαφές μοντέλο Sugeno με δυο εισόδους και εννέα κανόνες (β) ο χώρος εισόδου, χωρισμένος σε εννέα ασαφείς περιοχές. [1]

1.10.2 ΥΒΡΙΔΙΚΟΣ ΑΛΓΟΡΙΘΜΟΣ ΕΚΠΑΙΔΕΥΣΗΣ

Από την αρχιτεκτονική του ANFIS που βλέπουμε στο σχήμα 31β, παρατηρούμε ότι όταν οι τιμές των παραμέτρων από τις προϋποθέσεις είναι προκαθορισμένες (αμετάβλητες), η ολική έξοδος μπορεί να εκφραστεί σαν ο γραμμικός συνδυασμός των παραμέτρων από τα συμπεράσματα. Με σύμβολα, η έξοδος f στο σχήμα 31β μπορεί να ξαναγραφτεί ως εξής:

$$\begin{aligned} f &= \frac{w_1}{w_1 + w_2} f_1 + \frac{w_2}{w_1 + w_2} f_2 \\ &= \bar{w}_1 (p_1 x + q_1 y + r_1) + \bar{w}_2 (p_2 x + q_2 y + r_2) \\ &= (\bar{w}_1 x) p_1 + (\bar{w}_1 y) q_1 + (\bar{w}_1) r_1 + (\bar{w}_2 x) p_2 + (\bar{w}_2 y) q_2 + (\bar{w}_2) r_2 \end{aligned}$$

η οποία είναι γραμμική στις παραμέτρους των συμπερασμάτων p_1, q_1, r_1, p_2, q_2 και r_2 . Από αυτήν την παρατήρηση, έχουμε:

S = σύνολο όλων των παραμέτρων,

S_1 = σύνολο των (μη γραμμικών) παραμέτρων από τις προϋποθέσεις,

S_2 = σύνολο των (γραμμικών) παραμέτρων από τα συμπεράσματα

στην εξίσωση (14), και $H(\cdot)$ και $F(\cdot, \cdot)$ είναι η ταυτοτική (identity) συνάρτηση και η συνάρτηση του FIS, αντιστοίχως, στην εξίσωση (15). Έτσι εφαρμόζεται άμεσα ο υβριδικός αλγόριθμος εκπαίδευσης. Πιο συγκεκριμένα, στο εμπροσθόδρομο πέρασμα του υβριδικού αλγορίθμου εκπαίδευσης, οι έξοδοι των κόμβων πηγαίνουν προς τα μπρος μέχρι το επίπεδο 4 και οι παράμετροι των συμπερασμάτων ευρίσκονται με την μέθοδο των ελαχίστων τετραγώνων. Στο οπισθόδρομο πέρασμα, τα σήματα σφάλματος διαδίδονται προς τα πίσω και οι παράμετροι των προϋποθέσεων ενημερώνονται για τις νέες τιμές τους από βαθμωτή φθίνουσα μέθοδο. Στον παρακάτω πίνακα 1.2 συνοψίζονται οι δραστηριότητες του κάθε περάσματος.

	εμπροσθόδρομο πέρασμα	Οπισθόδρομο πέρασμα
παράμετροι των προϋποθέσεων	Προκαθορισμένες (αμετάβλητες)	βαθμωτή φθίνουσα μέθοδος
παράμετροι των συμπερασμάτων	εκτιμήτρια ελαχίστων τετραγώνων	προκαθορισμένες (αμετάβλητες)
σήματα	έξοδοι των κόμβων	σήματα σφάλματος

πίνακας 1.2. Δυο περάσματα στην υβριδική διαδικασία εκπαίδευσης για το ANFIS.

Οι παράμετροι των συμπερασμάτων που ευρίσκονται με αυτόν τον τρόπο, είναι βέλτιστες υπό την προϋπόθεση ότι οι παράμετροι των προϋποθέσεων είναι προκαθορισμένες (αμετάβλητες). Αναλόγως η υβριδική προσέγγιση συγκλίνει πολύ γρηγορότερα μιας και μειώνει τις διαστάσεις του διαστήματος αναζήτησης της αρχικής καθαρής μεθόδου οπισθόδρομης διάδοσης. Γι' αυτό πρέπει να βλέπουμε πρώτα απ' όλα αν είναι δυνατόν να αναλύσουμε το σύνολο των παραμέτρων στα συνθετικά του μέρη.

Ο μηχανισμός εκπαίδευσης δεν θα πρέπει να εφαρμόζεται για να επιλεγούν οι συναρτήσεις συμμετοχής στο ANFIS Sugeno, μιας και αυτές μεταφράζουν γλωσσικές και υποκειμενικές περιγραφές από πιθανώς κακώς ορισμένες αρχές. Η επιλογή πρέπει να αφήνεται στον χρήστη μιας και είναι διαδικασία που εξαρτάται από την κάθε περίπτωση. Κατ' αρχήν αν είναι μεγάλο το μέγεθος του διαθέσιμου συνόλου δεδομένων εισόδου-εξόδου, τότε χρειάζεται λεπτομερής ρύθμιση των συναρτήσεων συμμετοχής, μιας και αυτές, αφού καθορίζονται από ανθρώπους, σπανίως είναι βέλτιστες ως προς την ικανότητα τους να μας δώσουν τα επιθυμητά αποτελέσματα εξόδου. Παρ' όλ' αυτά, αν το σύνολο των δεδομένων είναι πολύ μικρό, τότε είναι πιθανό να μην περιέχει αρκετές πληροφορίες σχετικά με το σύστημα στόχος. Σ' αυτήν την περίπτωση, οι συναρτήσεις συμμετοχής που καθορίζονται από ανθρώπους, παρουσιάζουν σημαντικές πληροφορίες που όμως μπορεί να μην αντιπροσωπεύουν το σύνολο των δεδομένων. Γι' αυτό οι συναρτήσεις συμμετοχής θα πρέπει να παραμένουν οι προκαθορισμένες καθ' όλη τη διάρκεια της διαδικασίας εκπαίδευσης. [1]

Σε όλα τα προηγούμενα χρησιμοποιούσαμε προκαθορισμένες και κάπως αυθαίρετα επιλεγμένες συναρτήσεις συμμετοχής.

Πολλές φορές έχουμε μια συλλογή από δεδομένα εισόδου / εξόδου. Δεν μπορούμε όμως να χρησιμοποιήσουμε ένα προκαθορισμένο μοντέλο βασισμένο στα στοιχεία αυτά ώστε να ξέρουμε ποιες συναρτήσεις συμμετοχής και με ποιες παραμέτρους να χρησιμοποιήσουμε. Τότε μπορούμε να βοηθηθούμε από το ANFIS, ώστε να βρούμε τις καταλληλότερες συναρτήσεις συμμετοχής.

Οι Νεύρο-προσαρμοστικές τεχνικές είναι αρκετά απλές. Μας παρέχουν μια μέθοδο ώστε το ασαφές μοντέλο μας να εκπαιδευτεί με τις πληροφορίες που του δίνει το σύνολο των δεδομένων, ούτως ώστε να υπολογιστούν οι παράμετροι των συναρτήσεων συμμετοχής που με τον καλύτερο τρόπο επιτρέπουν στο Σύστημα εξαγωγής συμπερασμάτων να ανιχνεύει τα δεδομένα εισόδου / εξόδου. Είναι μια μέθοδος παρόμοια με αυτήν των νευρωνικών δικτύων.

Η προσαρμογή των παραμέτρων των συναρτήσεων συμμετοχής από ένα ANFIS γίνεται με αλγορίθμους εκμάθησης είτε μόνο backpropagation είτε με έναν υβριδικό αλγόριθμο, συνδυασμό backpropagation και μεθόδου ελαχίστων τετραγώνων.

Ο υπολογισμός των παραμέτρων αυτών, που αλλάζουν κατά την διάρκεια της διαδικασίας εκμάθησης, διευκολύνεται από ένα βαθμωτό διάνισμα που μας δείχνει πόσο καλά το FIS μοντελοποιεί τα δεδομένα εισόδου / εξόδου για το εν λόγω σύνολο παραμέτρων. Μόλις αποκτήσουμε το βαθμωτό διάνισμα μπορούμε πλέον να εφαρμόσουμε κάποιον από τους πολλούς αλγορίθμους βελτιστοποίησης, για να βελτιώσουμε αυτές τις παραμέτρους, ούτως ώστε να μειώσουμε το μέγεθος του σφάλματος, (συνήθως το άθροισμα των τετραγώνων των διαφορών μεταξύ πραγματικών και επιθυμητών τιμών). [3]

1.10.3 ΥΠΕΡΠΡΟΣΑΡΜΟΓΗ (OVERFITTING) ΔΕΔΟΜΕΝΩΝ

Σε συνήθεις περιπτώσεις αρκεί η εκπαίδευση για να πάρουμε τα επιθυμητά αποτελέσματα [3]. Όταν όμως οι μετρήσεις μας εμπεριέχουν και θόρυβο, ή τα προς εκπαίδευση δεδομένα χρησιμοποιήθηκαν πάρα πολύ ή αντίθετα όταν ήταν πάρα πολύ λίγα τότε παρατηρείται το φαινόμενο της **υπερπροσαρμογής (overfitting)**. Σ' αυτά τότε παρατηρούμε να αυξάνει συνεχώς η απόδοση του μοντέλου και αντίθετα να μειώνεται δραματικά η απόδοση σε δεδομένα καινούργια και άγνωστα μέχρι εκείνη τη στιγμή, με αύξηση του σφάλματος του μοντέλου. Χάνεται δηλαδή η δυνατότητα γενικής χρήσης του. [4]

Έτσι όμως τα προς εκπαίδευση δεδομένα δεν μπορούν να είναι αντιπροσωπευτικά των χαρακτηριστικών που θα θέλαμε να πάρουμε από αυτά. Γι' αυτό σ' αυτές τις περιπτώσεις θα πρέπει να επαληθεύσουμε το τελικό μας μοντέλο, χρησιμοποιώντας δεδομένα για έλεγχο. [3]

1.10.4 ΠΕΡΙΟΡΙΣΜΟΙ ΤΟΥ ANFIS

- Βασίζεται σε ένα FIS τύπου Sugeno
- Το FIS τύπου Sugeno πρέπει να είναι 0^{ου} ή 1^{ου} βαθμού
- Έχει μια έξοδο, που λαμβάνεται με τη μέθοδο αποασαφοποίησης σταθμισμένου μέσου. Όλες οι συναρτήσεις συμμετοχής εξόδου πρέπει να είναι ιδίου τύπου, είτε γραμμικές είτε σταθερές.
- Να μην γίνεται κοινή χρήση κανόνων. Διαφορετικοί κανόνες δεν μπορούν να έχουν την ίδια συνάρτηση συμμετοχής εξόδου.
- Να έχουν βάρη σε κάθε κανόνα.
- Δεν μπορεί να χρησιμοποιήσουμε οποιαδήποτε συνάρτηση συμμετοχής θέλουμε, πάρα μόνο προκαθορισμένες που επιβάλλουν οι ορισμοί του ANFIS.

1.10.5 ΠΡΑΚΤΙΚΑ ΖΗΤΗΜΑΤΑ

Το σφάλμα ελέγχου μειώνεται κατά την διάρκεια της εκπαίδευσης μέχρι ενός συγκεκριμένου σημείου και μετά ξανά αυξάνει. Το σημείο αυτό από το οποίο αυξάνει δηλώνει το σημείο από το οποίο αρχίζει η υπερπροσαρμογή. Το ANFIS έτσι επιλέγει για παραμέτρους του μοντέλου, εκείνες που σχετίζονται με το σημείο αυτό όπου και παρατηρείται και το μικρότερο σφάλμα ελέγχου. [3]

Το πρόβλημα που παρατηρούμε όταν τα δεδομένα για έλεγχο είναι αρκετά διαφορετικά από τα δεδομένα για εκπαίδευση και το FIS δεν έχει εντοπίσει τις παραμέτρους γι' αυτά τα δεδομένα πάρα πολύ καλά, μπορούμε να το αντιμετωπίσουμε, αναλύοντας το διάγραμμα με τα σφάλματα ελέγχου. Για να ελαχιστοποιήσουμε αυτά τα σφάλματα θα πρέπει, είτε να χρησιμοποιήσουμε περισσότερα δεδομένα για την εκπαίδευση, είτε να τροποποιήσουμε στις συναρτήσεις συμμετοχής που χρησιμοποιούμε, τον αριθμό τους ή τον τύπο τους.

Σε ένα συμβατικό FIS ο αριθμός των κανόνων καθορίζεται από κάποιον ειδικό που είναι εξοικειωμένος με το σύστημα στόχο που θα μοντελοποιήσουμε. Όταν όμως δεν υπάρχει ο ειδικός ο αριθμός των συναρτήσεων συμμετοχής που αντιστοιχούν σε κάθε μεταβλητή εισόδου, επιλέγεται εμπειρικά – που σημαίνει ότι απεικονίζουμε τα σύνολα των δεδομένων σε γραφικές αναπαραστάσεις και τα εξετάζουμε οπτικά, ή πιο απλά με δόκιμες και αποτυχίες. Για σύνολα δεδομένων με περισσότερες από τρεις εισόδους, η οπτική εξέταση δεν είναι πολύ αποτελεσματική και συνήθως βασιζόμαστε στην δοκιμή και αποτυχία. Η κατάσταση είναι παρόμοια με αυτή στα νευρωνικά δίκτυα, όπου δεν υπάρχει ένας εύκολος τρόπος να καθορίσουμε εκ των προτέρων τον ελάχιστο αριθμό των κρυφών μονάδων που απαιτούνται για να πετύχουμε το επιθυμητό επίπεδο απόδοσης. [1]

1.10.6 ΠΡΟΒΛΕΨΗ ΧΡΟΝΟΣΕΙΡΩΝ

[3] Κάθε σύνολο δεδομένων που χρησιμοποιούμε στο ANFIS, πρέπει να είναι σε μορφή **πίνακα**, με τα δεδομένα εισόδου να είναι διατεταγμένα σαν διανύσματα σε όλες τις στήλες, εκτός της τελευταίας την οποία καταλαμβάνει το διανύσματα με τα δεδομένα εξόδου.

[3] Στην πρόβλεψη χρονοσειρών, χρησιμοποιούμε γνωστές τιμές τους μέχρι ένα χρονικό σημείο t , για να προβλέψουμε την τιμή σε ένα σημείο στο μέλλον $t + P$. Η συνήθης μέθοδος για αυτού του είδους τις προβλέψεις, είναι να δημιουργήσουμε μια αντιστοίχιση από D δειγματικά σημεία δεδομένων, που απέχουν Δ μονάδες χρόνου μεταξύ τους,

$$x(t - (D - 1)\Delta), x(t - (D - 2)\Delta), \dots, x(t - \Delta), x(t),$$

για να προβλέψουμε μια μελλοντική τιμή $x(t + P)$.

Πχ. αν $D = 3$, $\Delta = P = 4$, θα δημιουργήσουμε ένα διάνυσμα τριών διαστάσεων για τα δεδομένα εκπαίδευσης στην είσοδο:

$$w(t) = [x(t - 8) \quad x(t - 4) \quad x(t)]$$

και η αντίστοιχη έξοδος θα είναι:

$$s(t) = x(t + 4)$$

Για να έχουμε καλές δυνατότητες γενικοποίησης, είναι σημαντικό να είναι ο αριθμός των δεδομένων προς εκπαίδευση αρκετές φορές μεγαλύτερος απ' ότι ο αριθμός των παραμέτρων που επιδιώκουμε να υπολογίσουμε. Μια καλή αναλογία είναι μεταξύ δεδομένων / παραμέτρων είναι το 5.

[1] Η απόδοση της πρόβλεψης, συγκρίνεται με τα αποτελέσματα μιας προσέγγισης ενός νευρωνικού δικτύου σειριακής συσχέτισης (cascade correlation neural network approach) και ενός συμβατικού μοντέλου αυτοπαλινδρόμησης (auto-regressive AR).

Ένα μέτρο σύγκρισης και ελέγχου των αποτελεσμάτων είναι η ρίζα του μέσου τετραγωνικού σφάλματος (root mean square error – RMSE). Κατά την διάρκεια της εκπαίδευσης συνήθως θα πρέπει το $RMSE_{trn}$ των δεδομένων για εκπαίδευση να είναι μικρότερο του $RMSE_{chk}$ των δεδομένων για έλεγχο ($RMSE_{chk} > RMSE_{trn}$).

ΜΕΡΟΣ 2^ο

2. ΠΡΑΚΤΙΚΗ ΕΦΑΡΜΟΓΗ

2.1 ΓΕΝΙΚΗ ΠΕΡΙΓΡΑΦΗ ΤΟΥ ΠΡΟΓΡΑΜΜΑΤΟΣ ΠΡΟΒΛΕΨΗΣ

Στο πρακτικό κομμάτι της εργασίας κατασκευάστηκε αλγόριθμος για να τρέχει σε Matlab, κάνοντας χρήση του fuzzy logic toolbox, το οποίο περιέχει όλα τα απαραίτητα εργαλεία για την ολοκληρωμένη ανάπτυξη και αναλυτική χρήση ενός ANFIS.

Ο αλγόριθμος δέχεται δεδομένα από μια χρονοσειρά που περιέχει στοιχεία για την κατανάλωση ηλεκτρικού φορτίου στην Κρήτη και με βάση τις τιμές του παρελθόντος, προβλέπει τις αναμενόμενες τιμές για το μέλλον. Η χρονοσειρά που χρησιμοποιήσαμε, περιέχει δεδομένα για κάθε μήνα από τον Ιανουάριο του 1974 έως τον Δεκέμβριο του 2000, ή αλλιώς:

$$12 \text{ μήνες} \times 27 \text{ χρόνια} = 324 \text{ μήνες.}$$

Τα δεδομένα τα ομαδοποιήσαμε με δυο τρόπους, για την περαιτέρω επεξεργασία τους:

- I.** την πρώτη φορά εισήχθησαν ξεχωριστά, τα δεδομένα του ιδίου μόνο μήνα κάθε φορά, για την περίοδο 1974–2000. Δηλαδή γίνανε 12 επαναλήψεις, και κάθε επανάληψη αντιστοιχούσε στον μήνα που είναι σε εκείνη τη θέση εφόσον αυτοί αριθμηθούν βάσει της χρονολογικής τους σειράς, πχ. η πρώτη επανάληψη στον Ιανουάριο, η δεύτερη στον Φεβρουάριο ... η δωδέκατη στον Δεκέμβριο.
- II.** την δεύτερη φορά εισήχθησαν τα 324 δεδομένα όλα μαζί και με την σωστή σειρά. Δηλαδή:
 - 1) πρώτο ήταν η κατανάλωση του Ιανουαρίου 1974,
 - 2) δεύτερο του Φεβρουαρίου 1974,
 - ...,
 - 13) δέκατο τρίτο ο Ιανουάριος του 1975,
 - ...,
 - 100) εκατοστό ο Απρίλιος του 1982,
 - ...,
 - 323) τριακοσιοστό εικοστό τρίτο ο Νοέμβριος του 2000, και
 - 324) τριακοσιοστό εικοστό τέταρτο ο Δεκέμβριος του 2000.

Η μεθοδολογία της επεξεργασία για την κάθε ομάδα {I, II} ήταν κοινή. Αλλάζε μόνο, όπως ήταν φυσικό, ο όγκος των δεδομένων που περιλαμβάνονταν σε κάθε επανάληψη, και ο συνολικός χρόνος που η CPU του υπολογιστή χρειαζόταν για να περατώσει την επεξεργασία για την εξαγωγή των αποτελεσμάτων.

Στην περίπτωση **I** χρησιμοποιήσαμε τα πρώτα 23 δεδομένα για την εκπαίδευση του συστήματος δημιουργώντας έναν πίνακα 22×2, όπου στην πρώτη στήλη είχαμε τα δεδομένα 1 – 22 και στη δεύτερη τα δεδομένα 2 – 23. Τα 5 τελευταία δεδομένα τα χρησιμοποιήσαμε για τον έλεγχο, δημιουργώντας έναν αντίστοιχο πίνακα 4×2, όπου στην πρώτη στήλη είχαμε τα δεδομένα 23 – 26 και στη δεύτερη τα δεδομένα 24 – 27.

Έπειτα δημιουργούμε ένα ANFIS, με τη βοήθεια του fuzzy logic toolbox, το οποίο θα προβλέψει τις καταλληλότερες παραμέτρους των συναρτήσεων συμμετοχής (MF parameters). Πειραματιστήκαμε αλλάζοντας τον αριθμό των εποχών (epochs), τον αριθμό των συναρτήσεων συμμετοχής, τον τύπο τους και το μέγεθος του βήματος. Για λόγους βοηθητικούς (για να έχουμε καλύτερη οπτική παρατήρηση) εξάγουμε ένα σχήμα στο οποίο περιέχονται δυο ξεχωριστές γραφικές παραστάσεις, η πρώτη με το όνομα “**error curves**” και η δεύτερη με το όνομα “**step sizes**”. Στο πρώτο γράφημα απεικονίζονται τα σφάλματα εκπαίδευσης και ελέγχου (ως τη ρίζα του μέσου τετραγωνικού σφάλματος–RMSE) συναρτήσει του αριθμού των επαναλήψεων (epochs). Στο δεύτερο γράφημα για λόγους αναφοράς σχεδιάζουμε την καμπύλη του μεγέθους βήματος συναρτήσει των εποχών. Αυτό μας βοηθάει να προσαρμόσουμε το αρχικό μέγεθος βήματος στο βέλτιστο, βλέποντας τους αντίστοιχους ρυθμούς μείωσης και αύξησης, πάνω στο σχήμα.

Εν συνέχεια συγκρίνουμε τα αποτελέσματα του ANFIS ως προς τα επιθυμητά αποτελέσματα του μήνα για τον οποίον γίνεται η επανάληψη συναρτήσει των ετών και οπτικοποιούμε το αποτέλεσμα με κοινή παρουσίαση τους στο τρίτο γράφημα με το όνομα “**actual values and ANFIS prediction**”.

Παίρνοντας τη διαφορά (δηλαδή το σφάλμα της πρόβλεψης) των δυο προηγούμενων αποτελεσμάτων (πραγματικά – ANFIS) δημιουργούμε ένα νέο τέταρτο γράφημα των σφαλμάτων συναρτήσει των ετών με το όνομα “**prediction errors**”. Αυτό μας βοηθάει να αντιληφθούμε με οπτικό τρόπο πότε παρουσιάζεται το φαινόμενο της υπερπροσαρμογής.

Τις τελικές συναρτήσεις συμμετοχής που προέκυψαν από την πρόβλεψη των παραμέτρων τους με τη βοήθεια του ANFIS τις παρουσιάζουμε στο πέμπτο γράφημα με τίτλο “**final MFs**”.

Ακολούθησε η εξαγωγή τεσσάρων τύπων σφαλμάτων ώστε να μπορέσει να γίνει η σύγκριση μεταξύ των μοντέλων με τις διάφορες συναρτήσεις συμμετοχής, αριθμού τους, εποχών, κτλ.. Τα σφάλματα αυτά είναι τα:

- το μέσο τετραγωνικό σφάλμα (MSE):

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

- η ρίζα του μέσου τετραγωνικού σφάλματος (RMSE):

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

- το μέσο απόλυτο σφάλμα (MAE):

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

- το μέσο απόλυτο ποσοστιαίο σφάλμα (MAPE):

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

όπου A_t είναι οι πραγματικές τιμές και το F_t .

Το τελευταίο σφάλμα είναι αυτό που μας απασχόλησε και περισσότερο για τη σύγκριση μεταξύ των μοντέλων. Σε αυτό το σημείο τελειώνει το μέρος της πρόβλεψης και ακολουθεί το μέρος της σύγκρισης με το συντελεστή παλινδρόμησης (ή αλλιώς correlation = συσχέτιση), καθώς και με τα μοντέλα AR, ARMA με τη βοήθεια του System Identification Toolbox της Matlab, από τα οποία δυο τελευταία μοντέλα εξάγουμε τους ίδιους τέσσερις τύπους σφαλμάτων όπως και στο ANFIS.

Στην περίπτωση **II** χρησιμοποιήσαμε τα πρώτα 276 δεδομένα για την εκπαίδευση του συστήματος δημιουργώντας έναν πίνακα 275×2 όπου στην πρώτη στήλη είχαμε τα δεδομένα 1 – 275 και στη δεύτερη τα δεδομένα 2 – 276. Τα 49 τελευταία δεδομένα τα χρησιμοποιήσαμε για τον έλεγχο, δημιουργώντας έναν αντίστοιχο πίνακα 48×2 όπου στην πρώτη στήλη είχαμε τα δεδομένα 276 – 323 και στη δεύτερη τα δεδομένα 277 – 324.

Έπειτα δημιουργούμε ένα ANFIS, με τη βοήθεια του fuzzy logic toolbox, το οποίο θα προβλέψει τις καταλληλότερες παραμέτρους των συναρτήσεων συμμετοχής (MF parameters). Πειραματιστήκαμε ομοίως με την περίπτωση **I** και εξάγαμε τα ίδια αντίστοιχα γραφήματα.

Εν συνέχεια συγκρίνουμε τα αποτελέσματα για τον κάθε μήνα ξεχωριστά που μας ενδιαφέρει, όπως προέκυψαν από το ANFIS (με το σύνολο των 324 μηνών) ως προς τα επιθυμητά αποτελέσματα συναρτήσεων των ετών και οπτικοποιούμε το αποτέλεσμα με κοινή παρουσίαση τους στο τρίτο γράφημα με το όνομα “**actual values and ANFIS prediction**”. Επιπλέον σε αυτήν την περίπτωση **II**, συγκρίνουμε τα αποτελέσματα για όλους τους μήνες μαζί, και πάλι όπως πριν, του ANFIS προς τα επιθυμητά για το «ολικό» γράφημα “**actual values and ANFIS prediction**”.

Παίρνοντας και πάλι τη διαφορά (δηλαδή το σφάλμα της πρόβλεψης) των δυο προηγούμενων αποτελεσμάτων (πραγματικά – ANFIS) δημιουργούμε για τον κάθε μήνα ξεχωριστά που μας ενδιαφέρει, το νέο τέταρτο γράφημα των σφαλμάτων συναρτήσεων των ετών με το όνομα “**prediction errors**”. Επιπλέον και πάλι σε αυτήν την περίπτωση **II**, για όλους πάλι τους μήνες μαζί, δημιουργούμε και εδώ το «ολικό» γράφημα “**prediction errors**”. Αυτό, όπως προείπαμε, μας βοηθάει να αντιληφθούμε με οπτικό τρόπο πότε παρουσιάζεται το φαινόμενο της υπερπροσαρμογής.

Ομοίως με την **I** για τις τελικές συναρτήσεις συμμετοχής που προέκυψαν από την πρόβλεψη των παραμέτρων τους με τη βοήθεια του ANFIS τις παρουσιάζουμε στο νέο πέμπτο γράφημα με τίτλο “**final MFs**”.

Ακολούθησε η εξαγωγή των τεσσάρων γνωστών τύπων σφαλμάτων για τις περαιτέρω συγκρίσεις.

Και σε αυτό το σημείο τελειώνει το μέρος της πρόβλεψης και ακολουθεί το μέρος της σύγκρισης ομοίως για όλους μαζί τους μήνες, όπως και στην **I**, με το συντελεστή παλινδρόμησης (ή αλλιώς correlation = συσχέτιση), καθώς και με τα μοντέλα AR, ARMA, από τα οποία εξάγουμε και πάλι τους ίδιους τέσσερις τύπους σφαλμάτων όπως και στο ANFIS.

2.2 ΠΑΡΟΥΣΙΑΣΗ ΚΑΙ ΣΥΓΚΡΙΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ

Στις δόκιμες που κάναμε χρησιμοποιήσαμε τις εξής συναρτήσεις συμμετοχής στο πρόγραμμα της matlab: την τριγωνική, την τραπεζοειδή, την Γκαουσσισιανή, την διπλή Γκαουσσισιανή, τη γενικευμένη καμπανοειδή, τη σιγμοειδή, τη διαφορά σιγμοειδών, το γινόμενο σιγμοειδών, καθώς και την πι.

Δοκιμάσαμε για αριθμό συναρτήσεων που χρησιμοποιήθηκαν στις επαναλήψεις του αλγορίθμου, ίσο με δύο, τρία, ... μέχρι και οκτώ.

Επίσης κάναμε δόκιμες για 20.000 εποχές συνήθως, και για ενδεικτικούς λόγους κάποιες φορές για 180.000, για 250.000, για 1.000.000 και για 2.000.000 εποχές. Για τις εποχές από τις 180.000 και άνω η βελτίωση που παίρναμε ήταν είτε αμελητέα είτε ανύπαρκτη, ενώ ο χρόνος που απαιτούνταν για την ολοκλήρωση του αλγορίθμου αυξανόταν δραματικά. Ενδεικτικά να αναφέρουμε ότι για 2.000.000 εποχές για να βγάλει τα αποτελέσματα για ένα μόνο μήνα απαιτήθηκε ο εντυπωσιακά μεγάλος χρόνος των 7 λεπτών και 12 δευτερολέπτων, ενώ για 20.000 εποχές για να βγάλει τα αποτελέσματα και για τους 12 μήνες χρειαζόντουσαν περίπου μόνο 80 δευτερόλεπτα. Οι δόκιμες γίνονταν σε πλατφόρμα PC με επεξεργαστή Athlon XP στα 1.916MHz και διαθέσιμη μνήμη PC2700 512MB. Αφού λοιπόν δεν υπήρχε προφανές κέρδος κρατήσαμε μόνο τα αποτελέσματα που αφορούσαν τις 20.000 εποχές.

Αναφορικά με το αρχικό μέγεθος βήματος αυτό που χρησιμοποιήθηκε για όλες τις δόκιμες ήταν ίσο με 0,1. Λόγω της μεγάλης υπολογιστικής ισχύος των σημερινών CPU δεν μπορούσαν να φανούν διαφορές ως προς τον χρόνο περάτωσης των δοκιμών λόγω ταχύτερης ή βραδύτερης σύγκλισης για διάφορα αλλά μεγέθη βήματος που δοκιμάστηκαν (μικρότερα ή μεγαλύτερα από το 0,1). Έτσι δεν υπάρχει και νόημα να τα αναφέρουμε.

2.2.1 ΕΙΔΙΚΟΤΕΡΑ:

Τα περισσότερα καλύτερα αποτελέσματα τα πήραμε από την περίπτωση **I**, όμως από αυτήν πήραμε και το ένα απολύτως χειρότερο. Τα σφάλματα στην **II** ήταν γενικώς πολύ μεγαλύτερα, όμως παρουσίαζαν μικρότερες διακυμάνσεις συνεπώς είχαμε σε μικρότερο βαθμό υπερπροσαρμογή των δεδομένων σε σχέση με την **I**. Στην **I** σε ορισμένες

περιπτώσεις συναρτήσεων συμμετοχής και ειδικότερα εκεί όπου χρησιμοποιήσαμε τρεις συναρτήσεις τον αριθμό παρουσιάστηκαν εντυπωσιακά μεγάλες διακυμάνσεις σφαλμάτων μεταξύ των μηνών, συνεπώς είχαμε σε μεγαλύτερο βαθμό υπερπροσαρμογή των δεδομένων. Η υπερπροσαρμογή οφείλεται στο σχετικά περιορισμένο χρονικό διάστημα από το οποίο είχαμε δεδομένα προς επεξεργασία.

Πάντως δεν μπορούμε να πούμε ότι κάποια συνάρτηση συμμετοχής είναι συγκριτικά καλύτερη των άλλων και ότι ο βέλτιστος αριθμός συναρτήσεων είναι κάποιος συγκεκριμένος. Αναλόγως της περίπτωσης (I ή II), και της συναρτήσεως συμμετοχής ο βέλτιστος αριθμός ποικίλει αφού μπορεί να είναι είτε δυο είτε τρία είτε έξι κτλ..

2.3 ΠΙΝΑΚΕΣ

Όλα τα αποτελέσματα των σφαλμάτων MAPE των δοκιμών συγκεντρώθηκαν στους παρακάτω πίνακες. Η συμπλήρωση του για μεγαλύτερη ευκολία πραγματοποιήθηκε με τη βοήθεια του Add-in “excel link” της matlab. Αυτό το Add-in μας βοηθάει να μεταφέρουμε αυτόματα τα περιεχόμενα ενός πίνακα της matlab σε κελιά του excel. Έτσι συντομεύτηκε εντυπωσιακά και ο χρόνος συγκέντρωσης των τιμών για τις περαιτέρω συγκρίσεις, από ότι αν μεταφέραμε τα δεδομένα με «χειροκίνητο» τρόπο.

Η σύμβαση που κάναμε για την ονοματολογία των κελιών που αναφέρονται σε κάποια συγκεκριμένη συνάρτηση, αριθμό εποχών και αριθμό συναρτήσεων είναι η εξής: xxxxx ##,# για το οποίο όπου xxxxx = όνομα συνάρτησης ανάλογο με αυτό που χρησιμοποιείται και στο fuzzy logic toolbox της matlab, όπου ## = αριθμός εποχών σε χιλιάδες και # = ο αριθμός των συναρτήσεων που χρησιμοποιήθηκαν. Π.χ. gbell 20,3 αντιστοιχεί στα αποτελέσματα που πήραμε από την γενικευμένη καμπανοειδή συνάρτηση, για 20.000 εποχές και 3 τον αριθμό συναρτήσεων. Η αντιστοιχία συντομεύσεων και πλήρων ονομάτων των συναρτήσεων είναι οι παρακάτω:

- tri = τριγωνική
- trap = τραπεζοειδής
- gauss = Γκαουσιανή
- gauss2 = διπλή Γκαουσιανή
- gbell = γενικευμένη καμπανοειδής
- sig = σιγμοειδής
- dsig = διαφορά σιγμοειδών
- psig = γινόμενο σιγμοειδών
- pi = π

Στον πρώτο πίνακα 2.1 συγκεντρώθηκαν τα MAPE για τον κάθε μήνα ξεχωριστά, όπως προέκυψαν από την περίπτωση I (ξεχωριστά κάθε μήνας), και επιπλέον για λόγους σύγκρισης ανεγράφησαν και οι τιμές της δοκιμής με το μικρότερο MAPE από την περίπτωση II για κάθε μήνα, που ήταν για διαφορά σιγμοειδών με δυο συναρτήσεις και είκοσι χιλιάδες εποχές (dsig 20,2b όπου βάλαμε το b για να την διαχωρίζουμε από την αντίστοιχη dsig

20,2 της περίπτωσης Ι). Μπήκαν ακόμα και τα ελάχιστα MAPE που προκύπτουν για κάθε μήνα από το AR ή το ARMA.

Για την ευκολότερη ανάγνωση του μεγάλου πίνακα κάναμε τις εξής τυποποιήσεις:

- στα κελιά χωρίς χρώμα βλέπουμε τις χειρότερες τιμές σφαλμάτων (τις μεγαλύτερες)
- στα κελιά με αχνό χρώμα βλέπουμε τις «συνήθειες» τιμές σφαλμάτων
- στα κελιά με σκούρο χρώμα βλέπουμε τις καλύτερες τιμές σφαλμάτων (τις μικρότερες), και επιπλέον η γραμματοσειρά είναι εντονότερη.

Πίνακας 2.1.

MAPE anfis	1	2	3	4	5	6	7	8	9	10	11	12
min(AR,ARMA)	4,746	4,772	4,993	5,700	4,891	2,462	2,943	2,649	2,401	4,088	5,808	4,918
gbell 20,2	4,170	4,124	4,608	5,587	23,864	17,199	2,647	2,130	4,642	6,156	5,913	4,712
dsig 20,2	4,664	4,124	4,231	5,175	4,831	9,372	2,764	2,138	8,183	7,047	6,316	4,925
tri 20,2	6,340	6,483	6,522	11,256	15,514	3,593	2,946	3,097	3,028	8,670	8,722	6,327
gauss 20,2	8,136	4,124	4,634	5,797	15,708	3,510	2,819	2,178	1,697	7,060	6,118	4,316
gauss2 20,2	4,670	4,124	4,634	5,797	16,522	5,728	2,764	2,150	2,369	5,240	6,091	5,664
psig 20,2	4,644	4,124	4,439	5,868	4,831	8,278	2,764	2,136	2,230	6,662	6,316	4,925
pi 20,2	4,649	4,125	4,634	5,552	5,182	2,398	2,764	1,981	2,251	5,248	6,195	5,423
trap ~20,2	4,643	4,149	4,654	5,983	5,373	3,187	2,835	2,145	1,860	5,030	6,172	5,031
trap ~20,3()	4,475	5,665	4,822	5,478	13,795	3,536	3,147		1,889	7,135	6,466	5,784
tri 20,3	7,481	6,336	33,607	36,537	7,446	5,786	5,596	3,786	1,624	6,628	5,457	8,055
gauss 20,3	6,409	6,450	4,093	4,866	4,991	3,196	3,011	2,327	1,945	6,469	6,431	4,943
gauss2 20,3	7,290	5,375	4,788	5,152	6,763	2,814	3,098	2,494	1,820	6,191	7,354	5,982
gbell 20,3	6,696	19,371	13,466	9,910	3,501	9,036	2,677	1,929	1,840	5,275	6,476	4,363
dsig 20,3	4,662	44,252	7,718	8,702	5,890	4,585	2,633	2,107	1,539	6,970	5,108	46,758
psig 20,3	58,824	969,241	16,855	6,045	5,523	3,640	2,486	20,157	2,579	8,087	8,453	5,587
pi 20,3	4,160	5,377	4,788	5,505	14,202	3,531	3,098		1,721	6,971	6,698	5,425
dsig 20,2b	4,171	4,525	5,020	7,487	7,735	6,256	5,657	3,334	1,915	6,110	9,557	14,211
minimum mf	4,160	4,124	4,093	4,866	3,501	2,398	2,486	1,929	1,539	5,030	5,108	4,316
check	1	1	1	1	1	1	1	1	1	0	1	1

Τέλος στην προτελευταία γραμμή βάλαμε το ελάχιστο MAPE που προέκυπτε από όλες τις δοκιμές ANFIS για τον κάθε μήνα. Στην τελευταία γραμμή με το όνομα check συγκρίναμε σε ποιους μήνες μπορούμε να πάρουμε το ελάχιστο MAPE από οποιαδήποτε δοκιμή ANFIS ή από τα AR ή ARMA. Όπου η τιμή είναι 1 σημαίνει ότι υπερτερεί κάποιο ANFIS και όπου είναι 0 υπερτερεί το AR ή το ARMA. Βλέπουμε ότι οι στατιστικές μέθοδοι AR ή ARMA υπερτερούν μονάχα σε έναν μήνα του έτους, τον Οκτώβριο.

Στους άλλους δυο μικρούς πίνακες βλέπουμε ταξινομημένο με σειρά από το μικρότερο προς το μεγαλύτερο MAPE, τον μέσο όρο του κάθε δωδεκάμηνου (1974 – 2000) για την περίπτωση Ι στον πίνακα 2.2, και στον πίνακα 2.3 το MAPE για το σύνολο της περιόδου 1974 – 2000 (εδώ στην σύμβαση της ονοματολογίας που αναφέραμε και πριν, υπάρχει επιπλέον το γράμμα k για να συμβολίζει τις χιλιάδες: xxxx ##k, # π.χ. gbell 20k,2 για 20.000 εποχές. Και αυτό διότι χρειάστηκε να γίνει μια δοκιμή με 142 εποχές και συμβολίστηκε με pi 142,2 για να μη γίνει κάποιο μπέρδεμα και ληφθεί το 142 στην περίπτωση αυτή ως χιλιάδες).

Πίνακας 2.2

Μήνες 1-12	MAPE_anfis
minimum mf	3,6291
pi 20,2	4,2002
trap ~20,2	4,2552
gauss 20,3	4,5943
psig 20,2	4,7680
gauss2 20,3	4,9267
dsig 20,2	5,3141
gauss2 20,2	5,4794
gauss 20,2	5,5082
pi 20,3	5,5888
trap ~20,3()	5,6539
dsig 20,2b	6,3315
tri 20,2	6,8748
gbell 20,3	7,0450
gbell 20,2	7,1461
tri 20,3	10,6948
dsig 20,3	11,7436
psig 20,3	92,2898

Πίνακας 2.3

Έτη 1974-2000	MAPE_anfis
dsig 20k,2	6,2862
dsig 250k,2	6,2862
psig 20k,2	6,2915
pi 142,2	6,3124
trap 20k,2	6,3444
dsig 20k,3	6,4153
psig 20k,3	6,5747
gauss 20k,6	6,6331
gauss 20k,5	6,8947
psig 20k,4	6,9315
gauss2 20k,6	7,0861
gauss 20k,4	7,1484
gauss2 20k,5	7,2170
gbell 20k,3	7,2251
tri 20k,2	7,2344
gbell 20k,4	7,2928
gauss 20k,7	7,6712
gbell 20k,5	8,2799
tri 20k,3	8,8782
gbell 20k,6	9,6667
gbell 20k,2	10,4023
gauss2 20k,7	10,8310
gauss 20k,8	11,5490

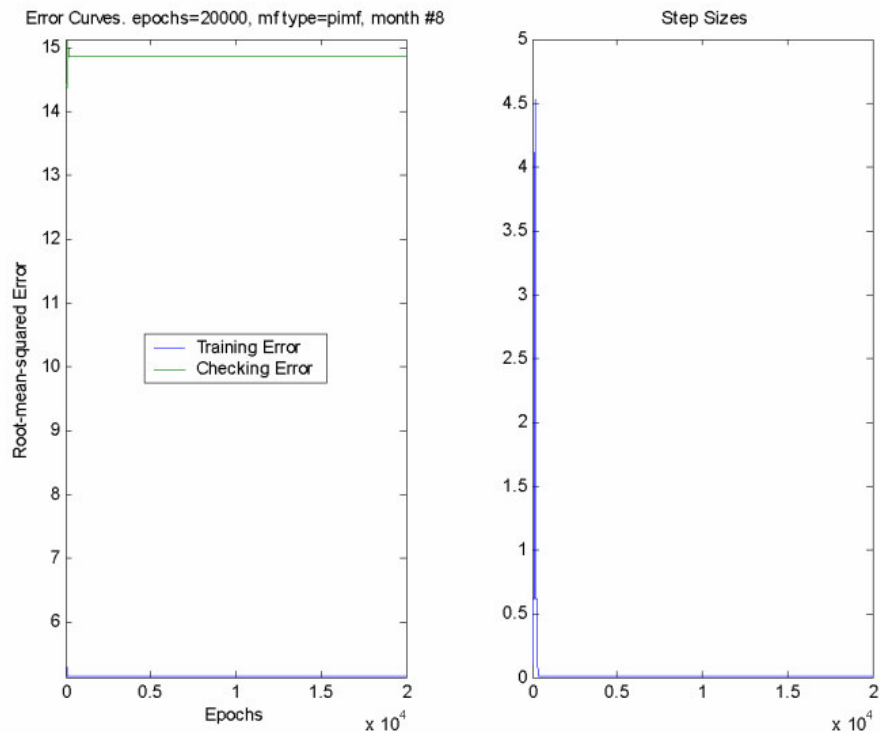
Πάντως δεν φαίνεται να υπερτερεί κάποια επιλογή σε πάρα πολύ σημαντικό βαθμό έναντι των άλλων. Υπάρχουν περιπτώσεις που, όπως είδαμε, δίνουν καλά αποτελέσματα σχεδόν σε όλες, χωρίς να είναι τα κορυφαία, άλλες περιπτώσεις όπου δίνουν εξαιρετικά καλά αποτελέσματα σε κάποιους μήνες και λιγότερα καλά σε άλλους και τέλος περιπτώσεις που θα ήταν καλό να αποφεύγονται, και τρανταχτό παράδειγμα είναι το γινόμενο σιγμοειδών με 3 τον αριθμό συναρτήσεων, το οποίο σε εκπαίδευση 20.000 εποχών έδωσε κατά μέσο όρο MAPE 92,29%. Όπως παρατηρούμε και από τους πίνακες σχεδόν σε κάθε μήνα έχουμε και κάποιο άλλο μοντέλο που να υπερτερεί.

Ενδεχομένως στο μέλλον, όταν θα έχουμε ακόμα περισσότερη επεξεργαστική ισχύ, που σημαίνει ότι **ο χρόνος περάτωσης των δοκιμών θα είναι αμελητέος σε σχέση με το παρόν**, θα μπορούμε με έναν αντίστοιχο αλγόριθμο να κάνουμε δοκιμές ταυτοχρόνως για όσες περισσότερες συναρτήσεις συμμετοχής, αριθμού τους και εποχών θέλουμε και να επιλέγουμε έτσι ένα πολύ περισσότερο πολύπλοκο μοντέλο το οποίο θα έχει διαφορετικές παραμέτρους για κάθε μήνα, αντί των ίδιων για όλους. Έτσι θα μπορούμε να εκμεταλλευτούμε στο έπακρον όλες τις δυνατότητες που μας προσφέρει το ANFIS.

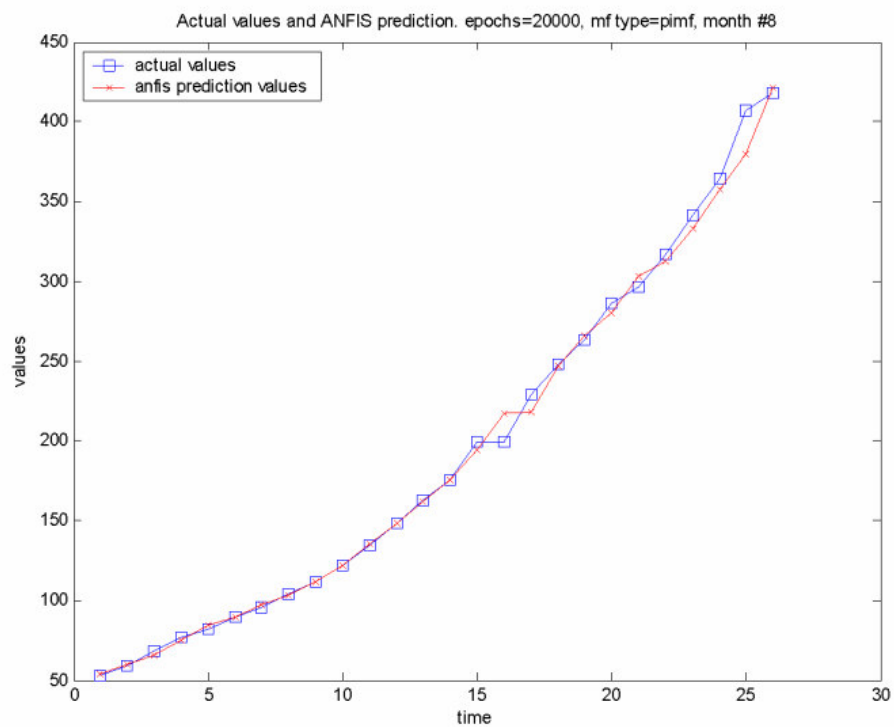
2.4 ΓΡΑΦΗΜΑΤΑ

Παρακάτω απεικονίζονται ενδεικτικά τα γραφήματα από τα τελικά αποτελέσματα για τις διάφορες δόκιμες που πραγματοποιήσαμε.

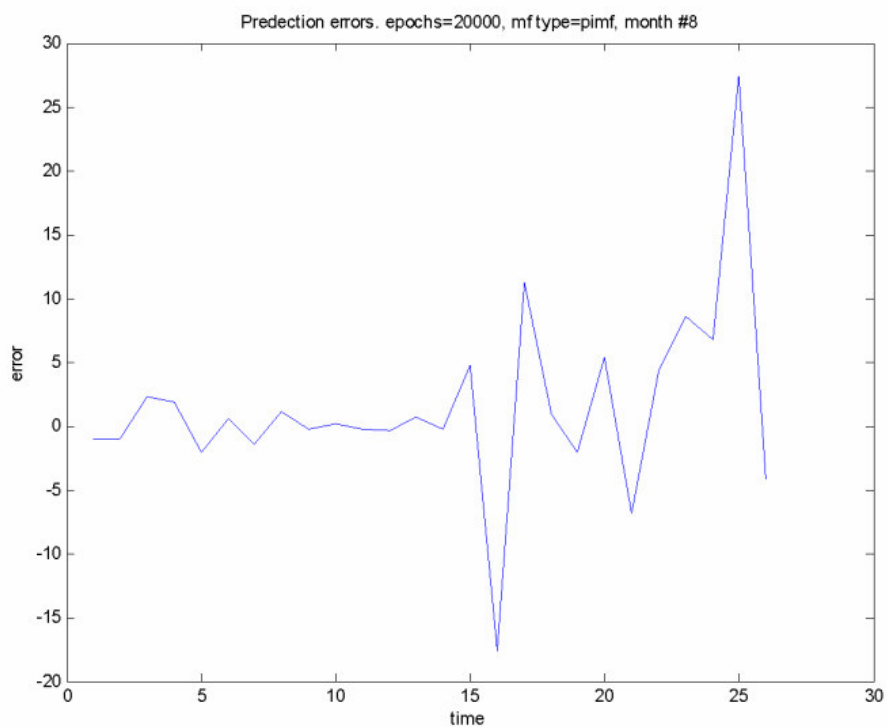
Παραδείγματος χάριν για το μήνα Αύγουστο έχουμε τα εξής γραφήματα που πάρθηκαν μέσω της περίπτωσης **I** με τη συνάρτηση συμμετοχής π για 20.000 εποχές και 2 τον αριθμό συναρτήσεων.



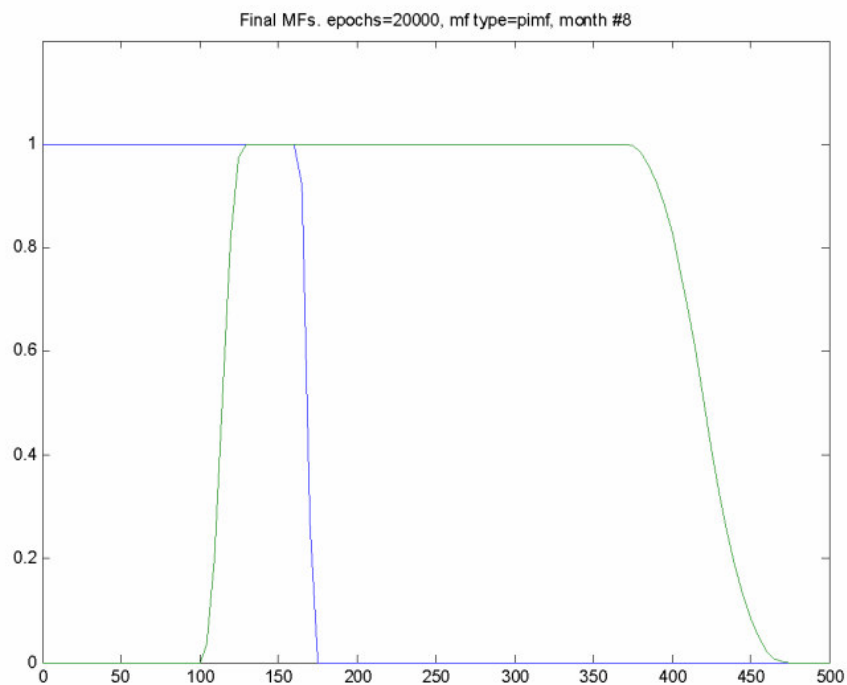
Γράφημα (1). 1) Καμπύλες σφαλμάτων και 2) καμπύλες μεγέθους βήματος της δοκιμής π 20,2.



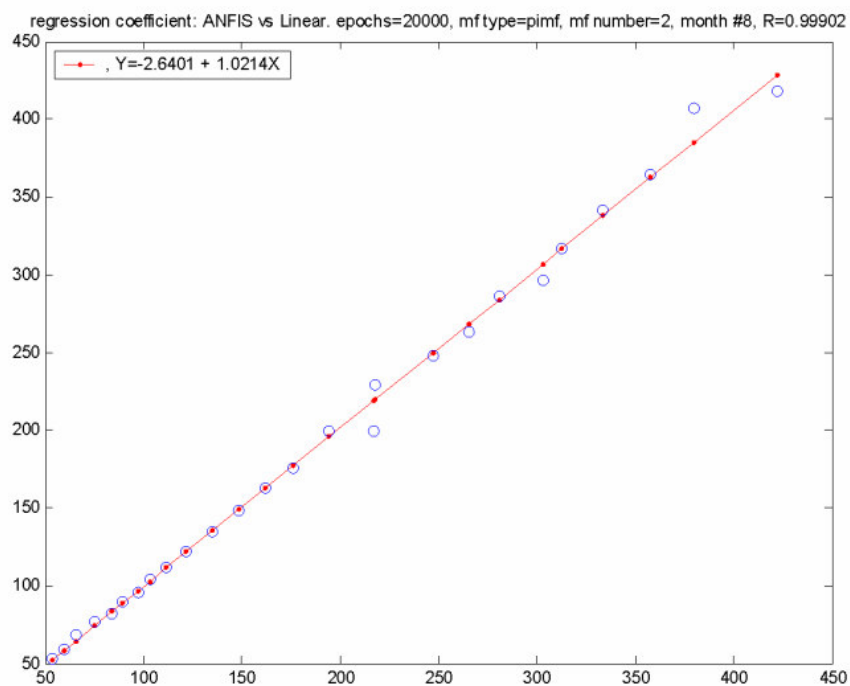
Γράφημα (2). Σύγκριση προβλέψεων του ANFIS έναντι των επιθυμητών τιμών της δοκιμής ri 20,2.



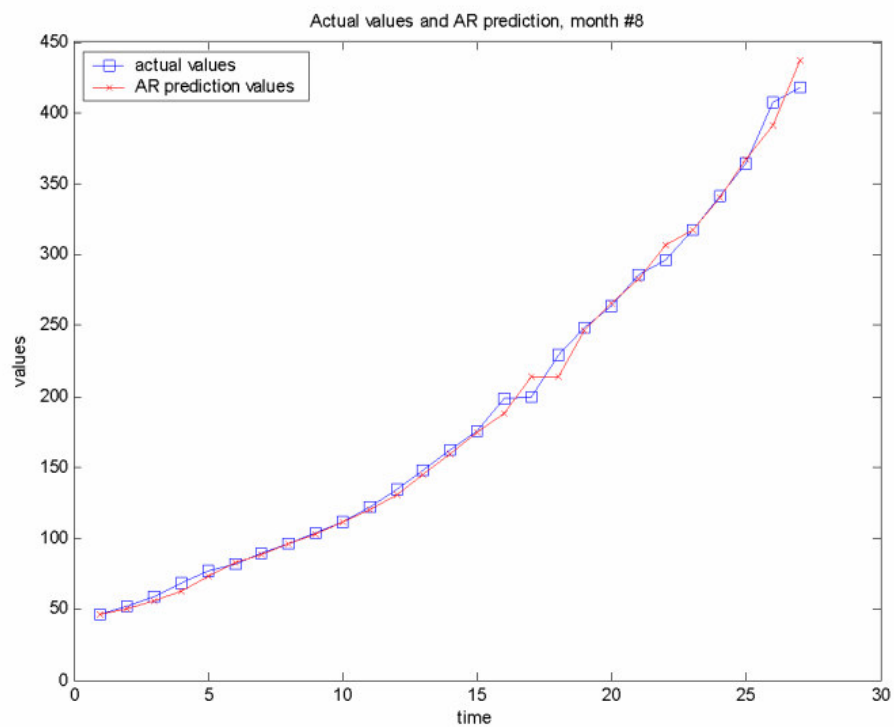
Γράφημα (3). Σφάλματα πρόβλεψης του ANFIS της δοκιμής ri 20,2.



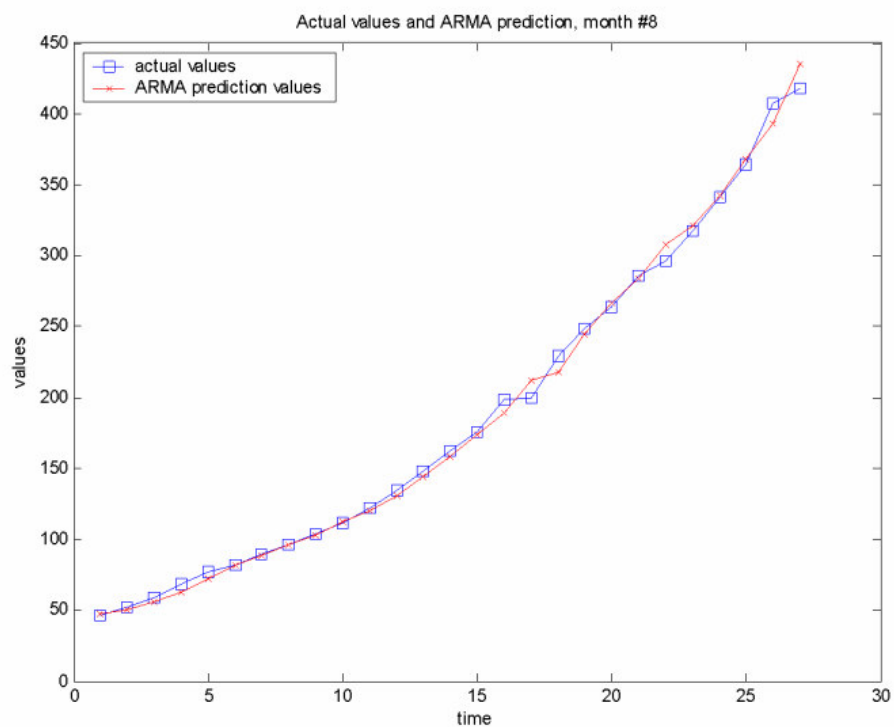
Γράφημα (4). Τελικές συναρτήσεις συμμετοχής της δοκιμής $\rho_i 20,2$.



Γράφημα (5). Σύγκριση προβλέψεων του ANFIS έναντι των επιθυμητών τιμών της δοκιμής $\rho_i 20,2$ και ο συντελεστής συσχέτισης τους (regression coefficient). [5]



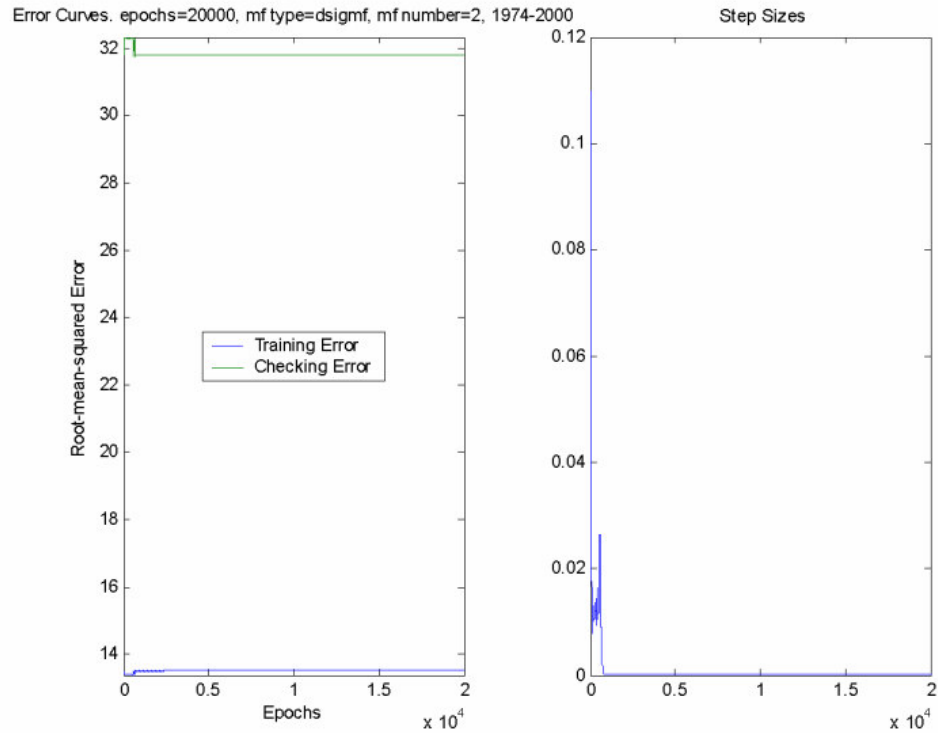
Γράφημα (6). Σύγκριση προβλέψεων του μοντέλου AR έναντι των επιθυμητών τιμών.



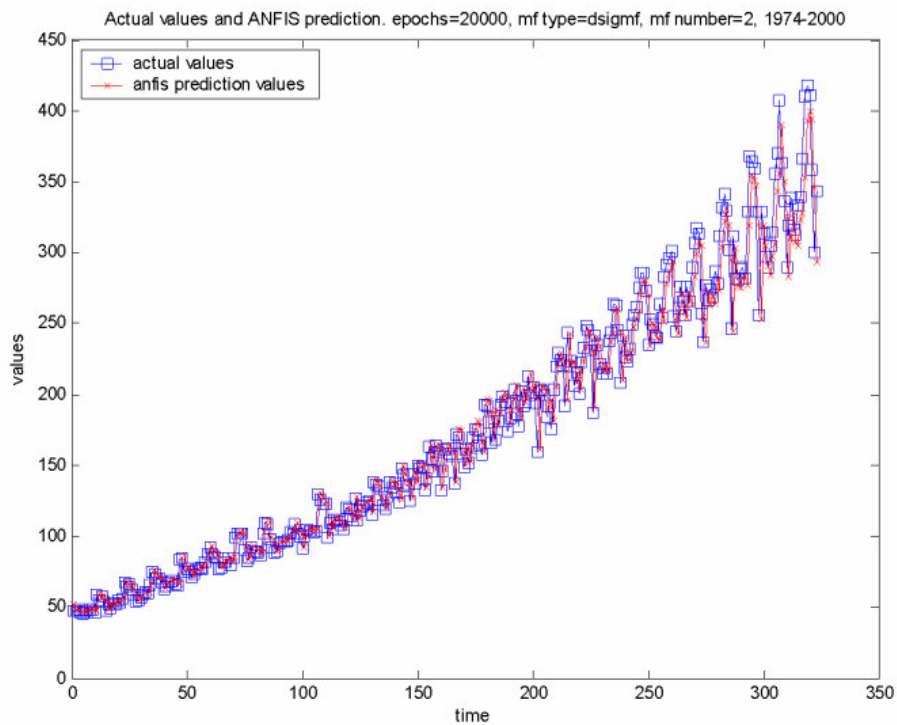
Γράφημα (7). Σύγκριση προβλέψεων του μοντέλου ARMA έναντι των επιθυμητών τιμών.

Αντίστοιχα γραφήματα πήραμε και για τους υπόλοιπους μήνες, με άλλους τύπους και αριθμό συναρτήσεων συμμετοχής και εποχών.

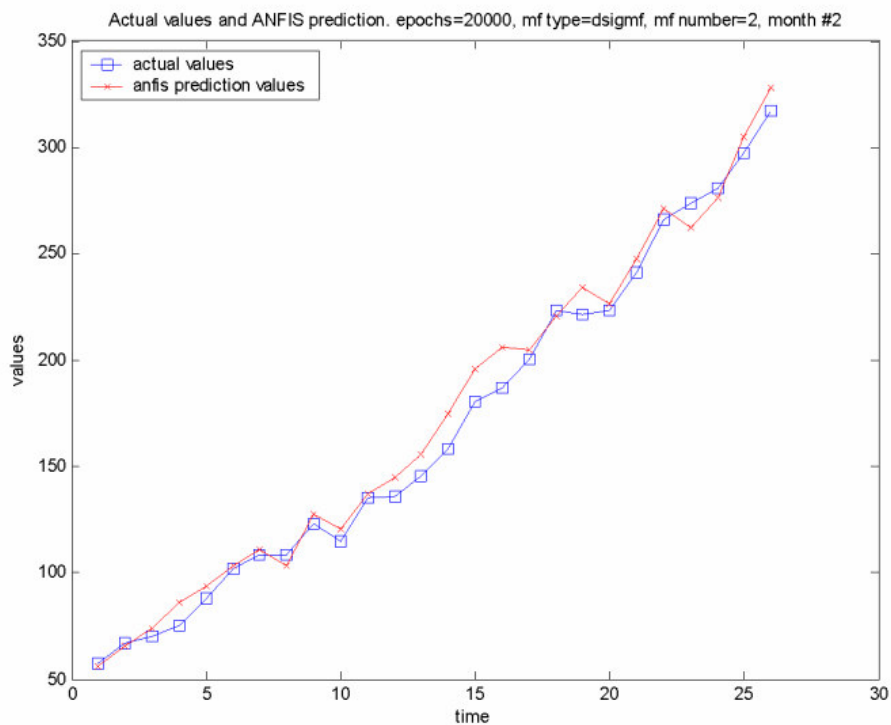
Μέσω της περίπτωσης **II**, με δυο διαφορές σιγμοειδών και 20.000 εποχές πήραμε τα παρακάτω γραφήματα:



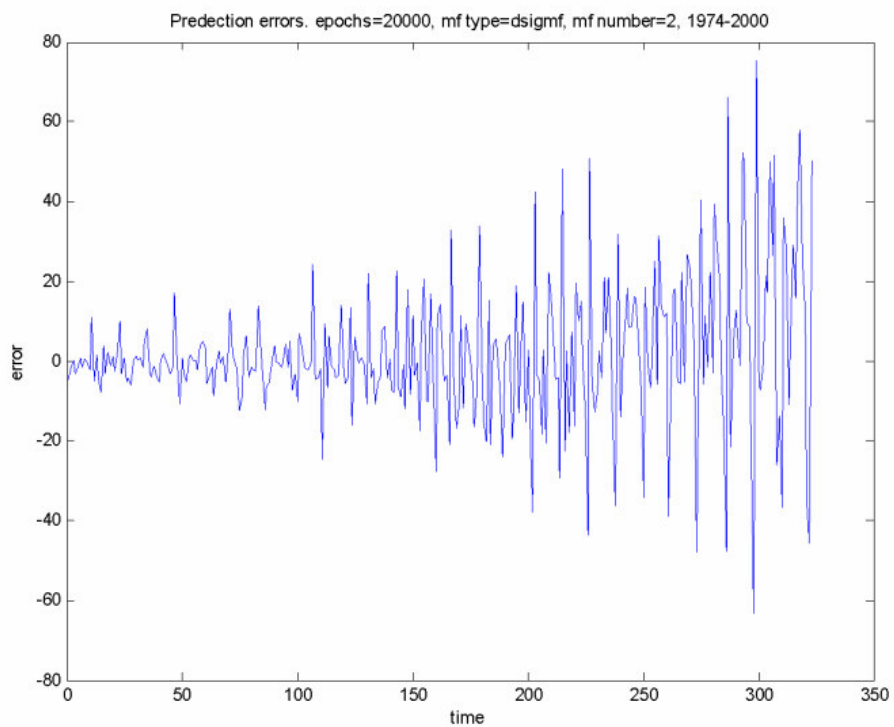
Γράφημα (8). 1) Καμπύλες σφαλμάτων και 2) καμπύλες μεγέθους βήματος της δοκιμής dsig 20,2 για τη χρονική περίοδο συνολικά 1974 – 2000.



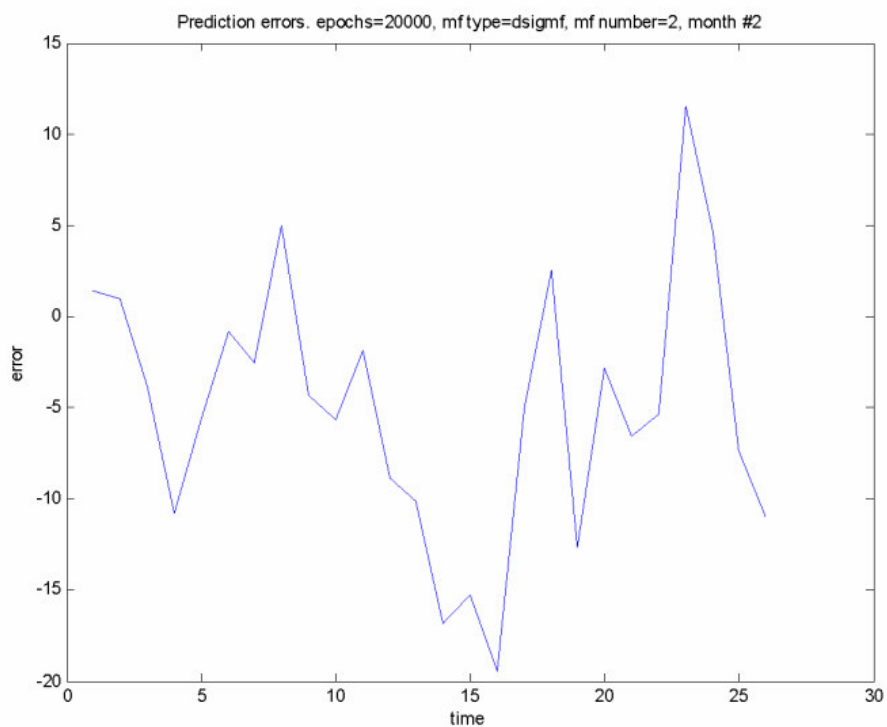
Γράφημα (9). Σύγκριση προβλέψεων του ANFIS έναντι των επιθυμητών τιμών της δοκιμής dsig 20,2 για τη χρονική περίοδο συνολικά 1974 – 2000.



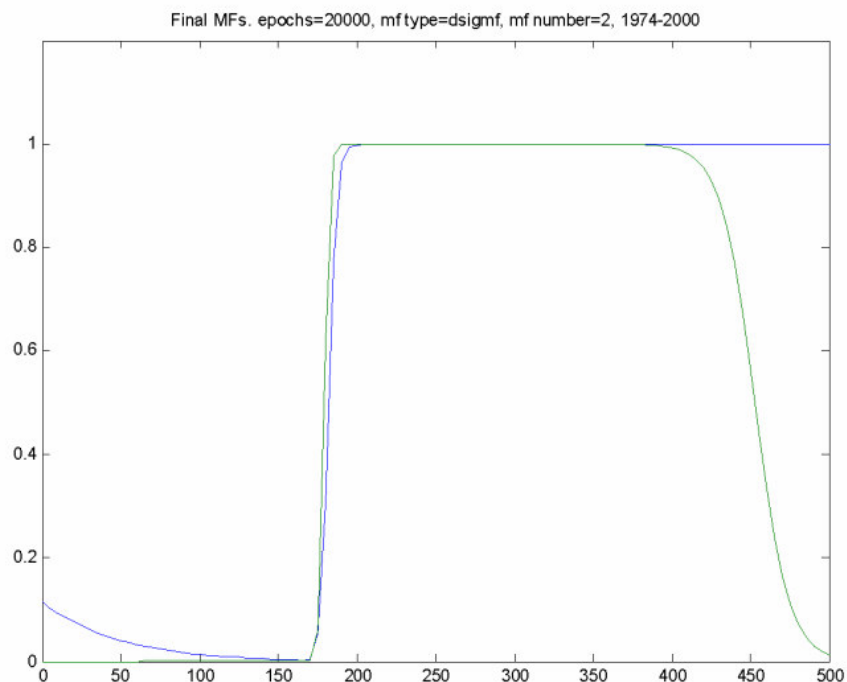
Γράφημα (10). Σύγκριση προβλέψεων του ANFIS έναντι των επιθυμητών τιμών της δοκιμής dsig 20,2 για το μήνα Φεβρουάριο.



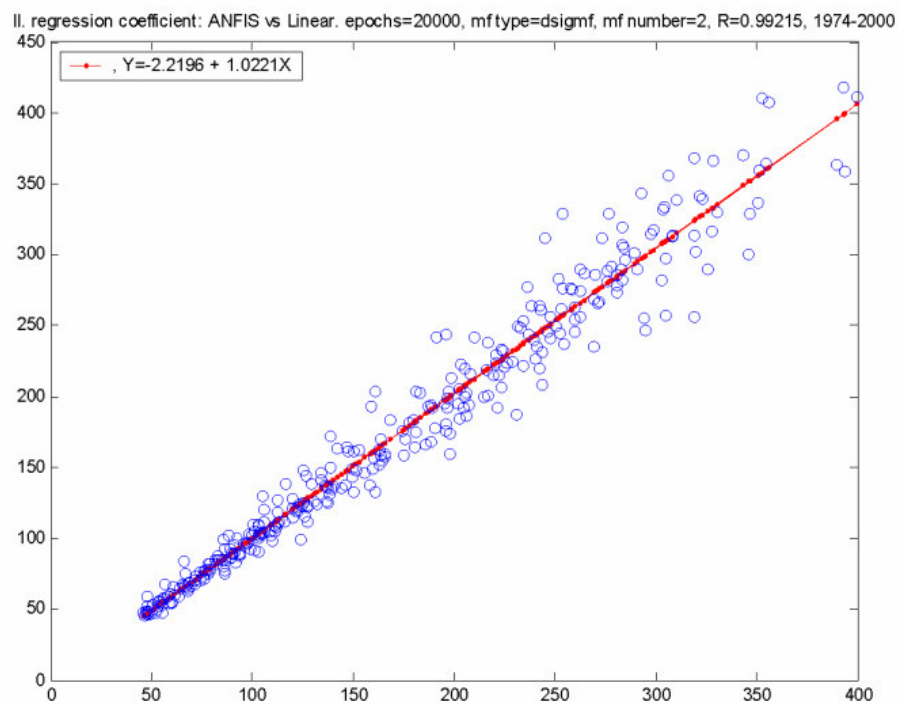
Γράφημα (11). Σφάλματα πρόβλεψης του ANFIS της δοκιμής dsig 20,2 για τη χρονική περίοδο συνολικά 1974 – 2000.



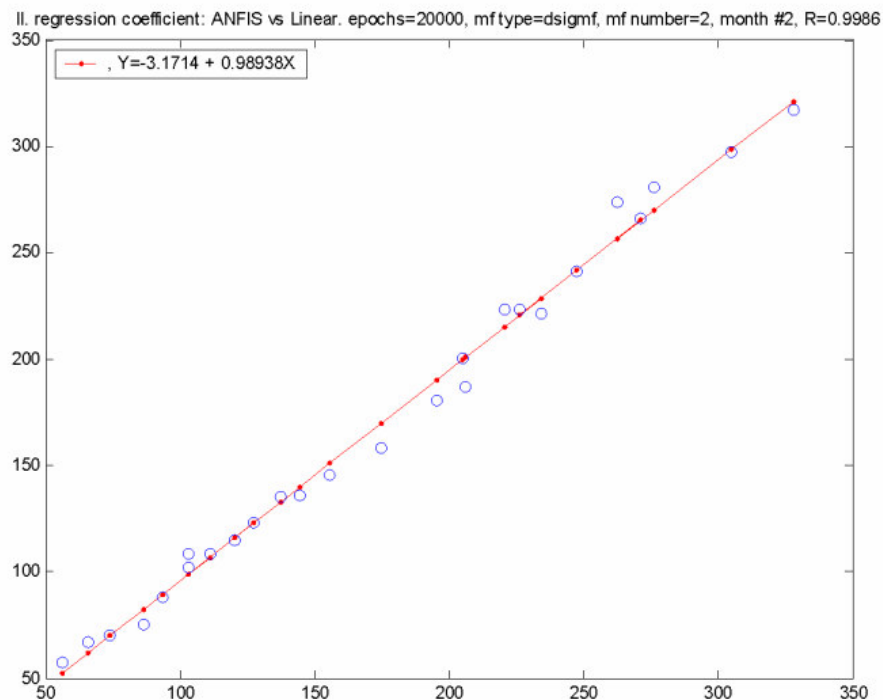
Γράφημα (12). Σφάλματα πρόβλεψης του ANFIS της δοκιμής dsig 20,2 για το μήνα Φεβρουάριο.



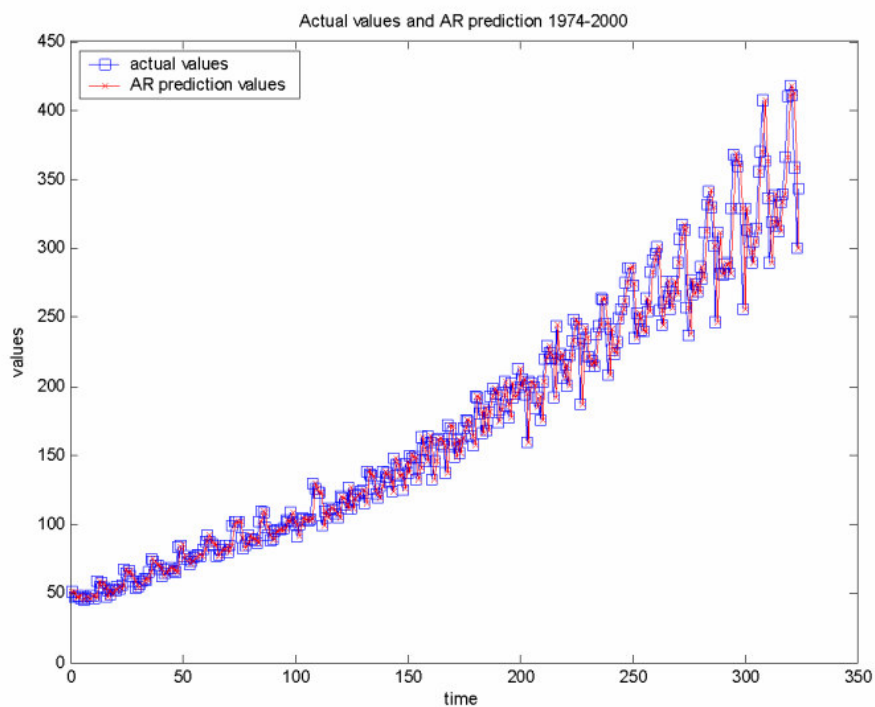
Γράφημα (13). Τελικές συναρτήσεις συμμετοχής της δοκιμής dsig 20,2 για τη χρονική περίοδο συνολικά 1974 – 2000.



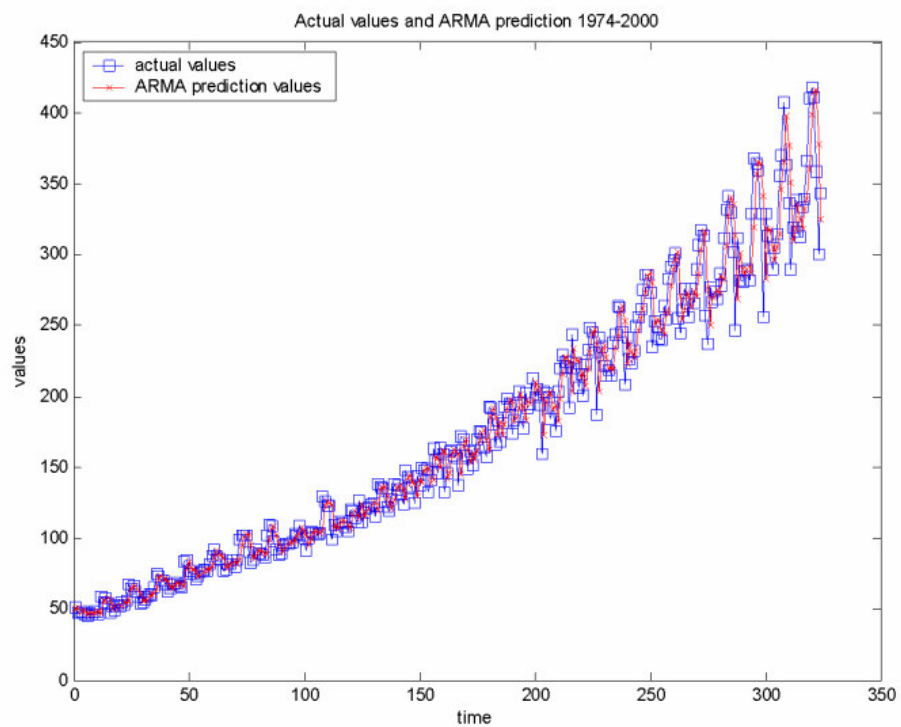
Γράφημα (14). Σύγκριση προβλέψεων του ANFIS έναντι των επιθυμητών τιμών της δοκιμής dsig 20,2 και ο συντελεστής συσχέτισης τους (regression coefficient) για τη χρονική περίοδο συνολικά 1974 – 2000. [5]



Γράφημα (15). Σύγκριση προβλέψεων του ANFIS έναντι των επιθυμητών τιμών της δοκιμής dsig 20,2 και ο συντελεστής συσχέτισης τους (regression coefficient) για το μήνα Φεβρουάριο. [5]



Γράφημα (16). Σύγκριση προβλέψεων του μοντέλου AR έναντι των επιθυμητών τιμών για τη χρονική περίοδο συνολικά 1974 – 2000.



Γράφημα (17). Σύγκριση προβλέψεων του μοντέλου ARMA έναντι των επιθυμητών τιμών για τη χρονική περίοδο συνολικά 1974 – 2000.

ΕΠΙΛΟΓΟΣ

Κλείνοντας θα μπορούσαμε να ανακεφαλαιώσουμε στα εξής σημαντικά σημεία, που καλό είναι να τα λαμβάνει υπόψη του οποίος θα επιθυμούσε να ασχοληθεί με ένα νεύρο-ασαφές σύστημα.

Ένα ANFIS για να μπορέσει να μελετηθεί, απαιτεί από τον χρήστη να έχει ένα ανεπτυγμένο γνωστικό υπόβαθρο, το οποίο θα περιλαμβάνει τις στατιστικές μεθόδους, την αριθμητική ανάλυση, τα αλγεβρικά σύστημα και τους πίνακες. Επίσης να γνωρίζει και τι είναι τα νευρωνικά δίκτυα και πως μπορούν αυτά να εκπαιδευτούν και να γίνουν προσαρμοστικά.

Έτσι ξεκινώντας από τα θετικά, μπορούμε να πούμε ότι αυτή η πολυπλοκότητα του, του δίνει την δύναμη να μπορεί να προσαρμόζεται καλύτερα στις διάφορες απαιτήσεις των όποιων προβλημάτων. Δίνει έτσι μικρότερα σφάλματα, άρα και καλύτερα αποτελέσματα. Η εκπαίδευση ενός τέτοιου συστήματος μπορεί να γίνει αρκετά εύκολα, καθώς μπορούμε άμεσα να αντιληφθούμε τυχόντα προβλήματα και δυσκολίες, μέσα από γραφήματα που εξάγουμε από αυτήν την μέθοδο. Οπτική ανάγνωση των αποτελεσμάτων που μας παρέχουν αρκεί τις πολλές φορές, να μας διδάξει προς τα που θα πρέπει να βρίσκεται η επόμενη κίνηση μας ώστε να βαδίζουμε κάθε φορά και σε ένα καλύτερο μονοπάτι.

Συνεχίζοντας στα αρνητικά, θα λέγαμε ότι πιο σωστό θα ήταν να χαρακτηριζτούν ως «παρενέργειες» των θετικών του στοιχείων. Ενώ η πολυπλοκότητα του σημαίνει και καλύτερες δυνατότητες προσαρμογής, αυτόματα σημαίνει και μεγαλύτερες ανάγκες για επεξεργαστική ισχύ. Το γεγονός αυτό, σαν σε ντόμινο, οδηγεί την ίδια στιγμή σε άλλες δυο παρενέργειες. Η πρώτη μας λέει ότι χρειάζεται περισσότερο χρόνο για να ολοκληρώσει την εξαγωγή αποτελεσμάτων από ένα μοντέλο, σε σχέση με ένα απλούστερο στατιστικό. Η δεύτερη, ως ένα ξεχωριστό παρακλάδι, απαιτεί μεγαλύτερες οικονομικές επενδύσεις σε τεχνολογικό εξοπλισμό, αφού ως γνωστόν, μεγαλύτερη επεξεργαστική ισχύς = νέες τεχνολογίες = ακριβότερο κόστος αγοράς, αν θέλουμε να εξακολουθούμε να έχουμε το πλεονέκτημα έναντι των αντιπάλων και του ιδίου του χρόνου. Ένα τελευταίο πρόβλημα είναι και το ότι χρειαζόμαστε πολλά δεδομένα για την εκπαίδευση του συστήματος. Σε περίπτωση που αυτά είναι λίγα ή δεν αντιπροσωπεύουν σωστά την περίπτωση που μας ενδιαφέρει, παραμονεύει ο κίνδυνος της υπερπροσαρμογής των δεδομένων, με τις γνωστές ανεπιθύμητες ιδιότητες της.

Γεγονός είναι όμως, ότι τα οποία αρνητικά στοιχεία δεν πρέπει να μας απογοητεύουν και να σκιάζουν τα θετικά. Τα τελευταία είναι σημαντικότερα αφού τα οφέλη που κερδίζουμε είναι ένας πειρασμός από τον οποίο μακροπρόθεσμα θα έχουμε κέρδη. Είναι σίγουρο δε, ότι δεν θα απογοητεύουμε, χρησιμοποιώντας τα, και σίγουρα στο εγγύς μέλλον τα νεύρο-ασαφή συστήματα θα είναι ο κανόνας, πάρα η εξαίρεση στις επιλογές των αναλυτών δεδομένων.

Βιβλιογραφία:

- [1]. **J.-S. R. JANG, C.-T. SUN, E. MIZUTANI,**
*Neuro-Fuzzy and Soft Computing, A Computational Approach
to Learning and Machine Intelligence,*
εκδόσεις Prentice Hall, 1997
- [2]. **ΣΟΦΙΑ ΔΗΜΕΛΗ,**
Σύγχρονες Μέθοδοι Ανάλυσης Χρονολογικών Σειρών,
εκδόσεις Κριτική, Σεπτέμβριος 2003
- [3]. *Fuzzy Logic Toolbox For Use with MATLAB® User's
Guide Version 2,*
εκδόσεις The MathWorks, Ιούλιος 20002
- [4]. www.wikipedia.org
- [5]. **H.F. KWOK, D.A. LINKENS, M. MAHFOUF, G.H.
MILLS,**
*Rule-bas derivation for intensive care ventilator control using
ANFIS,*
εκδόσεις Elsevier, 2002