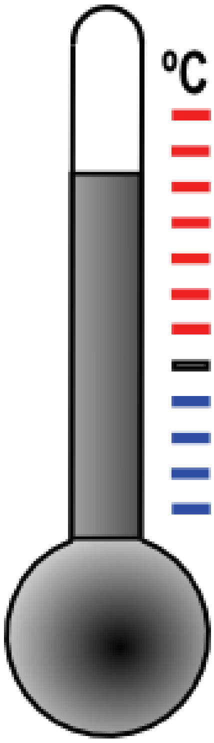
ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΑΡΑΓΩΓΗΣ ΚΑΙ ΔΙΟΙΚΗΣΗΣ

ΕΡΓΑΣΤΗΡΙΟ ΑΝΑΛΥΣΗΣ ΔΕΔΟΜΕΝΩΝ & ΠΡΟΒΛΕΨΗΣ

ΤΙΤΛΟΣ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ:

“ΠΡΟΒΛΕΨΗ ΕΞΩΤΕΡΙΚΗΣ ΘΕΡΜΟΚΡΑΣΙΑΣ ΜΕ ΧΡΗΣΗ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ”



Ζερβουδάκης Ιωάννης

Επιβλέπων Καθηγητής Ατσαλάκης Γεώργιος

Χανιά

2010

Η εργασία αυτή είναι αφιερωμένη στη μνήμη των γονιών μου.

Με το πέρας αυτής της εργασίας ολοκληρώνεται η πορεία σπουδών μου στο Πολυτεχνείο Κρήτης. Θέλω να ευχαριστήσω την οικογένεια μου για την υπομονή που έδειξαν όλα αυτά τα χρόνια. Επίσης θέλω να ευχαριστήσω τον επιβλέποντα καθηγητή μου κύριο Ατσαλάκη Γεώργιο για την άψογή συνεργασία μας , την παροχή των δεδομένων, τις συμβουλές που μου έδωσε και τον χρόνο που μου αφιέρωσε, θέλω επίσης να ευχαριστήσω τους φίλους μου για τη βοήθεια που μου έδωσαν. Τέλος θέλω να ευχαριστήσω τη σύντροφο μου Ευρυδίκη Παπαχρήστου για την στήριξη και τη συμπαράσταση που μου έδειξε σε αυτή μου την προσπάθεια.

Περιεχόμενα

ΕΙΣΑΓΩΓΗ

Η ακριβής πρόβλεψη της θερμοκρασίας αποτελεί ένα απαραίτητο συστατικό στην λήψη αποφάσεων το οποίο έχει να κάνει με πολλούς τομείς όπως η γεωργική παραγωγή, η διαχείριση ενέργειας, η διαχείριση κατασκευών (μονώσεις κτιρίων ,κλιματισμός κουφώματα) (Ruano, Crispim, Conceicao, & Lucio, 2006) ,η διαχείριση υδάτινων πόρων και η υγεία τόσο των ανθρώπων , όσο και των οικόσιτων ζώων. Ακραίες θερμοκρασίες δύναται να προκαλέσουν ζημιές στη σοδειές, οι οποίες θα οδηγήσουν τόσο σε μείωση της απόδοσης της παραγωγής, όσο και μείωση της ίδιας της παραγωγής, όπως και στη δημιουργία επικίνδυνων συνθηκών εργασίας για αυτούς που εργάζονται εκτεθειμένοι στα στοιχεία της φύσης. Περίπου το 50% της σοδειάς των ροδάκινων, αλλά και το 87% της σοδειάς των μούρων στη πολιτεία της Georgia των ΗΠΑ καταστράφηκε εξαιτίας καταστροφικών παγετών στις αρχές του Απριλίου του έτους 2007 (Chevalier, Hoogenboom, McClendon, & Paz, 2010) . Επίσης στη χώρα μας μεγάλο μέρος της παραγωγής χάνεται ή ζημιώνεται από απρόσμενες χαλαζοπτώσεις και παγετούς ,όπως και μεγάλο μέρος της γεωργικής παραγωγής μπορεί να καταστραφεί λόγο καύσωνα. Συγκεκριμένα, στην περιοχή του νομού Χανίων το καλοκαίρι του 2010 λόγο κακοδιαχείρισης των υδάτινων πόρων μεγάλο μέρος των ελαιώνων παρέμειναν χωρίς νερό για αρκετές ημέρες με αποτέλεσμα την πρόκληση ζημιών στον καρπό. Οι ακραίες θερμοκρασίες έχουν δυσμενείς επιπτώσεις και στην κτηνοτροφία. Η θερμοπληξία σε αγελάδες γαλακτοπαραγωγής μπορεί να προκαλέσει μείωση στην παραγωγή γάλακτος (ακόμα και 30%), όπως επίσης μπορεί να επιφέρει και θάνατο στα κοπάδια. Κρίνεται απαραίτητο να αναφερθεί ότι περίπου κάθε χρόνο στις ΗΠΑ καταγράφονται περίπου 400 θάνατοι. Το καλοκαίρι του 2010 από τους ασυνήθιστους καύσωνες στη Ρωσία καταγράφηκαν χιλιάδες θάνατοι, γεγονός το οποίο συνέβη και σε ολόκληρη την Ευρώπη. Η πρόβλεψη της θερμοκρασίας του αέρα αποτελεί ένα σημαντικό παράγοντα στην παροχή μιας πρώτης προειδοποίησης τόσο για παγετούς , όσο και για πολύ υψηλές θερμοκρασίες. Με επαρκή προειδοποίηση , οι απώλειες στις σοδειές λόγω παγετού μπορεί να μειωθούν κάνοντας χρήση μέτρων πρόληψης όπως θερμαντήρες οπωρώνων και άρδευσης. Ανάλογα μέτρα πρόληψης και προστασίας μπορούν να ληφθούν και για καταστάσεις πολύ υψηλών θερμοκρασιών με στόχο την προστασία των εργαζομένων, των καλλιεργειών, αλλά και των οικόσιτων ζώων.

Τα νευρωνικά δίκτυα είναι ευρέως χρησιμοποιούμενα στον τομέα της λήψης και της υποστήριξης αποφάσεων, αλλά και στον τομέα των προβλέψεων (Li & Cheng, 2007) (Aliev, Fazlollahi, Aliev, & Guirimov, 2006) (Huarng, 2001) (Chaudhuri, 2009) (Fdez-riverola, Corchado, & Torres, 2002). Έχουν γίνει εκτεταμένες εφαρμογές νευρωνικών δικτύων στην πρόβλεψη τιμών μετοχών, χρηματιστηριακών δεικτών, τιμών συναλλάγματος, εγγραφές σε πανεπιστημιακά ιδρύματα και άλλα (Huarng, 2001) (Li & Cheng, 2007) (Κοσματόπουλος & Αρβανιτάκη, 2005) . Επίσης έχουν εκπονηθεί εργασίες με θέμα προβλέψεις με τη βοήθεια νευρωνικών δικτύων (Ατσαλάκης & Λαμπουσάκης,2009), (Ατσαλάκης & Μινουδάκη, 2007), (Ατσαλάκης & Σταματάκης,2008) . Στόχος της παρούσας εργασίας είναι να διερευνήσει την ικανότητα ενός νεύρο-ασαφούς μοντέλου (ANFIS) στο να δώσει ακριβείς προβλέψεις για το πρόβλημα της πρόβλεψης της θερμοκρασίας, να συγκρινεί το συγκεκριμένο νευρωνικό δίκτυο με άλλες μεθόδους πρόβλεψης (μοντέλα χρονικής σειράς, μοντέλα αυτοπαλινδρόμισης) και συνάμα να προτείνει μια μοντέρνα και ακριβή μέθοδο για το πρόβλημα πρόβλεψης της θερμοκρασίας.

Η δομή της εργασίας έχει ως εξής. Η εργασία ξεκινά με μια ανασκόπηση της βιβλιογραφίας όπου παρουσιάζονται δημοσιευμένα άρθρα με εφαρμογές νευρωνικών δικτύων και γενετικών αλγόριθμων σε προβλέψεις γενικά, αλλά και εφαρμογές νευρωνικών δικτύων στο πρόβλημα της πρόβλεψης της θερμοκρασίας, αλλά και άλλες μεθόδους πρόβλεψης θερμοκρασίας, ενώ στο τέλος της ανασκόπησης θα παρουσιαστούν άλλες εφαρμογές του νεύρο-ασαφούς συστήματος (ANFIS Adaptive Neuro-Fuzzy Inference System.) το οποίο χρησιμοποιείται για την πρόβλεψη της θερμοκρασίας. Αμέσως μετά η εργασία συνεχίζεται με την εισαγωγή στην έννοια της ασαφούς λογικής και την έννοια των ασαφών συστημάτων εξαγωγής συμπεράσματος (ΑΣΕΣ Fuzzy Inference System-FIS). Μετά θα γίνει μία εισαγωγή στα τεχνητά νευρωνικά δίκτυα(ΤΝΔ ANN-Artificial Neural Networks). Στο επόμενο κεφάλαιο ακολουθεί μια αναφορά στο νευρωνικό μοντέλο (περιγραφή αλγόριθμος εκπαίδευσης) του οποίο χρησιμοποιείται για την εξαγωγή προβλέψεων. Στη συνέχεια ακολουθεί το κεφάλαιο με τα αποτελέσματα από τη χρήση του συστήματος πρόβλεψης, αλλά και την αξιολόγηση της αξιοπιστίας του μοντέλου, γίνεται σύγκριση με άλλες μεθόδους πρόβλεψης και η εργασία ολοκληρώνεται με τα συμπεράσματα τα οποία εξάγονται από την εφαρμογή του μοντέλου.

ΚΕΦΑΛΑΙΟ 1:**ΑΝΑΣΚΟΠΗΣΗ ΒΙΒΛΙΟΓΡΑΦΙΑΣ**

Στο παρόν κεφάλαιο γίνεται μια ανασκόπηση στη βιβλιογραφία. Επίσης αναφέρονται άρθρα τα οποία παρουσιάζουν μοντέρνες μεθόδους πρόβλεψης, άρθρα στα οποία γίνεται μελέτη και εφαρμογή μεθόδων πρόβλεψης θερμοκρασίας σε εξωτερικούς, αλλά και σε εσωτερικούς χώρους. Επίσης γίνεται και μια αναφορά, σε διάφορες χρήσεις του νεύρο-ασαφούς μοντέλου (ANFIS), γεγονός που αποδεικνύει την ευελιξία και την αποτελεσματικότητα εφαρμογής του μοντέλου σε διάφορους τομείς πρόβλεψης.

Οι Sheng-Tun Li και Yi-Chung Cheng το 2005 ανέπτυξαν ένα ντετερμινιστικό μοντέλο ασαφούς χρονικής σειράς για προβλέψεις με εφαρμογές στις προβλέψεις εγγραφών, θερμοκρασίας και άλλες, μελέτησαν την ακρίβεια του και έκαναν σύγκριση με άλλες μεθόδους.

Το 2006 οι Rafik Aliev, Bizan Fazlollahi, Rashad Aliev και Babek Guirimov παρουσίασαν σε άρθρο τους μια μέθοδο πρόβλεψης ασαφούς χρονικής σειράς βασισμένη σε ασαφή επαναλαμβανόμενα νευρωνικά δίκτυα. Το συγκεκριμένο μοντέλο κρίθηκε ικανό να δώσει προβλέψεις σε διάφορους τομείς. Συγκεκριμένα στο άρθρο τους οι συγγραφείς παρουσίασαν εφαρμογές του μοντέλου σε πρόβλεψη για την κατανάλωση ηλεκτρικής ενέργειας, την πρόβλεψη θερμοκρασίας, ακόμη και την πρόβλεψη των εγγραφών σε ένα πανεπιστήμιο σ.σ. Alabama.

Τον Μάιο του 2000 ο Kunhuang Huarng ανέπτυξε ευρετικά μοντέλα ασαφούς χρονικής σειράς για προβλέψεις και πρότεινε εφαρμογές στην πρόβλεψη τιμών οικονομικών δεικτών , εγγραφών σε πανεπιστήμια , πρόβλεψη θερμοκρασίας και άλλα.

Το 2002 οι Riverola Corchado και Torres ανέπτυξαν ένα υβριδικό μοντέλο CBR (Case-based reasoning model),βασισμένο στη κατά Sugeno συνεπαγωγή, δηλαδή ένα μοντέλο νεύρο-συμβολισμού για την επίλυση προβλημάτων με σκοπό να προβλέψουν παραμέτρους σε περίπλοκο και δυναμικό περιβάλλον χωρίς εποπτεία. Αυτό λύνει τη δυσκολία πρόβλεψης όταν οι κανόνες ορισμού ενός συστήματος είναι άγνωστοι και καθίσταται δύσκολη η απόφαση των χαρακτηριστικών του συστήματος. Το σύστημα κάνει χρήση ενός case-based μοντέλου συλλογισμού, το οποίο ενσωματώνει ένα δίκτυο με δομή αυξανόμενων κυττάρων και ένα κατά Sugeno ασαφές μοντέλο για την εξαγωγή προβλέψεων . Οι ερευνητές εφάρμοσαν το μοντέλο για να προβλέψουν το φαινόμενο των κόκκινων παλιρροιών που εμφανίζονται στις ακτές της βορειοδυτικής Ιβηρικής χερσονήσου.

Στο άρθρο των Arindam Chaudhuri και Kajal De του 2009 πρότειναν την πρόβλεψη χρονικής σειράς με χρήση ενός νεύρο-ασαφούς μοντέλου παλινδρόμησης. Το συγκεκριμένο μοντέλο προτάθηκε με στόχο τη μείωση του όγκου των δεδομένων που χρειάζεται ένα νευρωνικό μοντέλο για να δώσει ακριβείς προβλέψεις.

Το 2000 ο Karel Dejmal πρότεινε μία στατιστική μετά-επεξεργασία για ένα ήδη ανεπτυγμένο μοντέλο αριθμητικής πρόβλεψης του καιρού. Το συγκεκριμένο μοντέλο έδινε έξοδο την πρόβλεψή της μέγιστης και της ελάχιστης θερμοκρασίας για κάθε ημέρα.

Το 2000 οι Shyi-Ming Chen και Jeng-Ren Huang στο άρθρο τους αναλύουν ένα μοντέλο ασαφούς χρονικής σειράς μονής εισόδου (θερμοκρασία) για την πρόβλεψη της θερμοκρασίας, με στόχο να εξαλείψουν τα μειονεκτήματα των συμβατικών μεθόδων πρόβλεψης και να προτείνουν μία μοντέρνα και πιο ακριβή μέθοδο πρόβλεψης. Ανέπτυξαν και πρότειναν 2 νέους αλγόριθμους ασαφούς χρονικής σειράς, με είσοδο θερμοκρασία και πυκνότητα νεφών οι οποίοι έδωσαν ικανοποιητικά αποτελέσματα με μεγάλη ακρίβεια.

Το 2002 οι Sohn Rha και Sheo ανέπτυξαν δυναμικά γραμμικά μοντέλα για την πρόβλεψη θερμοκρασίας με εφαρμογή την πρόβλεψη της θερμοκρασίας του εδάφους στη Νότια Κορέα. Αξίζει να σημειωθεί ότι στη συγκεκριμένη εφαρμογή τα δεδομένα λαμβάνονταν ανά τρεις ώρες και η χρήση του δυναμικού μοντέλου είχε στόχο την εξάλειψη(ελαχιστοποίηση) του σφάλματος το οποίο δίνουν τα συμβατικά μοντέλα πρόβλεψης(αριθμητικά μοντέλα και άλλα) .

Το 2009 οι Smith , Hoogenboom και McClendon πρότειναν τη χρήση τεχνητών νευρωνικών δικτύων και ανέπτυξαν ένα νευρωνικό μοντέλο για την αυτοματοποιημένη πρόβλεψη της θερμοκρασίας σε ορίζοντα ενός έτους. Το μοντέλο που αναπτύχθηκε έδωσε προβλέψεις ξεχωριστά για το χειμώνα και ξεχωριστά για το καλοκαίρι με σχετικά μικρό σφάλμα. Το μοντέλο έπαιρνε εισόδους θερμοκρασία , βροχοπτώσεις, ποσοστά υγρασίας .

Το 1995 οι Figuerola και Mazeo ανέπτυξαν ένα αναλυτικό μοντέλο πρόβλεψης και το εφάρμοσαν στην πρόβλεψη της νυκτερινής , αλλά και της πρωινής θερμοκρασίας υπό συνθήκες ήρεμου, καθαρού ουρανού στις αγροτικές περιοχές της Αργεντινής.

Το 2002 οι Tasadduq , Rehnman και Bubshait ανέπτυξαν ένα νευρωνικό μοντέλο για την πρόβλεψη της μέσης ωριαίας θερμοκρασίας στη Σαουδική Αραβία.

Το 2002 οι Kocak Saylan και Eitzinger ανέπτυξαν ένα μη γραμμικό μοντέλο πρόβλεψης θερμοκρασίας μέσω μονοδιάστατης και πολυδιάστατης ενσωμάτωσης χρονικής σειράς. Το μοντέλο που αναπτύχθηκε εφαρμόστηκε (εξετάστηκε η ακρίβεια του) με δεδομένα αξιολόγησης από την περιοχή της βορειοδυτικής Αυστρίας τη χρονική περίοδο του έτους 1994 (Ιανουάριος του 1994 με Δεκέμβριος του 1994). Το μοντέλο που δημιουργήθηκε έδωσε ακριβή βραχυπρόθεσμα και μακροπρόθεσμα αποτελέσματα.

Το 2009 οι Dombayaci και Golcu ανέπτυξαν ένα τεχνητό νευρωνικό δίκτυο για την πρόβλεψη της μέσης θερμοκρασίας με εφαρμογή και μελέτη περίπτωσης στην Τουρκία. Το μοντέλο έκανε χρήση δεδομένων στη χρονική περίοδο 2003-2006 με δεδομένα εκπαίδευσης της μετρήσεις από τα έτη 2003-2005 και δεδομένα για την αξιολόγηση του μοντέλου της μετρήσεις του έτους 2006.Το μοντέλο που αναπτύχθηκε ήταν ένα μοντέλο τριών εισόδων έξι κρυφών στρωμάτων και μίας εξόδου. Για την εκπαίδευση του μοντέλου έγινε χρήση του αλγόριθμου οπισθόδρομης διάδοσης Levenberg-Marquardt. Το μοντέλο όταν δοκιμάστηκε έδωσε ικανοποιητικά αποτελέσματα πρόβλεψης, αλλά και μικρά σφάλματα σε σχέση με κλασσικές μεθόδους πρόβλεψης με τις οποίες συγκρίθηκε.

Το 2005 οι Ruano, Crispim, Conceicao και Lucio ανέπτυξαν ένα νευρωνικό δίκτυο για την πρόβλεψη της εσωτερικής θερμοκρασίας σε κτίρια και το εφάρμοσαν στην ενεργειακή μελέτη ενός σχολικού κτιρίου στην Πορτογαλία. Κατόπιν έγινε σύγκριση του μοντέλου με ένα φυσικό μοντέλο στην οποία το νευρωνικό δίκτυο έδωσε πολύ καλά αποτελέσματα. Η συγκεκριμένη εφαρμογή είχε στόχο τον έλεγχο του συστήματος κλιματισμού και θέρμανσης του συγκεκριμένου σχολικού κτιρίου.

Το 2008 οι Yang Kun και Wang JieMin πρότειναν μια μέθοδο πρόβλεψης – διόρθωσης για την εκτίμηση της ροής θερμότητας στο έδαφος με δεδομένα τη θερμοκρασία του εδάφους και την υγρασία εδάφους. Το μοντέλο έδωσε το προφίλ θερμότητας με χρήση της εξίσωσης θερμικής διάδοσης και προσαρμόζει το προφίλ αυτό με διαφορές από την παρατήρηση πραγματικών και υπολογισμένων τιμών. Η σύγκριση του μοντέλου μέσω μελέτης υποθέσεων οι οποίες βασίστηκαν σε παρατηρήσεις , συνθετικά δεδομένα , δίνει ικανοποιητικά αποτελέσματα και δεν παρουσιάζει υψηλό σφάλμα**.**

Το 2009 οι Tao Lu και Marti Viljaren παρουσίασαν σε άρθρο τους νευρωνικά μοντέλα για την πρόβλεψη της εσωτερικής θερμοκρασίας και της σχετικής υγρασίας σε εσωτερικούς χώρους. Η έρευνα έδωσε ακριβή αποτελέσματα τόσο στη χρήση των μοντέλων μη-γραμμικής αυτοπαλινδρόμισης με εξωτερική είσοδο όσο και με τη χρήση γενετικών αλγορίθμων. Επίσης προτάθηκαν τρόποι βελτίωσης των αλγόριθμων εκπαίδευσης.

Το 1984 ο Kostoff παρουσίασε μια μέθοδο πρόβλεψης θερμοκρασίας και ελέγχου, με χρήση νευρωνικών πολλαπλασιαστών, για τον έλεγχο της θερμοκρασίας σε μοντέλα πυρηνικής σύντηξης Αναπτύχθηκε ένας αριθμητικός αλγόριθμος για τη μελέτη της ασταθής συμπεριφοράς της μονοδιάστατης εξίσωσης θερμότητας, με στόχο να μοντελοποιήσει με νευρωνικούς πολλαπλασιαστές ένα σύστημα σύντηξης. Η μέθοδος έδωσε ικανοποιητικά αποτελέσματα τα οποία διέφεραν ελάχιστα από της τιμές που πρόεκυψαν από πειράματα και από την αναλυτική επίλυση της εξίσωσης της μετάδοσης της θερμότητας. Συγκεκριμένα η διαφορά ήταν λιγοστές ποσοστιαίες μονάδες, ενώ η χειρότερη διαφορά ήταν περίπου 10-15%

Το 2009 οι Chevalier, Hoogenboom, McClendon και Paz ανέπτυξαν σε άρθρο τους ένα μοντέλο παλινδρόμησης διανύσματος υποστήριξης για την πρόβλεψη θερμοκρασίας του αέρα το οποίο είχε μειωμένα σετ δεδομένων εκπαίδευσης. Επίσης παρουσίασαν και μία συγκριτική μελέτη του μοντέλου έναντι σε νευρωνικά δίκτυα. Το μοντέλο ήταν εξαιρετικά ακριβές και στη σύγκριση του έδωσε πιο ακριβείς προβλέψεις από το νευρωνικό μοντέλο.

Στην εργασία των Ατσαλάκη και Σταματάκη που παρουσιάστηκε το 2008 μελετήθηκαν νευρωνικά δίκτυα και γενετικοί αλγόριθμοι για την πρόβλεψη της κατανάλωσης ηλεκτρικού ρεύματος.

Στην εργασία των Ατσαλάκη και Λαμπουσάκη που εκπονήθηκε το 2009 ένα μοντέλο ANFISχρησιμοποιήθηκε για τη πρόβλεψη τιμών του ευρωπαϊκού διατραπεζικού επιτοκίου Euribor (Euro Interbank Offered Rate).

Στην εργασία των Ατσαλάκη και Τζωρτζάκη οποία εκπονήθηκε το 2007 έγινε χρήση ενός νευρο-ασαφούς προσαρμοστικού συστήματος εξαγωγής συμπεράσματος με στόχο την πρόβλεψη τιμών ακινήτων.

Στην εργασία των Κοσματόπουλου και Αρβανιτάκη, η οποία παρουσιάστηκε το 2005 αναπτύχθηκαν γραμμικά μοντέλα και τεχνητά νευρωνικά δίκτυα για την εξαγωγή μακροοικονομικών προβλέψεων.

Στην εργασία των Ατσαλάκη και Μινουδάκη που εκπονήθηκε το 2007 έγινε πρόβλεψη τις κατανάλωσης νερού στο νομό Χανιών με χρήση του νεύρο-ασαφούς συστήματος ANFIS.

ΚΕΦΑΛΑΙΟ 2:ΑΣΑΦΗΣ ΛΟΓΙΚΗ

2.1: Εισαγωγή Στην Ασαφή Λογική

Η θεωρία της ασαφούς λογικής αναπτύχθηκε για πρώτη φορά το 1965 από τον Zadeh. Χρειάστηκε να περάσει μια δεκαετία ώστε να αναγνωριστεί ευρέως από την επιστημονική κοινότητα με κύρια εφαρμογή των έλεγχο συστημάτων. Οι κύριοι λόγοι που συνέβαλλαν στη αναγνώριση και αποδοχή των ασαφών συστημάτων έλεγχου σχετίζεται με την απλότητα στο σχεδιασμό τους , αλλά και με την ικανότητα τους να ελέγχουν πολύπλοκες και συνεχώς μεταβαλλόμενες διαδικασίες.

Η βασική διάφορα των ασαφών μοντέλων από τα κλασσικά εντοπίζεται στον τύπο πληροφοριών που επεξεργάζονται. Τα κλασσικά μοντέλα αποφεύγουν να χρησιμοποιούν αβέβαιες ακαθόριστες και ανακριβείς πληροφορίες, καθώς η χρήση τους θα οδηγούσε σε λήψη λανθασμένων αποφάσεων. Αντιθέτως τα ασαφή μοντέλα τα οποία δεν απαιτούν ακριβή αριθμητικά μοντέλα των μεταβλητών και των παραμέτρων του συστήματος ,έχουν την ικανότητα να χρησιμοποιήσουν τέτοιου είδους πληροφορίες. Οι ελεγκτές συστημάτων που σχεδιαστήκαν με βάση την ασαφή λογική (fuzzy logic) έχουν την ιδιότητα διαχείρισης ασαφής στοιχείων ή στοιχείων με θόρυβο. Αυτή η ικανότητα βασίζεται στην απλή ικανότητα του ανθρώπου να αντιλαμβάνεται τη συμπεριφορά ενός συστήματος με ποιοτικούς κανόνες. Η ασαφής λογική έχει την ικανότητα μέσω απλών λεκτικών κανόνων (linguistic rules) να μοντελοποιήσει την γνώση και την εμπειρία ενός πεπειραμένου χρηστή. Με αυτόν τον τρόπο δίνεται η ικανότητα διαμόρφωσης ενός μοντέλου βασισμένο στη γνώση (knowledge based model) το όποιο οδηγεί σε απλουστέρα μοντέλα τα όποια είναι και πιο εύχρηστα αλλά και πιο κοντά στην ανθρωπινή λογική. Με αυτό τον τρόπο καθίσταται δυνατός ο έλεγχος ακόμα και σε συστήματα στα οποία οι κλασσικές μέθοδοι έλεγχου αποτυγχάνουν.

2.2:Βασικές Έννοιες Ασαφούς Λογικής

Ένας από τους βασικούς όρους της ασαφής λογικής είναι το ασαφές σύνολο (fuzzy set), δηλαδή ένα σύνολο που τα όρια του δεν είναι πλήρως καθορισμένα και περιέχει στοιχεία με διαφορετικούς βαθμούς συμμετοχής. Για να γίνει πιο κατανοητή η έννοια του ασαφούς συνόλου ας δοθεί πρώτα ο ορισμός του κλασσικού συνόλου (crisp set). Έστω για παράδειγμα το σύνολο μέρες της εβδομάδας το οποίο περιέχει χωρίς αμφισβήτηση σαν στοιχεία της ημέρες Δευτέρα Τρίτη και Σάββατο, το όποιο δεν περιλαμβάνει με βεβαιότητα τα στοιχεία ταχύτητα θερμοκρασία. Ο παραπάνω ορισμός του συνόλου είναι σύμφωνος με την Αριστοτελική λογική η όποια υποστηρίζει ότι ένα στοιχείο Χ θα ανήκει είτε στο σύνολο «Α» , είτε στο σύνολο «όχι Α».

Σχήμα 2 1 :Απεικόνιση του κλασσικού συνόλου ημέρες της εβδομάδος.

Στη συνεχεία ας πάρουμε το σύνολο ημέρες του Σαββατοκύριακου. Στο παρακάτω σχήμα γίνεται μια προσπάθεια ταξινόμησης των ημερών του σαββατοκύριακου. Όλοι συμφωνούν ότι οι ημέρες Σάββατο και Κυριακή ανήκουν στο σύνολο αυτό. Η Παρασκευή όμως; Η διαίσθηση μας λέει να την εντάξουμε στο σύνολο αυτό. Από την άλλη όμως η λογική μας λέει να μην την εντάξουμε. Θα μπορούσαμε την ημέρα αυτή να την εντάξουμε λοιπόν στα όρια του συνόλου «Ημέρες του σαββατοκύριακου» με το σύνολο «όχι ημέρες του σαββατοκύριακου», λαμβάνοντας την έτσι σαν στοιχείο και

τον δυο συνόλων. Η λογική των κλασσικών συνόλων δεν θα το επέτρεπε αυτό , όμως η ανθρωπινή εμπειρία και διαίσθηση δέχεται την κατάταξη μιας έννοιας στο όριο δυο συνόλων σαν κομμάτι της καθημερινής πραγματικότητας. Η ασαφής λογική αποτελεί στην ουσία μια διεύρυνση της Boolean λογικής. Αν αντιστοιχήσουμε στην κατάσταση αληθές την λογική τιμή 1 και στην κατάσταση ψευδές τη λογική τιμή 0 τότε η ασαφής λογική μας επιτρέπει την ύπαρξη ενδιάμεσων τιμών όπως 0.5, 0.8, 0.9.
Για παράδειγμα:

Σχήμα 2.2: Απεικόνιση του ασαφούς συνόλου ήμερες του σαββατοκύριακου.

Ανήκει το Σάββατο στο παραπάνω σύνολο;

1 ή «αληθές»

Ανήκει η Κυριακή στο παραπάνω σύνολο;

0.9 ή ανήκει αλλά όχι όσο το Σάββατο.

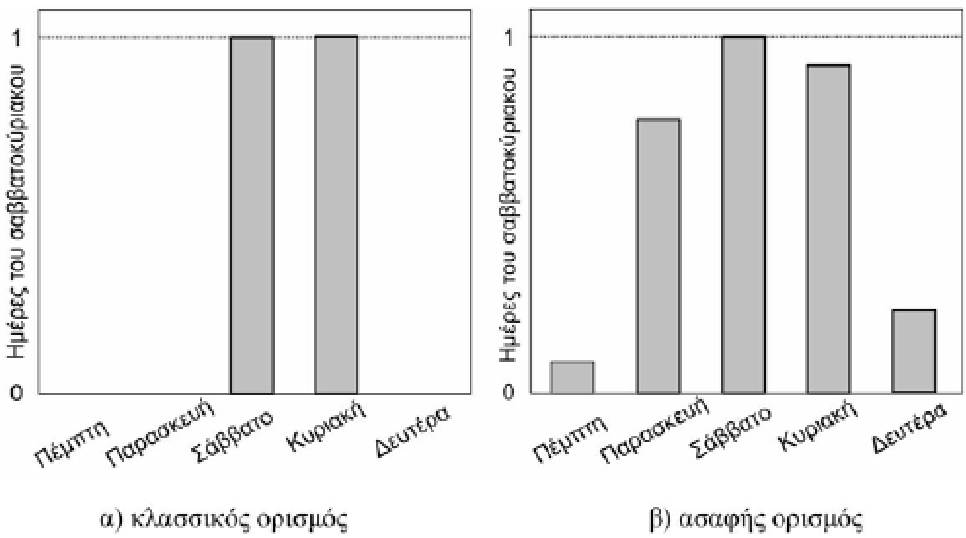
Ανήκει η Παρασκευή;

0.8 κατά κύριο λογά ανήκει αλλά όχι όλη.

Ανήκει η Τρίτη ;

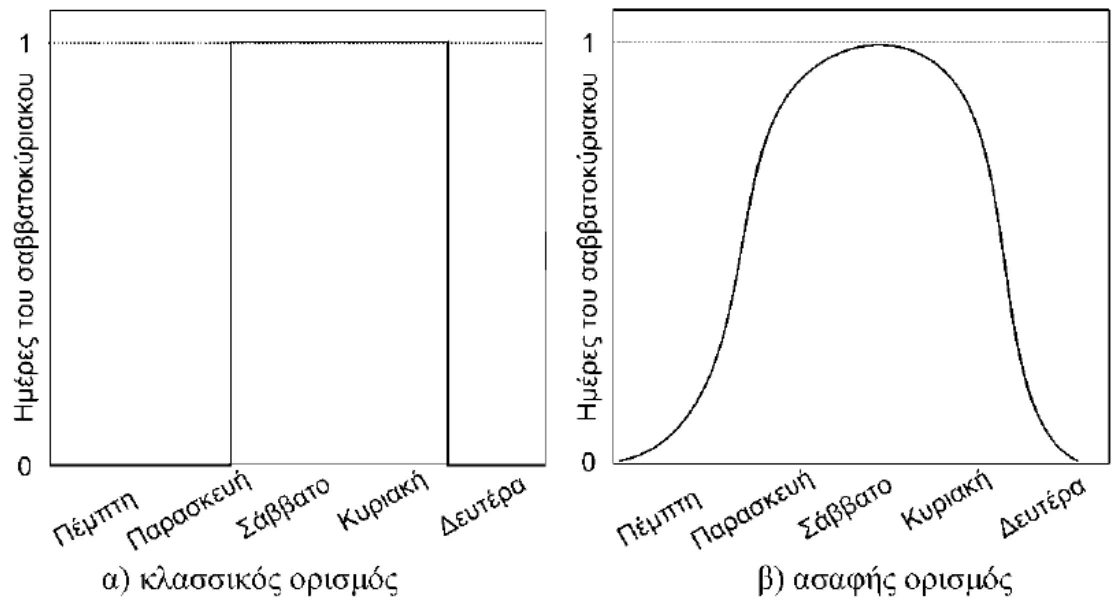
0 ή ψευδές.

Στο παρακάτω σχήμα δίνεται με μορφή διαγράμματος η συμμετοχή κάποιων ημερών στο σύνολο «ημέρες του σαββατοκύριακου» με τον κλασσικό ορισμό και με τον ασαφή ορισμό.



Σχήμα2.3 :Διάγραμμα συμμετοχής στο σύνολο «Ημέρες του σαββατοκύριακου»

Αν στα παραπάνω σχήματα ο άξονας των ημερών πάρει συνέχεις τιμές τότε τα διαγράμματα θα πάρουν τη μορφή του σχήματος που ακολουθεί .Έτσι είναι δυνατό να οριστεί ο βαθμός στον οποί ο κάθε δεδομένη στιγμή ανήκει στο σαββατοκύριακο. Στο αριστερό σχήμα φαίνεται ότι τα μεσάνυχτα της παρασκευής ο βαθμός συμμετοχής στο σύνολο «ημέρες του σαββατοκύριακου» αλλάζει νηματικά από το ‘0’ στο ‘1’.Αυτο αποτελεί ένα τρόπο ορισμού του σαββατοκύριακου χωρίς όμως να είναι ο ορισμός που η πλειοψηφία των ανθρώπων έχουν στο μυαλό τους σαν σαββατοκύριακο.

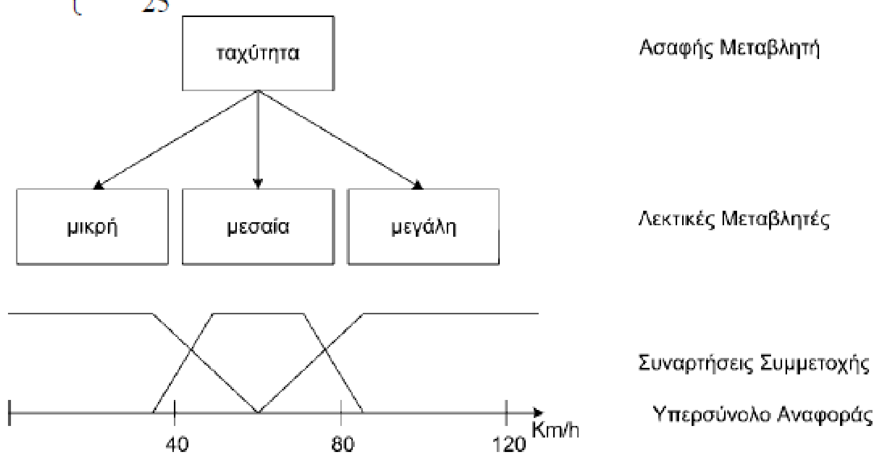


Σχήμα 2.4:Διαγραμμα συμμετοχής στο σύνολο «Ημέρες του σαββατοκύριακου» (συνεχής μορφή).

Η καμπύλη στο δεξιό κομμάτι παρουσιάζει πιο ομαλές μεταβολές και μας δείχνει ότι και η Παρασκευή και κατά ένα μικρότερο βαθμό η Δευτέρα έχουν συμμετοχή στην έννοια του σαββατοκύριακου και αξίζουν ένα μικρό βαθμό συμμετοχής στο σύνολο «Ημέρες του σαββατοκύριακου». Η παραπάνω καμπύλη είναι γνωστή σαν συνάρτηση συμμετοχής (membership function). Η συνάρτηση η όποια συμβολίζεται συνήθως με το γράμμα μ είναι η καμπύλη η οποία ορίζει σε ποιο βαθμό κάθε στοιχείο ενός ασαφούς συνόλου αντιστοιχίζεται σε ένα βαθμό συμμετοχής στο σύνολο ανάμεσα στο ‘0’ και στο ‘1’. Το σύνολο στο οποίο παίρνει τιμές η είσοδος ονομάζεται σαν το υπερσύνολο αναφοράς (universe of discourse). Το διάστημα για το οποίο η συνάρτηση παίρνει θετικές τιμές ονομάζεται σύνολο στήριξης (support set).

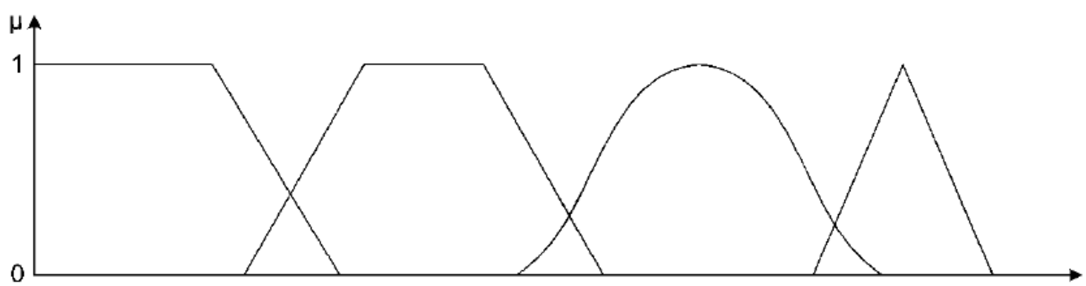
Παρακάτω δίνεται ένα παράδειγμα κλασσικών και ασαφών συνόλων ώστε να γίνουν πιο κατανοητές οι βασικές έννοιες της ασαφούς λογικής. Έστω η ασαφής μεταβλητή ταχύτητα. Η μεταβλητή αυτή μπορεί να περιγράφει από της λεκτικές μεταβλητές (linguistic variables) «χαμηλή» «μεσαία» «υψηλή», όπως φαίνεται στο παρακάτω σχήμα. Για κάθε μια από της γλωσσικές μεταβλητές αντιστοιχίζεται ένα ασαφές σύνολο το οποίο περιγράφεται από τη συνάρτηση συμμετοχής του. Για παράδειγμα η συνάρτηση συμμετοχής για το ασαφές σύνολο «μικρή» θα μπορούσε να δοθεί από την παρακάτω εξίσωση:

$$\mu_{\mu\iota\kappa\rho\eta}(\nu) = \begin{cases} 1, & \nu \in [0,35] \\ 1 - \dfrac{\nu - 35}{25}, & \nu \in (35,60] \end{cases}$$



Σχήμα 2.5: Οροί της ασαφούς λογικής.

Οι τιμές των συναρτήσεων συμμετοχής ανήκουν στο [0,1]. Η μορφή των συναρτήσεων αυτών ποικίλει ανάλογα με την εφαρμογή. Οι συνηθέστερες συναρτήσεις είναι η τριγωνική , η τραπεζοειδής και η γκαουσιαννή



Σχήμα 2.6: Μορφές συναρτήσεων συμμετοχής

2.3:Ασαφής Συλλογισμός (Fuzzy Reasoning)

Ασαφείς κανόνες (Fuzzy rules)

Ο ασαφής συλλογισμός πραγματοποιείται με κανόνες της μορφής:

«Αν..............τοτε……………..»

Ένας ασαφής κανόνας προσδιορίζει τη σχέση μετάξι των ασαφών μεταβλητών εισόδου και εξόδου και αποτελείται από δυο μέρη: Την προϋπόθεση (antecedent) και το επακόλουθο (consequence). Ένα τυπικό παράδειγμα ενός ασαφούς κανόνα είναι το εξής:

$$\underbrace{\text{Αν η είσοδος\_1 είναι } \textit{μικρή} \text{ και η είσοδος\_2 είναι } \textit{μεγάλη αρνητική},}_{\text{προϋπόθεση}} \;\underbrace{\text{τότε η έξοδος είναι } \textit{μηδέν}}_{\text{επακόλουθο}}$$

Στον παραπάνω κανόνα οι είσοδοι 1 και 2 και οι έξοδος είναι οι ασαφείς μεταβλητές ενώ η μικρή ,η μεγάλη αρνητική και η μηδέν είναι οι λεκτικές μεταβλητές.

2.3.1:Λογικές πράξεις με ασαφή σύνολα

Λογά της χρήσης της πλειότιμης λογικής στην ασαφή λογική οι πράξεις της ασαφούς λογικής διαφέρουν από αυτές της δίτιμης Boolean λογικής. Αν στης ασαφείς μεταβλητές αποδοθούν οι ακραίες τιμές τους ‘1’ «εντελώς αληθές» και ‘0’ «εντελώς ψευδές» θα ισχύουν οι μέχρι τώρα γνώστες πράξεις όπως η AND, η OR και η NOT όπως φαίνεται στον πίνακα που ακολουθεί.

A	B	A AND B
0	0	0

A	B	A OR B
0	0	0

A	NOT A
0	1

0	1	0
1	0	0
1	1	1

0	1	1
1	0	1
1	1	1

1	0

Πινάκας 2.1:Λογικες πράξεις της δίτημης (Boolean) λογικής.

Λαμβάνοντας υπόψη ότι στην ασαφή λογική οι τιμές εισόδου είναι πραγματικοί αριθμοί ανάμεσα στο ‘0’ και το ‘1’, θα πρέπει να βρεθούν κάποιες συναρτήσεις που θα διατηρούν τα αποτελέσματα των παραπάνω πινάκων , για όλους τους αριθμούς ανάμεσα στο ‘0’ και το ‘1’. Μια απάντηση θα ήταν η πράξη AND να αντικατασταθεί από την πράξη MIN (ελάχιστο) ή από το γινόμενο. Ομοίως θα μπορούσε να αντικατασταθεί η πράξη OR από την πράξη MAX (μέγιστο) και η πράξη NOT A από το 1-A. Στον παρακάτω πίνακα φαίνεται πως οι σχέσεις αλήθειας παραμένουν ίδιες αν γίνουν οι παραπάνω αντιστοιχίσεις.

A	B	Min(A,B)
0	0	0
0	1	0
1	0	0
1	1	1

A	B	Max(A,B)
0	0	0
0	1	1
1	0	1
1	1	1

A	1-A
0	1
1	0

Πίνακας 2.2: Λογικές πράξεις ασαφούς λογικής.

Συνεπώς η λογική πράξη AND που ταυτίζεται με την τομή δύο συνόλων ορίζεται ως εξής:

$$AND(\mu_A,\mu_B)=\mu_{A\cap B}=\min(\mu_A,\mu_B) \qquad (2.2)$$

Συμφώνα με το καρτεσιανό γινόμενο τομής, ενώ ορίζεται ως εξής:

$$AND(\mu_A,\mu_B) = \mu_{A\cap B} = \qquad (2.3)$$

Συμφώνα με το αλγεβρικό γινόμενο .

Αντίστοιχα η λογική πράξη Η (OR) που ταυτίζεται με την ένωση (Unit) δύο συνόλων ορίζεται ως εξής:

$$AND(\mu_A,\mu_B)=\mu_{A\cap B}=\max(\mu_A,\mu_B) \qquad (2.4)$$

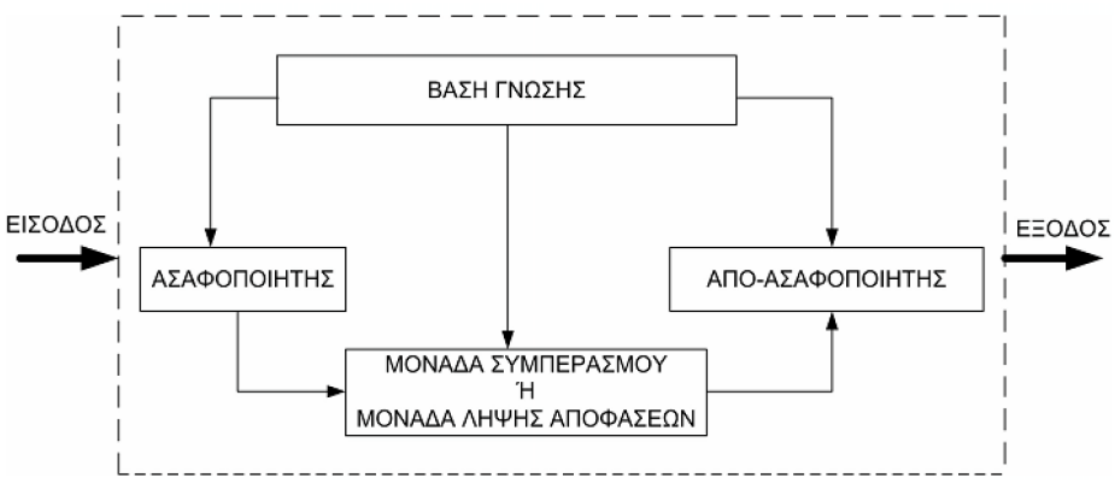
Τέλος το συμπλήρωμα (Complement) του ασαφούς συνόλου Α δίνεται από τη σχέση:

$$NOT(\mu_A)=1-\mu_A \qquad (2.5)$$

Όπου μ_A και μ_B είναι οι τιμές των συναρτήσεων συμμετοχής των ασαφών συνόλων Α και Β αντίστοιχα.

2.4. Ασαφές Σύστημα Εξαγωγής Συμπερασμάτων

Το ασαφές σύστημα εξαγωγής συμπερασμάτων –ΑΣΕΣ , {Fuzzy inference system-FIS} είναι υπεύθυνο για την διεξαγωγή του ασαφούς συλλογισμού. Το ΑΣΕΣ αποτελείται από τέσσερα βασικά μέρη όπως φαίνεται και στο σχήμα που ακολουθεί. Αυτά είναι ο Ασαφοποιητής{Fuzzyfier}. Η βάση γνώσης {Knowledge Base}. Η μονάδα λήψης αποφάσεων ή Μονάδα Συμπερασμάτων {Inference Unit} και ο Αποασαφοποιητής {Defuzzyfier}. Στη βάση γνώσης βρίσκονται καταχωρημένοι με τη μορφή κώδικα οι κανόνες έλεγχου και τα ασαφή σύνολα, «η γνώση δηλαδή του συστήματος». Η μονάδα λήψης αποφάσεων εκτελεί τους κανόνες παίρνοντας σαν είσοδο της ασαφοποιημένες εισόδους , παραδίδει το ασαφές αποτέλεσμα στον αποασαφοποιητή, ο οποίος παράγει με σαφή μορφή την έξοδο του ΑΣΕΣ.



Σχήμα 2.7: Δομή του ασαφούς συστήματος εξαγωγής συμπερασμάτων.

Η λειτουργία του ΑΣΕΣ μπορεί να διαιρεθεί σε τρία στάδια . Αρχικά οι κλασσικές είσοδοι εισάγονται στο σύστημα και ασαφοποιούνται , στο δεύτερο στάδιο οι βασικές είσοδοι επεξεργάζονται στη βάση γνώσης και στο τελικό στάδιο, όλα τα αποτελέσματα όλων των κανόνων συνδυάζονται και αποασαφοποιούνται. Παρακάτω παρουσιάζεται αναλυτικά ο τρόπος λειτουργίας κάθε σταδίου ξεχωριστά.

Στάδιο ασαφοποίησης

Σε αυτό το στάδιο ο ασαφοποιητής μετατρέπει της φυσικές μεταβλητές που εισάγονται στο σύστημα σε ασαφείς.

Εξαγωγή συμπερασμάτων

Σε αυτό το στάδιο οι λειτουργίες εξαγωγής συμπερασμάτων στους ασαφείς κανόνες διεκπεραιώνονται από τη μονάδα λήψης αποφάσεων. Πραγματοποιείται μια εξονυχιστική ανίχνευση όλων των κανόνων με στόχο των υπολογισμό του βαθμού συμμετοχής. Οι κανόνες με μικρό βαθμό συμμετοχής παίζουν δευτερεύοντα ρόλο στη λήψη απόφασης , ενώ οι κανόνες με υψηλό βαθμό συμμετοχής αποτελούν τους κυρίαρχους στη λήψη της απόφασης. Το τελικά ασαφή σύνολα εξόδου εξαρτώνται από τον τύπο της συνεπαγωγής που θα χρησιμοποιηθεί .Οι δημοφιλέστερες συνεπαγωγές είναι η συνεπαγωγή κατά Mamdani , η κατά Sugeno, και η συνεπαγωγή κατά Larsen. Οι συνεπαγωγές κατά Mamdani κάνουν χρήση του καρτεσιανού γινόμενου τομής, ενώ οι συνεπαγωγές κατά Larsen κάνουν χρήση του καρτεσιανού αλγεβρικού γινόμενου.

Συνεπαγωγή κατά Mamdani:

$$\mu_R(x,y) = \mu_A(\chi) \cap \mu_B(y) = \boldsymbol{min}(\mu_A(\chi), \mu_B(y)) \qquad (2.6)$$

Ο συνδυασμός Ν εξαρτημένων γίνεται με το συνδετικό Ή.

Συνεπαγωγή κατά Larsen:

$$\mu_R(x,y) = \mu_A(\chi) * \mu_B(y) \tag{2.8}$$

Ο συνδυασμός N εξαρτημένων γίνεται και εδώ με το συνδετικό Ή.

$$\mu_R^N(x,y) = \cup_k \left(\mu_A^k(\chi) * \mu_B^k(y)\right) \tag{2.9}$$

Αποασαφοποίηση

Το τελευταίο βήμα του αλγορίθμου ενός ασαφούς ελεγκτή είναι το στάδιο της αποασαφοποίησης (defuzzyfication) της ασαφούς εξόδου σε σαφή τιμή . Υπάρχουν πολλοί μέθοδοι αποασαφοποίησης . Κρίνεται απαραίτητο να αναφερθεί ότι οι μέθοδοι αποασαφοποίησης που περιγράφονται στη με θέμα την ασαφή λογική βιβλιογραφία διαφέρουν από συγγραφέα σε συγγραφέα . Για αυτό το λογά στης μεθόδους αποασαφοποίησης που περιγράφονται παρακάτω θα πρέπει να δοθεί περισσότερη σημασία στον ορισμό της κάθε μεθόδου και όχι στο όνομα της . Οι πιο γνωστές μέθοδοι αποασαφοποίησης είναι οι εξής:

Αποασαφοποίηση μεγίστου (Maximum defuzzyfier).

Στην τεχνική αυτή εξετάζεται η σύνθετη συνάρτηση συμμετοχής της εξόδου του ελεγκτή και επιλέγεται ως έξοδος η τιμή της μεταβλητής y όπου μΥ(γ) είναι μέγιστο.

Συνεπώς:

$$\hat{y}_{MAX} = max\ \mu_Y(y) \tag{2.10}$$

Η παραπάνω μέθοδος δεν παρέχει ικανοποιητικά αποτελέσματα ιδιαιτέρα όταν υπάρχουν πολλαπλά μέγιστα.

Αποασαφοποίηση με μέσο όρο των μεγίστων (Mean of Maxima defuzzyfier):

Σε αυτή τη μέθοδο εξετάζεται η συνάρτηση συμμετοχής μΥ(γ) για να βρεθούν εφόσον υπάρχουν οι τιμές του γγια της οποίες η μΥ(γ) είναι μέγιστη . Στη συνέχεια υπολογίζεται ο μέσος όρος των τιμών των εξόδων που αντιστοιχούν σε αυτή τη συνθήκη, δηλαδή :

$$\hat{y}_{MOM} = \frac{1}{m}\sum_{j=1}^{m} \max \mu_y(y_i) \tag{2.11}$$

Σε περίπτωση που υπάρχει μόνο ένα μέγιστο η παραπάνω μέθοδος συμπίπτει με την πρώτη μέθοδο .

Αποασαφοποίηση κεντρώου (Center of Area "COA" defuzzyfier)

Σε αυτή τη μέθοδο υπολογίζεται το κέντρο του εμβαδού της συνθέτης συνάρτησης συμμετοχής της εξόδου:

$$\hat{y}_{COA} = \frac{\int_S y\mu_Y(y)}{\int_S \mu_Y(y)} \tag{2.12}$$

Όπου S το σύνολο στήριξης της συνάρτησης μΥ(γ) σε περίπτωση που το παραπάνω σύνολο στήριξης είναι διακριτό η παραπάνω σχέση γίνεται:

$$\hat{y}_{COA} = \frac{\sum_{i=1}^{I} y_i \mu_Y^i(y_i)}{\sum_{i=1}^{I} \mu_Y^i(y_i)} \tag{2.13}$$

Στην τεχνική αυτή τα σχήματα των συναρτήσεων συμμετοχής παίζουν μεγάλο ρόλο στο τελικό αποτέλεσμα. Η μέθοδος αυτή εφαρμόζεται κατά κύριο λόγο στον ασαφή έλεγχο διεργασιών λόγω των λογικών αποτελεσμάτων που προκύπτουν.

Αποασαφοποίηση κέντρου βάρους (Center of Gravity COG defuzzyfication)

Σε αυτή την τεχνική υπολογίζεται το κέντρο βάρους των συναρτήσεων συμμετοχής όλων των κανόνων που έχουν ενεργοποιηθεί , σταθμισμένο με το βαθμό εκπλήρωσης σ κάθε κανόνα. Το τελικό αποτέλεσμα , στην περίπτωση που οι συναρτήσεις είναι συμμετρικές είναι:

$$\widehat{y}_{COG} = \frac{\sum_{i=1}^{I} \sigma_i \mu_Y^i(y_i)}{\sum_{i=1}^{I} \mu_Y^i(y_i)} \qquad (2.14)$$

Όλες οι παραπάνω τεχνικές έχουν στόχο την εύρεση σαφούς εξόδου του ελεγκτή , δηλαδή της μεταβλητής ελέγχου από τη σύνθετη συνάρτηση συμμετοχής της εξόδου.

2.4.1:Παράδειγμα

Στην παρακάτω παράγραφο δίνεται ένα απλό παράδειγμα για το τι και ποια βήματα ακολουθεί ένας ασαφής ελεγκτής για τη λήψη μιας απόφασης.

Ο ασαφής ελεγκτής που μελετάται έχει δύο εισόδους την Είσοδο 1 και την Είσοδο 2 και μια έξοδο.

Στην Είσοδο 1και στην Έξοδο αντιστοιχούν από πέντε λεκτικές μεταβλητές οι:

LO= (Low) χαμηλός

ZO=(Zero) κανονικός

LH= (Little High) Λίγο ψηλός

MH=(Medium High) Ψηλός

VH= (Very High) Πολύ Ψηλός

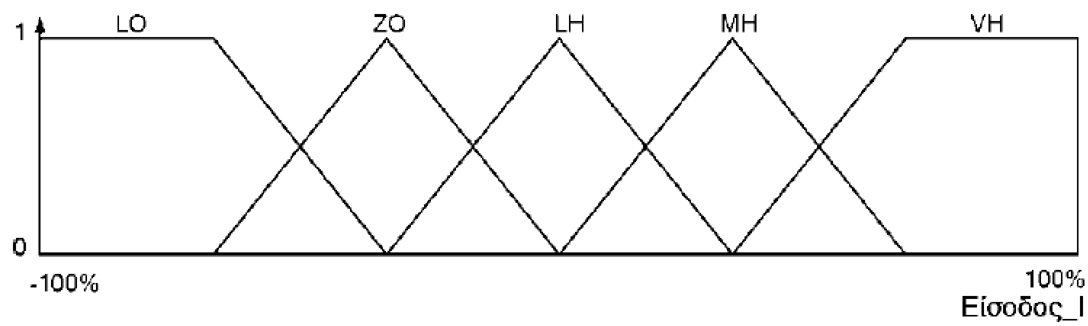
Ενώ στην Είσοδο 2 τρεις λεκτικές μεταβλητές

VL= (Very Low) πολύ χαμηλός

ZO=(Zero) κανονικός

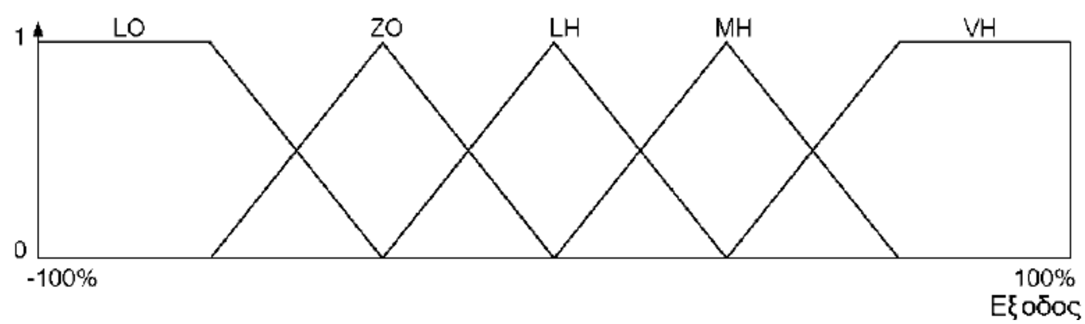
VH= (Very High) Πολύ Ψηλός

Οι συναρτήσεις συμμετοχής των εισόδων και της εξόδου έχουν τη μορφή που φαίνεται στα παρακάτω σχήματα.

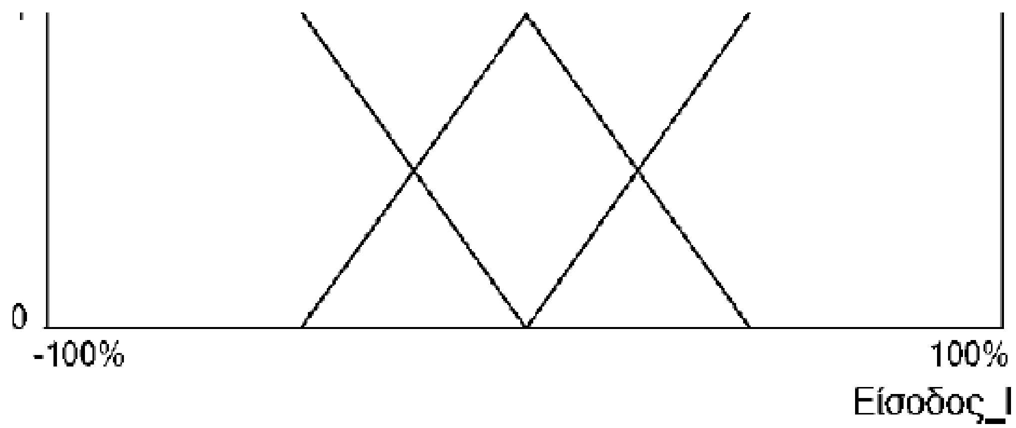


Σχήμα 2.8: Συναρτήσεις συμμετοχής της εισόδου: Είσοδος 1

.



Σχήμα 2.9: Συναρτήσεις συμμετοχής της εξόδου



Σχήμα 2.10:Συναρτησεις συμμετοχής της εισόδου Είσοδος 2.

Δεκαπέντε κανόνες προσδιορίζουν της απαιτούμενες ενέργειες του ασαφούς ελεγκτή:

R^1: Εάν η *Είσοδος* 1 είναι LO και η *Είσοδος 2* είναι VL τότε η Έξοδος είναι LO

R^2: Εάν η *Είσοδος 1* είναι ZO και η *Είσοδος 2* είναι VL τότε η Έξοδος είναι ZO

R^3: Εάν η *Είσοδος 1* είναι LH και η *Είσοδος 2* είναι VL τότε η Έξοδος είναι LH

R^4: Εάν η *Είσοδος 1* είναι MH και η *Είσοδος 2* είναι VL τότε η Έξοδος είναι LH

R^5: Εάν η *Είσοδος 1* είναι VH και η *Είσοδος 2* είναι VL τότε η Έξοδος είναι LH

R^6: Εάν η *Είσοδος 1* είναι LO και η *Είσοδος 2* είναι ZO τότε η Έξοδος είναι LO

R^7: Εάν η *Είσοδος 1* είναι ZO και η *Είσοδος 2* είναι ZO τότε η Έξοδος είναι LH

R^8: Εάν η *Είσοδος 1* είναι LH και η *Είσοδος 2* είναι ZO τότε η Έξοδος είναι MH

R^9: Εάν η *Είσοδος 1* είναι MH και η *Είσοδος 2* είναι ZO τότε η Έξοδος είναι MH

R^{10}: Εάν η *Είσοδος 1* είναι VH και η *Είσοδος 2* είναι ZO τότε η Έξοδος είναι VH

R^{11}: Εάν η *Είσοδος 1* είναι LO και η *Είσοδος 2* είναι VH τότε η Έξοδος είναι LH

R^{12}: Εάν η *Είσοδος 1* είναι ZO και η *Είσοδος 2* είναι VH τότε η Έξοδος είναι MH

R^{13}: Εάν η *Είσοδος 1* είναι LH και η *Είσοδος 2* είναι VH τότε η Έξοδος είναι MH

R^{14}: Εάν η *Είσοδος 1* είναι MH και η *Είσοδος 2* είναι VH τότε η Έξοδος είναι VH

R^{15}: Εάν η *Είσοδος 1* είναι VH και η *Είσοδος 2* είναι VH τότε η Έξοδος είναι VH

Οι παρακάτω κανόνες παρουσιάζονται σε συνοπτική μορφή ως πίνακας κανόνων :

Είσοδος 1 / *Είσοδος 2*	LO	ZO	LH	MH	VH
VL	LO	ZO	LH	LH	LH
ZO	LO	LH	MH	MH	VH
VH	LH	MH	MH	VH	VH

Πίνακας 2.3. Κανόνες του ασαφούς ελεγκτή.

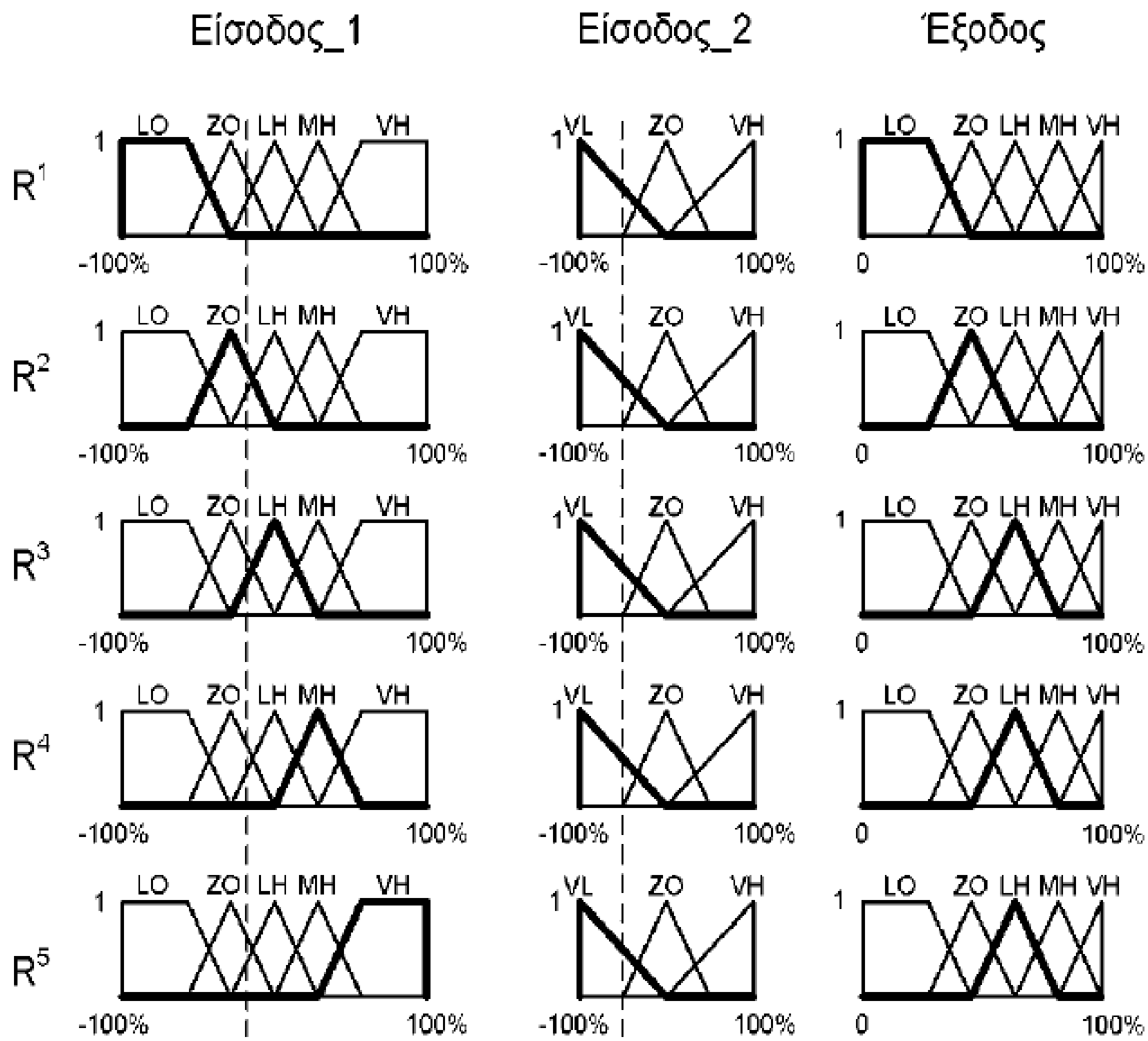
Σε γραφική μορφή οι κανόνες παρουσιάζονται στο παρακάτω σχήμα . Υποτίθεται ότι την χρονική στιγμή των μετρήσεων οι μετρήσεις των εισόδων 1 και 2 έχουν τιμές -20% και -50% αντίστοιχα.

Βήμα πρώτο: Σάρωση όλων των κανόνων συστηματικά ώστε να βρεθεί ο βαθμός συμμετοχής κάθε κανόνα στην τελική απόφαση του ελεγκτή .Ο χρόνος διεξαγωγής της απόφασης εξαρτάται άμεσα από το πλήθος των κανόνων στη βάση γνώσης. Άρα για την ταχεία λήψη απόφασης το πλήθος των κανόνων πρέπει να είναι μικρό. Συνήθως είναι αρκετό ένα πλήθος από 20 με 30 κανόνες.

Τη δεδομένη χρονική στιγμή ισχύει.

$\mu_1^1 = 0,\ \mu_1^2 = 0.66,\ \mu_1^3 = 0.33,\ \mu_1^4 = 0,\ \mu_1^5 = 0$ και

$\mu_2^1 = 0.5,\ \mu_2^2 = 0.5,\ \mu_2^3 = 0.5,\ \mu_2^4 = 0.5,\ \mu_2^5 = 0.5$



Σχήμα 2.11:αποικονηση των πέντε πρώτων κανόνων

Ακλουθώντας τη διαδικασία min (μ_1^1, μ_2^1)για κάθε κανόνα υπολογίζονται οι βαθμοί συμμετοχής σj για κάθε κανόνα (εάν οι ασαφείς στο αίτιο των κανόνων συνδεόταν με το λογικό τελεστή «ή» αντί του «και» θα υπολογιζόταν το max (μ_1^1, μ_2^1) .

R^1: $\sigma_1 = \min\,(\mu_1^1, \mu_2^1) = \min(0{,}0.5) = 0$

R^2: $\sigma_2 = \min\,(\mu_1^2, \mu_2^2) = \min(0.66{,}0.5) = 0.5$

R^3: $\sigma_3 = \min\,(\mu_1^3, \mu_2^3) = \min(0.33{,}0.5) = 0.33$

R^4: $\sigma_4 = \min\,(\mu_1^4, \mu_2^4) = \min(0{,}0.5) = 0$

R^5: $\sigma_5 = \min\,(\mu_1^5, \mu_2^5) = \min(0{,}0.5) = 0$

Οι υπόλοιποι κανόνες που δεν υπολογίζονται έχουν μηδενικό βαθμό συμμετοχής , καθώς ισχύει $\mu^i{}_2$ =0 για i =6 έως 15.

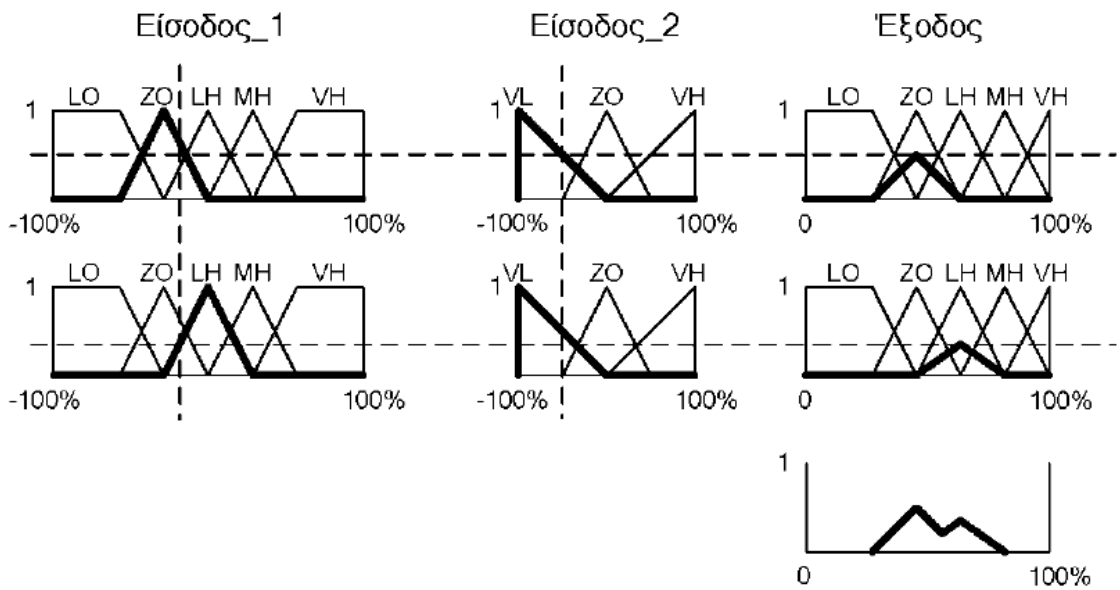
Βήμα δεύτερο: Το δεύτερο βήμα είναι η εύρεση της συνάρτησης εξόδου του ελεγκτή. Η διαδικασία αυτή εξαρτάται από των κανόνα συνεπαγωγής που εφαρμόζεται. Έστω ότι έχουμε τη συνεπαγωγή του Larsen τότε η συνάρτηση συμμετοχής της εξόδου είναι η ένωση των συναρτήσεων συμμετοχής κάθε κανόνα σταθμισμένων με τον αντίστοιχο βαθμό συμμετοχής, δηλαδή:

$$\mu(y) = \sigma_1\mu^1(y) \cup \sigma_2\mu^2(y) \cup \sigma_3\mu^3(y) \cup \sigma_4\mu^4(y) \cup \sigma_5\mu^5(y) =$$

$$= \sigma_2\mu^2(y) \cup \sigma_3\mu^3(y) =$$

$$= 0.5 \cdot \mu^2(y) \cup 0.33 \cdot \mu^3(y)$$

το αποτέλεσμα είναι η καμπύλη που φαίνεται στο κατώτερο μέρος του παρακάτω σχήματος.

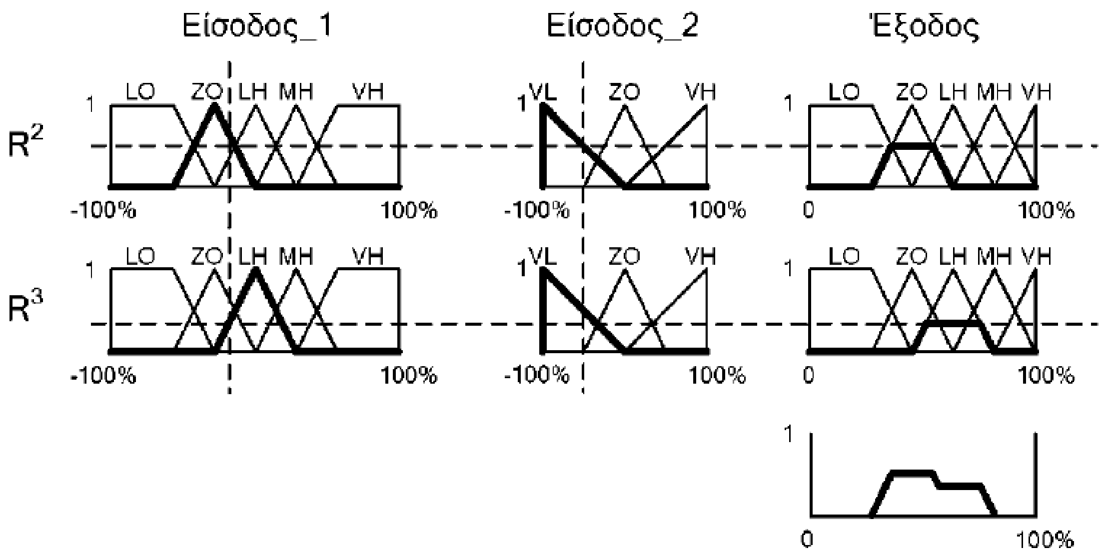


Σχήμα 2.12 διαδικασία υπολογισμού της σύνθετης συνάρτησης συμμετοχής με τη χρήση του κανόνα της συνεπαγωγής Larsen

Στην περίπτωση της συνεπαγωγής Mamdani η συνάρτηση συμμετοχής της εξόδου είναι η ένωση των ελαχίστων των βαθμών συμμετοχής και της συνάρτησης συμμετοχής της εξόδου κάθε κανόνα , δηλαδή:

$$\mu(y) = min(\sigma_1\mu^1(y)) \cup min(\sigma_2\mu^2(y)) \cup min(\sigma_3\mu^3(y)) \cup min(\sigma_4\mu^4(y)) \cup min(\sigma_5\mu^5(y)) =$$
$$= min(\sigma_2\mu^2(y)) \cup min(\sigma_3\mu^3(y)) = min(0.5\mu^1(y)) \cup min(0.33\mu^1(y))$$

Το αποτέλεσμα είναι η καμπύλη που φαίνεται στο κάτω μέρος του σχήματος που ακολουθεί:



Σχήμα 2.13:Διαδικασια υπολογισμού της σύνθετης συνάρτησης συμμετοχής της εξόδου με τον κανόνα του Mamdani.

Τρίτο βήμα: Το τελευταίο βήμα του αλγόριθμου του ασαφούς ελεγκτή είναι η Αποασαφοποίηση της ασαφούς εξόδου σε Σάφη τιμή.

Ανάλογα με την τεχνική αποασαφοποίησης που χρησιμοποιείται η έξοδος του ελεγκτή παίρνει τις παρακάτω τιμές:

Τεχνική αποασαφοποίησης MOM: 0

Τεχνική αποασαφοποίησης COA: -0.49

Τεχνική αποασαφοποίησης COG: -0.51

Θεωρώντας ότι η συνεπαγωγή έχει γίνει με τη μέθοδο Min του Mamdani.

2.5:Η Συνεπαγωγή Sugeno

Η συνεπαγωγή Sugeno ή Takagi-Sugeno-Kang προτάθηκε για πρώτη φόρα το 1985. Αυτή η συνεπαγωγή παρουσιάζει αρκετές ομοιότητες με τη συνεπαγωγή τύπου Mamdani. Η κυρία διάφορα τους σχετίζεται με τις συναρτήσεις συμμετοχής της εξόδου. Στο μοντέλο Sugeno οι συναρτήσεις μεταφοράς της εξόδου είναι γραμμικές ή σταθερές. Ένας τυπικός κανόνας ασαφούς συστήματος Sugeno μηδενικής τάξεως έχει τη μορφή:

Αν το Χ είναι Α και το Υ είναι Β τότε το ζ=k οπού Α και Β ασαφή σύνολα και k σταθερά.

Η γενική μορφή ενός ασαφούς συστήματος Sugeno πρώτης τάξεως είναι η εξής:

Αν το x είναι Α και το y είναι Β τότε ζ=ρ*x+r . Οπού Α και Β ασαφή σύνολα και ρ και rσταθερά.

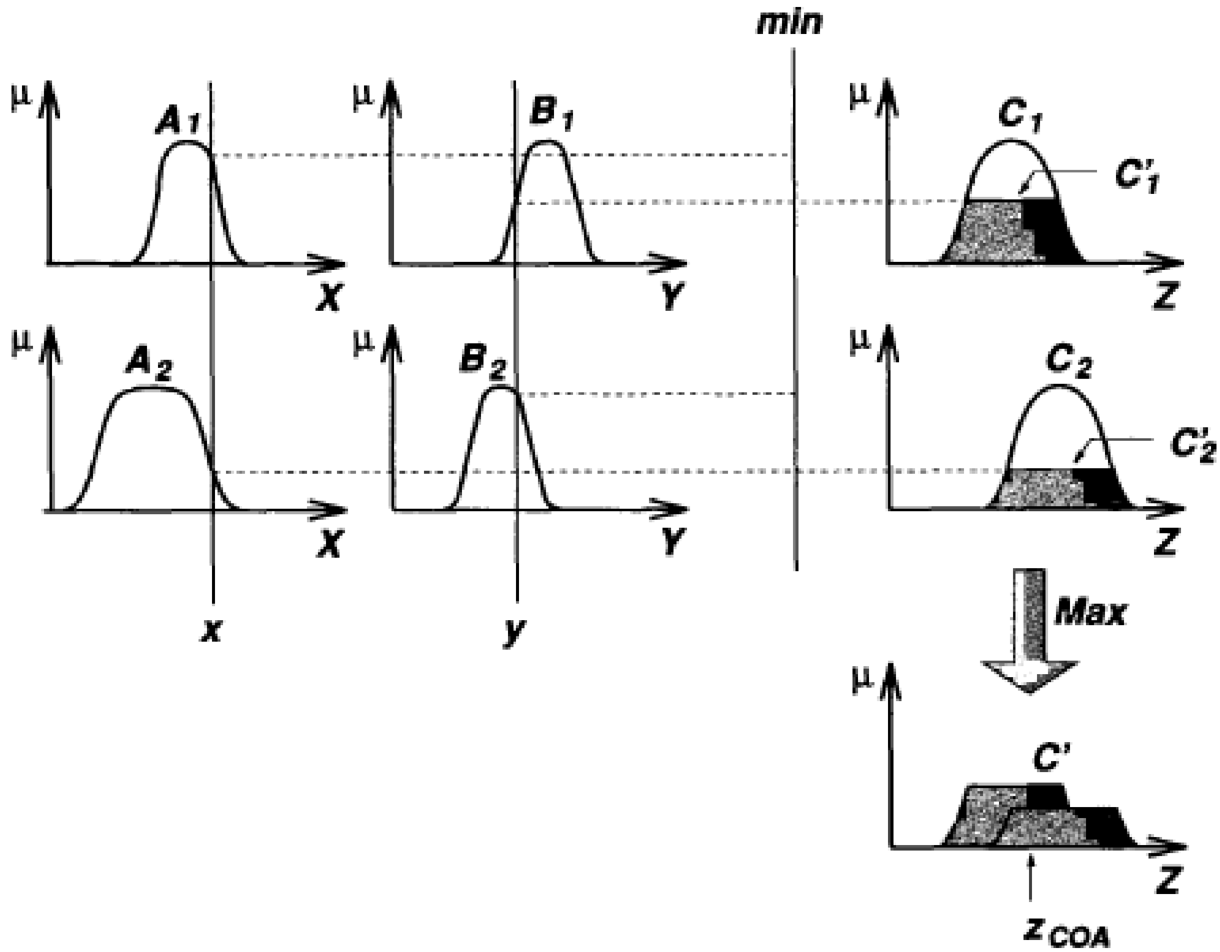
Παρακάτω ακολουθούν παραδείγματα με ασαφή συστήματα εξαγωγής συμπερασμάτων (ΑΣΕΣ-FIS) τύπου Mamdani και Sugeno και μια περιγραφή τους.

2.6: ΑΣΕΣ Τύπου Mamdani

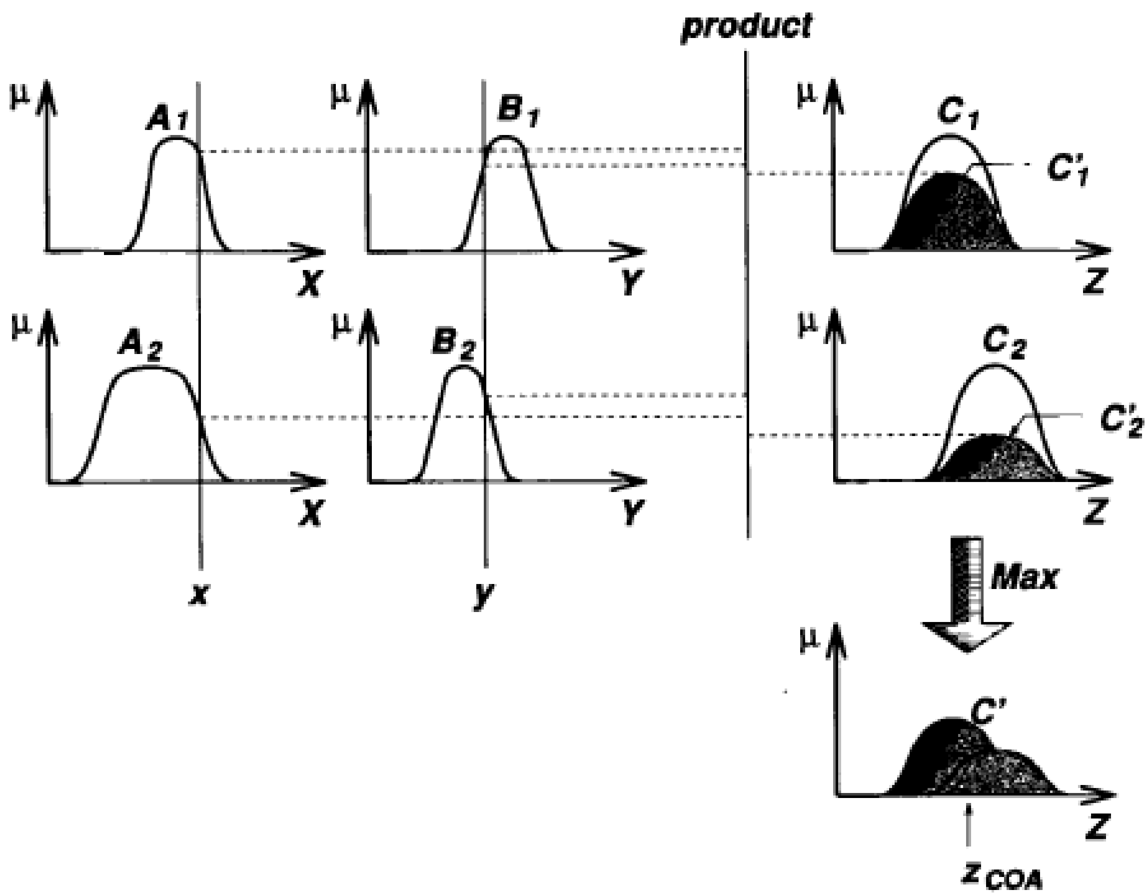
Το ΑΣΕΣ τύπου Mamdani είχε προταθεί σε μια πρώτη προσπάθεια να ελεγχθεί ένας συνδυασμός μιας ατμομηχανής και ενός μπόιλερ με ένα σύνολο γλωσσικών κανόνων ελέγχου αντλούμενων από έμπειρους χειριστές (ανθρώπους). Στο παρακάτω σχήμα φαίνεται πως ένα ΑΣΕΣ Mamdani δυο κανόνων μας δίνει το τελικό αποτέλεσμα z όταν υπόκειται σε δυο εισόδους x και y.

Εάν υιοθετήσουμε μέγιστο και αλγεβρικό γινόμενο σαν επιλογή μας για την T-norm και T-conorm, και χρησιμοποιήσουμε σύνθεση μέγιστου παραγόμενου αντί για σύνθεση μεγίστου-ελαχίστου τότε το αποτέλεσμα ασαφούς συλλογιστικής φαίνεται στο σχήμα που ακόλουθεί όπου η έξοδος του κάθε κανόνα είναι ένα ασαφές σύνολο που κλιμακώνεται προς τα κάτω μέσω αλγεβρικού γινομένου. Άλλες παραλλαγές είναι πιθανές αν χρησιμοποιήσουμε διαφορετικές T-norm και T-conorm .
(Παράρτημα: Ασαφείς πράξεις με norms) .

Στην εφαρμογή Mamdani δυο ΑΣΕΣ χρησιμοποιούνται σαν δυο ελεγκτές για να παράγουν την τιμή της θερμότητας ως είσοδο στον μπόιλερ και να ελέγξουν το άνοιγμα του κυλίνδρου της μηχανής , αντίστοιχα, για να ρυθμίσουν την πίεση του ατμού στο μπόιλερ και την ταχύτητα της μηχανής. Από την στιγμή που σαν είσοδοι στο σύστημα είναι καθορισμένες τιμές (crisp values) πρέπει να χρησιμοποιήσουμε έναν αποασαφοποιητή (defuzzifier) για την μετατροπή ενός ασαφούς συνόλου σε μια καθορισμένη τιμή.



Σχήμα2.14 : Το ΑΣΕΣ Mamdani που χρησιμοποιεί min και max για T-norm και T-conorm αντίστοιχα. . (Παράρτημα Β ασαφείς πράξεις με norms)

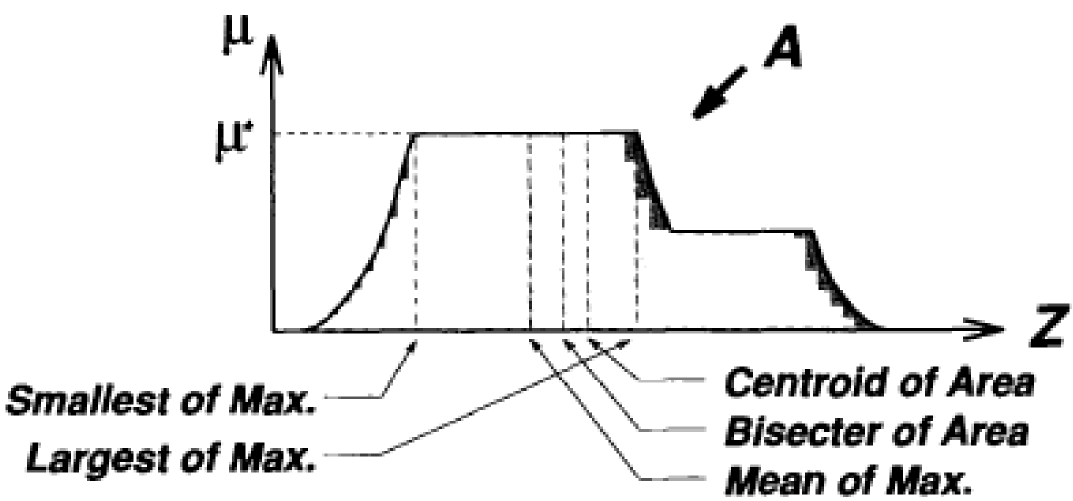


Σχήμα2.15: Το ΑΣΕΣ Mamdani που χρησιμοποιεί παραγόμενο(product) και max για T-norm και T-conorm αντίστοιχα. (Παράρτημα :Ασαφείς πράξεις με norms)

Αποασαφοποίηση (defuzzification):

Η αποασαφοποίηση, όπως προαναφέρθηκε, είναι ένας τρόπος αναπαράστασης μιας καθορισμένης τιμής (crisp) από ένα ασαφές σύνολο ως μια αντιπροσωπευτική τιμή.

Παρακάτω αναπαρίστανται σχηματικά οι πέντε προαναφερθείσες μέθοδοι αποασαφοποίησης :



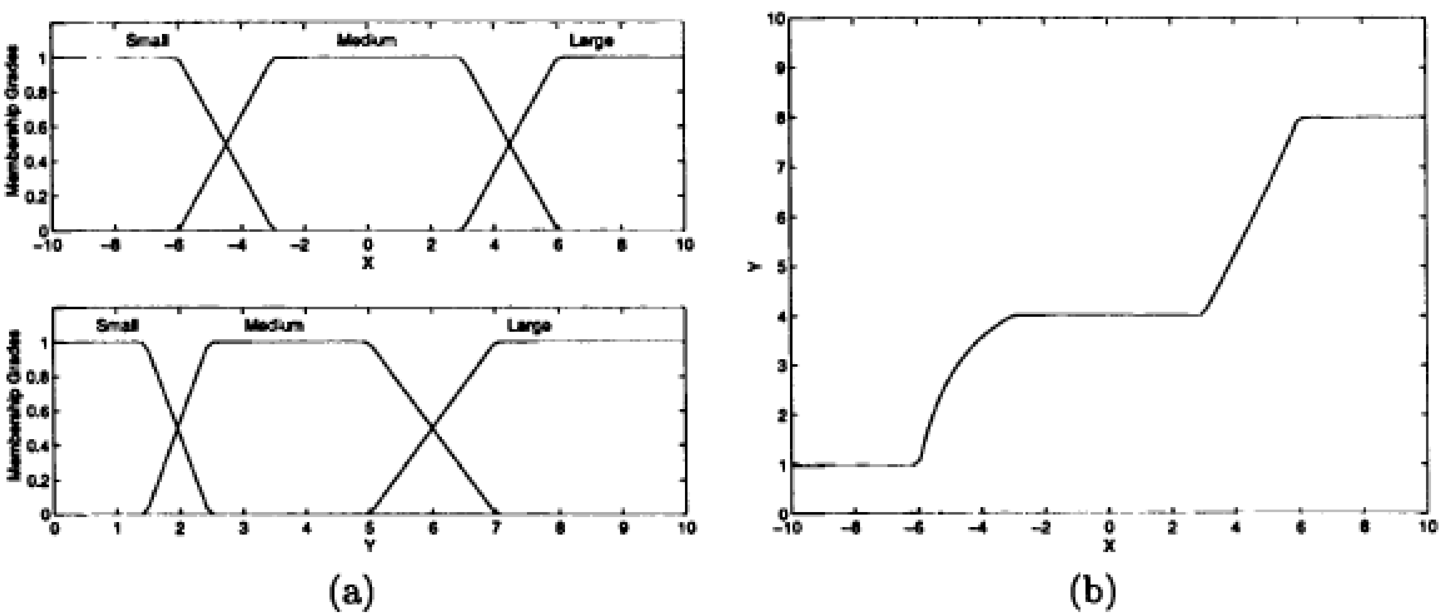
Σχήμα2.16 : Διάφορα σχήματα αποασαφοποίησης για την εξαγωγή ενός crisp συνόλου.

Παρακάτω ακολουθούν δύο σχηματικά παραδείγματα ΑΣΕΣ τύπου Mamdani με τους κανόνες τους και τις συναρτήσεις συμμετοχής.

Παράδειγμα 1: ΑΣΕΣ Mamdani μονής-εισόδου μονής εξόδου

Παράδειγμα με τρεις κανόνες (στο παρακάτω σχήμα ακολουθούν οι κανόνες και το διάγραμμα συναρτήσεων συμμετοχής :

$$\begin{cases} \text{If } X \text{ is small then } Y \text{ is small.} \\ \text{If } X \text{ is medium then } Y \text{ is medium.} \\ \text{If } X \text{ is large then } Y \text{ is large.} \end{cases}$$

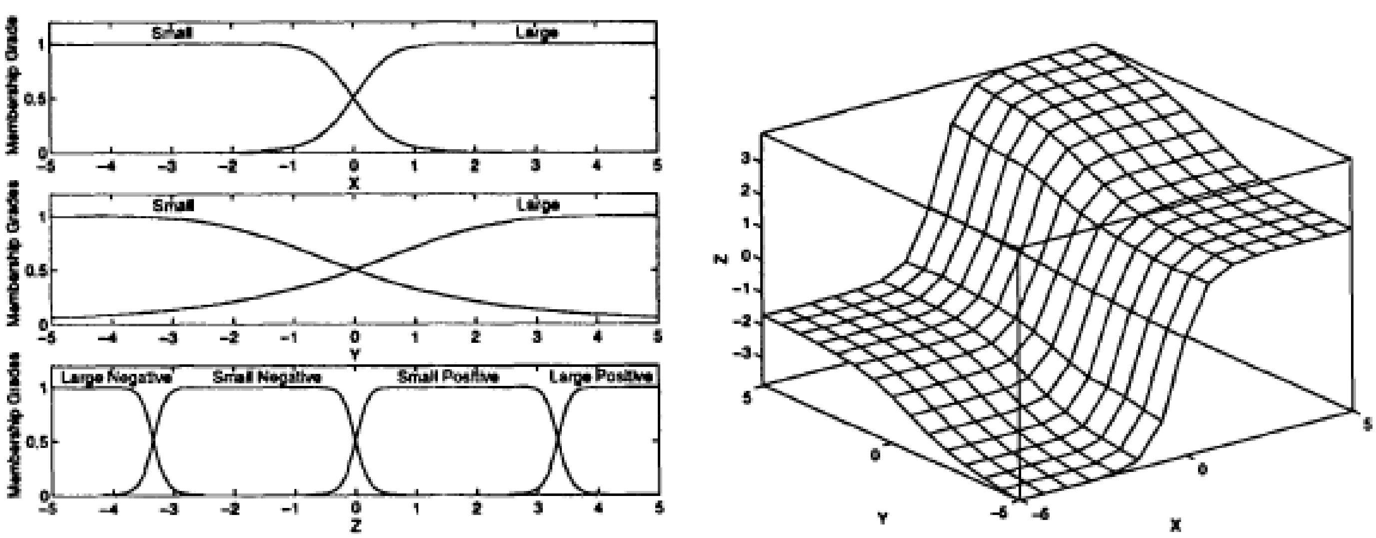


Σχήμα 2.17 : ΑΣΕΣ Mamdani μονής-εισόδου μονής εξόδου
A) antecedent and consequent MFs
B) Συνολική καμπύλη εισόδου-εξόδου

Παράδειγμα 2: ΑΣΕΣ Mamdani δυο-εισόδων μονής- εξόδου .

Ένα παράδειγμα ΑΣΕΣ Mamdani δυο-εισόδων μονής- εξόδου με τέσσερεις κανόνες μπορεί να όπως φαίνεται παρακάτω και στο σχήμα που ακολουθεί :

$$\begin{cases} \text{If } X \text{ is small and } Y \text{ is small then } Z \text{ is negative large.} \\ \text{If } X \text{ is small and } Y \text{ is large then } Z \text{ is negative small.} \\ \text{If } X \text{ is large and } Y \text{ is small then } Z \text{ is positive small.} \\ \text{If } X \text{ is large and } Y \text{ is large then } Z \text{ is positive large.} \end{cases}$$

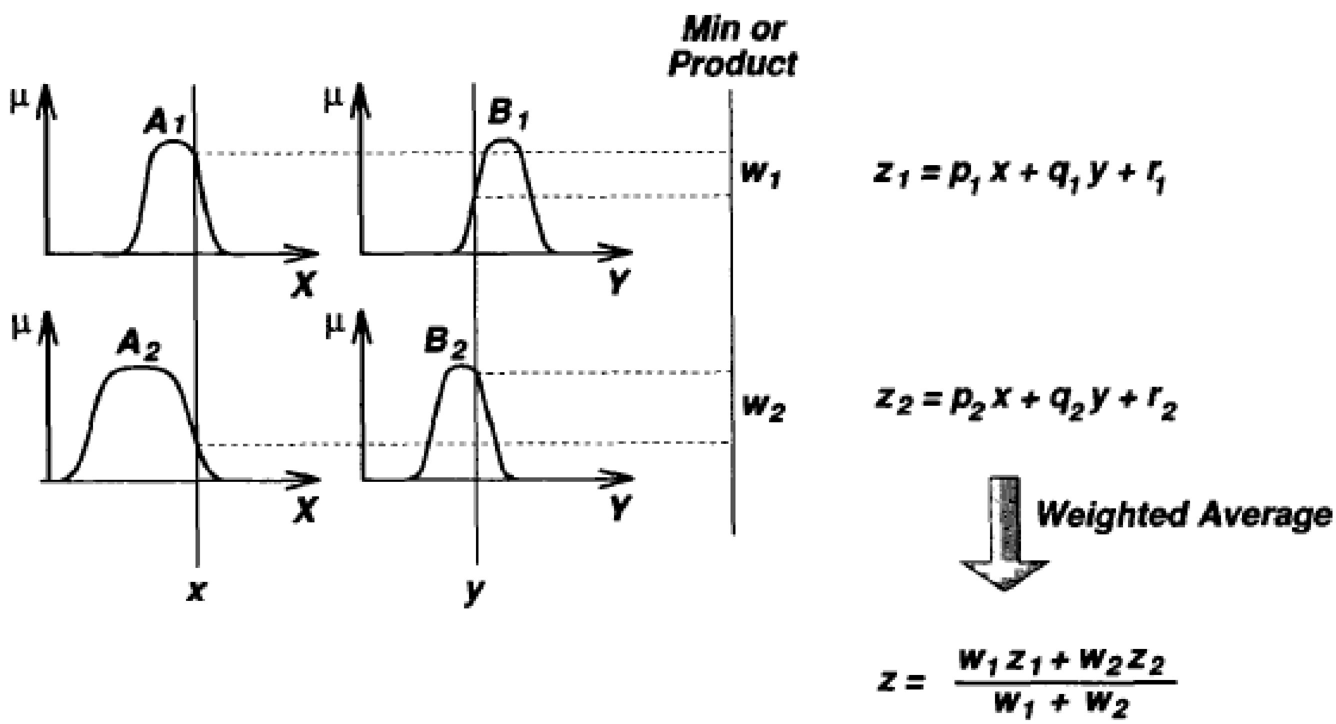


(a) (b)

Σχήμα 2.18: ΑΣΕΣ τύπου Mamdani δυο-εισόδων μονής- εξόδου Α) antecedent and consequent MFs Β) συνολική καμπύλη εισόδου-εξόδου

2.7: ΑΣΕΣ Τύπου Sugeno

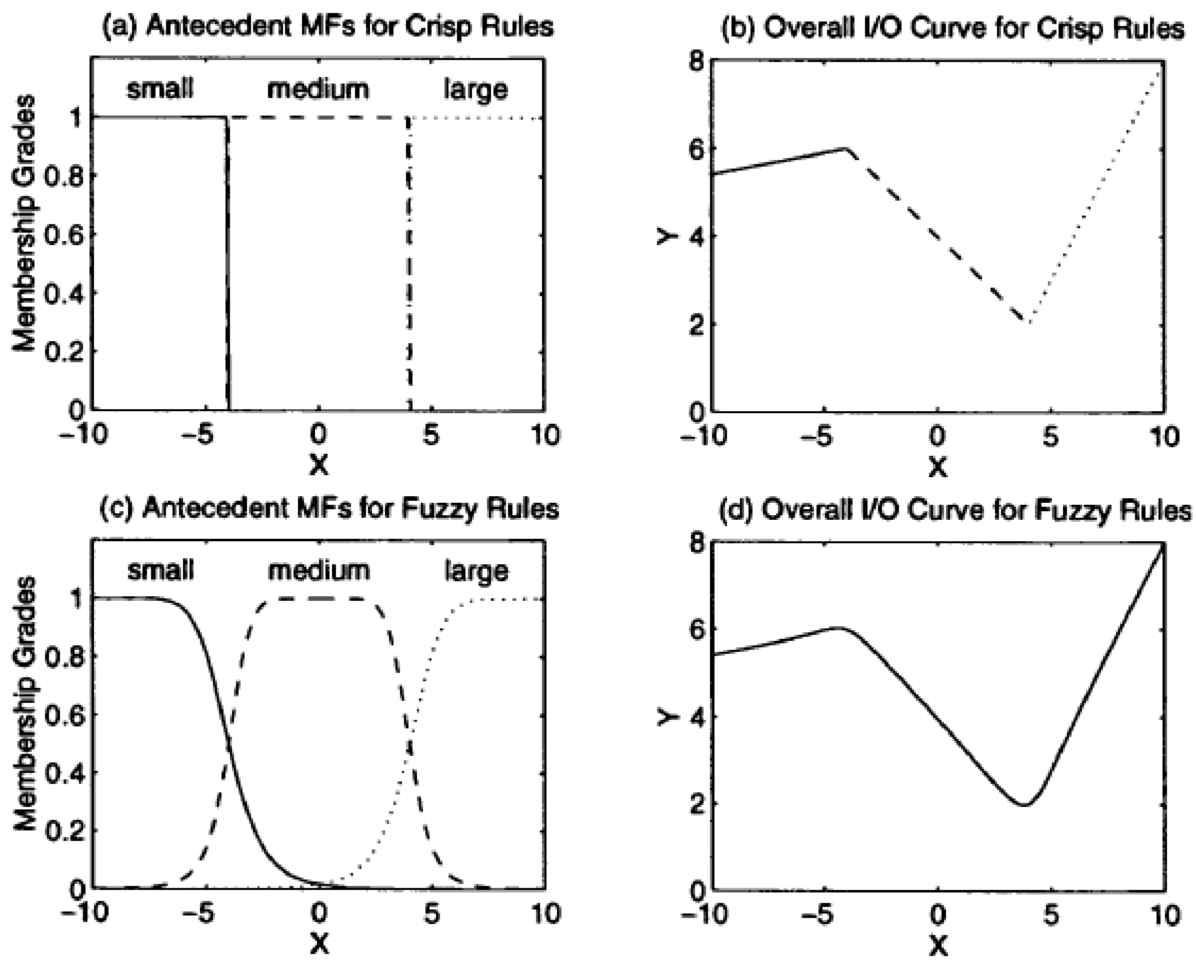
Το ασαφές μοντέλο του Sugeno (γνωστό και ως TSK fuzzy model) προτάθηκε από τους Takagi, Sugeno,και Kang σε μια προσπάθεια να αναπτυχθεί μια συστηματική προσέγγιση για την παραγωγή ασαφών κανόνων από ένα δοσμένο σύνολο εισόδου-εξόδου.



Σχήμα2.19: Το ασαφές μοντέλο Sugeno

Παράδειγμα2.7.1: Ασαφές και μη ασαφές σύνολο κανόνων –μια σύγκριση
Ένα παράδειγμα μονής εισόδου ασαφούς μοντέλου τύπου Sugeno μπορεί να εκφραστεί ως (και σε σχήμα παρακάτω) :

$$\begin{cases} \text{If } X \text{ is small then } Y = 0.1X + 6.4. \\ \text{If } X \text{ is medium then } Y = -0.5X + 4. \\ \text{If } X \text{ is large then } Y = X - 2. \end{cases}$$

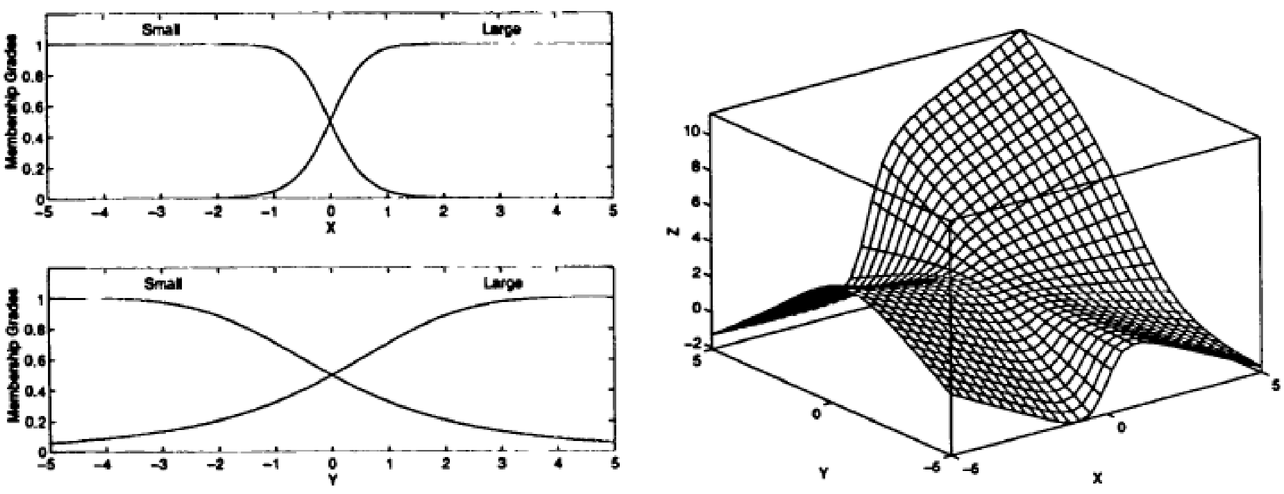


Σχήμα 2.20:Σύγκριση μεταξύ ασαφών και μη ασαφών κανόνων:
a) Συναρτήσεις συμμετοχής που προηγούνται b) καμπύλες εισόδου εξόδου για τους μη ασαφής κανόνες c) Συναρτήσεις συμμετοχής που προηγούνται d) καμπύλες εισόδου εξόδου για τους ασαφής κανόνες

Παράδειγμα 2.7.2: Δύο –εισόδων μονής –εξόδου ασαφές μοντέλο Sugeno

Ένα παράδειγμα δύο –εισόδων μονής –εξόδου ασαφούς μοντέλου Sugeno με τέσσερεις κανόνες μπορεί να εκφραστεί όπως φαίνεται παρακάτω και με τη μορφή διαγράμματος στο σχήμα που ακολουθεί.

$$\begin{cases} \text{If } X \text{ is small and } Y \text{ is small then } z = -x + y + 1. \\ \text{If } X \text{ is small and } Y \text{ is large then } z = -y + 3. \\ \text{If } X \text{ is large and } Y \text{ is small then } z = -x + 3. \\ \text{If } X \text{ is large and } Y \text{ is large then } z = x + y + 2. \end{cases}$$



(a) (b)

Σχήμα 2.21:Ασαφές μοντέλο Sugeno δύο –εισόδων μονής –εξόδου
a) Συναρτήσεις συμμετοχής εισόδου (antecedent) και εξόδου (consequent)
b) συνολική επιφάνεια εισόδου-εξόδου

ΚΕΦΑΛΑΙΟ 3: ΤΕΧΝΗΤΑ ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

3.1:Εισαγωγή Στα τεχνητά Νευρωνικά Δίκτυα.

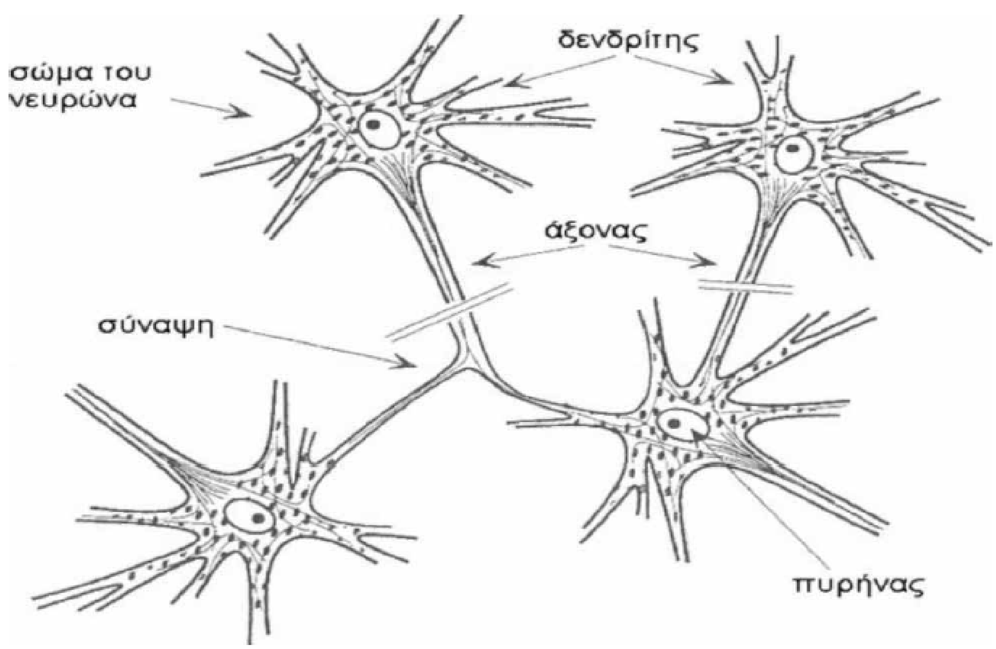
Το ανθρώπινο μυαλό , ο ανθρώπινος νους σε αντίθεση με τους σύγχρονους υπολογιστές (P.C.) , δεν είναι τόσο ικανό στην ταχεία επίλυση περίπλοκων μαθηματικών πράξεων. Από την άλλη όμως ο ανθρώπινος εγκέφαλος έχει ιδιότητες που ένας σύγχρονος οικιακός ηλεκτρονικός υπολογιστής δεν έχει. Όπως για παράδειγμα την αναγνώριση και ταυτοποίηση προσώπων , εικόνων αλλά και αντικειμένων ακόμα και όταν αυτά δεν είναι ιδιαιτέρα ευδιάκριτα ή ακόμα και όταν είναι παραμορφωμένα. Όπως για παράδειγμα σε συνθήκες χαμηλής ορατότητας, κακού φωτισμού.

Από το παραπάνω πολύ απλό παράδειγμα γίνεται ξεκάθαρο ότι ο ανθρώπινος εγκέφαλος έχει την ικανότητα της μάθησης λόγω εμπειρίας και Αυτό αποτελεί ένα από τα κύρια χαρακτηριστικά της ανθρώπινης ευφυΐας . Είναι ξεκάθαρο ότι η ανθρώπινη μνήμη έχει την ικανότητα να αποθηκεύει μεγάλη ποσότητα, αλλά και ποικιλία γνώσης και να συσχετίζει πληροφορίες με μεγάλη ευκολία., χωρίς ιδιαίτερη προσπάθεια, αλλά και αρκετά γρήγορα. Από την άλλη ο υπολογιστής έχει την ικανότητα να αποθηκεύει τεράστιες ποσότητες δεδομένων (οι μοντέρνοι οικιακοί υπολογιστές έχουν πια χωρητικότητες μνήμης που ξεπερνούν το 1 Terabyte), χωρίς όμως να μπορεί να της εκμεταλλευτεί ανάλογα. Μια πιθανή εξήγηση για αυτή την διαφορά έχει να κάνει με τον διαφορετικό τρόπο με τον οποίο λειτουργούν ο εγκέφαλος και ο ηλεκτρονικός υπολογιστής. Ο άνθρωπος είναι πιο έξυπνος λογά του τρόπου με τον οποί λειτουργεί ο εγκέφαλος του. Ο εγκέφαλος χρησιμοποιεί μια αρχιτεκτονική λειτουργίας που είναι πιο κατάλληλη στη φυσική επεξεργασία πληροφοριών. Ο ανθρώπινος εγκέφαλος δρα σαν ένας πολύ ισχυρός υπολογιστής ο οποίος αποτελείται από παρά πολλούς άλλους μικρότερους (απλούστερους) υπολογιστές ή αλλιώς νευρώνες (neurons), καθένας από τους οποίους επικοινωνεί με χιλιάδες άλλους. Η διάταξη των νευρώνων είναι τέτοια ώστε να μπορούν να επεξεργάζονται πληροφορίες ταυτόχρονα.

Η παραπάνω ιδιότητα των βιολογικών δικτύων ενέπνευσε την έρευνα γύρω από τα Τεχνητά Νευρωνικά Δίκτυα (ΤΝΔ), τα οποία εφαρμόζονται στης μέρες μας με επιτυχία τόσο στην επεξεργασία εικόνας , στον έλεγχο βιομηχανικών διεργασιών , στην αναγνώριση φωνής , αλλά και στις επικοινωνίες. Πρώτου γίνει αναφορά στα ΤΝΔ ας γίνει μια συνοπτική περιγραφή των βιολογικών νευρωνικών δικτύων.

3.2:Βιολογικά Νευρωνικά Δίκτυα

Ο ανθρώπινος εγκέφαλος αποτελείται περίπου από εκατό δισεκατομμύρια κύτταρα τους νευρώνες. Καθένας από αυτούς παρουσιάζει περίπου 1000-10000 συνδέσεις με άλλους νευρώνες. Το σχήμα που ακολουθεί παρουσιάζει μια απλή ,τυπική δομή ενός βιολογικού νευρώνα. Ο νευρώνας αποτελείται από τρία κύρια μέρη. Το κυτταρικό σώμα (cell body), πολλές δενδροειδής αποφύσεις που ονομάζονται δενδρίτες (dendrites) και τον άξονα.



Σχήμα 3.1: Απλοποιημένη απεικόνιση τεσσάρων νευρώνων.

Η λειτουργία ενός νευρώνα είναι αρκετά πολύπλοκη , αλλά οι βασικές της αρχές είναι απλές και αρκετά κατανοητές. Οι νευρώνες αλληλοεπηρεάζονται μέσω ηλεκτρικών σημάτων. Οι δενδρίτες λαμβάνουν αυτά τα σήματα ενώ ο άξονας τα μεταφέρει σε άλλους νευρώνες. Το σημείο σύνδεσης του άξονα με τον δενδρίτη ενός νευρώνα καλείται σύναψη (synapse). Τα σήματα εισόδου από τους νευρώνες μεταδίδονται είτε μέσω ενισχυτικών είτε μέσω κατασταλτικών συνάψεων. Οι ενισχυτικές συνάψεις αυξάνουν την πιθανότητα διέγερσης ενός νέρωνα ενώ οι κατασταλτικές την μειώνουν. Οι δενδρίτες προωθούν τα σήματα στο σώμα οπού και αθροίζονταν το άθροισμα αυτό είναι μεγαλύτερο από το δυναμικό ενεργοποίησης του άξονα τότε ο νευρώνας διεγείρεται. Τότε τα άκρα του άξονα στέλνουν σήμα σε άλλες συνάψεις. Μετά από την ενεργοποίηση και την αποστολή του σήματος ο νευρώνας απενεργοποιείται έτσι ώστε να είναι σε θέση ενεργοποίησης από επόμενο σήμα.

Τα βασικά χαρακτηρίστηκα των νευρωνικών δικτύων είναι η Αρχιτεκτονική τους και οι λειτουργικές τους ιδιότητες. Η αρχιτεκτονική έχει να κάνει με τη δομή του δικτύου , δηλαδή τον αριθμό των τεχνητών νευρώνων και τον τρόπο σύνδεσης τους. Τα νευρωνικά δίκτυα αποτελούνται από πολλούς διασυνδεδεμένους νευρώνες οι οποίοι αποτελούν τα στοιχεία επεξεργασίας (processing elements) με παρόμοια χαρακτηριστικά όπως εισόδους , εξόδους, συνοπτικά βάρη , ενεργοποίηση και τάση πόλωσης (bias). Η νευροδυναμική των νευρωνικών δικτύων καθορίζει τις ιδιότητες τους , όπως τον τρόπο με τον οποίο μαθαίνει, θυμάται, συσχετίζει και συγκρίνει συνεχώς τη νέα πληροφορία με την υπάρχουσα γνώση, ταξινομεί τη νέα πληροφορία και δημιουργεί νέες ταξινομήσεις αν αυτό είναι απαραίτητο.

Τα νευρωνικά δίκτυα δεν επεξεργάζονται την πληροφορία σειριακά αλλά με τη μέθοδο της Παράλληλης Ανάλυσης της σύνθετης πληροφορίας στα βασικά της στοιχεία .

Για παράδειγμα ένα βασικό χρώμα μπορεί να διαχωριστεί στις βασικές του συχνότητες και πλάτη και ανά πάσα στιγμή να δημιουργηθεί ξανά από αυτές. Όμοιος ένα νευρωνικό δίκτυο αποσυνθέτει την σύνθετη πληροφορία στα βασικά της στοιχεία . Αυτά τα στοιχεία και η μεταξύ τους συσχέτιση αποθηκεύονται σε ένα μέρος του εγκεφάλου. Για παράδειγμα όταν ένας άνθρωπος βλέπει μια εικόνα ο εγκέφαλος του δεν αποθηκεύει στη μνήμη του ένα πίνακα από εικονοστοιχεία όπως γίνεται στην περίπτωση του ηλεκτρονικού υπολογιστή, αλλά αποθηκεύει στοιχειώδη χαρακτηριστικά όπως γραμμές σχήματα και χρώματα.

3.3:Τεχνητά Νευρωνικά Δίκτυα

Τα τεχνητά νευρωνικά δίκτυα (ΤΝΔ) είναι στην ουσία απλοποιημένα ισοδύναμα μοντέλα του κεντρικού ανθρώπινου νευρικού συστήματος. Είναι δίκτυα που αποτελούνται από απλά μη γραμμικά στοιχεία τους νευρώνες , οι οποίοι αλληλοσυνδέονται με τέτοιο τρόπο ώστε να έχουν την ικανότητα επεξεργασίας πληροφοριών ταυτόχρονα. Κάθε νευρώνας στέλνει σήματα διέγερσης ή αποδιέγερσης σε άλλους νευρώνες. Η κατάσταση ενεργοποίησης ενός νευρώνα εξαρτάται από τα σήματα τα οποία δέχεται ο νευρώνας αυτός από τους νευρώνες με τους οποίους βρίσκεται συνδεδεμένος.

Τα ΤΝΔ χωρίζονται σε δυο κατηγορίες:

-Στα στατικά δίκτυα τα οποία δεν έχουν στόχευα με μνήμα αλλά μπορούν να έχουν ως εισόδους προηγούμενες τιμές εισόδων.

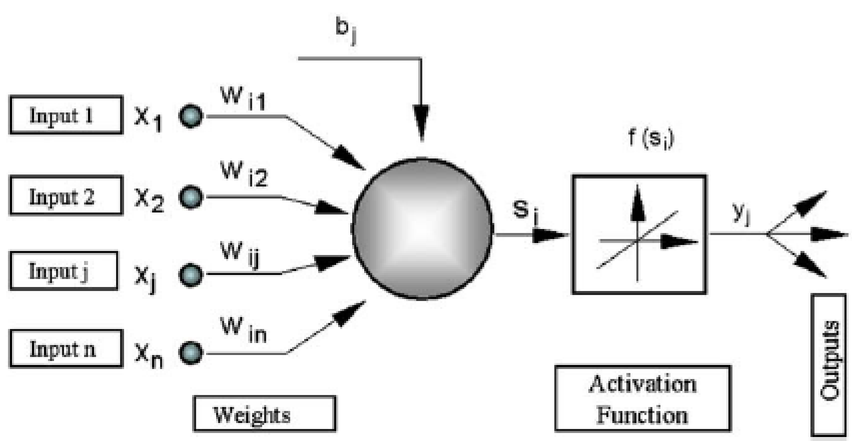
-Στα δυναμικά δίκτυα με στοιχεία μνήμης που είναι κατάλληλα για την πραγματοποίηση δυναμικών συστημάτων.

3.3.1:Βασικό Μοντέλο Ενός Νευρώνα.

Ας εξεταστεί η βασική μονάδα ενός ΤΝΔ ο νευρώνας (κόμβος) ο οποίος παρουσιάζεται στο σχήμα που ακολουθεί. Η συμπεριφορά ενός στατικού νευρώνα μπορεί να δοθεί από την παρακάτω εξίσωση.

$$f(\sum_{i=1}^{n} w_i x_i + \beta) \quad ή \quad f(\sum_{i=1}^{n+1} w_i x_i) \quad με \; x_{n+1} = 1 \; και \; \beta = w_{n+1}$$

Όπου w_i είναι τα συναπτικά βάρη (synaptic weights) χ οι είσοδοι του νευρώνα και β η τάση πόλωσης (bias). Θετική τιμή ενός βάρους αναπαριστά διέγερση της εισόδου του νευρώνα ενώ αντίθετα αρνητική τιμή ενός βάρους αναπαριστά αποδιέγερση της εισόδου του νευρώνα. Άρα η απόλυτη τιμή του βάρους καθορίζει την ισχύ της σύνδεσης.



Σχήμα3.2: Αναπαράσταση ενός τεχνητού νευρώνα. (Dombayaci & Golcu, 2009)

Το σταθμισμένο άθροισμα των βαρών ενεργοποιεί μία συνάρτηση μεταφοράς η οποία είναι συνήθως μη γραμμική. Μια συνήθης μορφή της συνάρτησης μεταφοράς είναι η μορφή με το χαρακτήρα της λογικής μονάδος κατωφλιού (Threshold logic unit TLU). Ο(πότε και η έξοδος σκανδαλίζεται όταν η τιμή του σταθμισμένου βάρους υπερβεί την τιμή πόλωσης (bias) β . Άλλες παραλλαγές της συνάρτησης μεταφοράς είναι:

Η γραμμική (Linear):

f(σ)=σ

Η δυαδική (binary):

$$f(\sigma)=\begin{cases}1, & \varepsilon\alpha\nu\ \sigma>0\\ 0, & \varepsilon\alpha\nu\ \sigma<0\end{cases}$$

Η σιγμοειδής ή Λογιστική (sigmoid):

$$f(\sigma)=\frac{1}{1+e^{-\sigma}}\in[0,1]$$

Η σχέση με υπερβολική εφαπτομένη :

$$f(\sigma)=\frac{1-e^{-\sigma}}{1+e^{-\sigma}}\cdot\frac{1}{2}(1+\tanh\sigma)$$

Η σχέση τύπου Percepton :

$$f(\sigma) = \begin{cases} \sigma, & \varepsilon\alpha\nu\ \sigma > 0 \\ 0, & \varepsilon\alpha\nu\ \sigma < 0 \end{cases}$$

Αν ο νευρώνας είναι δυναμικός , με μνήμη, τότε το σταθμισμένο άθροισμα παίρνει την παρακάτω μορφή:

$$\sigma(k) = \sigma(k-1) + \sum_{i=1}^{n+1} w_i x_i(k)$$

Όπου στην περίπτωση αυτή εισέρχεται στη σχέση ο δείκτης διακριτού χρόνου k και πρέπει να αποθηκευτεί στη μνήμη το προηγούμενο άθροισμα σ(k-1). Τότε η έξοδος u του νευρώνα αποτελεί είσοδο για κάποιον άλλο νευρώνα ή κάποιους άλλους νευρώνες .

3.4:Τοπολογίες Νευρωνικών Δικτύων

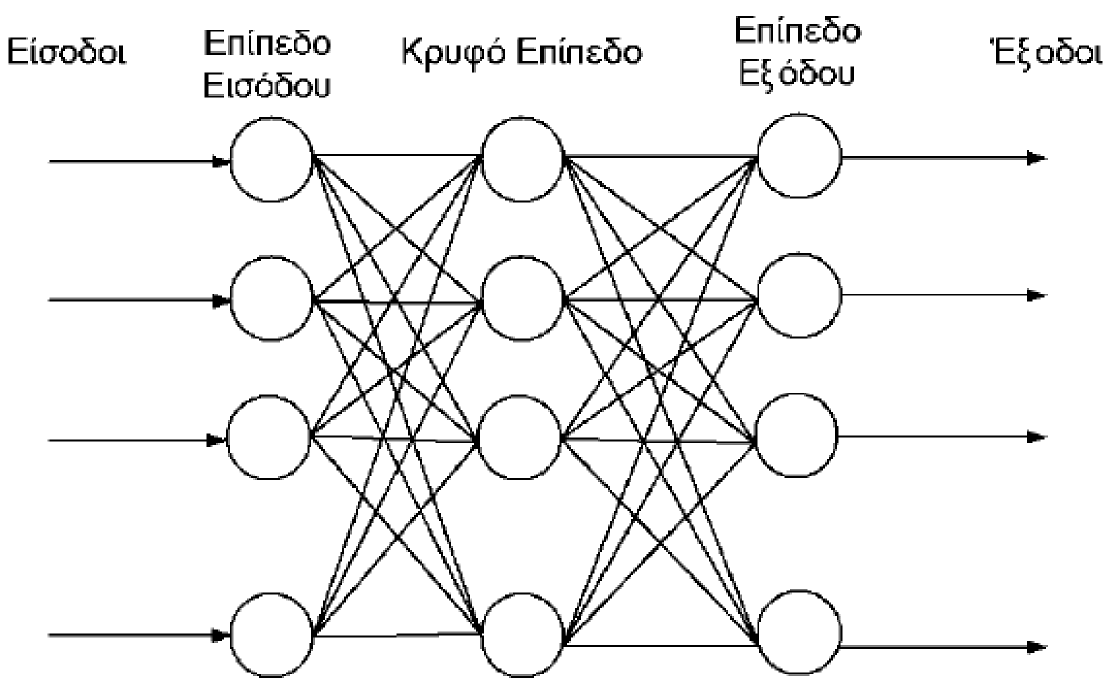
Τα τεχνητά νευρωνικά δίκτυα αποτελούνται από συμπλέγματα νευρώνων δομημένους κατά στρώματα ή επίπεδα (layers). Μπορεί να αποτελούνται είτε από ένα στρώμα (μονοστρωματικά δίκτυα) είτε από περισσότερα (πολυστρωματικά δίκτυα). Στην περίπτωση των πολυστρωματικών δικτύων το κατώτερο στρώμα περιλαμβάνει τους κόμβους εισόδων ενώ το υψηλότερο στρώμα περιλαμβάνει τους κόμβους της εξόδου. Τα ενδιάμεσα στρώματα ονομάζονται κρυφά στρώματα (hidden layers) και δεν έχουν άμεση επικοινωνία με το περιβάλλον.

3.4.1:Μονοστρωματικό Δίκτυο Τύπου Percepton

Το πιο απλό είδος νευρωνικού δικτύου είναι το μονοστρωματικό δίκτυο τύπου Percepton. Σε αυτό τον τύπο δικτύων υπολογίζεται το σταθμισμένο άθροισμα των εισόδων σε κάθε κόμβο. Αν η τιμή του αθροίσματος ξεπεράσει την τιμή ενός κατωφλίου , η οποία έχει οριστεί εκ των προτέρων , τότε ο νευρώνας ενεργοποιείται. Στη βιβλιογραφία κατά κύριο λόγο ο παραπάνω όρος αναφέρεται σε δίκτυα με έναν κόμβο.

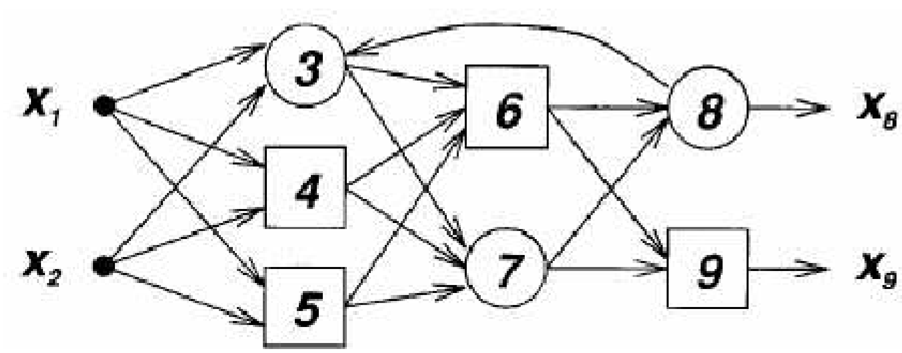
3.4.2: Πολυστρωματικά Δίκτυα (Multy-layer Networks).

Η πιο απλή τοπολογία πολυστρωματικών δικτύων είναι αυτή των δικτύων εμπρόσθιας τροφοδότησης (feed-forward networks) . Στο συγκεκριμένο τύπο δικτύων η ροη της πληροφορίας γίνεται μόνο προς τα εμπρός , δηλαδή από το στρώμα εισόδου προς το στρώμα εξόδου δια μέσω των κρυφών-ενδιάμεσων στρωμάτων, αν υπάρχουν τέτοια, χωρίς καμία ανατροφοδότηση. Στο σχήμα που ακολουθεί παρουσιάζεται μια απεικόνιση ενός τέτοιου δικτύου.



Σχήμα 3.3:Δίκτυο εμπρόσθιας τροφοδότησης.

Μία άλλη κατηγορία πολυστρωματικών δικτύων είναι τα δίκτυα με ανατροφοδότηση (feedback Networks). Σε τέτοιου τύπου δίκτυα υπάρχει τουλάχιστον ένας κλειστός βρόχος (closed loop) από τον οποίο δεδομένα από ανώτερα στρώματα τροφοδοτούνται στα κατώτερα στρώματα. Τέτοιου είδους δίκτυα είναι τα δίκτυα Hopfield στα οποία όλες οι συνδέσεις είναι συμμετρικές. Ένα τυπικό σχήμα νευρωνικού δικτύου με ανατροφοδότηση δίνεται παρακάτω.



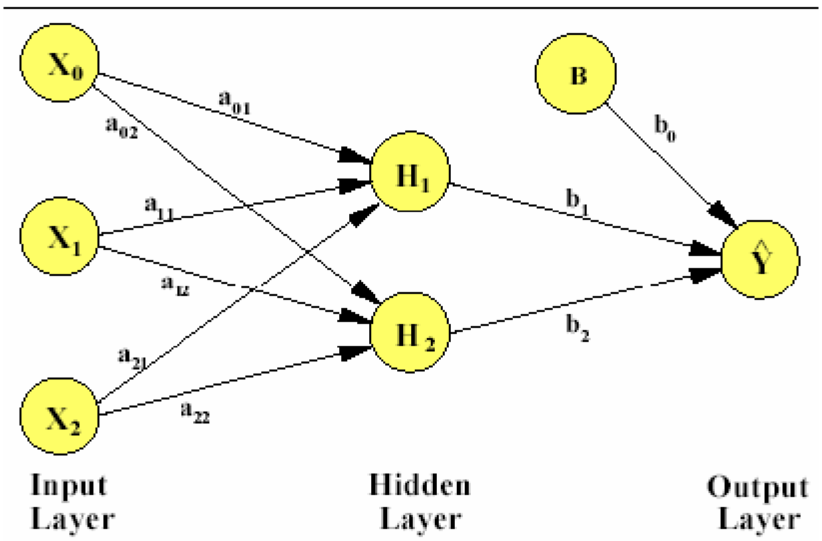
Σχήμα 3.4:Δίκτυο με ανατροφοδότηση

3.4.3: Νευρωνικά Δίκτυα Με Κρυμμένα Επίπεδα

Τα δίκτυα που περιγράφηκαν μέχρι στιγμής είχαν μια πολύ απλή δομή δύο επιπέδων που συνέδεαν τις εισόδους με τις εξόδους. Στις πραγματικές εφαρμογές, η αρχιτεκτονική του δικτύου είναι πιο πολύπλοκη. Οι ερευνητές σχεδόν πάντα σχεδιάζουν δομές που περιλαμβάνουν ένα ή περισσότερα *κρυμμένα επίπεδα*, όπως στο σχήμα 2.6. Σ' αυτό το σχήμα τα aij δηλώνουν τα βάρη που συνδέουν την είσοδο *i* με την κρυμμένη μονάδα *j*. Υποθέτουμε ότι Χο είναι ο όρος σταθερού βάρους (δηλαδή ο συντελεστής ανάσχεσης) για τις κρυμμένες μονάδες, ενώ το Β είναι ο όρος σταθερού βάρους για τη μονάδα εξόδου. Σε αντίθεση με τις μονάδες εισόδου και εξόδου, οι κρυμμένες μονάδες δεν αναπαραστούν κάποια πραγματική έννοια. Δεν έχουν ούτε ερμηνεία ούτε νόημα. Είναι απλά το ενδιάμεσο αποτέλεσμα στη διαδικασία υπολογισμού της τιμής εξόδου. Συνεπώς δεν έχουν αντίστοιχη έννοια στην οικονομετρία. Οι κρυμμένες μονάδες συμπεριφέρονται όπως οι μονάδες εξόδου, δηλαδή υπολογίζουν το σταθμικό άθροισμα των μεταβλητών εισόδου και έπειτα επεξεργάζονται το αποτέλεσμα χρησιμοποιώντας τη συνάρτηση ενεργοποίησης, που είναι σχεδόν πάντα η συνάρτηση logistic. Στο δίκτυο που απεικονίζεται στο σχήμα 6, το αποτέλεσμα που παράγουν οι κρυμμένες μονάδες θα είναι:

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

$$H_2 = f(a_{02} + a_{21}X_1 + a_{22}X_2) = \frac{1}{1 + e^{-(a_{02} + a_{21}X_1 + a_{11}X_1)}} \tag{9}$$



Σχήμα 3.5: *Το feedforward νευρωνικό δίκτυο με ένα κρυμμένο επίπεδο*

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

$$Y = b_0 + b_1H_1 + b_2H_2$$

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

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

$$Y = f(\, b_0 + b_1H_1 + b_2H_2)$$

$$Y = \frac{1}{1+e^{-(b_0+b_1H_1+b_2H_2)}}$$

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

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

Δεν υπάρχει θεωρητική βάση για τον προσδιορισμό του αριθμού των κρυμμένων μονάδων ή επιπέδων στο δίκτυο. Βάσει της ιδιότητας γενικής εκτιμήτριας που περιγράφηκε παραπάνω, είναι λογική η χρήση μεγάλου αριθμού κρυμμένων μονάδων. Εντούτοις, εάν προστεθούν πολλά κρυμμένα επίπεδα, το δίκτυο μπορεί να γίνει επιρρεπές στο πρόβλημα της υπερεκπαίδευσης (*overfitting*) . Αυτό σημαίνει ότι το δίκτυο μπορεί να πετύχει μια υψηλότερη ακρίβεια πρόβλεψης για την περίοδο που εκτιμάται, αλλά θα παράγει φτωχές προβλέψεις όταν γίνει χρήση δεδομένων εκτός του δείγματος εκπαίδευσης. Επιπλέον, το πλήθος των βαρών του δικτύου αυξάνεται ραγδαία, καθώς προστίθενται περισσότερες κρυμμένες μονάδες, αυξάνοντας έτσι το χρόνο εκτίμησης του μοντέλου. Στην πράξη, η σχεδίαση της αρχιτεκτονικής του δικτύου είναι μια μονότονη διαδικασία δοκιμής και σφάλματος. Οι ερευνητές συνήθως εκτιμούν πολλά διαφορετικά δίκτυα και επιλέγουν εκείνο που οδηγεί στο μικρότερο σφάλμα πρόβλεψης.

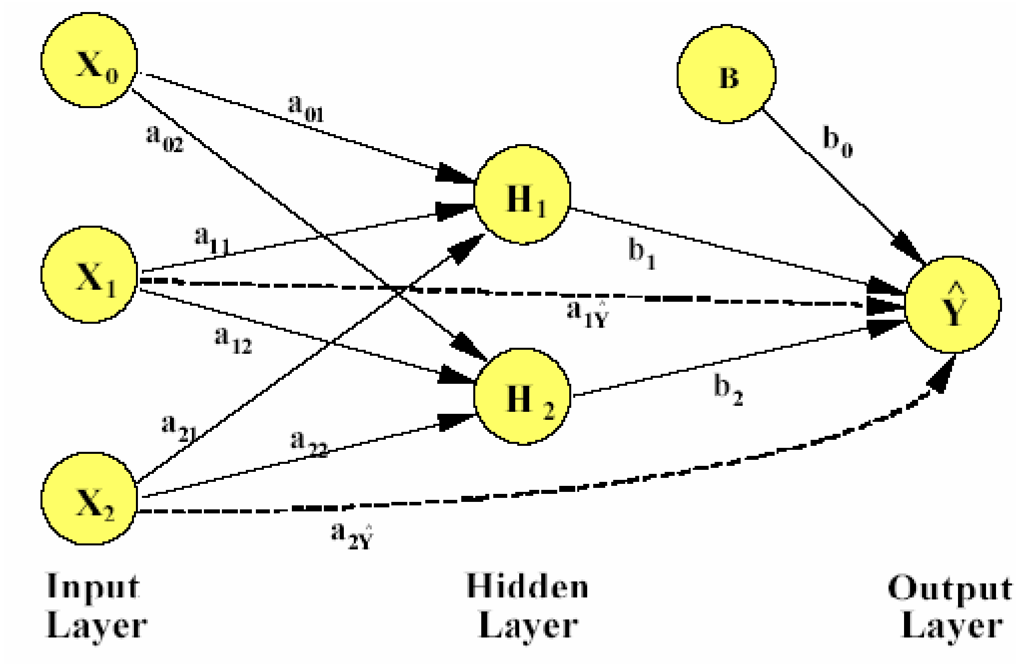
3.4.4: Αυξητικά Νευρωνικά Δίκτυα (Augmented Neural Networks)

Ένα ακόμη είδος νευρωνικών δικτύων μπορεί να κατασκευασθεί με την πρόσθεση άμεσων συνδέσεων από τις εισόδους προς την έξοδο (Σχήμα 2.7). Στην εν λόγω δομή, που καλείται *αυξητικό (augmented)* νευρωνικό δίκτυο**,** οι είσοδοι συνδέονται άμεσα με την έξοδο μέσω των βαρών a_{1Y} και a_{2Y}. Δεν είναι απαραίτητο ο όρος σταθερού βάρους Xo να συνδέεται με την έξοδο, καθώς η έξοδος έχει ήδη έναν όρο σταθερού βάρους B.

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

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

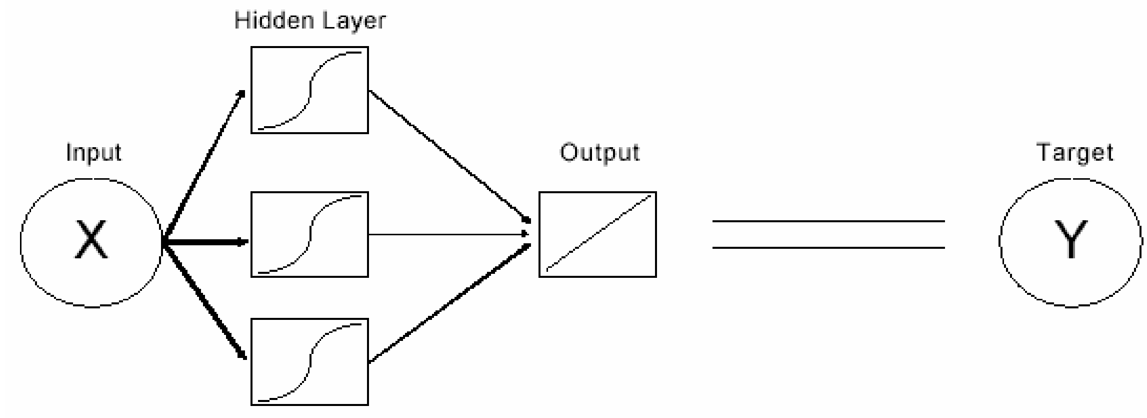
Επομένως το αυξητικό δίκτυο μπορεί να θεωρηθεί ως ένα πρότυπο μοντέλο γραμμικής παλινδρόμησης εμπλουτισμένο με μη γραμμικούς όρους. Εάν η εξαρτημένη μεταβλητή που μελετάμε δεν παρουσιάζει μη γραμμικές ιδιότητες, οι συντελεστές b1 και b2 ισοδυναμούν με 0, συνεπώς παίρνουμε ένα πρότυπο γραμμικό μοντέλο.



Σχήμα 3.6: *Το αυξητικό νευρωτικό δίκτυο*

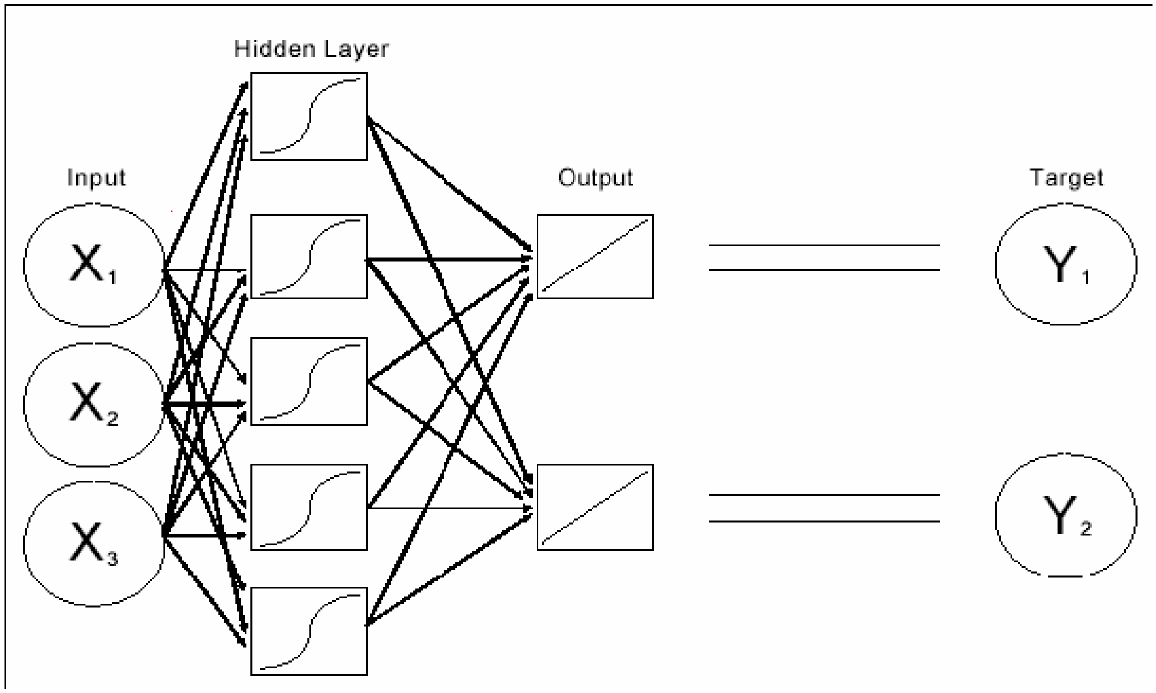
3.4.5: Νευρωνικά Δίκτυα Πολλαπλών Επιπέδων (MLP)

Μία από τις χρησιμότερες και πλέον επιτυχημένες εφαρμογές των νευρωνικών δικτύων στην ανάλυση δεδομένων και στην πρόβλεψη αποτελεί το μοντέλο πολλαπλών επιπέδων percepton (multilayer perceptron model) MLP. Τα μοντέλα πολλαπλών επιπέδων perceptron είναι μη γραμμικά μοντέλα νευρωνικών δικτύων που μπορούν να χρησιμοποιηθούν σχεδόν με όλες τις συναρτήσεις με υψηλό βαθμό ακρίβειας. Τα MLP περιέχουν ένα ή περισσότερα κρυμμένα επίπεδα νευρώνων που χρησιμοποιούν μη γραμμικές συναρτήσεις ενεργοποίησης, όπως συναρτήσεις logistic. Στο σχήμα 2.8 απεικονίζεται ένα MLP με ένα κρυμμένο επίπεδο και μία μόνο είσοδο και έξοδο. Το MLP στο σχήμα 2.8 αναπαριστά μια απλή μη γραμμική παλινδρόμηση.



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

Ο αριθμός των εισόδων και εξόδων στο MLP, μπορεί να διαχειριστεί κατάλληλα ώστε να αναλύει ποικίλους τύπους δεδομένων. Στο σχήμα 2.9 απεικονίζεται μια πολλαπλή μη γραμμική παλινδρόμηση πολλαπλών μεταβλητών.Η πολυπλοκότητα των MLP μπορεί να προσαρμοσθεί ανάλογα με τον αριθμό των κρυμμένων επιπέδων. Διαφορετικός αριθμός κρυμμένων επιπέδων δύναται να μετατρέψει ένα MLP από ένα απλό παραμετρικό μοντέλο σε ένα ευέλικτο μη παραμετρικό μοντέλο.



Σχήμα 3.8: *Perceptron πολλαπλών επιπέδων με πολλαπλές εισόδους και εξόδους*

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

Πριν ένα οποιοδήποτε ΤΝΔ μπορέσει να χρησιμοποιηθεί για την εκτέλεση κάθε επιθυμητής εργασίας, είναι αναγκαίο να εκπαιδευθεί. Βασικά, εκπαίδευση έιναι η διαδικασία του προσδιορισμού των βαρών τα οποία αποτελούν τα θεμελιώδη στοιχεία του ΤΝΔ. Η γνώση που αποκτάται από το δίκτυο αποθηκεύεται στα τόξα (arcs) και στους κόμβους με τη μορφή συναπτικών βαρών και των σταθερών βαρών (nodes biases) των κόμβων. Μέσω των συνδετικών τόξων (arcs) το ΤΝΔ μπορεί να εκτελέσει πολύπλοκες μη γραμμικές συναρτήσεις από τους κόμβους εισόδου προς τους κόμβους εξόδου. Η εκπαίδευση MLP είναι επιβλεπόμενη (supervised) κατά το ότι η επιθυμητή αντίδραση του δικτύου (target value) για κάθε υπόδειγμα εισόδου (παράδειγμα) είναι πάντα διαθέσιμη. Τα δεδομένα εισόδου εκπαίδευσης βρίσκονται σε μορφή διανυσμάτων των μεταβλητών εισόδου ή των υποδειγμάτων εκπαίδευσης. Σε κάθε στοιχείο του διανύσματος εισόδου αντιστοιχεί ένας κόμβος εισόδου στο επίπεδο εισόδου του δικτύου. Συνεπώς ο αριθμός των κόμβων εισόδου ισούται με τη διάσταση των διανυσμάτων εισόδου. Για ένα τυχαίο πρόβλημα πρόβλεψης, το πλήθος των κόμβων εισόδου είναι ορισμένο επακριβώς και είναι ο αριθμός των ανεξάρτητων μεταβλητών του προβλήματος. Ωστόσο, για ένα πρόβλημα πρόβλεψης χρονικών σειρών(time-series), ο κατάλληλος αριθμός των κόμβων εισόδου δεν είναι εύκολο να προσδιοριστεί.. Τα συνολικά διαθέσιμα δεδομένα συνήθως διανέμονται στο σύνολο εκπαίδευσης (in-sample data) και στο σύνολο δοκιμής (out-of-sample ή hold-out sample). Το σύνολο εκπαίδευσης χρησιμοποιείται για την εκτίμηση των συναπτικών βαρών, ενώ το σύνολο δοκιμής για τη μέτρηση της ικανότητας γενίκευσης του δικτύου. Η διαδικασία εκπαίδευσης έχει συνήθως ως εξής: Πρώτον, παραδείγματα των συνόλων εκπαίδευσης εισάγονται στους κόμβους εισόδου. Οι τιμές ενεργοποίησης των κόμβων εισόδου σταθμίζονται και συσσωρεύονται σε κάθε κόμβο στο πρώτο κρυμμένο επίπεδο. Έπειτα το σύνολο μορφοποιείται από τη συνάρτηση ενεργοποίησης στις τιμές ενεργοποίησης των κόμβων. Διαδοχικά αναπτύσσεται σε μια είσοδο μέσα στους κόμβους του επόμενου επιπέδου, έως ότου τελικά βρεθούν οι τιμές ενεργοποίησης των εξόδων. Ο αλγόριθμος εκπαίδευσης χρησιμοποιείται για τον προσδιορισμό των βαρών που ελαχιστοποιούν ορισμένα κριτήρια ολικού σφάλματος όπως το άθροισμα των τετραγωνικών σφαλμάτων (sum of squares errors SSE), ή των μέσων τετραγωνικών σφαλμάτων (mean squared errors MSE).

Συνεπώς, η εκπαίδευση του δικτύου είναι στην πράξη ένα μη γραμμικό πρόβλημα ελαχιστοποίησης χωρίς περιορισμούς. Στα περισσότερα πραγματικά προβλήματα, το νευρωνικό δίκτυο δεν είναι ποτέ 100% σωστό. Τα τεχνητά νευρωνικά δίκτυα είναι προγραμματισμένα να μαθαίνουν έως ένα δεδομένο όριο σφάλματος. Αφού το νευρωνικό δίκτυο μάθει έως το όριο σφάλματος, ο μηχανισμός προσαρμογής των βαρών τερματίζεται και το δίκτυο δοκιμάζεται σε γνωστές περιπτώσεις που δεν έχει ξανασυναντήσει. Η εφαρμογή των νευρωνικών δικτύων σε άγνωστες περιπτώσεις δίδει την πραγματική τιμή σφάλματος. Σε ένα καλά εκπαιδευμένο νευρωνικό δίκτυο το όριο σφάλματος και το πραγματικό σφάλμα πρέπει να ταυτίζονται.

3.4.6: Νευρωνικά Δίκτυα Υψηλής Τάξης

Τα νευρωνικά δίκτυα υψηλής τάξης έχει αποδειχθεί ότι έχουν εκπληκτικές ιδιότητες υπολογισμού, αποθήκευσης και μάθησης. Η απόδοσή τους οφείλεται στο γεγονός ότι η τάξη ή η δομή των νευρωνικών δικτύων υψηλής τάξης, μπορεί να προσαρμοσθεί στην τάξη ή τη δομή του προβλήματος. Συνεπώς το νευρωνικό δίκτυο που σχεδιάστηκε για συγκεκριμένη κατηγορία προβλημάτων εξειδικεύεται και γίνεται ακόμη πιο αποτελεσματικό στην επίλυση αυτών των προβλημάτων. Ένα νευρωνικό δίκτυο υψηλής τάξης ορίζεται ως μια υψηλής τάξης λογική μονάδα (High Order Threshold Logic Unit – HOTLU) η οποία περιλαμβάνει όρους με υψηλής τάξης βάρη. Συνήθως, αλλά όχι πάντα, η έξοδος του είναι (1,0) ή (+1,-1). Ένα υψηλής τάξης νευρωνικό δίκτυο στρωμάτων (slab) ορίζεται ως ένα σύνολο λογικών μονάδων υψηλής τάξης (HOTLUs).

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

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

όπου yi είναι η έξοδος της i-οστής νευρωνικής μονάδας υψηλής τάξης και S είναι η σιγμοειδής συνάρτηση. Tn (i) είναι ο όρος n-οστής τάξης για την i-οστή μονάδα και k είναι η τάξη της μονάδας. Ο όρος μηδενικής τάξης είναι ρυθμιζόμενο κατώφλι, που δηλώνεται ως Wo (i). Ο όρος n-οστής τάξης είναι ένα γραμμικό σταθμισμένο άθροισμα από τα n-οστής τάξης γινόμενα των εισόδων, παραδείγματα των οποίων είναι τα:

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

όπου x(j) είναι η j-οστή είσοδος στο νευρώνα i-οστής τάξης και Wn (i,j, . . .) είναι ένα ρυθμιζόμενο βάρος που συλλαμβάνει τη *n*-οστής τάξης συσχέτιση μεταξύ του *n*-οστής τάξης γινομένου των εισόδων και της εξόδου της μονάδας. Αρκετά στρώματα μπορούν να συνδεθούν σειριακά τροφοδοτώντας την έξοδο ενός στρώματος ως είσοδο ενός άλλου στρώματος, για την παραγωγή δικτύων πολλαπλών στρωμάτων. Τα sigma-pi νευρωνικά δίκτυα είναι πολυεπίπεδα δίκτυα που μπορούν να έχουν όρους υψηλής τάξης σε κάθε επίπεδο. Ο αλγόριθμος μάθησης για αυτά τα δίκτυα είναι ο γενικευμένος back-propagation. Ωστόσο, οι sigma-pi μονάδες, όπως αρχικά κατασκευάσθηκαν, δεν έχουν σταθερούς όρους βαρών, αν και είναι αρκετά απλό να ενσωματωθούν τέτοιες σταθερές στις εν λόγω μονάδες. Η διαδικασία μάθησης περιλαμβάνει την εφαρμογή μιας καθορισμένης συνάρτησης στο νευρωνικό δίκτυο μέσω επαναληπτικής προσαρμογής των βαρών βασισμένη σε έναν συγκεκριμένο κανόνα μάθησης και στην αντίδραση του δικτύου στα σύνολα εκπαίδευσης. Η συνάρτηση που θα μαθευτεί αναπαριστάται από ένα σύνολο παραδειγμάτων που αποτελείται από το πιθανό διάνυσμα εισόδου σε συνδυασμό με μία επιθυμητή έξοδο. Το σύνολο εκπαίδευσης είναι υποσύνολο του συνόλου όλων των πιθανών παραδειγμάτων της συνάρτησης. Η υλοποίηση της διαδικασίας μάθησης περιλαμβάνει διαδοχική παρουσίαση παραδειγμάτων σχεδίασης ως ζεύγη εισόδου - εξόδου στο δίκτυο, τα οποία προέρχονται από το σύνολο εκπαίδευσης. Έπειτα από κάθε παρουσίαση, τα βάρη του δικτύου προσαρμόζονται ώστε να συλλαμβάνουν τη συσχέτιση της δομής της σχεδίασης. Ένας συνηθισμένος κανόνας μάθησης μονού στρώματος είναι ο κανόνας perceptron, ο οποίος για τον κανόνα δεύτερης τάξης ενημέρωσης εκφράζεται ως:

$$W_2'(i,j,k) = W_2(i,j,k) + \left[t(i) - y(i)\right]x(j)x(k) \tag{3}$$

Εδώ t(i) είναι η επιθυμητή έξοδος και y(i) είναι η πραγματική έξοδος της i-οστής μονάδας για το x διάνυσμα εισόδου. Υπάρχουν παρόμοιοι κανόνες μάθησης για τους υπόλοιπους όρους Wi . Εάν το δίκτυο αποδώσει τη σωστή έξοδο για κάθε παράδειγμα εισόδου στο σύνολο εκπαίδευσης, λέμε ότι το δίκτυο συνέκλινε, ή ότι έμαθε το σύνολο εκπαίδευσης. Εάν, έπειτα από τη μάθηση του συνόλου εκπαίδευσης, το δίκτυο δώσει τη σωστή έξοδο για ένα σύνολο παραδειγμάτων του συνόλου εκπαίδευσης, τα οποία δεν έχει δει ακόμη, λέμε ότι το δίκτυο είναι σωστά γενικευμένο.

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

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

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

όπου το Np δηλώνει τον αριθμό των υποδειγμάτων του συνόλου εκπαίδευσης. Το σύνολο εκπαίδευσης εκφράζεται ως

$$\{(x^s, y^s),\ \ s \in (1, Np)\}$$

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

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

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

ή συνοπτικά:

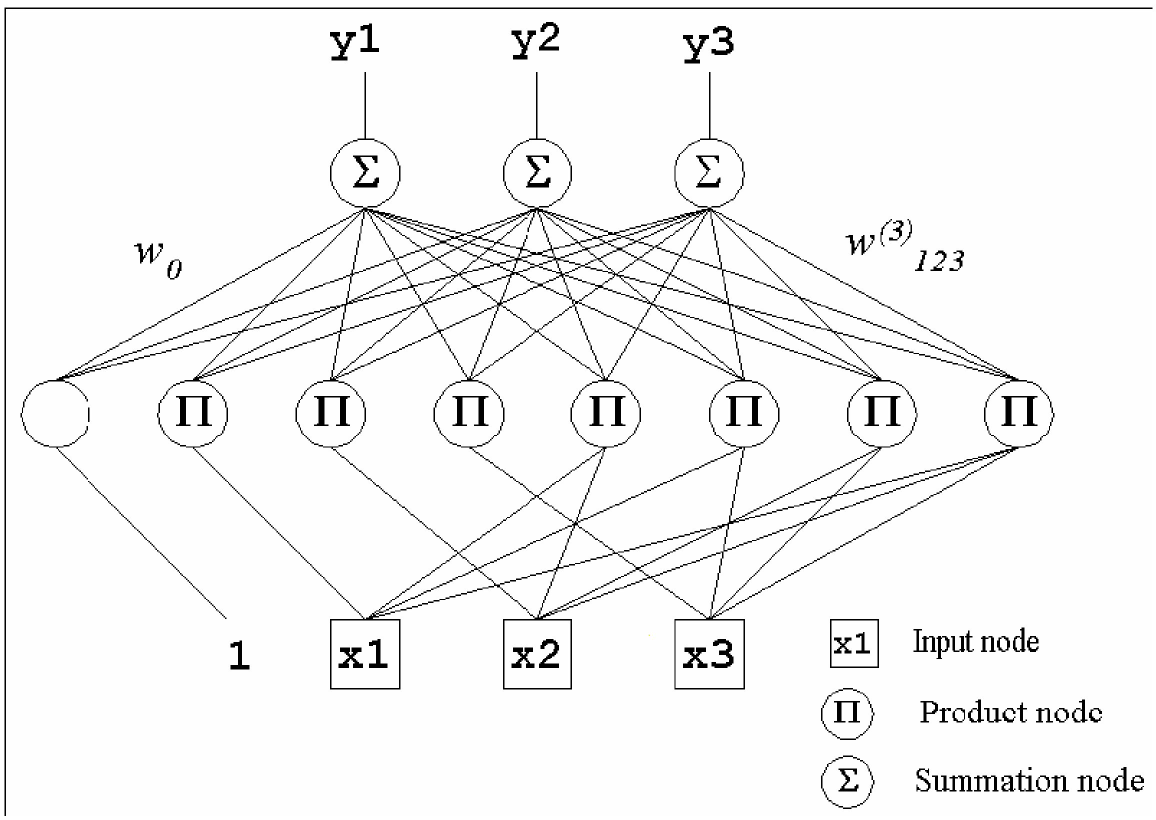
$$P(x) = w_0 + \sum_{i=1}^{M} w_i \prod_{j=1}^{d} x_j^r$$

όπου: wi είναι τα βάρη, M είναι ο μέγιστος αριθμός μονώνυμων, xi είναι τα χαρακτηριστικά του διανύσματος εισόδου, r = 0, 1, 2, . . . είναι η δύναμη με την οποία το j-οστό χαρακτηριστικό xi συμμετέχει στο i-οστό μονώνυμο. Η μη γραμμικότητα των συναρτήσεων δομείται από διαδοχικά επίπεδα τα οποία έχουν μονάδες sigma και pi. Στο σχήμα που ακολουθεί φαίνεται ένα νευρωνικό δίκτυο υψηλής τάξης το οποίο υπολογίζει τη συνάρτηση:

$P(x) = \sigma(w_0 + \sum_{i=1}^{M} w_i \sigma(\prod_{j=1}^{d} x_j^r))$ όπου x είναι το διάνυσμα εισόδου, w_i είναι τα βάρη, σ είναι η σιγμοειδής συνάρτηση:

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

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



Σχήμα 3.9: *Το πλήρως συνδεμένο νευρωνικό δίκτυο υψηλής τάξης*

3.5: Η Κατάλληλη Αρχιτεκτονική Για Τα Δίκτυα Υψηλής Τάξης

Το πλήθος των κρυμμένων μονάδων στα νευρωνικά δίκτυα υψηλής τάξης αυξάνεται εκθετικά με τον αριθμό των εισόδων. Ο αντικειμενικός σκοπός για την εύρεση της κατάλληλης αρχιτεκτονικής για τα HONN είναι η εξομάλυνση του προβλήματος της εκθετικής αύξησης και η γενίκευση του δικτύου, δηλαδή η μάθηση συναρτήσεων που παράγουν κατάλληλες αντιδράσεις σε άγνωστα παραδείγματα. Το πρόβλημα της μάθησης της τοπολογίας και των βαρών των HONN, για τη βελτίωση της ικανότητας γενίκευσης αντιμετωπίζεται με προσεγγίσεις που μπορούν να χωριστούν σε δύο ομάδες:

- *προσεγγίσεις κλαδέματος (pruning approache)s* που ξεκινούν με μεγάλα δίκτυα και κατά την εκπαίδευση διαγράφουν κρυμμένες μονάδες και βάρη.

- *προσεγγίσεις κατασκευής (constructive approaches)* που ξεκινούν με μικρά δίκτυα και καθώς αυξάνεται η ακρίβεια προσθέτουν κρυμμένες μονάδες και βάρη .

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

3.6: Σύγκριση Δικτύων Υψηλής Τάξης με Πολυεπίπεδα Δίκτυα

Τα νευρωνικά δίκτυα υψηλής τάξης χαρακτηρίζονται από διαφορετικές δυνατότητες ταξινόμησης:

•

Νευρωνικά Δίκτυα Υψηλής Τάξης (HONN)

- το *πρώτο επίπεδο* των HONN με τις μονάδες γινομένων δεύτερης τάξης έχει τη δυνατότητα να κατασκευάζει *κοίλες* και *κυρτές περιοχές (regions)*

Νευρωνικά Δίκτυα Πολλαπλών Επιπέδων (MLP)

- το *πρώτο επίπεδο* μαθαίνει ένα μόνο υπερεπίπεδο

- το *δεύτερο επίπεδο* κάνει τμηματικά γραμμικές και κυρτές *περιοχές*

- το *τρίτο επίπεδο* κάνει κοίλες και κυρτές *περιοχές*

3.7: Δίκτυα Sigma-Pi

Τα νευρωνικά δίκτυα sigma-pi (SPN) είναι feedforward δίκτυα στα οποία κάθε επίπεδο περιέχει όρους υψηλής τάξης. Τις περισσότερες φορές τα επίπεδα έχουν αθροιστικές μονάδες που τροφοδοτούνται μέσω σταθμικών συνδέσεων από τα ενδιάμεσα αποτελέσματα μονάδων γινομένων. Τα νευρωνικά δίκτυα sigma-pi αποτελούν σποραδικώς συνδεμένα νευρωνικά δίκτυα υψηλής τάξης. Οι ερευνητές περιορίζουν την πολυωνυμική τάξη (δηλαδή την τοπολογία του δικτύου) στην κατάλληλη διαμόρφωση ώστε να επιτυγχάνεται ο επιθυμητός βαθμός ακρίβειας χρησιμοποιώντας εκ των προτέρων γνώση για τη δεδομένη εργασία.

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

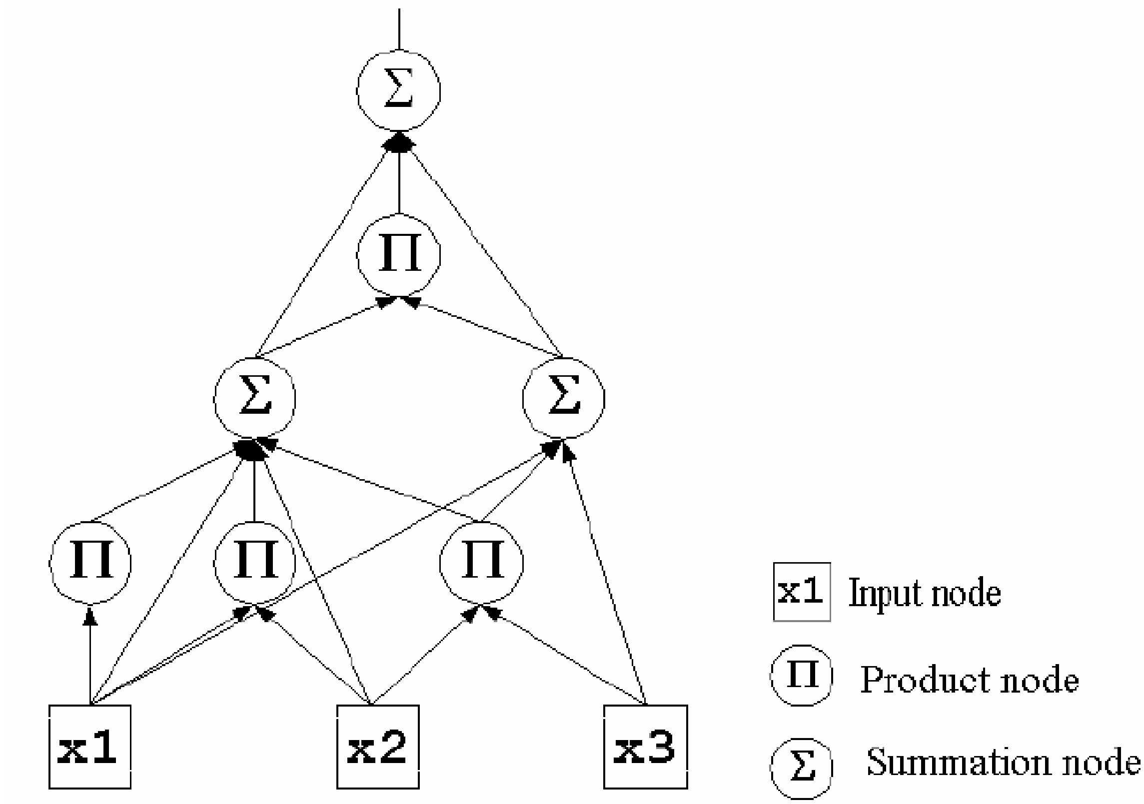
$$s = \sum_j w_i p_j$$

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

$$p_j = \prod_i w_i x_i$$

Η έξοδος των μονάδων διέρχεται μέσω της σιγμοειδούς συνάρτησης:

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



Σχήμα 3.10: *Το νευρωνικό δίκτυο sigma-pi*

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

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

όπου το s_k ορίζεται:

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

η παράγωγος του s_k είναι:

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

Συνεπώς το όφελος (benefit) είναι:

$$\beta_j = -\partial \mathrm{E}e / \partial s_j = o_j (1 - o_j) \sum_k \beta_k s_k'$$

3.7.1:**Αλγόριθμος Εκπαίδευσης backpropagation για τα sigma-pi δίκτυα**

Ακολουθεί μία σύντομη αναφορά στον αλγόριθμο εκπαίδευσης για τα παραπάνω δίκτυα..Σ .σ Ο αλγόριθμός θα αναλυθεί ενδελεχέστερα σε παρακάτω ενότητα.

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

Επανάληψη: Γ*ια κάθε* παράδειγμα εκπαίδευσης (x,y)

- πολλαπλασίασε τις εξόδους χρησιμοποιώντας τη σιγμοειδή συνάρτηση:

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

$$o_k = \sigma(s_k) = 1/(1+e_k^{-s}),$$

όπου:

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

-

πολλαπλασίασε το όφελος β_k στους κόμβους k στο επίπεδο εξόδου:

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

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

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

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

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

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

$$\beta_{j2} = o_j(1-o_j)[\textstyle\sum_k \beta_k s'_{j2k}]$$ // αποτελέσματα από πολλαπλούς κόμβους στο επόμενο επίπεδο

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

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

$$\Delta w_{ij1} = \eta \beta_{j1} o_{i1} \text{ και } \Delta w_{0j} = \eta \beta_j$$

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

$$\Delta w_{ij3} = \eta \beta_{j3} o_{i1} o_{i2} o_{i3}$$

- ενημέρωσε τα βάρη με τις μεταβολές που υπολόγισες:

$$w = w + \Delta w$$

ωσότου να ικανοποιηθεί η συνθήκη τερματισμού.

3.7.2: Δίκτυα Pi-Sigma

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

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

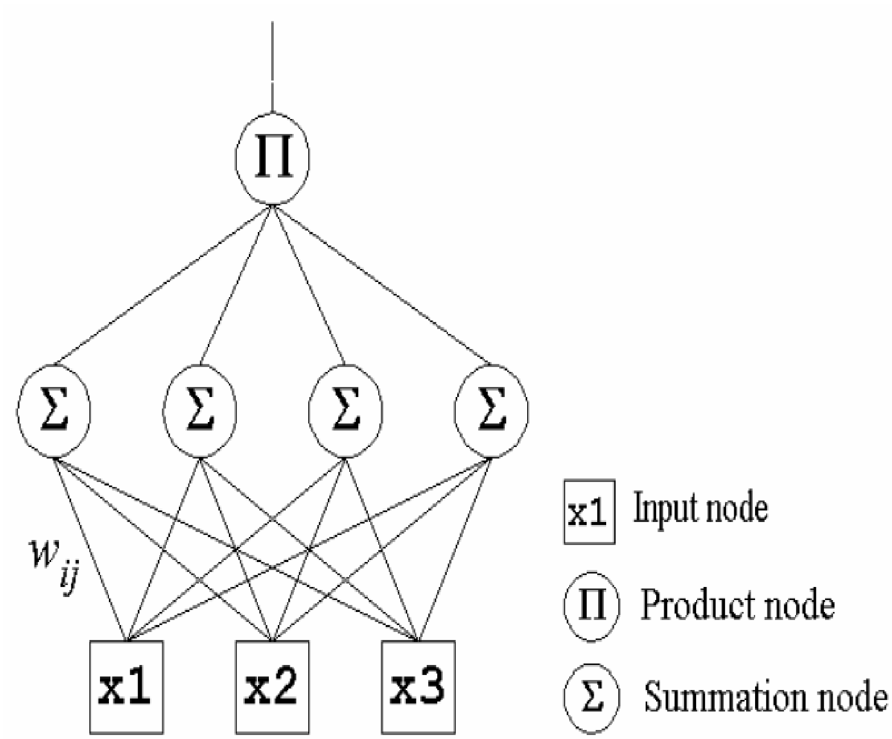
$$P(x) = \sigma(\prod_{i=1}{}^{M} (\sum_{j=1}{}^{d} w_j x_j + w_0))$$

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

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

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

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



Σχήμα3.11: *Το νευρωνικό δίκτυο Pi-Sigma*

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

$$\begin{aligned}\partial Ee/\partial s_j &= (\partial Ee/\partial o_k)(\partial o_k/\partial s_j)\\ &= (y_k - o_k)(\partial o_k/\partial p_j)(\partial p_j/\partial s_j)\\ &= (y_k - o_k)o_k(1-o_k)(p_i')\\ &= (y_k - o_k)o_k(1-o_k)(\prod_{ii=/=j}{}^{M} h_i)\end{aligned}$$

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

και η παράγωγος $\partial p_j/\partial s_j) = p_i'$ είναι: $p_j' = \prod_{i=/=j}{}^{M} h_i$

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

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

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

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

Επανάληψη:

για κάθε παράδειγμα εκπαίδευσης (x_e,y_e)

- πολλαπλασίασε την έξοδο:

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

- εάν ο perceptron δεν ανταποκριθεί σωστά *ενημέρωσε* τα βάρη *ωσότου* να ικανοποιηθεί η συνθήκη τερματισμού

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

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

3.8: Μεθοδολογία Κατασκευής Τεχνητών Νευρωνικών Δικτύων Για Πρόβλεψη

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

3.8.1: Η Αρχιτεκτονική του Δικτύου

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

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

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

3.8.2: Ο Αριθμός Των Κρυμμένων Επιπέδων και Κόμβων

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

Επηρεασμένοι από θεωρητικές εργασίες που καταδεικνύουν ότι ένα μόνο κρυμμένο επίπεδο είναι επαρκές ώστε το ΤΝΔ να εκτιμήσει οποιαδήποτε πολύπλοκη μη γραμμική συνάρτηση με την επιθυμητή ακρίβεια, οι περισσότεροι ερευνητές χρησιμοποιούν μόνο ένα κρυμμένο επίπεδο για προγνωστικούς σκοπούς. Ωστόσο, τα δίκτυα με ένα κρυμμένο επίπεδο μπορεί να απαιτούν πολύ μεγάλο πλήθος κρυμμένων κόμβων, το οποίο δημιουργεί αδυναμίες τόσο στο χρόνο εκπαίδευσης όσο και στην ικανότητα γενίκευσης. Δίκτυα με δύο κρυμμένα επίπεδα είναι πιθανόν να παρέχουν περισσότερα πλεονεκτήματα για ορισμένους τύπους προβλημάτων. Αρκετοί μελετητές ασχολήθηκαν με τέτοια προβλήματα και θεώρησαν περισσότερα από ένα κρυμμένα επίπεδα (συνήθως δύο κρυμμένα επίπεδα) κατά τη διαδικασία σχεδίασης του δικτύου. Η χρήση δύο κρυμμένων επιπέδων παρατηρήθηκε ότι σε ορισμένες περιπτώσεις είχε ως αποτέλεσμα πιο συμπαγή αρχιτεκτονική η οποία απέφερε μεγαλύτερη αποτελεσματικότητα στη διαδικασία εκπαίδευσης καθώς και την πραγματοποίηση προβλέψεων με μεγαλύτερη ακρίβεια απ’ ότι τα δίκτυα με ένα κρυμμένο επίπεδο. Ωστόσο στα δίκτυα με περισσότερα από δύο κρυμμένα επίπεδα δεν παρατηρείται κάποια περαιτέρω βελτίωση. Υπάρχουν αρκετοί λόγοι, για τους οποίους πρέπει να χρησιμοποιούμε όσο το δυνατό λιγότερα κρυμμένα επίπεδα:

1. Οι περισσότεροι αλγόριθμοι εκπαίδευσης είναι βαθμωτοί. Το επιπλέον επίπεδο μέσω του οποίου πρέπει να διορθωθούν προς τα πίσω τα σφάλματα προκαλεί αστάθεια στην κατάβαση. Η επιτυχία οποιουδήποτε βαθμωτού αλγόριθμου βελτιστοποίησης εξαρτάται από το βαθμό στον οποίο η κατάβαση παραμένει αμετάβλητη καθώς οι παράμετροι διαφοροποιούνται.

2. Ο αριθμός των τοπικών ελαχίστων αυξάνεται σημαντικά για πολλά κρυμμένα επίπεδα. Οι περισσότεροι βαθμωτοί αλγόριθμοι μπορούν να υπολογίσουν μόνο το τοπικό ελάχιστο, χάνοντας έτσι το γενικό. Ακόμη και αν ο αλγόριθμος είναι σε θέση να βρει το γενικό ελάχιστο, υπάρχει μεγάλη πιθανότητα έπειτα από πολλαπλές και χρονοβόρες επαναλήψεις, το αποτέλεσμα να κολλήσει στο τοπικό ελάχιστο.

Μολαταύτα, η χρήση δύο κρυμμένων επιπέδων μπορεί να αποφέρει καλύτερα αποτελέσματα για συγκεκριμένα προβλήματα, ιδιαίτερα όταν τα δίκτυα ενός επιπέδου αποτυγχάνουν να δώσουν ικανοποιητικό αποτέλεσμα λόγω του υπερβολικού πλήθους κρυμμένων κόμβων. Το ζήτημα του προσδιορισμού του βέλτιστου αριθμού των κρυμμένων κόμβων είναι πολύπλοκο και ιδιαίτερα σημαντικό. Γενικά, τα δίκτυα με λιγότερους κρυμμένους κόμβους προτιμούνται καθώς παρουσιάζουν συνήθως καλύτερη ικανότητα γενίκευσης και λιγότερα προβλήματα υπερεκπαίδευσης. Αλλά τα δίκτυα με πολύ λίγους κρυμμένους κόμβους ίσως να μην έχουν την απαιτούμενη δύναμη μοντελοποίησης και μάθησης των δεδομένων. Δεν υπάρχει θεωρητική βάση για την επιλογή της εν λόγω παραμέτρου, μολονότι έχουν αναφερθεί μερικές συστηματικές προσεγγίσεις (μέθοδος αποκοπής άχρηστων κρυμμένων κόμβων, μέθοδος πρόσθεσης κρυμμένων κόμβων, έρευνα πλέγματος). Ο συνηθέστερος τρόπος προσδιορισμού του πλήθους των κρυμμένων κόμβων είναι μέσω πειραμάτων ή με τη μέθοδο δοκιμής και σφάλματος. Αρκετοί εμπειρικοί κανόνες έχουν επίσης προταθεί, όπως ότι ο αριθμός των κρυμμένων κόμβων εξαρτάται από τον αριθμό των υποδειγμάτων εισόδου και κάθε βάρος πρέπει να έχει τουλάχιστον δέκα υποδείγματα εισόδου (μέγεθος δείγματος). Στην προσπάθεια αποφυγής του προβλήματος υπερεκπαίδευσης, ορισμένοι ερευνητές υιοθέτησαν εμπειρικούς κανόνες για τον περιορισμό του αριθμού των κρυμμένων κόμβων. Είναι επίσης γνωστοί ορισμένοι ευριστικοί περιορισμοί σχετικά με τον αριθμό των κρυμμένων κόμβων. Παραδείγματος χάριν στην περίπτωση των δημοφιλών δικτύων με ένα κρυμμένο επίπεδο, υπάρχουν αρκετές κατευθυντήριες γραμμές, όπως οι κανόνες “2n + 1”, “2n”, “n”, “n/2”, όπου n ο αριθμός των κόμβων εισόδου. Ωστόσο καμία από αυτές τις ευριστικές επιλογές δε λειτουργεί ικανοποιητικά για όλα τα προβλήματα.

3.8.3: **Ο Αριθμός των Κόμβων Εισόδου**

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

3.8.4: **Οι Αλληλοσυνδέσεις των Κόμβων**

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

3.8.5 :Η Συνάρτηση Ενεργοποίησης

Η συνάρτηση ενεργοποίησης καλείται επίσης και συνάρτηση μεταφοράς. Προσδιορίζει τη σχέση μεταξύ των εισόδων και εξόδων στους κόμβους και στο δίκτυο. Γενικά, η συνάρτηση ενεργοποίησης εισάγει ένα βαθμό μη γραμμικότητας που είναι πολύτιμος για τις περισσότερες εφαρμογές ΤΝΔ. Θεωρητικά και χωρίς να είμαστε ιδιαίτερα αυστηροί, κάθε διαφορίσιμη συνάρτηση μπορεί να λειτουργήσει ως συνάρτηση ενεργοποίησης. Στην πράξη, χρησιμοποιείται μόνο ένας μικρός αριθμός από συναρτήσεις ενεργοποίησης με «καλή συμπεριφορά» (πεπερασμένες, μονότονα αύξουσες και διαφορίσιμες).

Οι κυριότερες είναι:

1. Η σιγμοειδής (logistic) συνάρτηση:

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

2. Η συνάρτηση υπερβολικής εφαπτομένης (tanh):

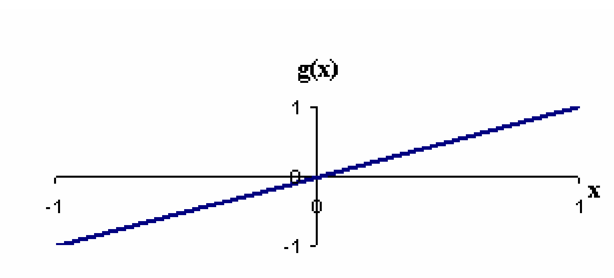
$$f(x) = (\exp(x) - \exp(-x)) / (\exp(x) + \exp(-x))$$

3. Η συνάρτηση ημίτονου ή συνημίτονου:

$f(x) = \sin(x)$ ή $f(x) = \cos(x)$

4. Η γραμμική συνάρτηση ή ταυτότητα:

$f(x) = x$



Σχήμα 3.12: *Η γραμμική συνάρτηση ή ταυτότητα*

Από τις παραπάνω, η συνάρτηση μεταφοράς logistic είναι η δημοφιλέστερη επιλογή. Υπάρχουν ορισμένοι ευρετικοί κανόνες για την επιλογή της συνάρτησης ενεργοποίησης. Για παράδειγμα, ορισμένοι ερευνητές προτείνουν τις logistic συναρτήσεις ενεργοποίησης για προβλήματα ταξινόμησης που περιλαμβάνουν μάθηση σχετικά με τη μέση συμπεριφορά και τις συναρτήσεις υπερβολικής εφαπτομένης εάν τα προβλήματα περιλαμβάνουν μάθηση σχετικά με αποκλίσεις από το μέσο όπως τα προβλήματα πρόβλεψης. Ωστόσο, δεν είναι σαφές κατά πόσον διαφορετικές συναρτήσεις ενεργοποίησης επιδρούν σημαντικά στην απόδοση του δικτύου. Γενικά, ένα δίκτυο μπορεί να έχει διαφορετικές συναρτήσεις ενεργοποίησης για διαφορετικούς κόμβους στο ίδιο ή σε διαφορετικά επίπεδα. Εντούτοις σχεδόν όλα τα δίκτυα χρησιμοποιούν τις ίδιες συναρτήσεις ενεργοποίησης ειδικά για τους κόμβους του ιδίου επιπέδου. Αν και στην πλειοψηφία των εφαρμογών χρησιμοποιείται η logistic συνάρτηση ενεργοποίησης για τους κρυμμένους κόμβους, δεν υπάρχει απόλυτη συμφωνία για το είδος της συνάρτησης ενεργοποίησης που πρέπει να χρησιμοποιείται στους κόμβους εξόδου. Αξιοσημείωτο είναι ότι όταν χρησιμοποιούνται μη γραμμικές συναρτήσεις στο επίπεδο εξόδου (logistic, υπερβολικής εφαπτομένης), οι επιθυμητές τιμές εξόδου συνήθως πρέπει να κανονικοποιηθούν ώστε να εναρμονίζονται με την κλίμακα των πραγματικών εξόδων του δικτύου, αφού ο κόμβος εξόδου με τη συνάρτηση logistic ή υπερβολικής εφαπτομένης έχει τυπικό πεδίο [0,1] ή [-1,1] αντίστοιχα. Συμβατικά, η logistic συνάρτηση ενεργοποίησης δείχνει κατάλληλη για τους κόμβους εξόδου για πολλά προβλήματα ταξινόμησης όπου οι επιθυμητές τιμές είναι συχνά δυαδικές. Ωστόσο, για ένα πρόβλημα πρόβλεψης που περιλαμβάνει συνεχείς επιθυμητές τιμές, είναι λογικό να χρησιμοποιηθεί η γραμμική συνάρτηση ενεργοποίησης για τους κόμβους εξόδου. Αξιοσημείωτο είναι ότι τα feedforward νευρωνικά δίκτυα με γραμμικούς κόμβους εξόδου υπόκεινται στον περιορισμό ότι δεν μπορούν να μοντελοποιήσουν χρονοσειρές που περιέχουν τάση (trend). Συνεπώς, για αυτού του είδους νευρωνικά δίκτυα, απαιτείται προ-διαφοροποίηση ώστε να εξαλειφθούν οι τάσεις. Μέχρι σήμερα δεν έχει διερευνηθεί η συγκριτική απόδοση της χρήσης γραμμικών και μη γραμμικών συναρτήσεων ενεργοποίησης για τους κόμβους εξόδου και δεν έχουν καταγραφεί εμπειρικά αποτελέσματα για την υποστήριξη ιδιαίτερης προτίμησης σε καμία εκ των δύο.

3.9: Αλγόριθμοι Εκπαίδευσης Τεχνητών Νευρωνικών Δικτύων

Στόχος της εκπαίδευσης είναι η συστηματική εύρεση των συντελεστών διασύνδεσης, δηλαδή των βαρών του δικτύου, τα οποία αρχικά είναι άγνωστα, με την ταυτόχρονη ελαχιστοποίηση κάποιων μέτρων του σφάλματος μεταξύ της επιθυμητής και της πραγματικής εξόδου του δικτύου. Μετά το περάς της εκπαίδευσης του δικτύου δεν επιτρέπονται περαιτέρω μεταβολές στα βάρη του.

Διακρίνονται δυο τεχνικές μάθησης:

Η μάθηση με εποπτεία ή επίβλεψη (Supervised learning), που ενσωματώνει μια εξωτερική πηγή (πχ τον εκπαιδευτή), ή γενική γνώση για το σύστημα και η μάθηση χωρίς επίβλεψη (unsupervised learning) οπού δεν υπάρχει εξωτερική γνώση αναφοράς, αλλά εξαρτάται από τοπικές πληροφορίες και εσωτερικά δεδομένα. Στην περίπτωση μάθησης με εποπτεία παρέχεται στο δίκτυο υπό μάθηση η επιθυμητή έξοδος για κάθε είσοδο ώστε το σύστημα να μαθαίνει της σωστές συσχετίσεις μεταξύ εισόδων και των αντίστοιχων εξόδων. Σε αυτή την κατηγορία χρησιμοποιείται κατά κύριο λόγω ο αλγόριθμος οπισθόδρομης διάδοσης (back propagation algorithm). Ο αλγόριθμος αυτός προσαρμόζει τα βάρη διασύνδεσης σε δίκτυα πολλαπλών στρωμάτων με βάση τη διάδοση ενός μέτρου του σφάλματος μεταξύ τις πραγματικής και της επιθυμητής εξόδου του δικτύου από την έξοδο προς την είσοδο αυτού. Για τη μάθηση χεις εποπτεία δεν παρέχεται στο σύστημα καμία πληροφορία σχετική με την επιθυμητή έξοδο που αντιστοιχεί σε κάθε είσοδο. Αντιθέτως το δίκτυο αυτοοργανώνεται και μαθαίνει να ανταποκρίνεται με διαφορετικό τρόπο σε διαφορετικά χαρακτηριστικά της εισόδου. Χαρακτηριστικά δίκτυα που κάνουν χρήση της παραπάνω μεθόδου εκμάθησης είναι τα δίκτυα ανίχνευσης χαρακτηριστικών (Feature detection) και τα δίκτυα ομαδοποίησης δεδομένων (Clustering).

3.9.1: Ο Κανόνας Μάθησης Δέλτα.

Έστω ένας μη-γραμμικός νευρώνας αποτελούμενος από δύο βαθμίδες (παραπάνω σχήματα). Η πρώτη βαθμίδα παράγει το σταθμισμένο άθροισμα σ=<w,x> (εσωτερικό γινόμενο και η δεύτερη , η μη γραμμική παράγει την έξοδο y=f(σ) , όπου f(.), σχέση εισόδου-εξόδου του μη-γραμμικού στοιχείου. Η ευαισθησία του σφάλματος e μεταξύ της επιθυμητής και της πραγματικής εξόδου του ΤΝΔ στο διάνυσμα βάρους ορίζεται από τη παρακάτω μερική παράγωγο:

$$\frac{\partial e}{\partial w} = -\frac{\partial y}{\partial w} = -\frac{\partial f(\sigma)}{\partial w} = -\frac{\partial f(\sigma)\partial \sigma}{\partial \sigma \partial w} = -f'(\sigma)\cdot x = -g(\sigma)\cdot x$$

Που υπολογίζεται με τον αλυσιδωτό κανόνα παραγώγισης. Είναι εύκολα κατανοητό ότι για να ορίζεται η ευαισθησία παντού η σχέση εισόδου-εξόδου του μη γραμμικού νευρώνα πρέπει να είναι παραγωγίσιμη , δηλαδή η παράγωγος g(σ)=f'(σ) θα πρέπει να ορίζεται για κάθε σ. Το σφάλμα μεταξύ της πραγματικής και της επιθυμητής τιμής ενός νευρώνα είναι:

$$e = d - f(\boldsymbol{\sigma}) = d - f(\langle w, x \rangle)$$

Ακολουθώντας τα ίδια βήματα, προσέγγιση της μερικής παραγώγου του τετραγώνου του σφάλματος είναι:

$$\hat{y} = \frac{\partial e^2}{\partial w} = 2e\frac{\partial e}{\partial w} = -2e\frac{\partial f}{\partial w} = -2e\frac{\partial f \partial \sigma}{\partial \sigma \partial w} = -2egx$$

Οπότε το wθα δίνεται από την παρακάτω επαναληπτική σχέση:

$w^{m+1} = w^m - \mu\gamma^m = w^m + \Delta w^m = w^m + 2me^m g^m x^m$

όπου m ο δείκτης επανάληψης.

Η επιλογή του συντελεστή μ είναι αυθαίρετη. Με την επιλογή μικρού συντελεστή ο αλγόριθμος θα έχει αργή σύγκλιση. Από την άλλη η επιλογή μεγάλου συντελεστή μ μπορεί να προκαλέσει απόκλιση ή και αστάθεια του αλγορίθμου.

Υπάρχουν πολλές τεχνικές για την αύξηση της ταχύτητας σύγκλισης του αλγόριθμου εκμάθησης, με την αυτόματη μεταβολή του συντελεστή μ, ανάλογα με την πορεία της μάθησης. Ένας απλός τρόπος είναι το να ξεκινήσει η μάθηση λαμβάνοντας τυχαίες τιμές και για τα βάρη του συστήματος και ένα συντηρητικό συντελεστή μ . Μετά της πρώτες επαναλήψεις του αλγόριθμου εξετάζεται η εξέλιξη της νόρμας του μεγέθους των διορθώσεων στα βάρη του συστήματος $\|\Delta wm\|$. Αν ο ρυθμός εξέλιξης της νόρμας είναι μικρότερος από ένα προκαθορισμένο όριο ε τότε η τιμή του συντελεστή μ διπλασιάζεται, ενώ αν ο ρυθμός εξέλιξης είναι μεγαλύτερος από ένα, επίσης προκαθορισμένο, όριο δ τότε η τιμή του συντελεστή μ μειώνεται στο μισό. Με τον παραπάνω απλό τρόπο επιτυγχάνεται μεταβαλλόμενος ρυθμός μάθησης.

3.9.2:Αλγόριθμοι Μάθησης Πολυστρωματικών ΤΝΔ

Ένα πολυστρωματικό ΤΝΔ αποτελείται από το στρώμα της εισόδου, το στρώμα της εξόδου και από ένα ή περισσότερα ενδιάμεσα στρώματα.. Ο αριθμός των νευρώνων που απαιτούνται σε κάθε στρώμα δεν είναι γνωστός εκ των προτέρων. Άρα για να βρεθεί ο ελάχιστος εκείνος αριθμός νευρώνων ο οποίος παρέχει ικανοποιητικό αποτέλεσμα απαιτείται κάποιος πειραματισμός (Tasadduq, Rehman, & Bubshait, 2002), (Dombayaci & Golcu, 2009). Για τον έλεγχο των διεργασιών είναι σαφές ότι βασικό κριτήριο είναι τόσο η ελαχιστοποίηση του αριθμού των στρωμάτων , όσο και η ελαχιστοποίηση του αριθμού των νευρώνων. Αν επιτευχθεί αυτό μειώνεται ο χρόνος εκπαίδευσης, αλλά και γίνεται πιο εύκολη η κωδικοποίηση του συστήματος σε υπολογιστή.

Το πρόβλημα μάθησης με εποπτεία ενός σύνθετου πολυστρωματικού ΤΝΔ ανάγεται στην εξεύρεση ενός ευσταθούς επαναληπτικού αλγόριθμου ο οποίος θα προσαρμόζει τα βάρη των διασυνδέσεων των νευρώνων στα διάφορα στρώματα συστηματικά, με δεδομένα τις επιθυμητές και τις πραγματικές εξόδους του δικτύου.

Ο αλγόριθμος μάθησης πολυστρωματικών ΤΝΔ είναι μία παραλλαγή του κανόνα μάθησης Δέλτα για ένα νευρώνα, ο οποίος παρουσιάστηκε παραπάνω, με διαφορά ότι τα διανύσματα στη επαναληπτική σχέση είναι πίνακες των οποίων οι διαστάσεις εξαρτώνται από των αριθμό των νευρώνων ανά στρώμα. Από τα παραπάνω είναι προφανές ότι όσο πιο πολύ νευρώνες σε ένα στρώμα και όσο πιο πολλά στρώματα έχει ένα ΤΝΔ τόσο χρονοβόρα γίνεται η εκπαίδευση του και τόσο αυξάνονται οι απαιτήσεις σε υπολογιστική ισχύ για παράδειγμα η χρήση ΤΝΔ στην αναγνώριση φωνής. Τα ΤΝΔ τα οποία χρησιμοποιούνται έχουν μερικές χιλιάδες νευρώνες ανά στρώμα και περίπου τρία με τέσσερα κρυφά στρώματα.

3.9.3:Ο Αλγόριθμος Μάθησης Με Οπισθόδρομη Διάδοση

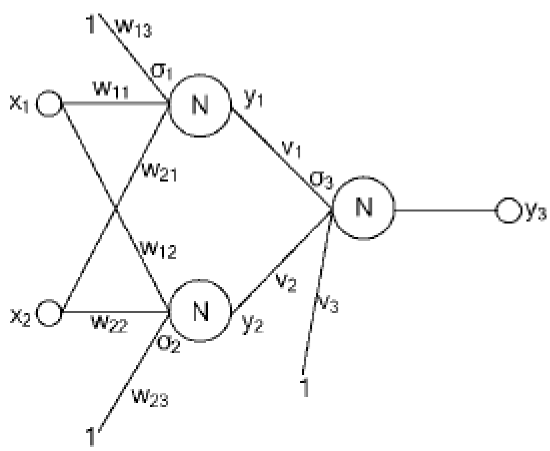
Για να γίνει κατανοητή η έννοια του αλγόριθμου οπισθόδρομης διάδοσης (backpropagation algorithm) ο οποίος είναι από τους πιο διαδεδομένους αλγορίθμους εκπαίδευσης δίδεται ένα απλό παράδειγμα. Έστω ένα δίκτυο δύο στρωμάτων με δύο νευρώνες στο στρώμα εισόδου και ένα νευρώνα στο στρώμα εξόδου (το δίκτυο παρουσιάζεται στο σχήμα που ακλουθεί). Η γενίκευση του αλγόριθμου είναι σχετικά εύκολη, αλλά την δυσκολεύει το γεγονός του ότι απαιτούνται πολλαπλοί δείκτες στα συναπτικά βάρη τα οποία καθιστούν τις εξισώσεις ιδιαίτερα δύσχρηστες. Από το σχήμα είναι προφανές ότι τα σήματα στις εισόδους των νευρώνων του πρώτου στρώματος είναι :

$\sigma_1 = w_{11}x_1 + w_{21}x_2 + w_{13}$ $\quad y_1 = f(\sigma_1)$

$\sigma_2 = w_{12}x1 + w_{22}x_2 + w_{23}$ $\quad y_2 = f(\sigma_2)$

Ομοίως η σταθμισμένη έξοδος του δευτέρου στρώματος είναι:

$\sigma_3 = v_1 y_1 + v_2 y_2 + v_3$ $\quad y_3 = f(\sigma_3)$



Σχήμα: Ένα απλό ΤΝΔ με 2 εισόδους και 1 έξοδο (2-1)

Το σφάλμα του δικτύου είναι η διαφορά μετάξι της πραγματικής και της επιθυμητής εξόδου του δικτύου, δηλαδή $e = d- y_3$. Σε κάθε επανάληψη το σφάλμα το οποίο υπολογίζεται είναι :

$e = d – f(y_3)$.

Άρα απαιτείται η μερική παράγωγος του τετραγώνου του σφάλματος για το κάθε βάρος. Για τα βάρη του πρώτου και του δεύτερου στρώματος η μερικές παράγωγοι του τετραγώνου του σφάλματος είναι :

$$\frac{\partial e^2}{\partial\left(w_{ij}\right)} = 2e\frac{\partial e}{\partial(\nu_i)} \qquad , i = 1,2 \; και \; j = 1,3$$

$$\frac{\partial e^2}{\partial(\nu_i)} = 2e\,\frac{\partial e}{\partial(\nu_i)} \quad , \qquad i = 1,3$$

Αρκεί τώρα να ακολουθήσουμε τα βήματα τα οποία έχουν παρουσιαστεί παραπάνω και έτσι εκφράζεται ο επαναληπτικός αλγόριθμος μάθησης:

$$W^{m+1} = W^m + \varDelta W^m \; = W^m + 2\mu e^m \; \frac{\partial e^m}{\partial W^m}$$

και

$$\nu^{m+1} = \nu^m + \varDelta \nu^{m\prime} + 2\mu e^m \frac{\partial e^m}{\partial \nu^m}$$

Ο αλγόριθμος απαιτεί τον υπολογισμό των ευαισθησιών του σφάλματος σε σχέση με τα βάρη των διασυνδέσεων των νευρώνων , δηλαδή της μερικές παραγώγους:

$$\frac{\partial e}{\partial w} \; και \; \frac{\partial \varepsilon}{\partial \nu_i}$$

Οι πρώτες τρείς μερικές παράγωγοι , δηλαδή οι παράγωγοι της ευαισθησίας του σφάλματος στα βάρη των νευρώνων, θα βρεθούν με χρήση του αλυσιδωτού κανόνα παραγώγισης και είναι:

$$\frac{\partial e}{\partial w_{11}} = -\frac{\partial y}{\partial w_{11}} = -\frac{\partial f(\sigma_3)\partial \sigma_3}{\partial \sigma_3 \partial w_{11}} = -\frac{\partial f(\sigma_3)\partial f(\sigma_1)\partial \sigma_1}{\partial \sigma_3 \partial f(\sigma_1)\partial \sigma_1 \partial w_{11}} = -g(\sigma_3)g(\sigma_1)v_1 x_1$$

$$\frac{\partial e}{\partial w_{12}} = -\frac{\partial y}{\partial w_{12}} = -\frac{\partial f(\sigma_3)\partial \sigma_3}{\partial \sigma_3 \partial w_{12}} = -\frac{\partial f(\sigma_3)\partial f(\sigma_1)\partial \sigma_1}{\partial \sigma_3 \partial f(\sigma_1)\partial \sigma_1 \partial w_{12}} = -g(\sigma_3)g(\sigma_1)v_1 x_2$$

$$\frac{\partial e}{\partial w_{13}} = -\frac{\partial y}{\partial w_{13}} = -\frac{\partial f(\sigma_3)\partial \sigma_3}{\partial \sigma_3 \partial w_{13}} = -\frac{\partial f(\sigma_3)\partial f(\sigma_1)\partial \sigma_1}{\partial \sigma_3 \partial f(\sigma_1)\partial \sigma_1 \partial w_{13}} = -g(\sigma_3)g(\sigma_1)v_1 x_3$$

Αντίστοιχα υπολογίζονται και οι υπόλοιπες μερικές παράγωγοι, οι οποίες αφορούν τα βάρη που αντιστοιχούν στη δεύτερη είσοδο. Αξίζει να σημειωθεί ότι οι αναλυτικές σχέσεις των μη-γραμμικών συναρτήσεων και των παραγώγων τους είναι γνώστες και άρα ο υπολογισμός τους γίνεται χωρίς ιδιαίτερη δυσκολία.

Στην περίπτωση που το μη-γραμμικό στοιχείο του νευρώνα είναι σιγμοειδές τότε η παράγωγος είναι:

$$f(\sigma) = \frac{1-e^{-\sigma}}{1+e^{-\sigma}} \; \kappa\alpha\iota \; g(\sigma) = f'(\sigma) = \frac{2\sigma}{(1+e^{-\sigma})^2}$$

Ομοίως υπολογίζονται οι ευαισθησίες στα βάρη του δεύτερου στρώματος.

$$\frac{\partial e}{\partial v_1} = -\frac{\partial y_3}{\partial v_1} = -\frac{\partial f(\sigma_3)\partial \sigma_3}{\partial \sigma_3 \partial v_1} = -g(\sigma_3)y_1$$

$$\frac{\partial e}{\partial v_2} = -\frac{\partial y_3}{\partial v_2} = -\frac{\partial f(\sigma_3)\partial \sigma_3}{\partial \sigma_3 \partial v_2} = -g(\sigma_3)y_2$$

$$\frac{\partial e}{\partial v_3} = -\frac{\partial y_3}{\partial v_3} = -\frac{\partial f(\sigma_3)\partial \sigma_3}{\partial \sigma_3 \partial v_3} = -g(\sigma_3)y_3$$

Οι παραπάνω εξισώσεις είναι απλές εφόσον διασχίζεται μόνο ένα στρώμα. Ο αλγόριθμος οπισθόδρομής διάδοσης παίρνει το όνομα του από το γεγονός ότι για να υπολογιστούν τα βάρη γίνεται ανάκληση στη ροή των υπολογισμών Έτσι με κάποιες τυχαίες αρχικές τιμές για τα βάρη των στρωμάτων και με δεδομένες τις τιμές των εισόδων, γίνεται η πρώτη υπολογιστική διαδρομή από την οποία υπολογίζονται οι μεταβλητές σ_1, σ_2 ,σ_3 και συνεπώς y_1 y_2, y_3. Με τη γνώση αυτή υπολογίζονται οι διάφορες συναρτήσεις των μη-γραμμικών σχέσεων εισόδων-εξόδου και οι παράγωγοι τους (f,g).

Σε κάθε επανάληψη του αλγορίθμου υπολογίζονται οι μεταβολές ή διορθώσεις ΔW_m στα βάρη του στρώματος της εισόδου και $\Delta \nu_m$ στα βάρη του στρώματος της εξόδου. Οι μεταβολές αυτές προστίθενται στις προηγούμενες τιμές των βαρών για να επαναληφθεί η διαδικασία μέχρι να επιτευχθεί το μέτρο του σφάλματος οπότε και τερματίζεται η διαδικασία μάθησης.. Για σύνθετα ΤΝΔ απαιτούνται δεκάδες χιλιάδες επαναλήψεις για τη μάθηση των βαρών οπότε η διαδικασία μάθησης είναι αρκετά χρονοβόρα και απαιτεί ταχύτατους υπολογιστές.

Αξίζει να σημειωθεί ότι ο συγκεκριμένος αλγόριθμος εποπτευόμενης μάθησης αν και είναι πολύ διαδεδομένος παρουσιάζει αργή σύγκλιση, οπότε η χρήση του γίνεται προβληματική για μεγάλα δίκτυα. Μια παραλλαγή του παραπάνω αλγόριθμου είναι ο αλγόριθμος μάθησης με ορμή (momentum) ο οποίος είναι ιδιαίτερα αποτελεσματικός στη μάθηση νευρωνικών ελεγκτών. Στην περίπτωση αυτή προστίθεται στη εξίσωση αναπροσαρμογής των βαρών ένας ακόμη όρος που είναι ανάλογος της μεταβολής των βαρών με αποτέλεσμα τη σημαντική επιτάχυνση του αλγορίθμου.

3.9.4:Δραστηριότητες Κατά Την Εφαρμογή Του Αλγορίθμου Εκπαίδευσης

Η εκπαίδευση του νευρωνικού δικτύου αποτελεί ένα μη γραμμικό πρόβλημα ελαχιστοποίησης χωρίς περιορισμούς, στο οποίο τα συναπτικά βάρη του δικτύου τροποποιούνται επαναληπτικά ώστε να ελαχιστοποιήσουν το ολικό μέσο τετραγωνικό σφάλμα, μεταξύ των επιθυμητών και των πραγματικών τιμών εξόδου, για όλους τους κόμβους εξόδου και όλα τα υποδείγματα εισόδου. Η ύπαρξη πολλών διαφορετικών μεθόδων βελτιστοποίησης παρέχει διάφορες επιλογές για την εκπαίδευση του νευρωνικού δικτύου. Την τρέχουσα περίοδο, δεν υπάρχει αλγόριθμος ο οποίος να εγγυάται τη γενική βέλτιστη λύση για ένα γενικό μη γραμμικό πρόβλημα βελτιστοποίησης σε λογική χρονική διάρκεια. Έτσι όλοι οι αλγόριθμοι βελτιστοποίησης στην πράξη αναπόφευκτα πάσχουν από τα τοπικά προβλήματα βελτιστοποίησης και ό,τι καλύτερο μπορεί να γίνει στον τομέα αυτό είναι η χρήση των διαθέσιμων μεθόδων βελτιστοποίησης που παρέχουν το “καλύτερο” τοπικό βέλτιστο, εάν η πραγματική γενική λύση είναι αδύνατο να επιτευχθεί. Η πιο δημοφιλής μέθοδος εκπαίδευσης είναι ο αλγόριθμος backprobagation ο οποίος αποτελεί ουσιαστικά μια μέθοδο μέγιστης κατάβασης (gradient steepest descent method). Για τον αλγόριθμο μέγιστης κατάβασης πρέπει να προσδιοριστεί ένα μέγεθος βήματος, το οποίο καλείται συντελεστής μάθησης (learning rate) στην ορολογία ΤΝΔ. Ο συντελεστής μάθησης είναι αποφασιστικής σημασίας για τον αλγόριθμο μάθησης backpropagation αφού καθορίζει το μέγεθος των μεταβολών στα βάρη. Είναι ευρέως γνωστό ότι η μέθοδος της μέγιστης κατάβασης έχει τα μειονεκτήματα της αργής σύγκλησης, ανεπάρκειας και έλλειψη ευρωστίας. Επιπλέον είναι πολύ ευαίσθητη στην επιλογή του συντελεστή μάθησης. Μικρότεροι συντελεστές τείνουν να επιβραδύνουν τη διαδικασία μάθησης, ενώ μεγαλύτεροι συντελεστές μπορεί να προκαλέσουν ταλάντωση του δικτύου στο χώρο των βαρών. Ένας τρόπος βελτίωσης της πρότυπης μεθόδου μέγιστης κατάβασης είναι να συμπεριληφθεί μια πρόσθετη παράμετρος *ορμής* (*momentum),* ώστε να είναι εφικτοί μεγαλύτεροι συντελεστές μάθησης, με αποτέλεσμα την ταχύτερη σύγκλιση και παράλληλα τη μείωση της τάσης για ταλάντωση. Το σκεπτικό της εισαγωγής της ορμής είναι να πραγματοποιηθεί η επόμενη αλλαγή βάρους στην ίδια κατεύθυνση με την προηγούμενη. Η πρότυπη τεχνική backpropagation με ορμή χρησιμοποιείται από τους περισσότερους ερευνητές. Εφόσον υπάρχουν λίγοι συστηματικοί τρόποι επιλογής του συντελεστή μάθησης και της ορμής ταυτόχρονα, οι “καλύτερες” τιμές αυτών των παραμέτρων μάθησης επιλέγονται συνήθως μέσω πειραματισμών. Καθώς ο ρυθμός μάθησης και η ορμή είναι δυνατό να πάρουν οποιαδήποτε τιμή μεταξύ 0 και 1, είναι πρακτικά αδύνατο να πραγματοποιηθεί εξαντλητική έρευνα για την εύρεση των καλύτερων συνδυασμών αυτών των παραμέτρων εκπαίδευσης. Μόνο προεπιλεγμένες τιμές μελετώνται από τους ερευνητές . Για παράδειγμα, οι Sharda και Patil (1992) δοκιμάζουν εννέα συνδυασμούς τριών ρυθμών εκπαίδευσης (0.1, 0.5, 0.9) και τριών τιμών ορμής (0.1, 0.5, 0.9). Γενικά οι παράμετροι εκπαίδευσης διαδραματίζουν αποφασιστικό ρόλο στην απόδοση των ΤΝΔ. Ωστόσο, υπάρχουν αντιφατικά συμπεράσματα σχετικά με τις καλύτερες παραμέτρους μάθησης, γεγονός το οποίο κατά τη γνώμη μας οφείλεται στην ανεπάρκεια και στην έλλειψη ευρωστίας του αλγόριθμου μέγιστης κατάβασης.

Εξαιτίας της ανεπάρκειας του συμβατικού αλγορίθμου backpropagation, έχουν προταθεί αρκετές διαφοροποιήσεις ή τροποποιήσεις του backpropagation, όπως η προσαρμοστική μέθοδος, η quickprop και μέθοδοι δεύτερης τάξης. Η τελευταία είναι η πιο αποτελεσματική μη γραμμική μέθοδος βελτιστοποίησης και χρησιμοποιείται στα περισσότερα πακέτα βελτιστοποίησης. Τα χαρακτηριστικά που διαθέτει όπως γρηγορότερη σύγκλιση, ευρωστία και η ικανότητα εύρεσης καλού τοπικού ελαχίστου την καθιστούν δημοφιλή στην εκπαίδευση των ΤΝΔ. Προσφάτως ερευνητές, πρότειναν τη χρήση ενός μη γραμμικού βελτιστοποιητή γενικού σκοπού, τον GRG2, στην εκπαίδευση των δικτύων. Τα πλεονεκτήματα του GRG2 έχουν προσδιοριστεί για διάφορα προβλήματα ΤΝΔ. Ο GRG2 αποτελεί ένα ευρέως διαθέσιμο λογισμικό βελτιστοποίησης το οποίο επιλύει μη γραμμικά προβλήματα βελτιστοποίησης χρησιμοποιώντας τη γενικευμένη μειωμένη μέθοδο της μέγιστης κατάβασης (reduced gradient method). Με τον GRG2 δεν είναι αναγκαία η επιλογή παραμέτρων μάθησης όπως ο συντελεστής μάθησης και η ορμή. Εν αντιθέσει, ένα διαφορετικό σύνολο παραμέτρων, όπως τα κριτήρια τερματισμού, η διαδικασία αναζήτησης κατεύθυνσης και τα όρια των μεταβλητών, πρέπει να προσδιοριστούν και είναι δυνατό να τεθούν ως προκαθορισμένες τιμές. Ένα ακόμη σχετικό ζήτημα στην εκπαίδευση των ΤΝΔ είναι ο προσδιορισμός μιας αντικειμενικής συνάρτησης ή συνάρτησης κόστους. Συνήθως χρησιμοποιούνται οι SSE και MSE, αφού ορίζονται σε σχέση με τα σφάλματα. Άλλες αντικειμενικές συναρτήσεις όπως του μεγίστου αποτελέσματος, κέρδους ή ωφέλειας μπορεί να είναι πιο κατάλληλες για προβλήματα οικονομικής πρόβλεψης. Η επιλογή μιας συνάρτησης κόστους μπορεί να επηρεάσει σημαντικά την απόδοση πρόβλεψης του δικτύου εάν ο αλγόριθμος μάθησης (backpropagation) και άλλες παράμετροι του δικτύου είναι αμετάβλητες. Συνεπώς, ένας πιθανός τρόπος άμεσης αντιμετώπισης της τελικής αντικειμενικής συνάρτησης είναι η αλλαγή του αλγόριθμου αναζήτησης από backpropagation σε γενετικούς αλγόριθμους, ή προσομοιωμένη ανόπτηση (simulated annealing) και άλλες μεθόδους βελτιστοποίησης οι οποίες επιτρέπουν αναζήτηση σε αυθαίρετες συναρτήσεις χρησιμότητας.

Κανονικοποίηση των δεδομένων

Οι μη γραμμικές συναρτήσεις ενεργοποίησης όπως η συνάρτηση logistic περιορίζουν την πιθανή έξοδο ενός κόμβου στο διάστημα (0,1) ή (-1,1). Η κανονικοποίηση των δεδομένων συνήθως εκτελείται πριν την έναρξη της διαδικασίας εκπαίδευσης. Όπως προαναφέρθηκε, όταν μη γραμμικές συναρτήσεις μεταφοράς χρησιμοποιούνται στους κόμβους εξόδου, οι επιθυμητές τιμές εξόδου πρέπει να μετασχηματίζονται στο πεδίο τιμών των πραγματικών εξόδων του δικτύου. Ακόμη κι αν χρησιμοποιείται μια γραμμική συνάρτηση μεταφοράς εξόδου μπορεί να είναι ωφέλιμο να τυποποιηθούν οι έξοδοι όπως και οι είσοδοι για την αποφυγή υπολογιστικών προβλημάτων, την ικανοποίηση των απαιτήσεων του αλγορίθμου και τη διευκόλυνση της μάθησης του δικτύου.

Παρακάτω παρουσιάζονται συνοπτικά τέσσερις μέθοδοι για την κανονικοποίηση των εισόδων:

1.Κανονικοποίηση along channel: Ως δίοδος (channel) χαρακτηρίζεται ένα σύνολο στοιχείων στην ίδια θέση για όλα τα διανύσματα εισόδου στο σύνολο εκπαίδευσης ή δοκιμής. Δηλαδή, κάθε δίοδος μπορεί να θεωρηθεί ως "ανεξάρτητη" μεταβλητή εισόδου. Η κανονικοποίηση along channel εκτελείται από στήλη προς στήλη εάν τα διανύσματα εισόδου τοποθετηθούν σε μαθηματικό πίνακα. Με άλλα λόγια, κανονικοποιεί κάθε μεταβλητή εισόδου ατομικά.

2. Κανονικοποίηση across channel: Ο εν λόγω τύπος κανονικοποίησης εκτελείται για κάθε διάνυσμα εισόδου ατομικά, δηλαδή, η κανονικοποιήση πραγματοποιείται δια μέσου όλων των στοιχείων στο υπόδειγμα δεδομένων.

3. Κανονικοποίηση mixed channel: Αυτή η μέθοδος, όπως υποδηλώνει και η ονομασία της, χρησιμοποιεί συνδυασμό των κανονικοποιήσεων along και across.

4. Εξωτερική κανονικοποίηση: Όλα τα δεδομένα εκπαίδευσης κανονικοποιούνται σε συγκεκριμένο πεδίο τιμών.

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

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

- γραμμικός μετασχηματισμός στο [0,1]: $x_n = (x_0 - x_{\min})/(x_{\max} - x_{\min})$
- γραμμικός μετασχηματισμός στο $[a,b]$: $x_n = (b-a)(x_0 - x_{\min})/(x_{\max} - x_{\min}) + a$
- στατιστική κανονικοποίηση: $x_n = (x_0 - \bar{x})/s$
- απλή κανονικοποίηση: $x_n = x_0 / x_{\max}$

Όπου χ_n και χ_0 είναι τα κανονικοποιημένα και τα αρχικά δεδομένα, χ_{min},χ_{max}, $\bar{\chi}$ και s είναι η ελάχιστη, η μέγιστη, η μέση και η τυπική απόκλιση αντίστοιχα κατά μήκος των στηλών ή των σειρών. Δεν είναι σαφές εάν είναι αναγκαία η κανονικοποίηση των εισόδων, αφού τα συναπτικά βάρη μπορεί να ακυρώσουν την κανονικοποίηση. Υπάρχουν αρκετές μελέτες σχετικά με την επίδραση της κανονικοποίησης των δεδομένων στην μάθηση του δικτύου. Ορισμένοι ερευνητές συμπέραναν ότι γενικά η κανονικοποίηση των δεδομένων είναι ευεργετική, όσον αφορά το ρυθμό ταξινόμησης και το μέσο τετραγωνικό σφάλμα, αλλά τα πλεονεκτήματα ελαττώνονται καθώς αυξάνονται το μέγεθος του δικτύου και του δείγματος. Επιπλέον, η κανονικοποίηση των δεδομένων συνήθως επιβραδύνει τη διαδικασία εκπαίδευσης. Προσφάτως προτάθηκε μια αυτόματη μέθοδο κανονικοποίησης, ο κανόνας gamma learning, που επιτρέπει στο δίκτυο να αυτο-κανονικοποιείται κατά τη διαδικασία μάθησης και εξαλείφει την ανάγκη για κανονικοποίηση των δεδομένων πριν την εκπαίδευση. Η κανονικοποίηση των τιμών εξόδου (targets) είναι συνήθως ανεξάρτητη από την κανονικοποίηση των εισόδων.

Τα δείγματα εκπαίδευσης και δοκιμής (Training data Testing data)

Όπως προαναφέρθηκε, για τη δόμηση ΤΝΔ πρόβλεψης απαιτούνται δείγματα εκπαίδευσης και δοκιμής. **Το δείγμα εκπαίδευσης** χρησιμοποιείται για την ανάπτυξη του μοντέλου ΤΝΔ και **το δείγμα**

δοκιμής για την εκτίμηση της ικανότητας πρόβλεψης του μοντέλου. Ορισμένες φορές χρησιμοποιείται επίσης ένα τρίτο δείγμα, **το δείγμα αξιολόγησης (validation sample)**, ώστε να αποφευχθεί το πρόβλημα υπερεκπαίδευσης ή να προσδιοριστεί το σημείο τερματισμού της διαδικασίας εκπαίδευσης. Είναι συνήθης η χρήση ενός συνόλου δοκιμής εξίσου για σκοπούς ελέγχου αξιοπιστίας και δοκιμής, ιδιαίτερα για μικρά σύνολα δεδομένων. Κατά τη γνώμη μας, η επιλογή των δειγμάτων εκπαίδευσης και δοκιμής μπορεί να επηρεάσει την απόδοση των ΤΝΔ. Το κυρίαρχο ζήτημα, εδώ, είναι η υποδιαίρεση των δεδομένων στα σύνολα εκπαίδευσης και δοκιμής. Μολονότι δεν υπάρχει γενικός κανόνας λύσης στο εν λόγω πρόβλημα, πρέπει να ληφθούν υπόψη, στην απόφαση, αρκετοί παράγοντες, όπως τα χαρακτηριστικά του προβλήματος, καθώς και το είδος και το μέγεθος των δεδομένων. Είναι αποφασιστικής σημασίας τόσο το σύνολο εκπαίδευσης όσο και το σύνολο δοκιμής να είναι αντιπροσωπευτικά του δείγματος ή του υποκείμενου μηχανισμού. Ακατάλληλος διαχωρισμός των συνόλων εκπαίδευσης και δοκιμής θα επηρεάσει την επιλογή της βέλτιστης δομής ΤΝΔ και την εκτίμηση της προγνωστικής απόδοσης του ΤΝΔ. Οι έρευνες που έχουν διεκπεραιωθεί προσφέρουν ελάχιστη καθοδήγηση στην επιλογή των δειγμάτων εκπαίδευσης και δοκιμής. Οι περισσότεροι μελετητές επιλέγουν βάσει των κανόνων 90% έναντι 10%, 80% έναντι 20% ή 70% έναντι 30% κτλ. Ορισμένοι επιλέγουν βάσει του προβλήματος που μελετάται. Ένας ακόμη σχετικός παράγοντας είναι το μέγεθος του δείγματος. Δεν υπάρχει συγκεκριμένος κανόνας για τις απαιτήσεις του μεγέθους δείγματος για κάθε πρόβλημα. Η ποσότητα των δεδομένων για την εκπαίδευση του δικτύου εξαρτάται από τη δομή του δικτύου, τη μέθοδο εκπαίδευσης και την πολυπλοκότητα του εκάστοτε προβλήματος ή την ποσότητα του θορύβου στα διαθέσιμα δεδομένα. Γενικά, όπως σε κάθε στατιστική προσέγγιση, το μέγεθος δείγματος σχετίζεται στενά με την απαιτούμενη ακρίβεια του προβλήματος. Όσο μεγαλύτερο το μέγεθος δείγματος, τόσο πιο ακριβή θα είναι τα αποτελέσματα. Έχει καταγραφεί σε ερευνητικές μελέτες ότι καθώς αυξάνεται το μέγεθος, τα ΤΝΔ πρόβλεψης αποδίδουν καλύτερα.

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

Αξιοσημείωτο είναι ότι στα πραγματικά προβλήματα υπάρχουν όρια για την ακρίβεια που μπορεί να πετύχει κάθε μοντέλο. Για παράδειγμα, εάν θεωρήσουμε μόνο δύο παράγοντες: το θόρυβο των δεδομένων και το υποκείμενο μοντέλο, τότε το όριο ακριβείας ενός γραμμικού μοντέλου όπως το Box-Jenkins καθορίζεται από το θόρυβο στα δεδομένα και από το βαθμό στον οποίο η υποκείμενη συναρτησιακή μορφή είναι μη γραμμική. Με περισσότερες παρατηρήσεις, η ακρίβεια του μοντέλου δε δύναται να βελτιωθεί εάν δεν υπάρχει μη γραμμική δομή στα δεδομένα. Στα ΤΝΔ ο θόρυβος και μόνο καθορίζει το όριο της ακρίβειας εξαιτίας της ικανότητας εκτίμησης γενικής συνάρτησης που έχουν. Με ένα ικανοποιητικά μεγάλο δείγμα, τα ΤΝΔ είναι σε θέση να μοντελοποιήσουν οποιαδήποτε πολύπλοκη δομή των δεδομένων. Επομένως, τα ΤΝΔ ωφελούνται περισσότερο από μεγάλα δείγματα απ’ ό,τι τα γραμμικά στατιστικά μοντέλα. Ενδιαφέρον έχει ότι τα ΤΝΔ δεν απαιτούν απαραίτητα μεγαλύτερο δείγμα από εκείνο που χρειάζεται για την καλή απόδοση των γραμμικών μοντέλων. Έχει αποδειχθεί ότι τα μοντέλα πρόβλεψης ΤΝΔ αποδίδουν πολύ καλά ακόμη και με μεγέθη δείγματος μικρότερα του 50, ενώ τα μοντέλα Box-Jenkins απαιτούν συνήθως τουλάχιστον 50 σημεία δεδομένων προκειμένου να δώσουν επιτυχημένη πρόβλεψη.

Κριτήρια απόδοσης

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

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

όπου e_t είναι το ατομικό σφάλμα πρόβλεψης, y_t είναι η πραγματική τιμή και N είναι το σύνολο των όρων σφάλματος. Ενδεικτικά αναφέρουμε ότι στη βιβλιογραφία έχουν χρησιμοποιηθεί και άλλα κριτήρια ακριβείας, όπως το μέσο σφάλμα (mean error – ME), η στατιστική μεταβλητή U του Theil (Theil's U-statistic), το μέσο απόλυτο ποσοστό σφάλματος (median absolute percentage error – MdaPE), το γεωμετρικό μέσο σχετικά απόλυτο σφάλμα (geometric mean relative absolute error – GMRAE), η μέση σχετική διακύμανση (average relative variance – ARV) και το κριτήριο πληροφορίας του Bayes (Bayesian information criterion – BIC). Το MSE είναι το πιο συνηθισμένο μέτρο ακρίβειας, αν και τα πλεονεκτήματα της χρήσης του αμφισβητούνται όταν πρόκειται για την εκτίμηση της σχετικής ακρίβειας μεθόδων πρόβλεψης για διαφορετικά σύνολα δεδομένων. Επιπλέον, το MSE ίσως να μην είναι κατάλληλο για τη δόμηση ενός μοντέλου ΤΝΔ με δείγμα εκπαίδευσης, αφού αγνοεί τις σημαντικές πληροφορίες σχετικά με το νούμερο των παραμέτρων (συναπτικά βάρη), τις οποίες το μοντέλο πρέπει να υπολογίσει. Από τη σκοπιά της στατιστικής, καθώς ο αριθμός των υπολογισμένων παραμέτρων του μοντέλου αυξάνεται, οι βαθμοί ελευθερίας για το ολικό μοντέλο μειώνονται, αυξάνοντας έτσι την πιθανότητα υπερεκπαίδευσης στο δείγμα εκπαίδευσης. Μια βελτιωμένη έκδοση του MSE, για το μέρος της εκπαίδευσης, είναι το συνολικό άθροισμα των τετραγωνικών σφαλμάτων διαιρεμένο με τους βαθμούς ελευθερίας, που είναι ο αριθμός των παρατηρήσεων μείον τον αριθμό των συναπτικών βαρών και σταθερών βαρών.

3.10: Κύρια Χαρακτηριστικά των ΤΝΔ

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

Δεύτερον, τα ΤΝΔ έχουν τη δυνατότητα γενίκευσης. Έπειτα από τη διαδικασία μάθησης των δεδομένων, τα ΤΝΔ έχουν συχνά τη δυνατότητα εξαγωγής τεκμηριωμένων συμπερασμάτων των μη ορατών μερών του δείγματος ακόμη και εάν το δείγμα περιέχει πληροφορίες με θόρυβο. Καθώς η πρόβλεψη πραγματοποιείται, μέσω προγνωστικών της μελλοντικής συμπεριφοράς (το μη ορατό μέρος) από παραδείγματα παρελθοντικής συμπεριφοράς, αποτελεί ιδανικό τομέα εφαρμογής των νευρωνικών δικτύων.

Τρίτον, τα ΤΝΔ είναι προσεγγιστές γενικών συναρτήσεων. Έχει παρατηρηθεί ότι ένα δίκτυο είναι δυνατόν να προσεγγίσει κάθε συνεχή συνάρτηση με οποιαδήποτε επιθυμητή ακρίβεια. Τα ΤΝΔ μπορούν να χειρισθούν πιο γενικές και ευέλικτες μορφές συναρτήσεων από τις παραδοσιακές στατιστικές μεθόδους. Κάθε μοντέλο πρόβλεψης υποθέτει ότι υπάρχει μια υποκείμενη (γνωστή ή άγνωστη) σχέση μεταξύ των εισόδων (οι παρελθοντικές τιμές των χρονοσειρών και/ ή των υπόλοιπων μεταβλητών) και των εξόδων (οι μελλοντικές τιμές). Συχνά, τα παραδοσιακά στατιστικά μοντέλα πρόβλεψης έχουν περιορισμούς στην εκτίμηση της υποκείμενης συνάρτησης εξαιτίας της πολυπλοκότητας των πραγματικών συστημάτων. Τα ΤΝΔ αποτελούν την καλύτερη εναλλακτική μέθοδο όσον αφορά την αναγνώριση τέτοιων συναρτήσεων.

Τέλος, τα ΤΝΔ είναι μη γραμμικά. Η πρόβλεψη αποτελούσε ανέκαθεν περιοχή εφαρμογής της γραμμικής στατιστικής. Οι παραδοσιακές προσεγγίσεις στην πρόγνωση χρονοσειρών (time series), όπως οι μέθοδοι Box-Jenkins και ARMA, προϋποθέτουν ότι οι υπό μελέτη χρονοσειρές προέρχονταν από γραμμικές διαδικασίες. Τα γραμμικά μοντέλα έχουν το πλεονέκτημα της κατανόησης και ανάλυσής τους με κάθε λεπτομέρεια καθώς και τη δυνατότητα εύκολης εξήγησης και υλοποίησης. Ωστόσο, μπορεί να αποδειχθούν εντελώς ακατάλληλα εάν ο υποκείμενος μηχανισμός είναι μη γραμμικός. Είναι παράλογη η εκ των προτέρων υπόθεση ότι οι συγκεκριμένες χρονοσειρές παράγονται από γραμμικές διαδικασίες. Στην πραγματικότητα, τα σύγχρονα συστήματα και ιδιαίτερα τα μακροοικονομικά είναι κατά πλειοψηφία μη γραμμικά. Ωστόσο αυτά τα μη γραμμικά μοντέλα αντιμετωπίζουν κάποιους περιορισμούς υπό την έννοια ότι είναι αναγκαίο να υποτεθεί μια ρητή σχέση για τα υπό μελέτη δεδομένα, όταν ελάχιστη γνώση του υποκείμενου νόμου είναι διαθέσιμη. Στην πραγματικότητα, η διαμόρφωση ενός μη γραμμικού μοντέλου για ένα συγκεκριμένο σύνολο δεδομένων αποτελεί ένα πολύ δύσκολο εγχείρημα, εφόσον υπάρχουν πολλά πιθανά μη γραμμικά πρότυπα και ένα προκαθορισμένο μη γραμμικό μοντέλο μπορεί να αποδειχθεί μη αρκετά γενικό για την σύλληψη όλων των σημαντικών χαρακτηριστικών. Τα τεχνητά νευρωνικά δίκτυα, που αποτελούν μη γραμμικές, οδηγούμενες από τα δεδομένα, προσεγγίσεις σε αντίθεση με τα παραπάνω μη γραμμικά μοντέλα, είναι ικανά για την εκτέλεση μη γραμμικής μοντελοποίησης χωρίς να απαιτούν a priori γνώση σχετικά με τη σχέση των μεταβλητών εισόδου και εξόδου. Συνεπώς συνιστούν ένα πιο γενικό και ευέλικτο εργαλείο μοντελοποίησης για την διεξαγωγή προβλέψεων.

3.11: Εφαρμογές των Τεχνητών Νευρωνικών Δικτύων

Τα νευρωνικά δίκτυα, λόγω του ότι βασίζονται στους νεύρωνες, στις διασυνδέσεις τους και στις συναρτήσεις μεταφοράς, εμφανίζουν ομοιότητες μεταξύ των διαφορετικών δομών και αρχιτεκτονικών τους. Η πλειάδα των παραλλαγών τους, πηγάζει από τους διαφορετικούς κανόνες μάθησης και τους τρόπους που αυτοί τροποποιούν την τοπολογία των δικτύων. Παρακάτω, παρατίθενται πέντε (5) κατηγορίες νευρωνικών δικτύων στις οποίες ανήκουν οι περισσότερες εφαρμογές. Ο διαχωρισμός αυτός δεν είναι αποκλειστικός, απλά βοηθά να ξεκαθαρίσουμε την σύγχυση σε ότι αφορά τις διαφορετικές αρχιτεκτονικές και την καταλληλότητα της κάθε μίας (για συγκεκριμένες εφαρμογές).

Έτσι, λοιπόν, έχουμε νευρωνικά δίκτυα για:

-Προβλέψεις

-Ταξινόμηση δεδομένων

-Συσχετισμό δεδομένων

-Αντίληψη δεδομένων

-Φιλτράρισμα δεδομένων

Τα νευρωνικά δίκτυα, μεταβάλλονται καθώς περνούν από το πεδίο της ακαδημαϊκής έρευνας στον πραγματικό κόσμο, όπου οι χρήστες θέλουν απλά να κάνουν την δουλειά τους με αυτά τα εργαλεία. Πολλά από τα δίκτυα που χρησιμοποιούνται είναι αρκετά ακριβή, αλλά οι χρήστες δεν είναι απόλυτα ικανοποιημένοι, τις περισσότερες φορές, γιατί επιθυμούν να έχουν διαθέσιμα εργαλεία τα οποία θα λύνουν τα προβλήματα στο έπακρο. Τα δίκτυα που υπάρχουν, όμως, έχουν ακρίβεια 85- 90% και οι εφαρμογές που ανέχονται τόσο μεγάλο σφάλμα είναι λίγες.

Οι κυριότεροι τομείς στους οποίους γίνεται εφαρμογή τεχνητών νευρωνικών δικτύων (ΤΝΔ) είναι:

-Η επεξεργασία γλώσσας (μετατροπή προφορικού λόγου σε γραπτό, φωνητικές εντολές, αυτόματη μετάφραση ή αντιγραφή, αναγνώριση φωνής, βοηθήματα σε κωφούς ή σε άτομα με ειδικές ικανότητες)

-Αναγνώριση χαρακτήρων (μετατροπή χειρόγραφων, υπογραφών κ.ά. σε ηλεκτρονική μορφή)

-Συμπίεση εικόνων/ δεδομένων (συμπίεση και αποσυμπίεση δεδομένων σε πραγματικό χρόνο)

-Αναγνώριση προτύπων (επεξεργασία - προγραμματισμός αισθητήρων/ ανιχνευτών, έλεγχος ποιότητας, ιατρικές διαγνώσεις)

-Επεξεργασία σήματος

-Έλεγχος πολύπλοκων συστημάτων

-Οικονομία (χρηματιστήριο, επιτόκια, τραπεζικοί δείκτες)

Αφού τα νευρωνικά δίκτυα είναι πολύ καλά στο να αναγνωρίζουν τάσεις και πρότυπα στα δεδομένα, είναι πολύ καλά σε τομείς πρόβλεψης όπως:

-Πρόβλεψη πωλήσεων/ ζήτησης

-Βιομηχανικός έλεγχος διεργασιών

-Έρευνα πελατών- ερευνά αγοράς

-Επικύρωση στοιχείων

-Διαχείριση κινδύνου (χρηματοοικονομικές προβλέψεις)

-Marketing

Υπάρχει και η δυνατότητα χρήσης των νευρωνικών δικτύων στον εντοπισμό πόρων και στον προγραμματισμό. Επίσης, μπορούν να χρησιμοποιηθούν στην προσπέλαση βάσεων δεδομένων, δηλαδή στον εντοπισμό συγκεκριμένων στοιχείων από τα δεδομένα που υπάρχουν αποθηκευμένα σε μία βάση.

Υπάρχουν και κάποιες άλλες εφαρμογές τεχνητών νευρωνικών δικτύων σε τομείς της οικονομίας. Η διαδικασία έγκρισης δανείου σε δανειολήπτες, περιλαμβάνει την συμπλήρωση διάφορων εντύπων τα οποία θα βοηθήσουν τον υπεύθυνο να αποφασίσει. Τα δεδομένα από αυτά τα έντυπα, που έχουν χρησιμοποιηθεί κατά το παρελθόν, χρησιμοποιούνται για να εκπαιδευτεί ένα δίκτυο έτσι ώστε να είναι σε θέση να δώσει θετική ή αρνητική απάντηση για την χορήγηση ενός δανείου. Πιστωτικά ιδρύματα χρησιμοποιούν παρόμοιες τεχνικές οπισθόδρομης διάδοσης για να καθορίσουν πιστωτικά όρια και πιστωτικούς κινδύνους.

Στον τομέα του άμεσου marketing, εφαρμόζονται νευρωνικά δίκτυα σε βάσεις δεδομένων, έτσι ώστε να αυξηθούν οι τηλεφωνικές παραγγελίες, μειώνοντας τα τηλεφωνήματα σε ακατάλληλες ώρες, για συγκεκριμένες πληθυσμιακές ομάδες.

Ειδικότερα, τα προβλήματα πρόβλεψης μπορούν να χωριστούν σε δύο κατηγορίες. Σε προβλήματα ταξινόμησης και σε προβλήματα παλινδρόμησης. Στην ταξινόμηση, το ζητούμενο είναι να αποφανθούμε σε ποιά από ορισμένες διακριτές ομάδες ανήκουν οι διατάξεις δεδομένων εισόδου. Οι εφαρμογές έχουν να κάνουν με εκχώρηση ή όχι πίστωσης, με εντοπισμό ασθενειών, αναγνώριση υπογραφής. Σε όλες αυτές τις περιπτώσεις η έξοδος είναι μία καθαρά μονοσήμαντη μεταβλητή. Οι πιο γνωστές περιπτώσεις, είναι αυτές όπου οι διαφορετικές κλάσεις είναι δύο, αλλά η επίλυση προβλημάτων με περισσότερες κλάσεις είναι επίσης εφικτή.

Στα προβλήματα παλινδρόμησης, ο στόχος είναι να προβλέψουμε την τιμή μιας συνεχούς μεταβλητής, όπως είναι η τιμή μιας μετοχής για την επόμενη μέρα, η κατανάλωση καυσίμου ενός αυτοκινήτου, τα έσοδα του επόμενου χρόνου και άλλα. Σε αυτές τις περιπτώσεις η έξοδος είναι μία απλή αριθμητική τιμή.

ΚΕΦΑΛΑΙΟ 4:ΠΡΟΣΑΡΜΟΣΤΙΚΟ ΝΕΥΡΟ-ΑΣΑΦΕΣ ΣΥΣΤΗΜΑ ΕΞΑΓΩΓΗΣ ΣΥΜΠΕΡΑΣΜΑΤΩΝ

ADAPTIVE NEURO-FUZZY INFERENCE SYSTEM-(ANFIS) Ή ADAPTIVE NETWORK-BASED FUZZY INFERENCE SYSTEM

Το μοντέλο ANFIS αποτελεί κομμάτι της ευρέως χρησιμοποιούμενης κατηγορίας των υβριδικών νεύρο-ασαφών ελεγκτών. Στο μοντέλο εμπεριέχεται μια διαδικασία εκπαίδευσης όμοια με των τεχνητών νευρωνικών δικτύων , η οποία δίνει τη δυνατότητα να χρησιμοποιηθεί η πληροφορία από ένα σύνολο δεδομένων στη διαδικασία διαμόρφωσης των συναρτήσεων μεταφοράς ενός ασαφούς συστήματος. Συγκεκριμένα με ένα σύνολο δεδομένων εισόδου/εξόδου , ρυθμίζονται οι συναρτήσεις συμμετοχής ενός ασαφούς συστήματος ελέγχου, χρησιμοποιώντας είτε αποκλειστικά τον αλγόριθμο οπισθόδρομης διάδοσης (Back propagation algorithm), είτε με χρήση του παραπάνω σε συνδυασμό με τη μέθοδο των ελαχίστων τετράγωνων. Το μοντέλο ANFIS βασίζεται κατά κύριο λόγο σε σύστημα τύπου Sugeno, αν και είναι δυνατό να γίνει χρήση και άλλων μοντέλων όπως πχ μοντέλο Mamdani.Κρίνεται απαραίτητο να αναφερθεί ότι το μοντέλο είναι αποδοτικό όταν τα δεδομένα εκπαίδευσης αντιπροσωπεύουν ικανοποιητικά τα χαρακτηριστικά τα οποία πρόκειται να μοντελοποιήσουν, γεγονός που απαιτεί σχετικά μεγάλο μέγεθος δείγματος δεδομένων.

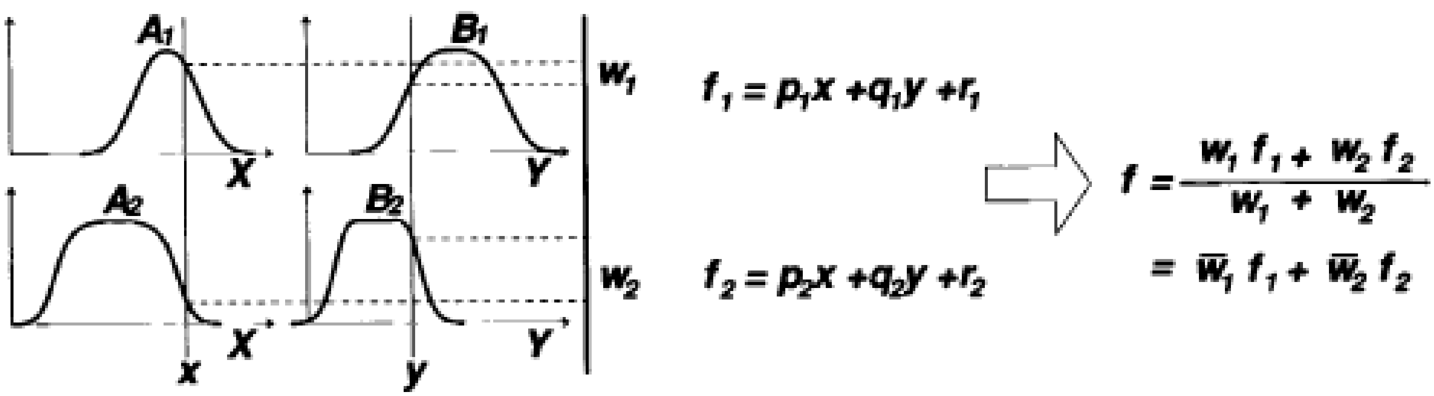
Παρακάτω θα αναλυθούν τα χαρακτηριστικά του ANFIS, η δομή του, αλλά και ο αλγόριθμος εκπαίδευσης που χρησιμοποιείται(αλγόριθμος οπισθόδρομης διάδοσης και εκτίμηση ελαχίστων τετραγώνων) .

4.1: Η Αρχιτεκτονική του ANFIS

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

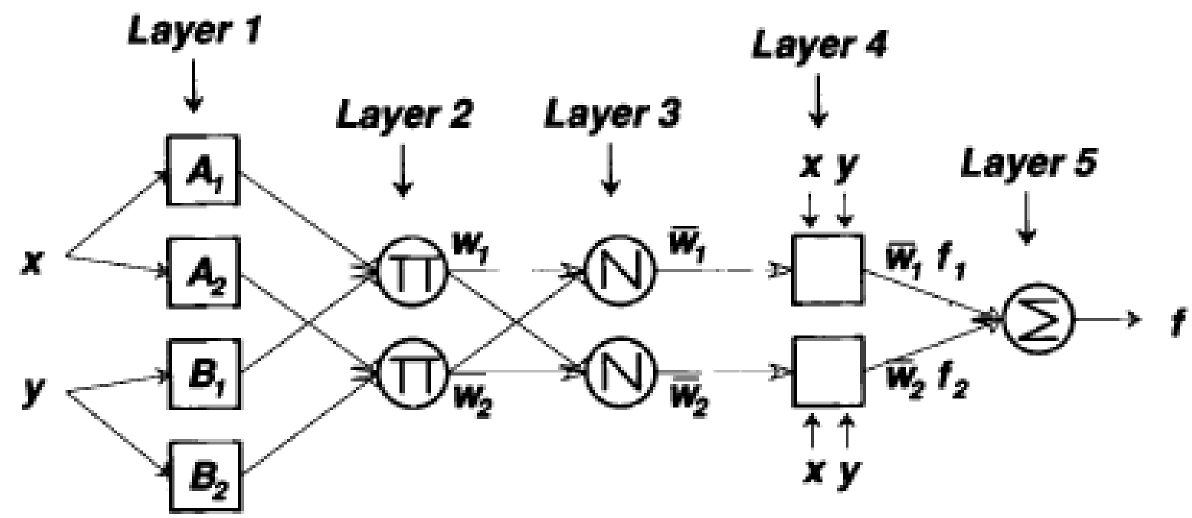
$$\text{Rule 1: If } x \text{ is } A_1 \text{ and } y \text{ is } B_1, \text{ then } f_1 = p_1x + q_1y + r_1,$$
$$\text{Rule 2: If } x \text{ is } A_2 \text{ and } y \text{ is } B_2, \text{ then } f_2 = p_2x + q_2y + r_2.$$

Το παρακάτω σχήμα δείχνει με απλό τρόπο τη διαδικασία εξαγωγής συμπεράσματος (inference procedure) του μοντέλου Sugeno, στην περίπτωση όπου για t-operator έχει επιλεγεί η τομή των δύο ασαφών συνόλων (A,B), οπότε μA∩B(x)= min[μA(x), μB(x)].



Σχήμα 4.1: Μιας διπλής-εισόδου πρώτης-εντολής (first order) ασαφές μοντέλο Sugeno με δυο κανόνες.

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



Σχήμα 4.2: Αρχιτεκτονική του ANFIS

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

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

$$O_{1,i} = \mu_{A_i}(x), \qquad \text{for } i = 1, 2, \text{ or}$$
$$O_{1,i} = \mu_{B_{i-2}}(y), \qquad \text{for } i = 3, 4,$$

όπου x (ή y)
– η είσοδος στον κόμβο και A_i(or B_{i-2})
- η γλωσσική μεταβλητή (small, large, κλπ.) που σχετίζεται με αυτή τη συνάρτηση του κόμβου.

Με άλλα λόγια, το $O_{1,i}$, είναι ο βαθμός συμμετοχής του

$A\ (= A_1,\ A_2,\ B_1 \text{ or } B_2)$

και καθορίζει το βαθμό στον οποίο η είσοδος x (ή y) ικανοποιεί τον ποσοτικοποιητή A . Εδώ η συνάρτηση συμμετοχής για το A μπορεί να είναι οποιαδήποτε κατάλληλη παραμετρική συνάρτηση συμμετοχής όπως η καμπανοειδής για παράδειγμα :

$$\mu_A(x) = \frac{1}{1 + \left|\frac{x - c_i}{a_i}\right|^{2b}},$$

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

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

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

Κάθε κόμβος-έξοδος αντιπροσωπεύει το βαθμό ενεργοποίησης ενός κανόνα (firing strength). Γενικά, οποιεσδήποτε άλλες T-norm (operators) που σημαίνουν τον ασαφή τελεστή AND μπορούν να χρησιμοποιηθούν σαν συνάρτηση κόμβων σε αυτό το επίπεδο.

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

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

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

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

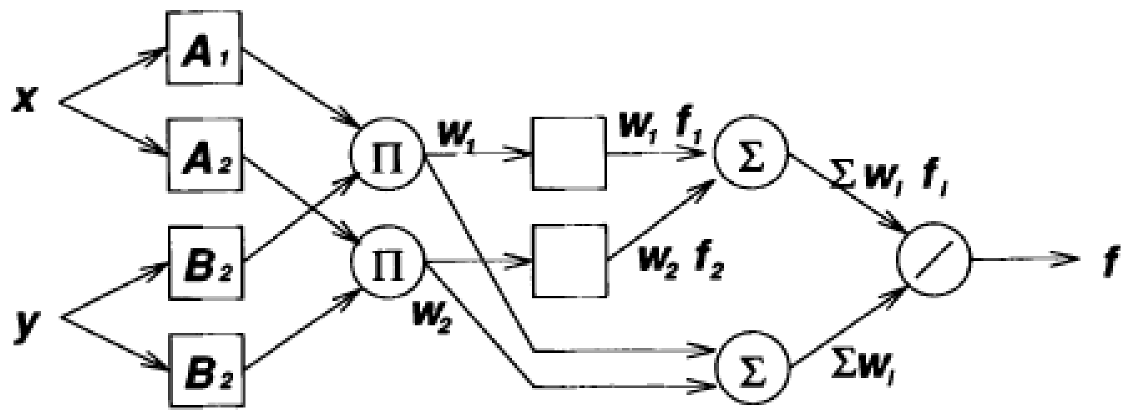
$$O_{4,i} = \overline{w}_i f_i = \overline{w}_i (p_i x + q_i y + r_i),$$

Όπου $\overline{w}_i$ είναι ο κανονικοποιημένος βαθμός ενεργοποίησης από το επίπεδο 3 , και $\{p_i,\ q_i,\ r_i\}$ το σύνολο των παραμέτρων. Οι παράμετροι σε αυτό το επίπεδο αναφέρονται ως επακόλουθοι (consequent parameters).

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

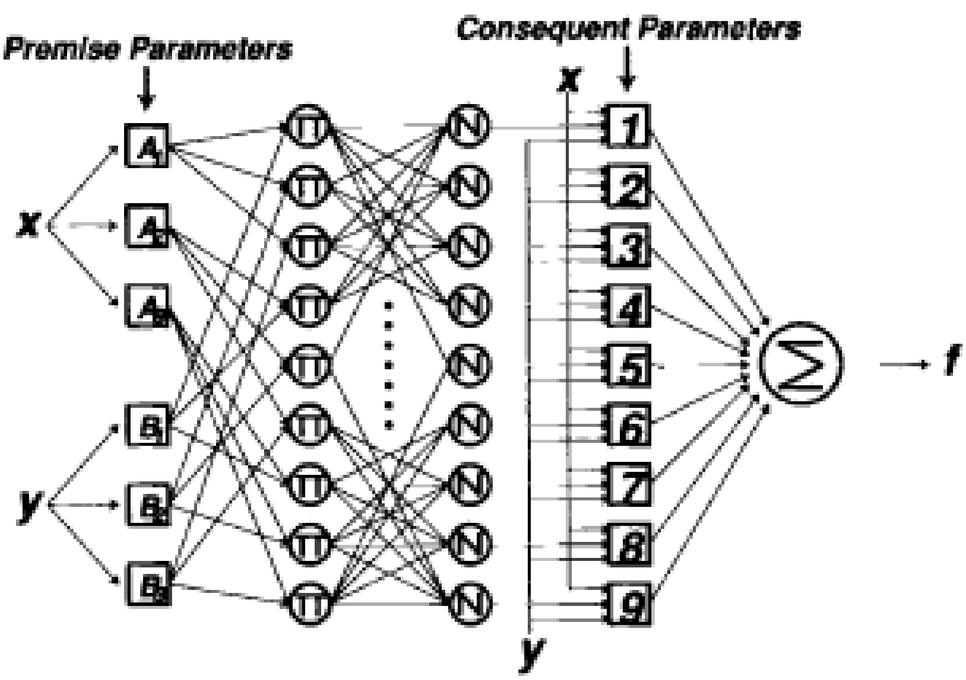
$$\text{overall output} = O_{5,1} = \sum_i \overline{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i}$$

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

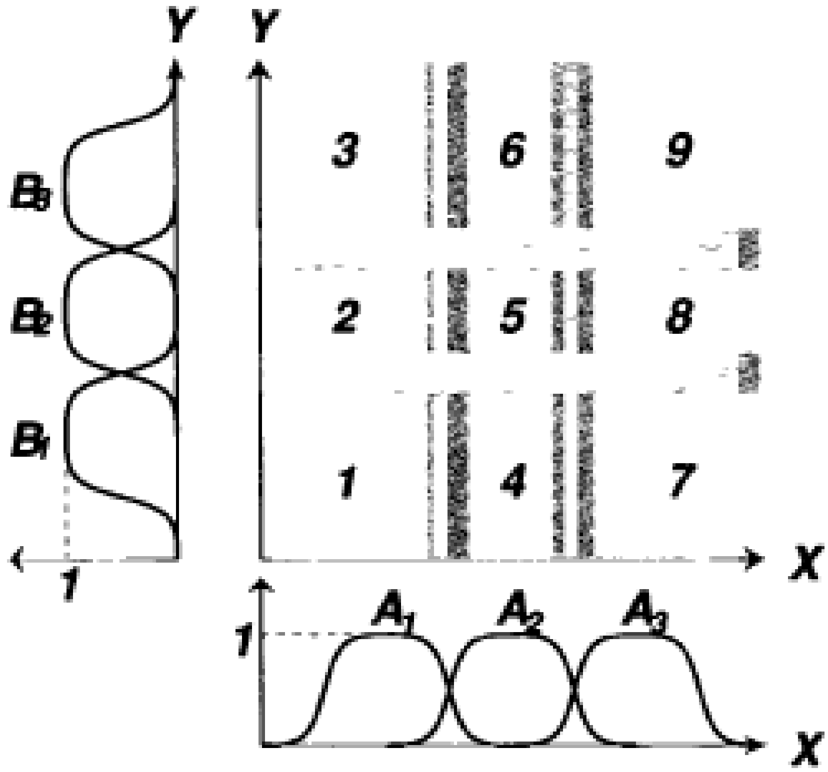


Σχήμα4 .3: Αρχιτεκτονική του ANFIS για το ασαφές μοντέλο Sugeno στο οποίο η κανονικοποίηση των βαρών γίνεται στο τελευταίο επίπεδο

Εκτός από το μηχανισμό εξαγωγής συμπεράσματος κατά Sugeno, μπορούμε να κατασκευάσουμε ένα ANFIS με χρήση του μοντέλου κατά Mamdani. Το μοντέλο κατά Sugeno χρησιμοποιείται περισσότερο, μιας και διακρίνεται για την διαφάνεια και την αποτελεσματικότητα του. Στο σχήμα 4.3(α) που ακολουθεί βλέπουμε μια αρχιτεκτονική ANFIS η οποία είναι ισοδύναμη με ένα ασαφές μοντέλο κατά Sugeno πρώτου βαθμού δυο εισόδων και εννέα κανόνων, σε κάθε είσοδο θεωρούμε ότι αντιστοιχούν τρεις συναρτήσεις συμμετοχής. Το σχήμα 4.3(β) παρακάτω απεικονίζει πως δύο διαστάσεων χώρος εισόδου είναι χωρισμένος σε εννέα υπέρθετες (overlapping) ασαφείς περιοχές όπου κάθε μια ελέγχεται από ένα ασαφή κανόνα if-then. Αυτό σημαίνει ότι το μέρος των προϋποθέσεων ενός κανόνα προσδιορίζει μια ασαφή περιοχή, ενώ το μέρος των συμπερασμάτων προσδιορίζει την έξοδο μέσα στην περιοχή.



Σχήμα 4.3.α: Η αρχιτεκτονική του ANFIS για το ασαφές μοντέλο κατά Sugeno με δύο εισόδους και εννέα κανόνες.



Σχήμα4.3.β: ο χώρος εισόδου, χωρισμένος σε εννέα ασαφείς περιοχές.

4.3: Υβριδικός Αλγόριθμος Εκπαίδευσης του ANFIS.

Το ANFIS χρησιμοποιεί έναν υβριδικό αλγόριθμο εκμάθησης για να προσδιορίσει τις παραμέτρους του ασαφούς συστήματος τύπου Sugeno. Εφαρμόζει έναν συνδυασμό της μεθόδου ελαχίστων τετραγώνων (least-squares method) και της μεθόδου οπισθόδρομης βαθμωτής ελαχιστοποίησης (backpropagation gradient descent) για την εκπαίδευση παραμέτρων των συναρτήσεων συμμετοχής του ασαφούς συστήματος εξαγωγής συμπεράσματος (ΑΣΕΣ) ώστε να μιμηθεί ένα δοσμένο σύνολο δεδομένων εκπαίδευσης

4.4: Εκτιμητής Ελαχίστων Τετραγώνων (Least-Squares Estimator)

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

$$y = \theta_1 f_1(\mathbf{u}) + \theta_2 f_2(\mathbf{u}) + \cdots + \theta_n f_n(\mathbf{u}), \qquad \text{εξ 4.1}$$

Όπου $\mathbf{u} = [u_1, \cdots, u_p]^T$ είναι το διάνυσμα εισόδου του μοντέλου , $f_1, \cdots, f_n$ είναι γνωστές συναρτήσεις του **u** και $\theta_1, \cdots, \theta_n$ είναι άγνωστες παράμετροι που θα υπολογιστούν. Η παραπάνω εξίσωση καλείται συνάρτηση παλινδρόμησης, και τα θi ονομάζονται συντελεστές παλινδρόμησης.

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

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

εξ.4.2

Σε μορφή πινάκων, οι προηγούμενες εξισώσεις μπορούν να γραφτούν σε συνοπτική μορφή:

$$\mathbf{A\theta} = \mathbf{y} \qquad \text{εξ 4.3.}$$

όπου A είναι ένας m× n πίνακας (πίνακας σχεδιασμού):

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

θ είναι ένα n ×1 διάνυσμα άγνωστων παραμέτρων:

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

και **y** είναι ένα m×1 διάνυσμα εξόδου:

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

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

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

Το μεγαλύτερο μέρος των υπολογισμών βασίζεται στους πίνακες A και y , μερικές φορές γίνεται αναφορά στο $(\mathbf{a}_i^T ; y_i)$, σαν το i-οστό ζευγάρι δεδομένων του συνόλου δεδομένων εκπαίδευσης.
Για να προσδιοριστεί μεμονωμένα το άγνωστο διάνυσμα θ , είναι απαραίτητο να ισχύει $m \geq n$. Αν ο A είναι τετραγωνικός ($m = n$) και αντιστρέψιμος, τότε η εξίσωση 4.3 μπορεί να λυθεί ως προς τον άγνωστο x και γίνεται:

$$\theta = \mathrm{A}^{-1} y$$

Ο m είναι συνήθως μεγαλύτερος από τον n , που σημαίνει ότι υπάρχουν περισσότερα ζευγάρια στοιχείων από τις παραμέτρους. Σε αυτήν την περίπτωση, μια ακριβής λύση που να ικανοποιεί όλες τις m εξισώσεις δεν είναι πάντα δυνατή, δεδομένου ότι τα στοιχεία μπορεί να μολυνθούν από θόρυβο, ή το μοντέλο μπορεί να μην είναι κατάλληλο για την περιγραφή του συστήματος στόχου. Κατά συνέπεια η εξίσωση 4.3 πρέπει να τροποποιηθεί με την ενσωμάτωση ενός διανύσματος λάθους e για να αποτελέσει το τυχαίο λάθος θορύβου ή το τυχαίο λάθος διαμόρφωσης ως εξής:

$$A\theta + e = y$$

Τώρα, αντί να βρεθεί η ακριβής λύση της εξίσωσης 4.3, πρέπει να βρεθεί το $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ το οποίο ελαχιστοποιεί το άθροισμα του τετραγωνικού σφάλματος και ορίζεται ως εξής :

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

εξ.4.4

όπου **e = y - Aθ** είναι το διάνυσμα λάθους που παράγεται από μια συγκεκριμένη επιλογή του **θ** . Πρέπει να σημειωθεί ότι το E(**θ**) είναι σε τετραγωνική μορφή και έχει ένα μοναδικό ελάχιστο $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$.

Το παρακάτω θεώρημα δηλώνει έναν απαραίτητο όρο που ικανοποιείται από τον εκτιμητή ελαχίστων τετραγώνων $\hat{\boldsymbol{\theta}}$.

Θεώρημα 4.1 Εκτιμητής ελαχίστων τετραγώνων:

Το τετραγωνικό σφάλμα στην εξίσωση 6 ελαχιστοποιείται όταν $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, ο οποίος καλείται εκτιμητής ελαχίστων τετραγώνων (LSE) και ο οποίος ικανοποιεί την κανονική εξίσωση

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

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

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

4.5 Οπισθόδρομη Εκμάθηση για Έμπροσθεν-Τροφοδοτούμενα Δίκτυα (ANFIS)

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

Έστω ότι δεδομένο έμπροσθεν-τροφοδοτούμενο προσαρμοστικό δίκτυο στην αναπαράστασή του σε επίπεδα έχει L επίπεδα και το επίπεδο l (όπου l = 0,1,..., L; l = 0 αντιπροσωπεύει το επίπεδο εισαγωγής) έχει N(l) κόμβους. Τότε η έξοδος και η συνάρτηση του κόμβου i [i = 1,..., N(l)] στο επίπεδο l μπορούν να αναπαρασταθούν σαν x_{ij} και f_{ij} αντίστοιχα. Ας υποτεθεί ότι δεν υπάρχει καμία σύνδεση μεταξύ μη συνεχόμενων στρωμάτων. Δεδομένου ότι η έξοδος ενός κόμβου εξαρτάται από τα εισερχόμενα σήματα και το σύνολο παραμέτρων του κόμβου, προκύπτει η ακόλουθη γενική έκφραση για τη συνάρτηση των κόμβων f_{ij} :

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

όπου α, β, γ κτλ. είναι οι παράμετροι αυτού του κόμβου.
Υποθέτοντας ότι το δοσμένο σύνολο δεδομένων εκπαίδευσης έχει P καταχωρήσεις, μπορεί να οριστεί ένα μέτρο σφάλματος για την p-οστή ($1 \leq p \leq P$) καταχώρηση των δεδομένων εκπαίδευσης σαν το άθροισμα των τετραγωνικών σφαλμάτων:

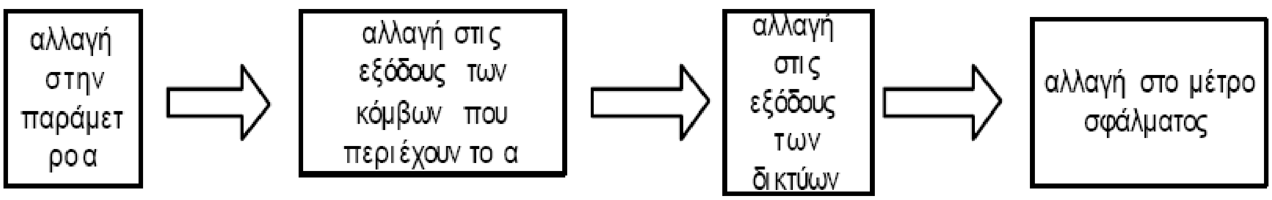
$$E_p = \sum_{k=1}^{N(l)} (d_k - x_{L,k})^2 \qquad \text{εξ .4.5.1}$$

Όπου d_k είναι το k-οστό συστατικό του p-οστού επιθυμητού διανύσματος εξόδου και x_{Lk} είναι το k-οστό συστατικό του πραγματικού διανύσματος εξόδου που

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

$$E = \Sigma_{\rho=1}^{P} E_P$$

Επιπλέον, ας υποτεθεί ότι το E_p εξαρτάται μόνο από τους κόμβους εξόδου. Για να χρησιμοποιηθεί η βαθμωτή ελαχιστοποίηση ώστε να ελαχιστοποιηθεί το μέτρο σφάλματος, πρέπει πρώτα να βρεθεί το διάνυσμα κλίσης. Πριν υπολογιστεί το διάνυσμα κλίσης, πρέπει να παρατηρηθούν οι ακόλουθες αιτιώδεις σχέσεις:



Τα βέλη συνεπαγωγής $\Rightarrow$ δείχνουν τις αιτιώδεις σχέσεις, δηλαδή μια μικρή αλλαγή σε μια παράμετρο επηρεάζει την έξοδο του κόμβου που περιέχει το a . Αυτό με τη σειρά του θα επηρεάσει την έξοδο του τελευταίου επιπέδου και συνεπώς το μέτρο σφάλματος.

Η βασική αρχή στον υπολογισμό του διανύσματος κλίσης είναι να περαστούν μια σειρά από πληροφορίες παραγώγων ξεκινώντας από το επίπεδο εξόδου και πηγαίνοντας ανάποδα από επίπεδο σε επίπεδο έως ότου καταλήξει η διαδικασία στο επίπεδο εισόδου.

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

$$\in_{i,i} = \frac{\partial^+ E_P}{\partial x_{i,i}}$$

Εξ. .4.5.4

Η έκφραση αυτή ονομάστηκε διατεταγμένη παράγωγος κατά τον Werbos. Το σήμα σφάλματος για τον i-οστό κόμβο εξόδου (στο επίπεδο L) μπορεί να υπολογιστεί απευθείας από την παρακάτω εξίσωση:

$$\in_{1,i} = \frac{\partial^+ E_p}{\partial x_{1,i}} = \frac{\partial E_p}{\partial x_{1,i}}$$

Εξ .4.5.5

Αυτό ισούται με $\in_{Lj}=-2(d_i-x_{Lj})$ αν το E_p ορίζεται όπως στην εξίσωση 4.5..2.. Για τον εσωτερικό κόμβο στην i-οστή θέση του επιπέδου l το σήμα σφάλματος μπορεί να βρεθεί από τον παρακάτω κανόνα αλυσίδας :

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

εξ:4.5.6

όπου $0 \le l \le L-1$. Άρα το σήμα σφάλματος ενός εσωτερικού κόμβου στο στρώμα l μπορεί να εκφραστεί σαν γραμμικός συνδυασμός των κόμβων στο στρώμα l +1. Επομένως, για οποιαδήποτε l και i [και $1 \le i \le N(l)$], μπορούν να βρεθούν τα:

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

εφαρμόζοντας πρώτα την εξίσωση (4.5.5) μία φορά για να υπολογιστούν τα σήματα σφάλματος στο επίπεδο εξόδου, και έπειτα εφαρμόζοντας την εξίσωση (4.5.6) επαναληπτικά έως ότου καταλήξει η διαδικασία στο επιθυμητό επίπεδο l . Η διαδικασία αυτή καλείται οπισθοδρόμηση

δεδομένου ότι τα σήματα σφάλματος λαμβάνονται διαδοχικά από το επίπεδο εξόδου προς το επίπεδο εισόδου(όπισθεν λήψη σφάλματος).
Το διάνυσμα κλίσης ορίζεται ως η παράγωγος του μέτρου σφάλματος ως προς κάθε παράμετρο και έτσι πρέπει να εφαρμοστεί ο κανόνας αλυσίδας ξανά για να βρεθεί το διάνυσμα κλίσης. Εάν a είναι μια παράμετρος του i-οστού κόμβου στο επίπεδο l , τότε ισχύει:

$$\frac{\partial^+ E_p}{\partial a} = \frac{\partial^+ E_p}{\partial x_{l,i}} \frac{\partial f_{l,i}}{\partial a} = \in_{l,i} \frac{\partial f_{l,i}}{\partial a}$$ εξ .4.5.7

Αξίζει να σημειωθεί ότι αν επιτραπεί στην παράμετρο α να μοιράζεται μεταξύ διαφορετικών κόμβων, τότε η παραπάνω εξίσωση (4.5.7) θα πρέπει να αλλαχθεί σε μια πιο γενική μορφή:

$$\frac{\partial^+ E_p}{\partial a} = \sum_{x^* \in S} \frac{\partial^+ E_p}{\partial x^*} \frac{\partial f^*}{\partial a}$$ εξ4.5 .8

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

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

$$\frac{\partial^+ E_P}{\partial a} = \sum_{p=1}^{P} \frac{\partial^+ E_p}{\partial a}$$ εξ 4.5.9

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

$$\Delta a = -\eta \frac{\partial^{+} E_{p}}{\partial a}$$

όπου το n είναι ο ρυθμός εκμάθησης, το οποίο μπορεί να εκφραστεί αναλυτικότερα ως:

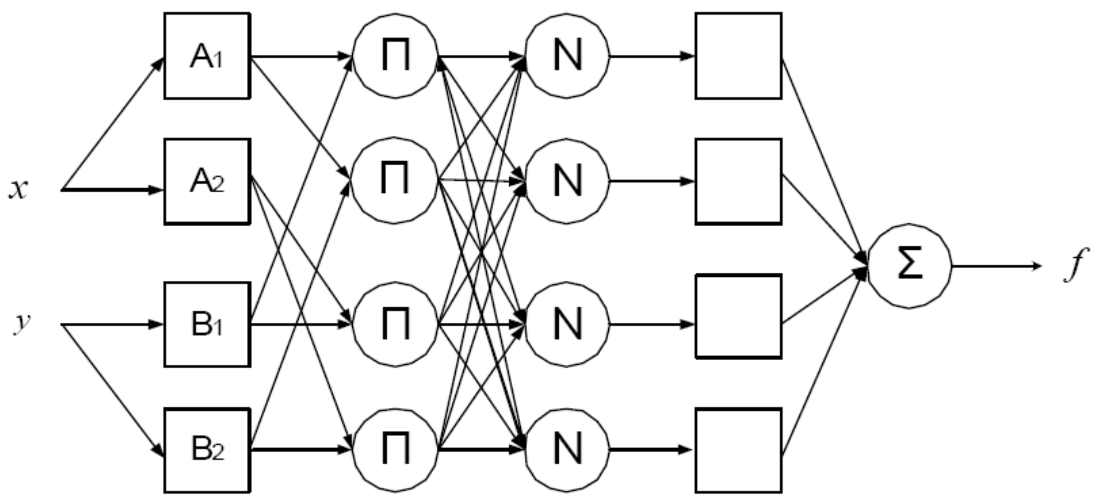
$$\eta = \frac{\kappa}{\sqrt{\sum_{\alpha} \left(\frac{\partial E}{\partial \alpha} \right)^{2}}}$$

Όπου k το μέγεθος βήματος, το μήκος δηλαδή κάθε μετάβασης κατά μήκος της κατεύθυνσης κλίσης στο διάστημα παραμέτρου. Συνήθως το μέγεθος βήματος μπορεί να αλλαχθεί για να μεταβληθεί η ταχύτητα της σύγκλισης. Όταν ένα έμπροσθεν-τροφοδοτούμενο δίκτυο n-κόμβων αναπαριστάται στην τοπολογική του διάταξη, μπορεί να υπολογιστεί το μέτρο σφάλματος E_p σαν την έξοδο ενός επιπλέον κόμβου με ένδειξη n +1, του οποίου η συνάρτηση κόμβου f n+1 μπορεί να οριστεί από τις εξόδους κάθε κόμβου με μικρότερη ένδειξη.(Επομένως, το E_p μπορεί να εξαρτάται άμεσα από οποιουσδήποτε κόμβους.) Εφαρμόζοντας πάλι τον κανόνα αλυσίδας, ισχύει ο ακόλουθος συνοπτικός τύπος για τον υπολογισμό του σήματος σφάλματος

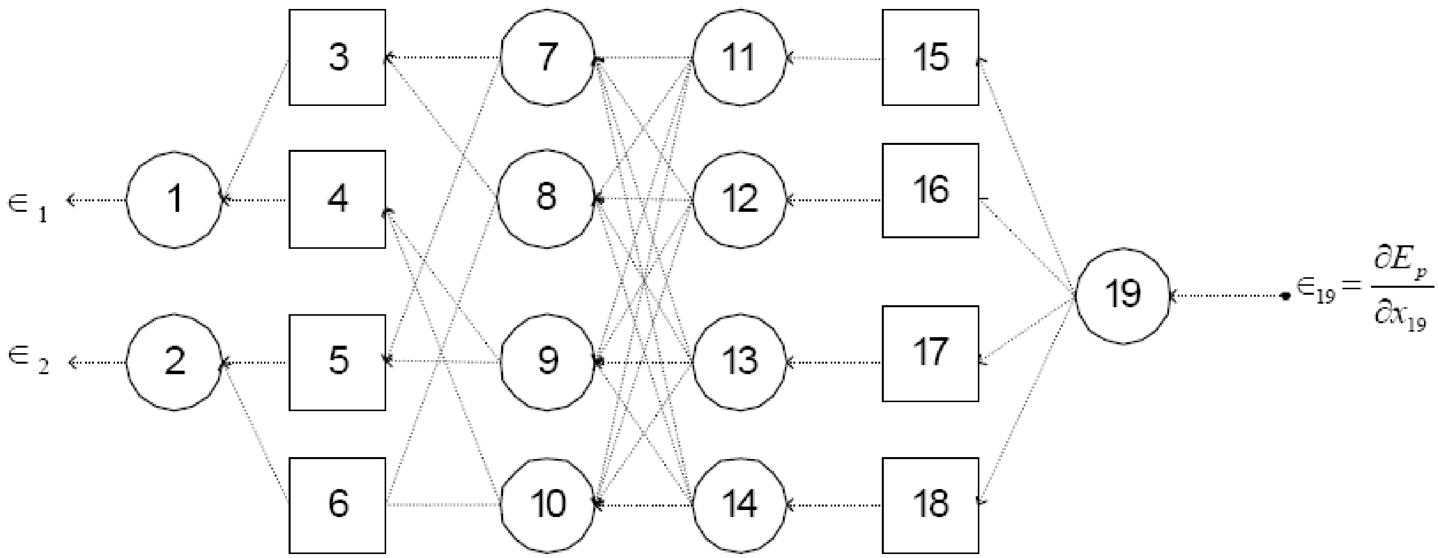
$$\in_{i} = {\partial E_{p}} / {\partial x_{i}}$$

$$\in_{i} = \frac{\partial f_{n+1}}{\partial x_{i}} + \sum_{i<j\leq n} \in_{j} \frac{\partial f_{j}}{\partial x_{i}}$$

Όπου ο πρώτος όρος δείχνει μια άμεση επίδραση του xi στο E_p μέσω της άμεσης διαδρομής από τον κόμβο i στον κόμβο n +1 και κάθε όρος παραγώγου στο άθροισμα δείχνει την έμμεση επίδραση του xi στο E_p μόλις βρεθεί το σήμα σφάλματος για κάθε κόμβο, τότε το διάνυσμα κλίσης για τις παραμέτρους παράγεται όπως πριν.
Ένας άλλος συστηματικός τρόπος να υπολογιστούν τα σήματα σφάλματος είναι μέσω της αναπαράστασης του δικτύου διάδοσης σφάλματος (ή του μοντέλου ευαισθησίας), το οποίο λαμβάνεται από το αρχικό προσαρμόσιμο δίκτυο με την αντιστροφή των συνδέσεων και την παροχή των σημάτων σφάλματος στο επίπεδο εξόδου ως είσοδοι στο νέο δίκτυο. Τα ακόλουθα σχήμα δείχνουν το μοντέλο ANFIS, το προσαρμοστικό δίκτυο και το δίκτυο διάδοσης σφάλματος :



Σχήμα4.4 : Το προσαρμοστικό δίκτυο



Σχήμα 4.5 : Το δίκτυο διάδοσης σφάλματος

Αμέσως μετά υπολογίζονται τα σήματα σφάλματος στους εσωτερικούς κόμβους. Χρησιμοποιούνται τα f_i και x_i για το συμβολισμό της συνάρτησης και της εξόδου του κόμβου i . Η έξοδος του κόμβου i είναι το σήμα σφάλματος αυτού του κόμβου στο πραγματικό προσαρμόσιμο δίκτυο. Σε σύμβολα, αν επιλεχθεί το τετραγωνικό μέτρο σφάλματος για E_p τότε ισχύει το ακόλουθό:

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

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

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

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

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

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

Αυτό είναι επειδή όλοι αυτοί οι κόμβοι εξαρτώνται από τον κόμβο 19.

Ομοίως ισχύει και για τους κόμβους 11, 12, 13 και 14 με τη διαφορά ότι κάθε ένας από αυτούς εξαρτάται από διαφορετικό κόμβο:

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

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

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

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

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

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

$$\in_{9} = \frac{\partial^+ E_p}{\partial x_{9}} = \in_{14} \frac{\partial f_{14}}{\partial x_{9}} + \in_{13} \frac{\partial f_{13}}{\partial x_{9}} + \in_{12} \frac{\partial f_{12}}{\partial x_{9}} + \in_{11} \frac{\partial f_{11}}{\partial x_{9}}$$

$$\in_{8} = \frac{\partial^+ E_p}{\partial x_{8}} = \in_{14} \frac{\partial f_{14}}{\partial x_{8}} + \in_{13} \frac{\partial f_{13}}{\partial x_{8}} + \in_{12} \frac{\partial f_{12}}{\partial x_{8}} + \in_{11} \frac{\partial f_{11}}{\partial x_{8}}$$

$$\in_{7} = \frac{\partial^+ E_p}{\partial x_{7}} = \in_{14} \frac{\partial f_{14}}{\partial x_{7}} + \in_{13} \frac{\partial f_{13}}{\partial x_{7}} + \in_{12} \frac{\partial f_{12}}{\partial x_{7}} + \in_{11} \frac{\partial f_{11}}{\partial x_{7}}$$

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

$$\in_6 = \frac{\partial^+ E_p}{\partial x_6} = \frac{\partial^+ E_p}{\partial x_{10}} \frac{\partial f_{10}}{\partial x_6} + \frac{\partial^+ E_p}{\partial x_8} \frac{\partial f_8}{\partial x_6} = \in_{10} \frac{\partial f_{10}}{\partial x_6} + \in_8 \frac{\partial f_8}{\partial x_6}$$

$$\in_5 = \frac{\partial^+ E_p}{\partial x_5} = \frac{\partial^+ E_p}{\partial x_9} \frac{\partial f_9}{\partial x_5} + \frac{\partial^+ E_p}{\partial x_7} \frac{\partial f_7}{\partial x_5} = \in_9 \frac{\partial f_9}{\partial x_5} + \in_7 \frac{\partial f_7}{\partial x_5}$$

$$\in_4 = \frac{\partial^+ E_p}{\partial x_4} = \frac{\partial^+ E_p}{\partial x_{10}} \frac{\partial f_{10}}{\partial x_4} + \frac{\partial^+ E_p}{\partial x_9} \frac{\partial f_9}{\partial x_4} = \in_{10} \frac{\partial f_{10}}{\partial x_4} + \in_9 \frac{\partial f_9}{\partial x_4}$$

$$\in_3 = \frac{\partial^+ E_p}{\partial x_3} = \frac{\partial^+ E_p}{\partial x_8} \frac{\partial f_8}{\partial x_3} + \frac{\partial^+ E_p}{\partial x_7} \frac{\partial f_7}{\partial x_3} = \in_8 \frac{\partial f_8}{\partial x_3} + \in_7 \frac{\partial f_7}{\partial x_3}$$

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

$$\in_2 = \frac{\partial^+ E_p}{\partial x_2} = \frac{\partial^+ E_p}{\partial x_6} \frac{\partial f_6}{\partial x_2} + \frac{\partial^+ E_p}{\partial x_5} \frac{\partial f_5}{\partial x_2} = \in_6 \frac{\partial f_2}{\partial x_2} + \in_5 \frac{\partial f_5}{\partial x_2}$$

$$\in_1 = \frac{\partial^+ E_p}{\partial x_1} = \frac{\partial^+ E_p}{\partial x_4} \frac{\partial f_4}{\partial x_1} + \frac{\partial^+ E_p}{\partial x_3} \frac{\partial f_3}{\partial x_1} = \in_4 \frac{\partial f_4}{\partial x_1} + \in_3 \frac{\partial f_3}{\partial x_1}$$

4.6: Συνδυασμός Βαθμωτής Ελαχιστοποίησης και Εκτιμητή Ελαχίστων Τετραγώνων

Η έξοδος ενός προσαρμοστικού δικτύου είναι γραμμική σε μερικές από τις παραμέτρους του δικτύου. Έτσι μπορούν να προσδιοριστούν αυτές οι γραμμικές παράμετροι με τη γραμμική μέθοδο ελαχίστων τετραγώνων που περιγράφτηκε παραπάνω. Αυτή η προσέγγιση οδηγεί σε ένα υβριδικό κανόνα εκμάθησης που συνδυάζει τη βαθμωτή ελαχιστοποίηση (SD) και τον εκτιμητή ελαχίστων τετραγώνων (LSE) για γρήγορο προσδιορισμό των παραμέτρων.

Από την αρχιτεκτονική δομή του ANFIS παρατηρούμε ότι 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. Σε σύμβολα, η έξοδος f ξαναγράφεται όπως φαίνεται παρακάτω:

$$f = \frac{w_1}{w_1 + w_2} f_1 + \frac{w_2}{w_1 + w_2} f_2$$

$$= \overline{w}_1 (p_1 x + q_1 y + r_1) + \overline{w}_2 (p_2 x + q_2 y + r_2)$$

$$= (\overline{w}_1 x) p_1 + (\overline{w}_1 y) q_1 + (\overline{w}_1) r_1 + (\overline{w}_2 x) p_2 + (\overline{w}_2 y) q_2 + (\overline{w}_2) r_2$$

Η οποία είναι γραμμική στις παραμέτρους των συμπερασμάτων p_1, q_1, r_1, p_2, q_2, και r_2.
Έστω ότι:
S= σύνολο όλων των παραμέτρων
S1= σύνολο των (μη γραμμικών) παραμέτρων από τις προϋποθέσεις
S2 =σύνολο των (γραμμικών) παραμέτρων από τα συμπεράσματα

Το προσαρμοστικό δίκτυο έχει μία έξοδο που αναπαριστάται ως εξής:

$$o = F(i, S)$$

όπου i είναι το διάνυσμα των μεταβλητών εισόδου, S είναι το σύνολο των παραμέτρων, και F είναι η συνολική συνάρτηση που εφαρμόζεται από το προσαρμόσιμο δίκτυο

Στην εξίσωση $S = S_1 \oplus S_2$

και H(.) και F(.) είναι η ταυτοτική συνάρτηση και η συνάρτηση του ΑΣΕΣ αντιστοίχως στην εξίσωση η H o F είναι γραμμική στα στοιχεία του 2 S ,

τότε εφαρμόζοντας την H στην εξίσωση $o = F(i, S)$, προκύπτει ότι

H(o) = H o F($B_ι$, S) . To $\oplus$ αντιπροσωπεύει την άμεση άθροιση (direct sum). Η H(.) είναι η ταυτοτική συνάρτηση και η F(.,.) είναι η συνάρτηση του συστήματος ασαφούς συμπεράσματος αντίστοιχα. Κατά συνέπεια, ο υβριδικός αλγόριθμος μάθησης που αναπτύχθηκε παραπάνω μπορεί να εφαρμοστεί άμεσα.

Αναλυτικότερα, όπως συμβαίνει και στα νευρωνικά δίκτυα, θα ξεχωρίσουν δύο φάσεις στη διαδικασία εκπαίδευσης:
1. Εμπροσθόδρομο πέρασμα (forward pass): Στη φάση αυτή το σήμα εισόδου διαδίδεται από το επίπεδο 1 μέχρι το επίπεδο 4 και οι παράμετροι (p_i, q_i, r_i) i = 1,2 εκτιμούνται με τη μέθοδο των ελαχίστων τετραγώνων.
2.Οπισθόδρομο πέρασμα (backward pass): Στη φάση αυτή έχοντας μία ένδειξη του λάθους πραγματοποιείται μία διόρθωση κατά μία ποσότητα που ορίζεται από τη μέθοδο της βαθμωτής κατάβασης (Gradient Descent), των μεταβλητών $α_i$, b_i, c_i.

Το ANFIS διασπά το σύνολο των παραμέτρων του σε δύο υποσύνολα, εκ των οποίων το ένα αποτελεί γραμμικό σύνολο παραμέτρων και, επομένως, είναι δυνατόν να εκπαιδευτεί με γραμμικούς αλγόριθμους, όπως η μέθοδος των ελαχίστων τετραγώνων (least squared). Οι γραμμικοί αλγόριθμοι μάθησης επιτυγχάνουν συνολικά ελάχιστα (global minimums) της συνάρτησης κόστους στο χώρο των παραμέτρων τους και είναι αποδοτικοί από πλευράς απαιτούμενου υπολογιστικού χρόνου. Το δεύτερο υποσύνολο παραμέτρων εκπαιδεύεται με αλγόριθμους που μπορούν να δημιουργήσουν μη-γραμμικές απεικονίσεις, όπως η βαθμωτή κατάβαση (gradient descent). Οι αλγόριθμοι αυτοί είναι απαιτητικοί από πλευράς απαιτούμενου χρόνου εκπαίδευσης και δεν υπάρχει εγγύηση για την πραγματοποίηση του συνολικού ελαχίστου της συνάρτησης κόστους στον χώρο των παραμέτρων τους. Ο αλγόριθμος μάθησης του ANFIS συνδυάζει περάσματα προς τα εμπρός και προς τα πίσω. Στο πέρασμα εμπρός γίνεται η μάθηση του συνόλου των γραμμικών παραμέτρων και στο πέρασμα προς τα πίσω γίνεται η προσαρμογή των μη-γραμμικών αντίστοιχα. Στον παρακάτω πίνακα 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 4.1: Δραστηριότητες κατά την εκπαίδευση

Το πλεονέκτημα της μεθόδου έγκειται στο γεγονός της ύπαρξης τόσο γραμμικών όσο και μη γραμμικών μεθόδων, που την καθιστά ταχύτερη από τα κλασικά νευρωνικά δίκτυα. 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.

Μέχρι τώρα χρησιμοποιούσαμε προκαθορισμένες και αυθαίρετα επιλεγμένες συναρτήσεις συμμετοχής. Πολλές φορές έχουμε μια συλλογή από δεδομένα εισόδου / εξόδου και δεν μπορούμε 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. Έτσι παρέχεται μια μέθοδος ώστε το ασαφές μοντέλο να εκπαιδευτεί με τις πληροφορίες που του δίνει το σύνολο των δεδομένων, ώστε να υπολογιστούν οι παράμετροι των συναρτήσεων συμμετοχής με σκοπό την εύρεση του καλύτερου τρόπο ώστε να επιτρέπει στο σύστημα εξαγωγής συμπερασμάτων να ανιχνεύει τα δεδομένα εισόδου/ εξόδου. Η μέθοδος αυτή είναι παρόμοια με αυτή των νευρωτικών δικτύων.

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

Ο υπολογισμός των παραμέτρων αυτών, που αλλάζουν κατά τη διάρκεια της διαδικασίας εκμάθησης, διευκολύνεται από ένα βαθμωτό διάνυσμα που μας δείχνει πόσο καλά το ΑΣΕΣ μοντελοποιεί τα δεδομένα εισόδου/ εξόδου για το εν λόγο σύνολο παραμέτρων. Μόλις αποκτήσουμε το βαθμωτό διάνυσμα μπορούμε να εφαρμόσουμε κάποιον από τους πολλούς αλγόριθμους 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).

4.7: Περιορισμοί του ANFIS

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

ΚΕΦΑΛΑΙΟ 5:ΑΠΟΤΕΛΕΣΜΑΤΑ ΑΠΟ ΤΗΝ ΧΡΗΣΗ ΤΟΥ ANFIS

5.1: Περιγραφή και Αποτελέσματα του Προγράμματος Πρόβλεψης.

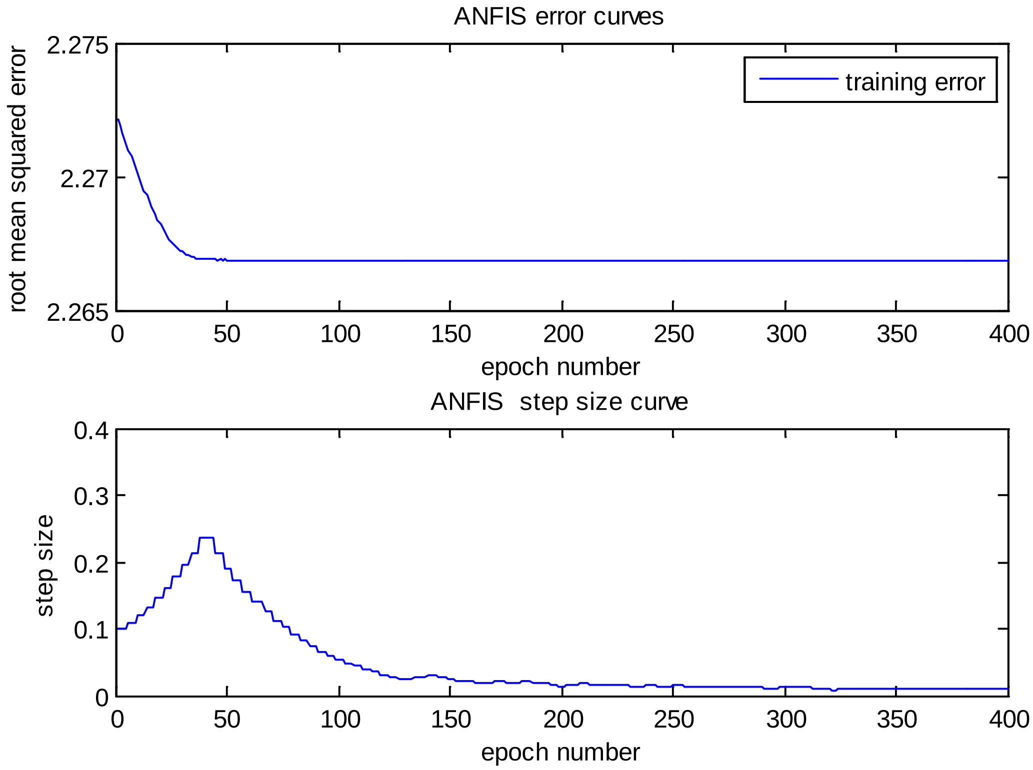
Σε αυτό το κομμάτι της εργασίας δημιουργήθηκε ο αλγόριθμος σε γλώσσα προγραμματισμού Matlab, κάνοντας χρήση του fuzzy logic toolbox, το οποίο περιέχει όλα τα απαραίτητα εργαλεία για την ολοκληρωμένη ανάπτυξη και αναλυτική χρήση ενός Προσαρμοστικού Νεύρο-Ασαφούς Συστήματος Εξαγωγής Συμπεράσματος (Adaptive Neuro-Fuzzy Inference System, ANFIS) για την πρόβλεψη της θερμοκρασίας στην περιοχή της Σούδας.

Ο αλγόριθμος δέχεται δεδομένα από μια χρόνο-σειρά που περιέχει τιμές θερμοκρασίας στην περιοχή της Σούδας με βάση τις τιμές του παρελθόντος, προβλέπει τις αναμενόμενες τιμές για το μέλλον. Η χρόνο-σειρά που χρησιμοποιήσαμε, περιέχει δεδομένα για συγκεκριμένες ημέρες και ώρες από το έτος 1971 έως και το έτος 2000. Αναλυτικότερα τα δεδομένα περιέχουν την εξωτερική θερμοκρασία στην περιοχή της Σούδας ανά τρείς ώρες για κάθε ημέρα από την 0 10/ 1/19 7 1στις 0 000 μέχρι και την 31/12/2000 στις 21:00.

Από το σύνολο των 87415 διαφορετικών τιμών που δημιουργήθηκε το 80% χρησιμοποιήθηκε για την εκπαίδευση του μοντέλου ενώ το υπόλοιπο 20% χρησιμοποιήθηκε για την αξιολόγηση της αξιοπιστίας του μοντέλου , δηλαδή πόσο αξιόπιστες ήταν οι προβλέψεις που μας έδωσε ο κώδικας μας.

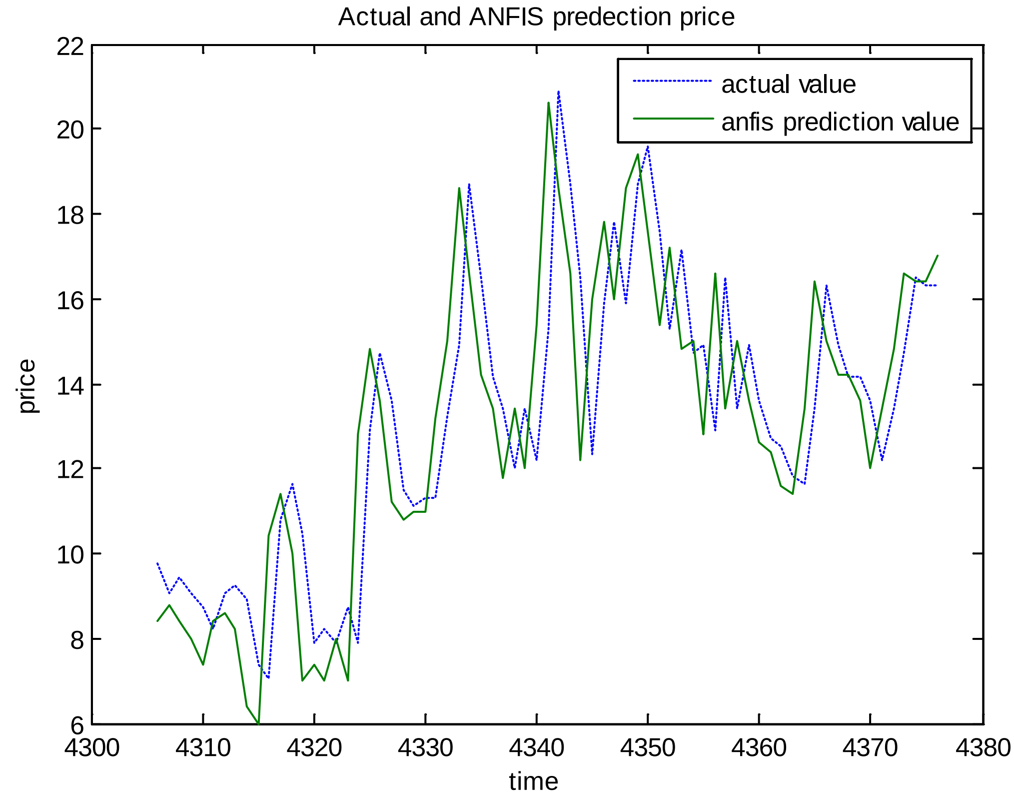
Η διαδικασία έχει ως εξής. Αρχικά συλλέχθηκαν τα δεδομένα και δημιουργήθηκε ένα αρχείο excel στο οποίο περιέχονται αναλυτικά σε κάθε στήλη το έτος , ο μήνας , η ημέρα του μήνα ,η τοπική ώρα και η εξωτερική θερμοκρασία. Έπειτα δημιουργήθηκε σε γλώσσα προγραμματισμού Matlab, με τη χρήση του fuzzy logic toolbox ένα ANFIS το οποίο υπολογίζει τις κατάλληλες παραμέτρους για της συναρτήσεις συμμετοχής (Membership Functions /MF parameters). Το παρακάτω σχήμα περιέχει δύο ξεχωριστές γραφικές παραστάσεις , η πρώτη ονομάζεται «error curves» και η δεύτερη ονομάζεται «step sizes» . Στην πρώτη γραφική παράσταση απεικονίζονται , το σφάλμα εκπαίδευσης και το σφάλμα ελέγχου, με τη μορφή της ρίζας του μέσου τετραγωνικού σφάλματος (RMSE: Root Mean Square Error) συναρτήσει του αριθμού επαναλήψεων (epochs).

Στο δεύτερο γράφημα παρουσιάζεται για λόγους αναφοράς η καμπύλη του μεγέθους βήματος συναρτήσει του αριθμού των επαναλήψεων. Το παραπάνω γίνεται για λόγους υπολογισμού, προσαρμογής του αρχικού μεγέθους βήματος στο βέλτιστο, με απλή παρατήρηση των ρυθμών αύξησης μείωσης από το σχήμα.



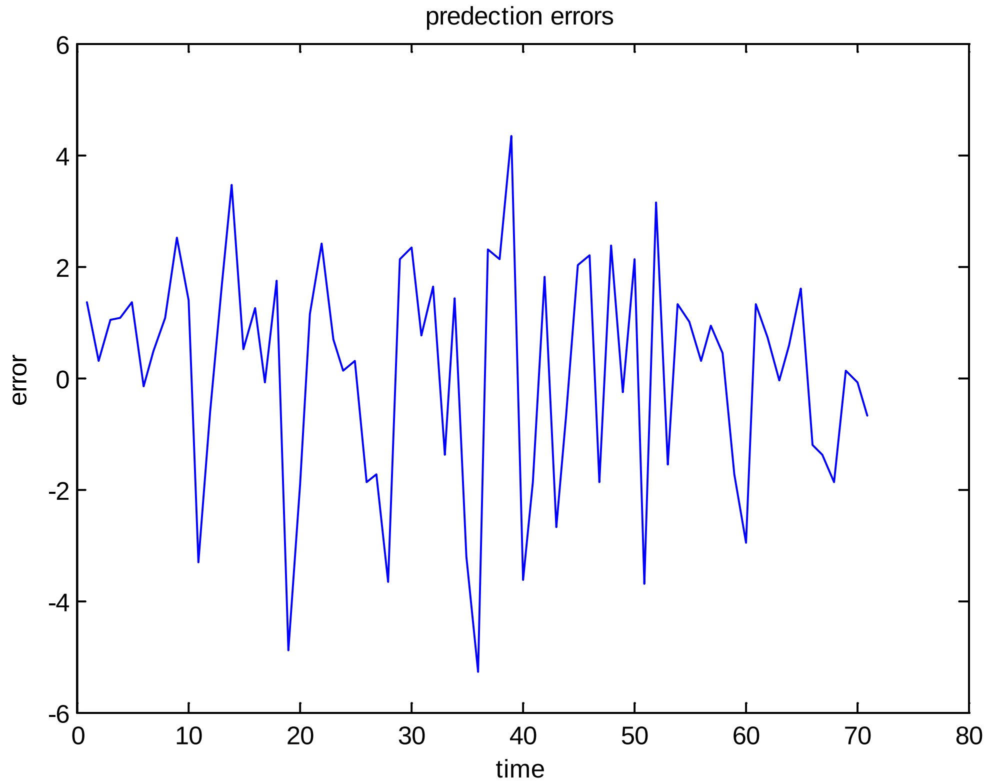
Σχήμα 5.1: Σφάλμα εκπαίδευσης και μέγεθος βήματος

Στη συνέχεια ακολουθεί διάγραμμα σύγκρισης τιμών θερμοκρασίας στην περιοχή της Σούδας που προβλέφθηκαν από το ANFIS με της πραγματικές τιμές από τα δεδομένα .



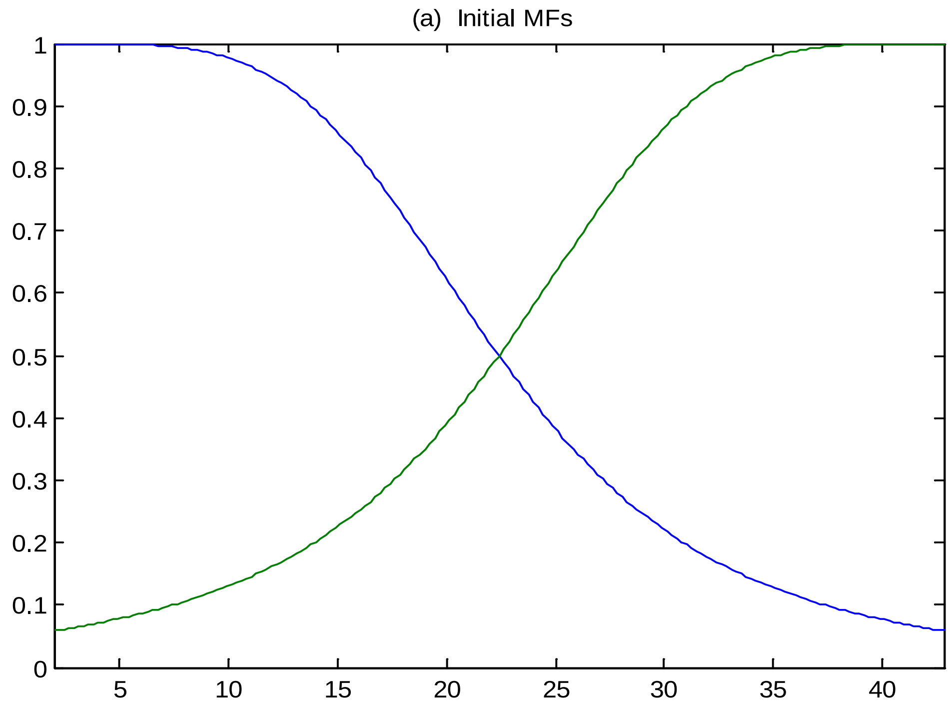
Σχήμα 5.2:Σύγκριση Πραγματικών τιμών με τιμές πρόβλεψης.

Από τη διαφορά των πραγματικών τιμών με της τιμές της πρόβλεψης του μοντέλου , δημιουργείται το σφάλμα πρόβλεψης , αλλά και το παρακάτω διάγραμμα στο οποίο απεικονίζονται τα σφάλματα συναρτήσει των παρατηρήσεων.

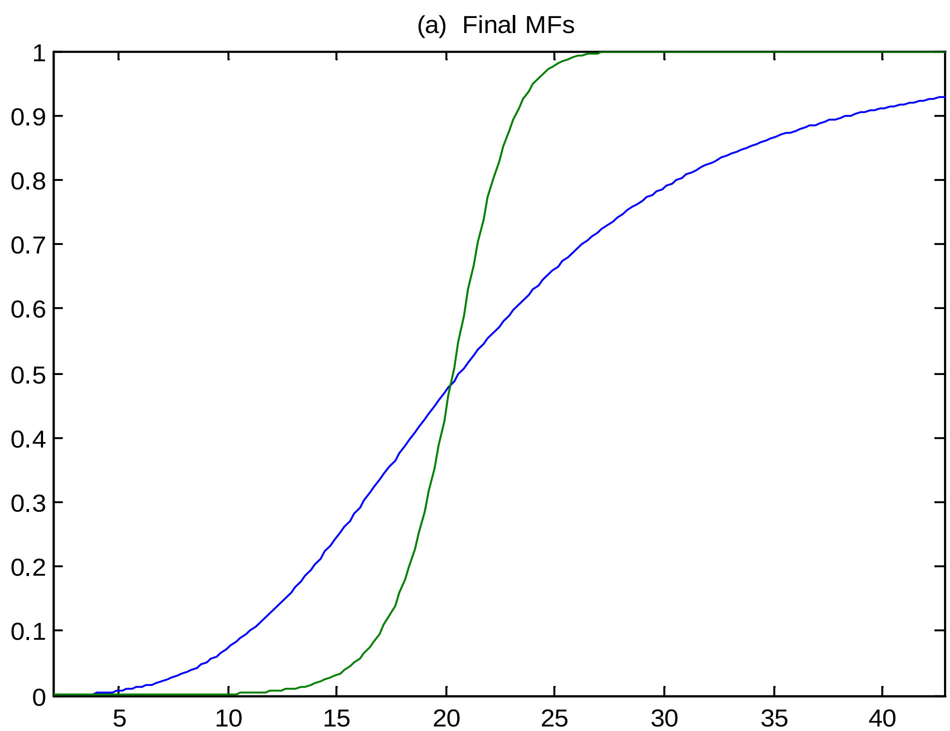


Σχήμα 5.3: Σφάλματα πρόβλεψης.

Αμέσως μετά ακολουθεί το διάγραμμα των αρχικών συναρτήσεων συμμετοχής οι οποίες προέκυψαν από την προσαρμογή των παραμέτρων που χρησιμοποίησε το ANFIS, αλλά και το διάγραμμα με της τελικές συναρτήσεις συμμετοχής , οι οποίες προέκυψαν μετά από την εκπαίδευση του μοντέλου και την τελική προσαρμογή των παραμέτρων των οποίων έκανε χρήση το μοντέλο που δημιουργήθηκε. Τα διαγράμματα έχουν τίτλο «initial MFs» και «final MFs» αντίστοιχα.

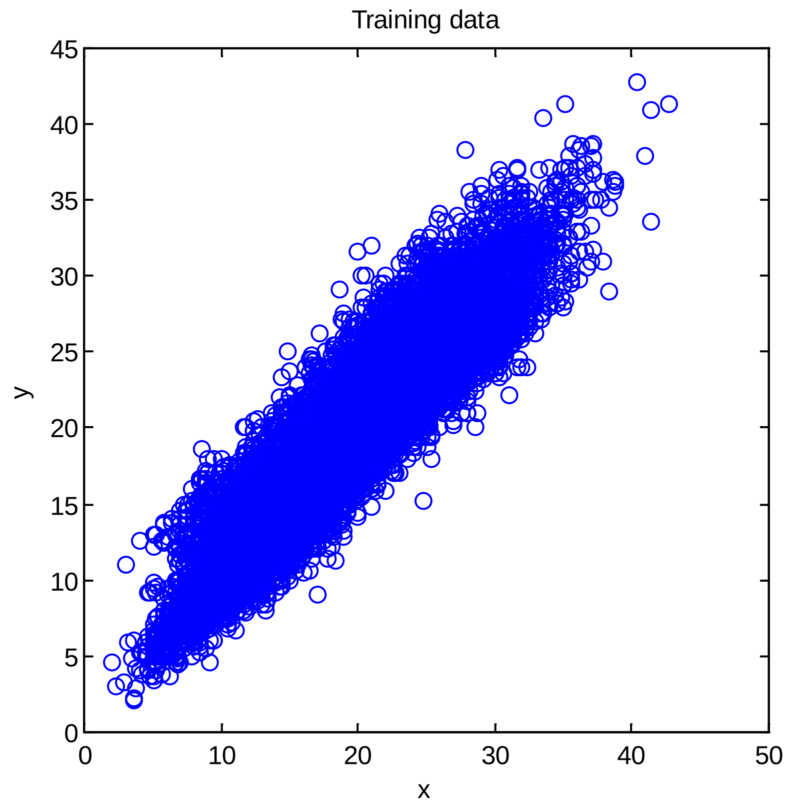


Σχήμα 5.4 : Αρχικές συναρτήσεις συμμετοχής

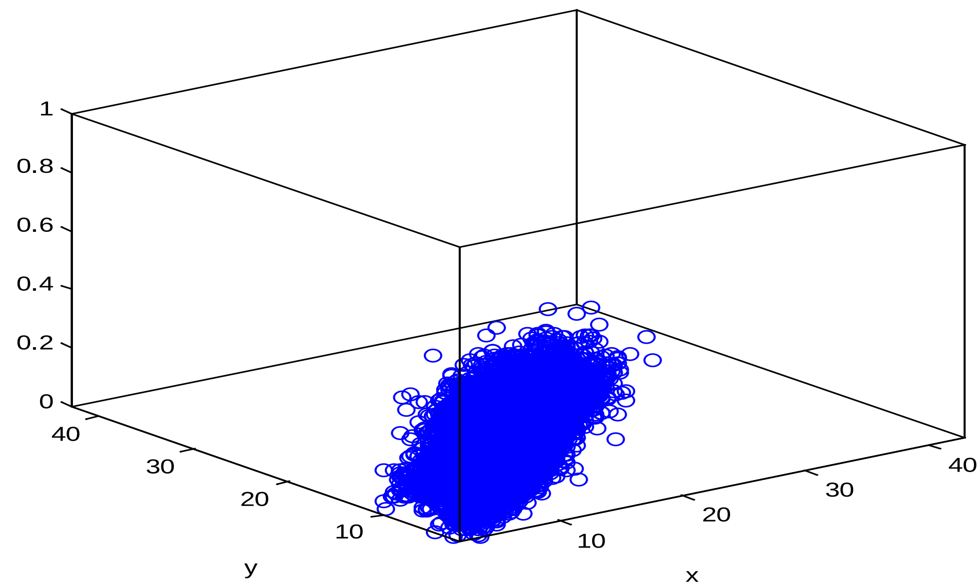


Σχήμα 5.5: Τελικές συναρτήσεις συμμετοχής

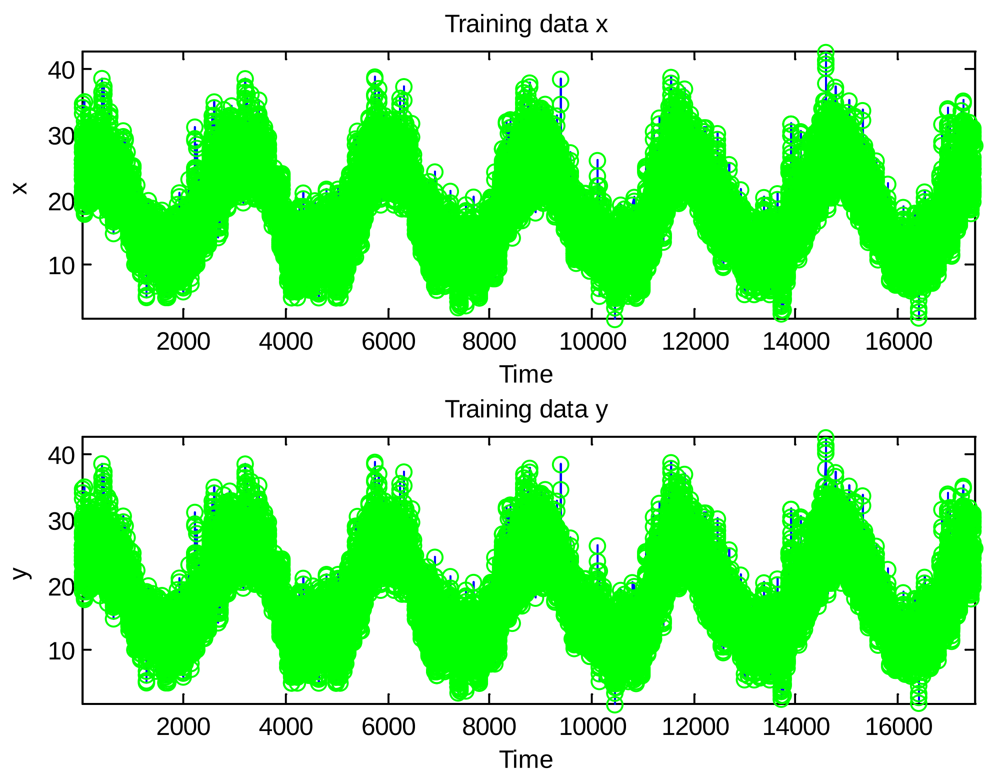
Ακολουθούν τα γραφήματα διασποράς των δεδομένων εκπαίδευσης σε δυσδιάστατη και τρισδιάστατη μορφή , όπως και σε διασπορά ανά άξονα.



Σχήμα 5.6: Δεδομένα εκπαίδευσης ως δυσδιάστατο γράφημα διασποράς.

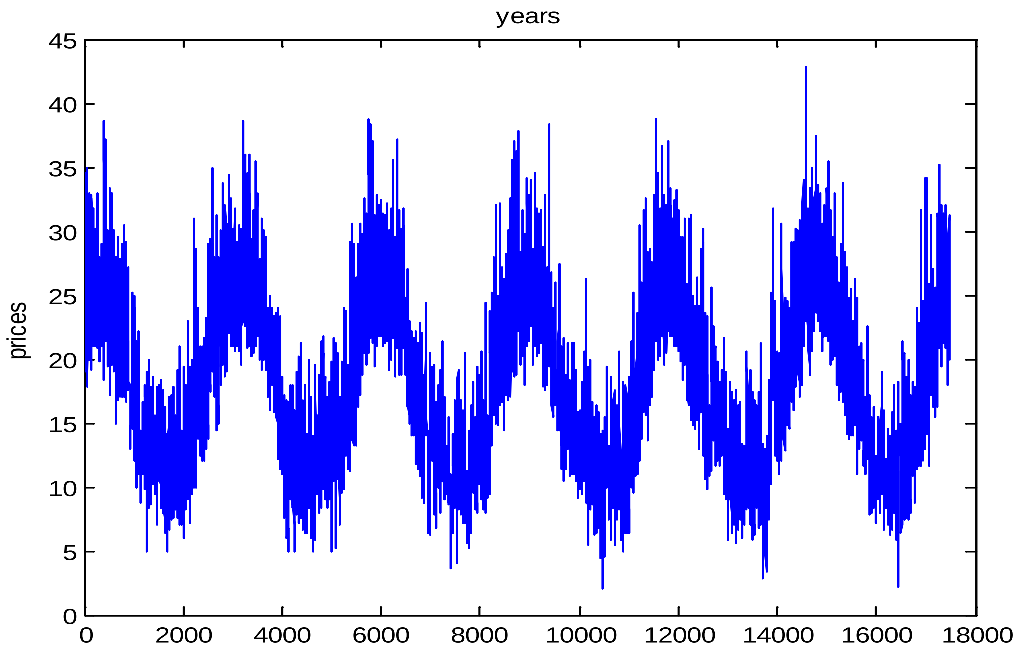


Σχήμα 5.7: Δεδομένα εκπαίδευσης ως τρισδιάστατο γράφημα διασποράς .



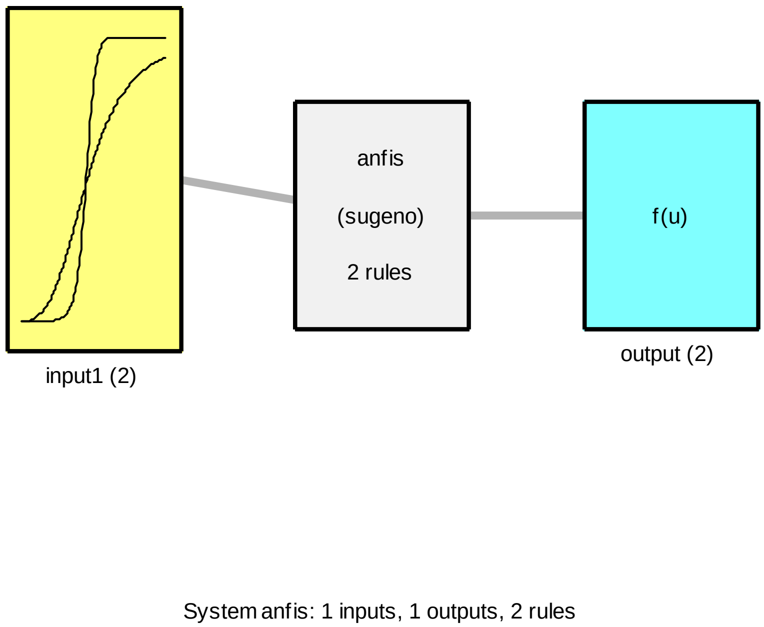
Σχήμα 5.8: Δεδομένα εκπαίδευσης.

Στο επόμενο διάγραμμα φαίνεται η διακύμανση των δεδομένων εκπαίδευσης σαν συνάρτηση της κάθε ημέρας



Σχήμα5.9:Διακύμανση δεδομένων εκπαίδευσης

Στο διάγραμμα που ακολουθεί παρακάτω παρουσιάζεται η δομή του ANFIS το οποίο χρησιμοποιήθηκε για την διεξαγωγή προβλέψεων. Διακρίνονται από τα αριστερά προς τα δεξιά η είσοδος με της συναρτήσεις συμμετοχής , ο μηχανισμός εξαγωγής συμπεράσματος, τύπου Sugeno, με δύο κανόνες και η έξοδος του συστήματος.



Σχήμα 5.10: Δομή του ANFIS.

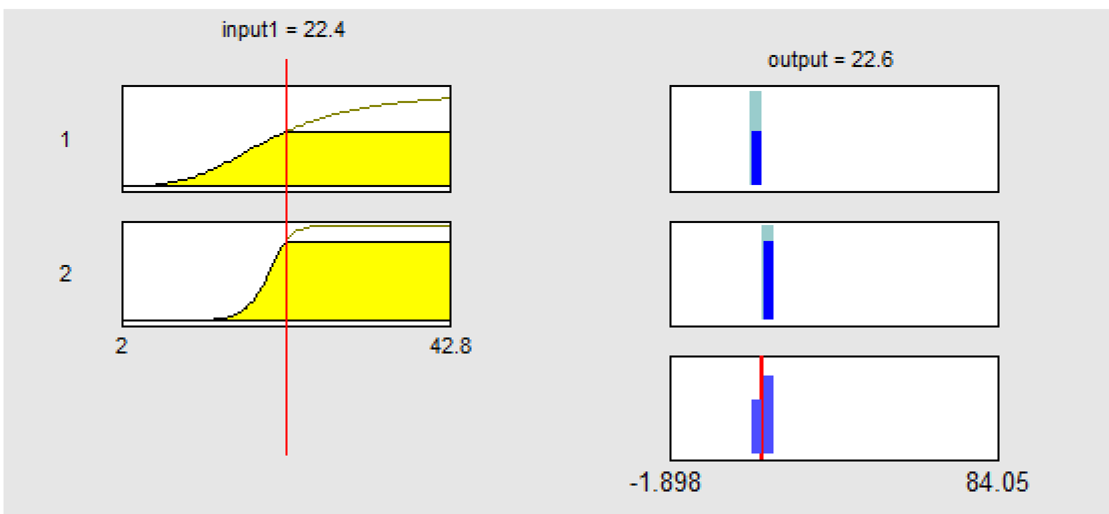
Στο σχήμα που ακολουθεί παρουσιάζονται συνοπτικά τα χαρακτηριστικά του ANFIS του οποίου γίνεται χρήση.

```
Name       = anfis
Type       = sugeno
NumInputs = 1
InLabels  =
      input1
NumOutputs = 1
OutLabels =
      output
NumRules = 2
AndMethod = prod
OrMethod = max
ImpMethod = prod
AggMethod = max
DefuzzMethod = wtaver
```

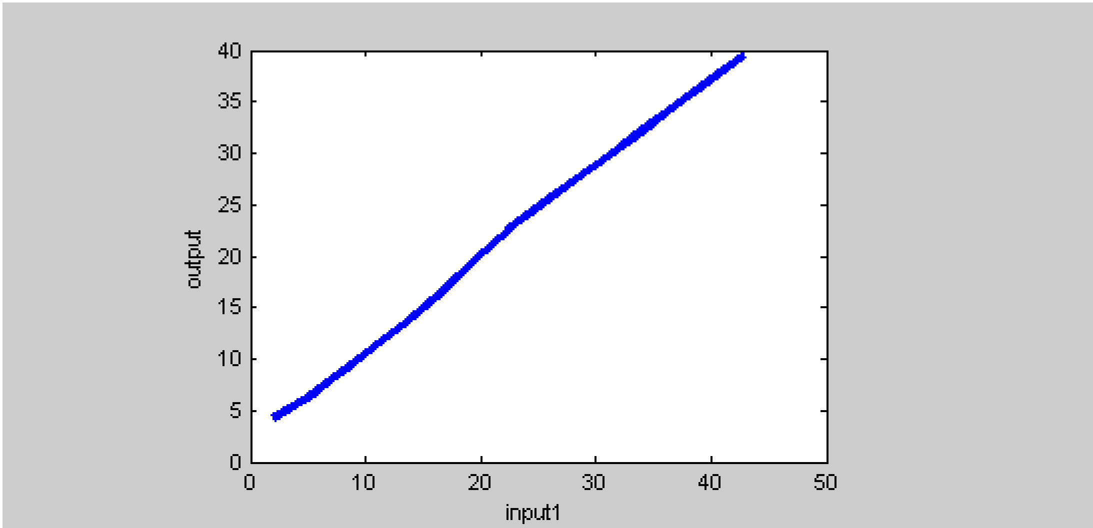
Σχήμα 5.11 :

Χαρακτηριστικά του ANFIS

Ακολουθούν τα διαγράμματα με τίτλο ANFIS Rule Viewer και ANFIS Surface Viewer, τα οποία προκύπτουν από το ANFIS Editor, το οποίο εξάγεται από τη Matlab, με χρήση του Matlab fuzzy toolbox. Τα παρακάτω διαγράμματα παρουσιάζουν τη μορφή των κανόνων κατά τη διαδικασία εκπαίδευσης, δηλαδή στην αρχή και στο τέλος της εκπαίδευσης , όπως και τη μορφή της εξόδου συναρτήσει της εισόδου. Από τα διαγράμματα φαίνεται καθαρά η αρχική και η τελική μορφή των κανόνων του μοντέλου, οι είσοδοι και οι αντίστοιχες έξοδοι, αλλά και η επιφάνεια που καλύπτει το μοντέλο κατά την διαδικασία εκπαίδευσης και η μορφή που περνούν οι αντίστοιχες συναρτήσεις συμμετοχής.



Σχήμα 5.12 : ANFIS Rule Viewer.



Σχήμα 5.13 :ANFIS Editor Surface viewer

5.2: Αξιολόγηση του Προγράμματος Πρόβλεψης

Για την περαιτέρω αξιολόγηση του μοντέλου έγινε χρήση ακόμα δυο μοντέλων χρονικής σειράς , ενός μοντέλου αυτοπαλινδρόμισης (Auto-Regression model, AR model) και ενός μοντέλου αυτοπαλινδρόμισης κινούμενου μέσου.(Auto-Regression Moving Average model, ARMA model), τα οποία κωδικοποιήθηκαν και αυτά σε γλώσσα προγραμματισμού Matlab.Για να γίνει η σύγκριση μεταξύ των τριών μοντέλων πρόβλεψης έγινε ο υπολογισμός των επιμέρους σφαλμάτων πρόβλεψης, δηλαδή υπολογίστηκε το μέσο τετραγωνικό σφάλμα (MSE, Mean Square Error), η ρίζα του μέσου τετραγωνικού σφάλματος, (RMSE Root Mean Square Error), το μέσο απόλυτο σφάλμα (MAE Mean Absolute Error), αλλά και το μέσο απόλυτο ποσοστιαίο σφάλμα (MAPE, Mean Absolute Percentage Error). Ο αριθμητικός υπολογισμός των παραπάνω σφαλμάτων γίνεται με τη χρήση των παρακάτω μαθηματικών τύπων:

- Για το μέσο τετραγωνικό σφάλμα –Mean Square Error-(MSE):

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

- Για τη ρίζα του μέσου τετραγωνικού σφάλματος-Root Mean Square Error-(RMSE):

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

-Για το μέσο απόλυτο σφάλμα –Mean Absolute Error-(MAE):

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

-Για το μέσο απόλυτο ποσοστιαίο σφάλμα –Mean Absolute Percent Error-(MAPE):

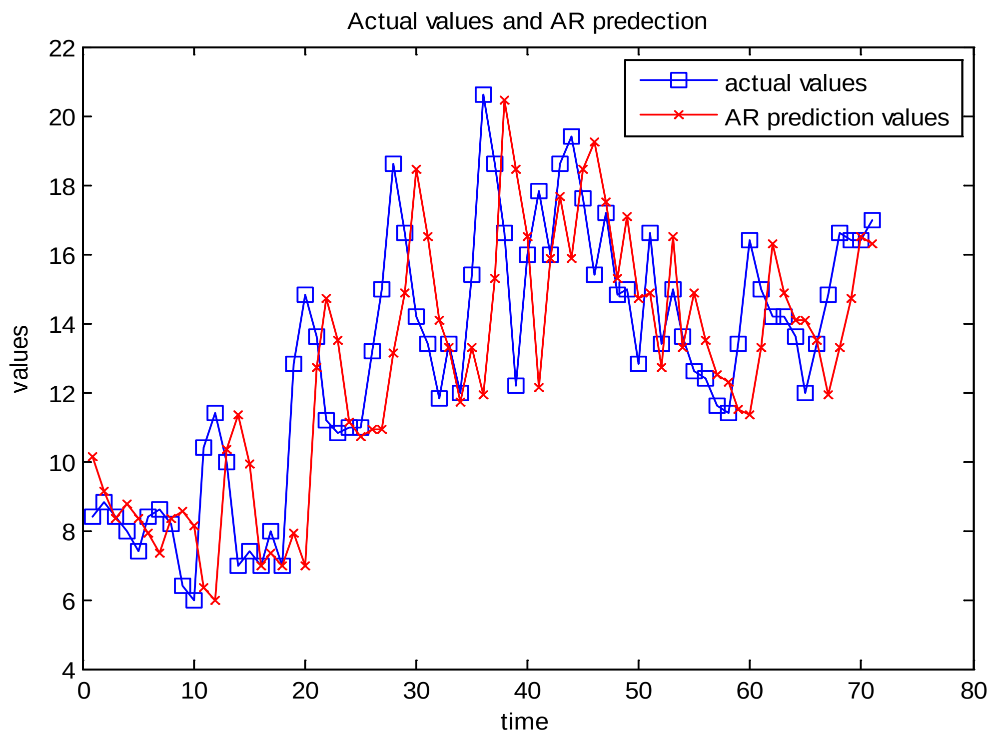
$$MAPE = \frac{100}{N} \cdot \sum_{t=1}^{N} |e_t|$$

$$\cdot\ e_t = \frac{F_t - A_t}{A_t}$$

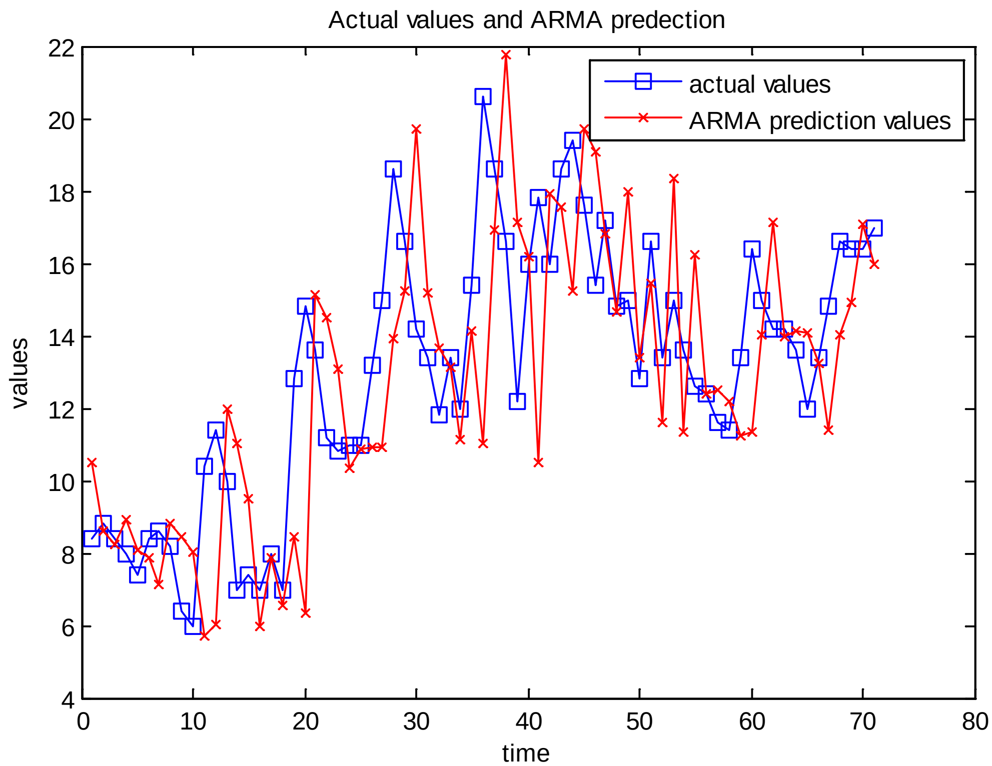
Όπου A_t ,F_t πραγματικές τιμές.

5.3: Σύγκριση Με Άλλες Μεθόδους Πρόβλεψης

Για τη σύγκριση του μοντέλου με της άλλες μεθόδους , αλλά και για περεταίρω πληροφόρηση ακολουθούν τα παρακάτω διαγράμματα τα οποία δείχνουν τη σύγκριση των αποτελεσμάτων των δύο άλλων μεθόδων (AR και ARMA) με της πραγματικές τιμές(οι πραγματικές τιμές εδώ , όπως και στο ANFIS αποτελούνται από το 20% των δεδομένων τα οποία χρησιμοποιήθηκαν για την αξιολόγηση του μοντέλου.).



Σχήμα 5.14 : Σύγκριση πραγματικών τιμών με τιμές πρόβλεψης από το μοντέλο AR



Σχήμα 5.15: Σύγκριση πραγματικών τιμών με τιμές πρόβλεψης από το μοντέλο ARMA

Με τη βοήθεια του System Identification Toolbox της Matlab, μέσω της κωδικοποίησης των αντιστοίχων μαθηματικών τύπων εξάγονται για όλα τα μοντέλα πρόβλεψης (ANFIS ,AR,ARMA) τα αντίστοιχα σφάλματα και έτσι προκύπτει ο ακόλουθος πίνακας με τα επιμέρους σφάλματα για κάθε μοντέλο πρόβλεψης:

	MSE	RMSE	MAE	MAPE
AR	15,0697	3,882	3,0388	15,9548
ARMA	14,3205	3,7842	2,9143	15,3907
ANFIS	5,3919	2,322	1,7687	9,7122

Πίνακας 5.1: Μοντέλα πρόβλεψης και αντίστοιχα σφάλματα.

Από τον παραπάνω πίνακα είναι φανερό ότι το μοντέλο ANFISπου αναπτύχτηκε είναι πιο αξιόπιστα από τα αλλά μοντέλα πρόβλεψης αφού παρουσιάζει μικρότερα επιμέρους σφάλματα. Στην κατάταξη αξιοπιστίας έρχεται δεύτερο το μοντέλο ARMA αφού έχει το μικρότερο σε όλες τις κατηγορίες σφάλμα από τα δυο μοντέλα παλινδρόμησης. Επίσης κρίνεται απαραίτητο να αναφερθεί ότι και οι τρείς μέθοδοι είναι αρκετά αξιόπιστες μιας και παρουσιάζουν αρκετά μικρό σφάλμα.

ΚΕΦΑΛΑΙΟ 6: ΣΥΜΠΕΡΑΣΜΑΤΑ.

Το νευρο-ασαφές μοντέλο το οποίο χρησιμοποιήθηκε για την εξαγωγή προβλέψεων στο πρόβλημα της θερμοκρασίας είναι αρκετά ακριβές και έδωσε ικανοποιητικά αποτελέσματα..Από τη χρήση του μοντέλου (ANFIS) στο συγκεκριμένο πρόβλημα, επαληθεύεται η ικανότητα του στο να παρέχει ακριβείς προβλέψεις σε διάφορους τομείς, και άρα επαληθεύεται το γεγονός του ότι το ANFIS έχει ένα ευρύ πεδίο εφαρμογών

Τα νευρωνικά δίκτυα μπορούν να χρησιμοποιηθούν στο να δώσουν προβλέψεις για αρκετά ζητήματα (θερμοκρασία, τιμές χρηματιστηριακών δεικτών, τιμές ακινήτων, κατανάλωση ισχύος) , με μοναδικές δυσκολίες την διαδικασία εκπαίδευσης τους, η οποία αποτελεί μια χρονοβόρα διαδικασία, αλλά και τη συγκέντρωση ικανοποιητικού δείγματος δεδομένων , για την εκπαίδευση, αλλά και την αξιολόγηση του μοντέλου. Επίσης μία άλλη δυσκολία είναι η δημιουργία της καταλληλότερης δομής ΤΝΔ για το εκάστοτε πρόβλημα (μέγεθος ΤΝΔ , αριθμός κρυφών στρωμάτων, αριθμός κόμβων, αριθμός εισόδων και λοιπά) Από την άλλη όμως το μοντέλο του οποίου έγινε χρήση έδωσε ακριβή αποτελέσματα με μοναδική είσοδο την θερμοκρασία , δεν απαιτούσε τη μέτρηση άλλων παραγόντων όπως υγρασία πυκνότητα νεφών βροχοπτώσεις γεγονός που το καθιστά ιδιαίτερα εύχρηστο.

Από την εφαρμογή του λοιπόν, συμπαιραίνουμε ότι ένα ANFIS , με τρην κατάλληλη εκπαίδευση και το κατάλληλο δείγμα δεδομένων είναι ικανό να αποτελέσει μια μοντέρνα και ακριβής μέθοδο πρόβλεψης.

ΒΙΒΛΙΟΓΡΑΦΙΑ

Aliev, R., Fazlollahi, b., Aliev, R., & Guirimov, B. (2006). Fuzzy time series prediction method based on fuzzy recurrent neural network. *ICONIP* , σσ. 860-869.

Chaudhuri, A. (2009). Time series forecasting using hybrid neyro-fuzzy regession model. *RFSDGrC 2009, LNAI 5908* , σσ. 369-381.

Chen, S.-M., & Hwang, J.-R. (2000, April). Temperature pprediction using fuzzy time series. *IEEE Trasactions on systems,man and cybernetics-PartB:Cybernetics Vol.30 No2* , σσ. 263-275.

Chevalier, R., Hoogenboom, G., McClendon, R. W., & Paz, J. A. (2010, April 22). Support Vector regression witj reduced training sets for air temperature prediction : a comparison with artificial neural networks. *Neural computing &apllications* .

Dejmal, K. (2000). Statistical postprocessing of NWP model outputs apllied to maximum and minimum temperature forecasts. *studia geoph. and geod* , σσ. 73-83.

Dombayaci, O. A., & Golcu, M. (2009). Daily means ambient temperature prediction using artificial neural nertwork method:a case study of Turkey. *Renewable Energy* , σσ. 1158-1161.

Fdez-riverola, F., Corchado, J. M., & Torres, H. M. (2002). An aytomated hybrid CBR system for forecasting. *ECCBR 2002 LNAI 2416* , σσ. 519-533.

Figuerola, P., & Mazzeo, N. (1997). An analytical model for the prediction of nocturnal and dawn surface temperatures under calm, clesr sky conditions. *Agricultural and forest meteorology* , σσ. 229-237.

Huarng, K. (2001). Heuristic models of fuzzy time series for forecasting. *Fuzzy sets and systems* , σσ. 369-386.

Kocak, K., Saylan, L., & Eitzinger, J. (2004). Nonlinear prediction of near-surface temperature via univariate and multyvariate time series embedding. *Ecological modeling* , σσ. 1-7.

Kostoff, R. (1985). Temperature prediction and control in pulsed fusion-hybrid Neutron multipliers. *Journal of fusion energy* .

Li, S.-T., & Cheng, Y.-C. (2007). Determenistic fuzzy timme series for forecasting enrollments. *International Journal computers&mathematics with applications* , σσ. 1904-1920.

Ruano, A., Crispim, E., Conceicao, E., & Lucio, M. (2006). Prediction of building's temperature using neural networks models. *Energy and buildings* , σσ. 682-694.

Smith, B. A., Hoogenboom, G., & McClendon, R. (2009). Artificial neural networksfor aytomated year-round temperature prediction. *Computers and electronics in agriculture* , σσ. 52-61.

Sohn, K.-T., Rha, D.-K., & Seo, Y.-K. (2003). The 3-hour-Interval predictionof ground-level temperature in South Korea using dynamic linear models . *Advances in atmospheric sciences Vol.20 No.4* , σσ. 575-582.

Tao, L., & Viljamen, M. (2009). Prediction of indoor temperature and relative humidity using neural network models: model comparison. *Neural computing & applications* , σσ. 345-357.

Tasadduq, I., Rehman, S., & Bubshait, K. (2002). Application of neural networks for the prediction of hourly mean surface temperatures in Saudi Arabia. *Renewable energy* , σσ. 545-554.

Yang, K., & Wang, J. (2008). A temperature prediction-correction method for estimating surface soil heat flux from soil temperature and moisture data. *Science in China:Earth Sciences* , σσ. 721-729.

Ατσαλάκης, Γ., & Λαμπουσάκης, Ι. (2009). ΠΡΟΒΛΕΨΗ ΕΠΙΤΟΚΙΩΝ ΜΕ ΧΡΗΣΗ ΝΕΥΡΟ-ΑΣΑΦΟΥΣ ΛΟΓΙΚΗΣ.

Ατσαλάκης, Γ., & Μινουδάκη, Χ. (2007). Πρόβλεψη κατανάλωσης νερού στο Ν.Χανίων χρησιμοιώντας το νεύρο-ασαφές σύστημα (ANFIS).

Ατσαλάκης, Γ., & Σταματάκης, Π. (2008, 1 17). ΠΡΟΒΛΕΨΗ ΚΑΤΑΝΑΛΩΣΗΣ ΗΛΕΚΤΡΙΚΟΥ ΦΟΡΤΙΟΥ ΜΕ ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ ΚΑΙ ΓΕΝΕΤΙΚΟΥΣΑΛΓΟΡΙΘΜΟΥΣ.

Ατσαλάκης, Γ., & Τζωρτζάκης, Α. (2007, 12). Πρόβλεψη Τιμών Ακινήτων με τη Χρήση ΝευρωνικώνΔικτύων και Γενετικών Αλγορίθμων.

Κοσματόπουλος, Η., & Αρβανιτάκη, Α.-Δ. (2005). ΜΑΚΡΟΟΙΚΟΝΟΜΙΚΗ ΠΡΟΒΛΕΨΗ ΜΕ ΧΡΗΣΗΓΡΑΜΜΙΚΩΝ ΜΟΝΤΕΛΩΝ ΚΑΙ ΤΕΧΝΗΤΩΝΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ.

ΠΑΡΑΡΤΗΜΑ:ΑΣΑΦΕΙΣ ΠΡΑΞΕΙΣ ΜΕ ΝΟΡΜΕΣ (NORMS)

Ασαφής τομή και ένωση (fuzzy intersection and union) T-norm T-conorm

Η τομή δυο ασαφών συνόλων A και B καθορίζεται από μια συνάρτηση $T:[0,1]\times[0,1]\rightarrow[0,1]$, η οποία προσθέτει δυο βαθμούς συμμετοχής ως εξής:

$$\mu_{A\cap B}(x) = T(\mu_A(x), \mu_B(x)) = \mu_A(x) \tilde{*} \mu_B(x),$$

Όπου το σύμβολο $\tilde{*}$ είναι ο δυαδικός τελεστής της συνάρτησης T. Αυτή η κλάση τελεστή ασαφούς τομής ,αναφέρεται συχνά ως T-norm τελεστής(τριγωνική νόρμα) έχει τους εξής περιορισμούς:

Ορισμός 1: T-norm

Μια T-norm είναι μια συνάρτηση δυο θέσεων (two place) $T(\cdot,\cdot)$ που ικανοποιεί :

$$\begin{array}{ll}
T(0,0)=0,\ T(a,1)=T(1,a)=a & \text{(boundary)}\\
T(a,b)\leq T(c,d) \text{ if } a\leq c \text{ and } b\leq d & \text{(monotonicity)}\\
T(a,b)=T(b,a) & \text{(commutativity)}\\
T(a,T(b,c))=T(T(a,b),c) & \text{(associativity)}.
\end{array}$$

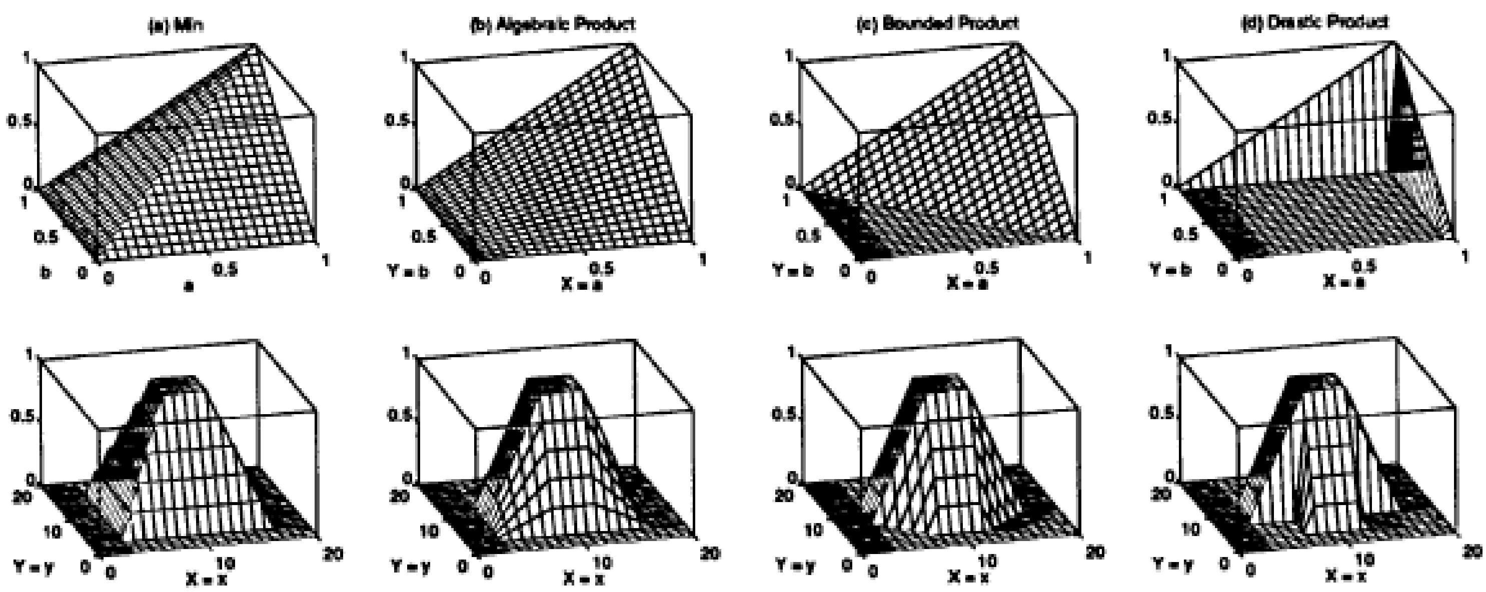
Παράδειγμα τεσσάρων T-norm

Οι τέσσερις πιο συχνά χρησιμοποιούμενες T-norm είναι :

$$\begin{array}{ll}
\textbf{Minimum:} & T_{min}(a,b)=\min(a,b)=a\wedge b.\\
\textbf{Algebraic product:} & T_{ap}(a,b)=ab.\\
\textbf{Bounded product:} & T_{bp}(a,b)=0\vee(a+b-1).\\
\textbf{Drastic product:} & T_{dp}(a,b)=\begin{cases} a, & \text{if } b=1.\\ b, & \text{if } a=1.\\ 0, & \text{if } a,b<1.\end{cases}
\end{array}$$

Από το σχήμα που ακολουθεί φαίνεται ότι :

$$T_{dp}(a,b) \leq T_{bp}(a,b) \leq T_{ap}(a,b) \leq T_{min}(a,b)$$



Σχήμα 1 : (πρώτη γραμμή) τέσσερεις T-norm $T_{min}(a,b)$, $T_{ap}(a,b)$, $T_{bp}(a,b)$, and $T_{dp}(a,b)$;
(δεύτερη γραμμή) οι αντίστοιχες επιφάνειες για α= $trapezoid(x,3,8,12,17)$ και $b = trapezoid(y,3,8,12,17)$.

Ορισμός 2: T-conorm (S-norm)

Η T-conorm (S-norm) είναι μια συνάρτηση δυο θέσεων $S(\cdot,\cdot)$ που ικανοποιεί :

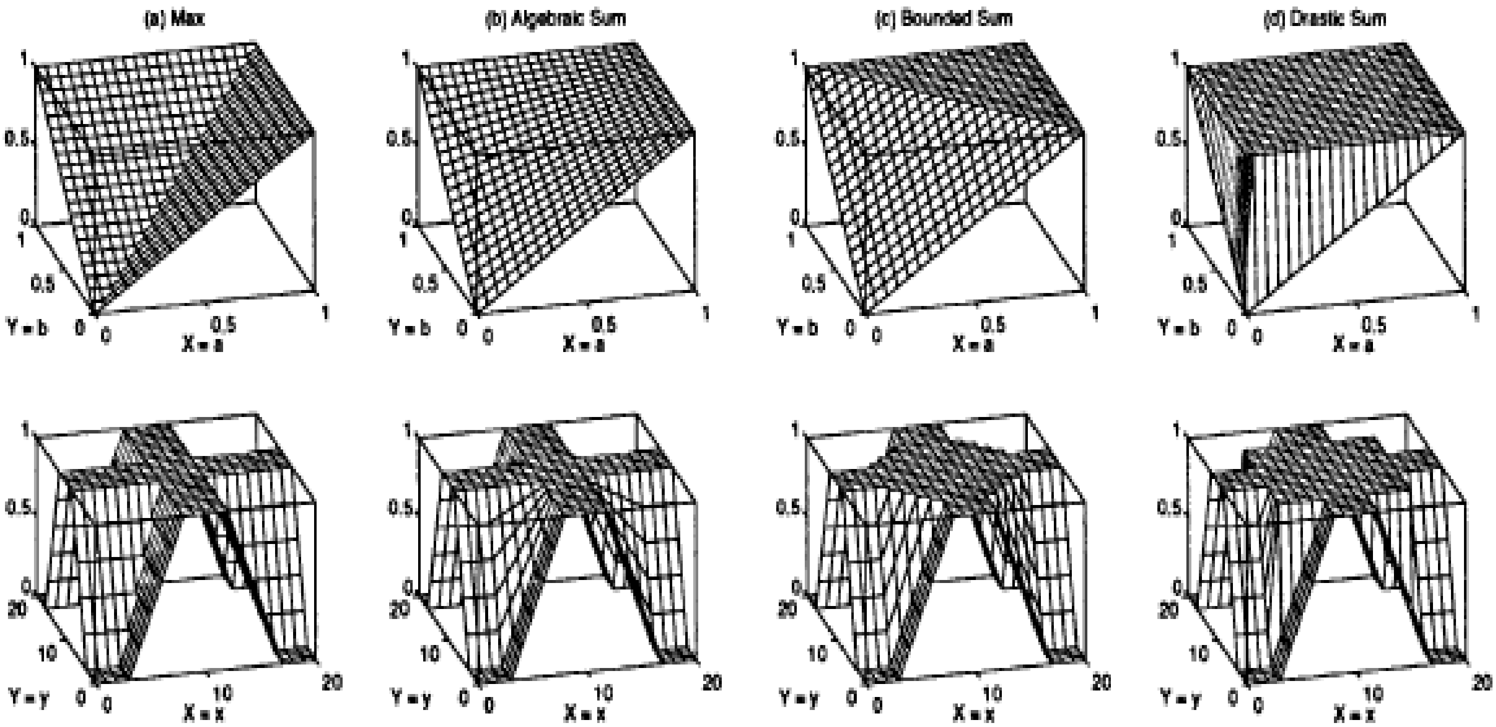
$$\begin{array}{ll} S(1,1)=1,\ S(0,a)=S(a,0)=a & \text{(boundary)} \\ S(a,b)\le S(c,d) \text{ if } a\le c \text{ and } b\le d & \text{(monotonicity)} \\ S(a,b)=S(b,a) & \text{(commutativity)} \\ S(a,S(b,c))=S(S(a,b),c) & \text{(associativity).} \end{array}$$

Παράδειγμα τεσσάρων T-conorm.
Αντίστοιχα με τις T-norm του προηγούμενου παραδείγματος έχουμε τέσσερεις T-conorm :

$$\begin{array}{ll} \textbf{Maximum:} & S(a,b)=\max(a,b)=a\vee b. \\ \textbf{Algebraic sum:} & S(a,b)=a+b-ab. \\ \textbf{Bounded sum:} & S(a,b)=1\wedge(a+b). \\ \textbf{Drastic sum:} & S(a,b)=\begin{cases} a, & \text{if } b=0. \\ b, & \text{if } a=0. \\ 1, & \text{if } a,b>0. \end{cases} \end{array}$$

Από το σχήμα που ακολουθεί φαίνεται ότι

$$S_{max}(a,b)\le S_{ap}(a,b)\le S_{bp}(a,b)\le S_{dp}(a,b).$$



Σχήμα2:(πρώτη γραμμή) τέσσερεις T-conorm
$S_{min}(a,b),\quad S_{ap}(a,b)\quad S_{bp}(a,b)\quad and\quad S_{dp}(a,b);$
(δεύτερη γραμμή) οι αντίστοιχες επιφάνειες για
α= $trapezoid(x,3,8,12,17)$ και
$b \;=\; trapezoid(y,3,8,12,17).$