
Τ.Ε.Ι ΠΑΤΡΩΝ
ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ ΚΑΙ ΟΙΚΟΝΟΜΙΑΣ
ΤΜΗΜΑ ΕΠΙΧΕΙΡΗΜΑΤΙΚΟΥ ΣΧΕΔΙΑΣΜΟΥ ΚΑΙ
ΠΛΗΡΟΦΟΡΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ



ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ:

«ΤΕΧΝΗΤΑ ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ ΚΑΙ ΠΡΟΒΛΕΨΗ
ΟΙΚΟΝΟΜΙΚΩΝ ΧΡΟΝΟΣΕΙΡΩΝ»

Ταμπάκης Παναγιώτης (Α.Μ. 45)

Επιβλέπων Καθηγητής: Ψαράκης Εμμανουήλ

ΠΑΤΡΑ

Ιούνιος 2004

ΑΡΙΘΜΟΣ ΕΙΣΑΓΩΓΗΣ	5775
----------------------	------

Τ.Ε.Ι ΠΑΤΡΩΝ
ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ ΚΑΙ ΟΙΚΟΝΟΜΙΑΣ
ΤΜΗΜΑ ΕΠΙΧΕΙΡΗΜΑΤΙΚΟΥ ΣΧΕΔΙΑΣΜΟΥ ΚΑΙ
ΠΛΗΡΟΦΟΡΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ:

*«ΤΕΧΝΗΤΑ ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ ΚΑΙ ΠΡΟΒΛΕΨΗ
ΟΙΚΟΝΟΜΙΚΩΝ ΧΡΟΝΟΣΕΙΡΩΝ»*

Ταμπάκης Παναγιώτης (Α.Μ. 45)

Επιβλέπων Καθηγητής: Ψαράκης Εμμανουήλ

ΠΑΤΡΑ

Ιούνιος 2004

Περιεχόμενα

Περιεχόμενα.....	i
Πρόλογος.....	I
Κεφάλαιο 1.....	1
Τεχνητή Νοημοσύνη.....	1
1.1 Εισαγωγή.....	1
1.2 Πεδία εφαρμογής της Τεχνητής Νοημοσύνης.....	2
1.3 Έμπειρα Συστήματα.....	4
1.3.1 Δομή Έμπειρων Συστημάτων.....	6
1.4 Γενετικοί Αλγόριθμοι.....	6
Κεφάλαιο 2.....	8
Τεχνητά Νευρωνικά Δίκτυα.....	8
2.1 Εισαγωγή.....	8
2.2 Γενικά.....	9
2.2.1 Μονάδες Επεξεργασίας (Processing Units).....	10
2.2.2 Συνάψεις Μονάδων.....	10
2.2.3 Κανόνες Ενεργοποίησης και Εξόδου (Activation and Output Rules).....	11
2.3 Τοπολογίες Δικτύων (Network Topologies).....	12
2.4 Εκπαίδευση Τ.Ν.Δ.....	12
2.5 Τρόποι Μάθησης ενός Τ.Ν.Δ.....	12
2.6 Ρύθμιση των βαρών των συνάψεων.....	13
2.7 Perceptron και Adaline.....	13
2.7.1 Perceptron.....	14
2.7.1.1 Ο Κανόνας Μάθησης που χρησιμοποιείται στο Perceptron και το Θεώρημα Σύγκλισης.....	16
2.7.1.2 Το Adaline (Adaptive Linear Element).....	18
2.7.2 Δίκτυα με Γραμμικές Συναρτήσεις Ενεργοποίησης: Κανόνας Δέλτα.....	20
2.8 Ανάδρομη Διάδοση σφάλματος.....	21
2.9 Πολυεπίπεδα “προς τα εμπρός” Δίκτυα.....	21
2.10 Πολυεπίπεδο Perceptron.....	22
2.11 Πολυπλοκότητα Δικτύου (Network Complexity).....	23
2.12 Θεώρημα Καθολικής Προσέγγισης.....	25
2.13 Ο Γενικευμένος Κανόνας Δέλτα.....	26
2.14 Κατανόηση της Ανάδρομης Διάδοσης Σφάλματος.....	28
2.15 Δουλεύοντας με τον Back-Propagation.....	29
2.16 Προσαρμογή των βαρών με σιγμοειδή συνάρτηση ενεργοποίησης.....	29
2.17 Ρυθμός Μάθησης και Συντελεστής Ορμής.....	30
2.17.1 Ρυθμός Μάθησης (Learning Rate).....	31
2.17.2 Συντελεστής Ορμής.....	31
2.18 Παράγοντες που επηρεάζουν την απόδοση ενός Πολυεπίπεδου “προς τα εμπρός” Δικτύου.....	32
Κεφάλαιο 3.....	33
Αλγόριθμοι Διερεύνησης της Επιφάνειας Σφάλματος.....	33
3.1 Εισαγωγή.....	33
3.2 Προσαρμοζόμενος Γραμμικός Συνδυαστής.....	33
3.2.1 Σήμα Εισόδου και Διάνυσμα Βαρών.....	34

3.2.1.1 Στασιμότητα.....	35
3.2.1.2 Μερικός Χαρακτηρισμός μιας Στοχαστικής Διαδικασίας Διακριτού Χρόνου .	36
3.2.1.3 Εργοδικότητα	37
3.2.2 Επιθυμητή Απόκριση και Ακολουθία Σφαλμάτων	39
3.2.3 Ανάδελτα και Ελάχιστο MSE	41
3.2.4 Αυτοσυσχέτιση του Σφάλματος και της Εισόδου.....	41
3.3 Ιδιότητες της Τετραγωνικής Επιφάνειας Σφάλματος	42
3.3.1 Κανονική Μορφή του Πίνακα Αυτοσυσχέτισης.....	43
3.3.2 Ιδιοτιμές και Ιδιοδιανύσματα του Πίνακα Συσχέτισης Εισοδου	43
3.3.3 Γεωμετρική Σημασία των Ιδιοτιμών και Ιδιοδιανυσμάτων	44
3.4 Ερευνώντας την Επιφάνεια Σφάλματος.....	46
3.4.1 Μέθοδοι Εξερεύνησης της Επιφάνειας Σφάλματος.....	47
3.4.2 Βασικές Αρχές της Μεθόδου Αναζήτησης Βασισμένης στο Ανάδελτα	47
3.4.3 Ένας Απλός Αλγόριθμος Αναζήτησης Βασισμένος στο Ανάδελτα.....	49
3.4.4 Σταθερότητα και Ρυθμός Σύγκλισης.....	49
3.4.5 Η Καμπύλη Μάθησης	51
3.4.6 Η Μέθοδος του Newton.....	51
3.4.7 Η Μέθοδος του Newton στην Πολυδιάστατη Περίπτωση.....	53
3.4.8 Αναζήτηση με τη Μέθοδο Απότομης Κατάβασης.....	54
3.4.9 Σύγκριση των Καμπυλών Μάθησης	56
3.5 Υπολογισμός του Ανάδελτα και η Επίδρασή του στη Διαδικασία Προσαρμογής	58
3.5.1 Εκτίμηση του Ανάδελτα με Μέτρηση της Παραγώγου στην Περίπτωση ενός Βάρους	58
3.5.2 Performance Penalty	60
3.5.3 Μέτρηση της Παραγώγου και Διαταραχή στην Περίπτωση πολλών Βαρών	61
3.5.4 Διακύμανση του Εκτιμώμενου Ανάδελτα.....	62
3.5.5 Επιδράσεις στο Διάνυσμα Βαρών κατά τη Διαδικασία Προσαρμογής	66
3.5.6 Πλεονάζον Μέσο Τετραγωνικό Σφάλμα και η Επίδραση του στη Διαδικασία Προσαρμογής.....	69
3.5.7 Misadjustment.....	76
3.5.8 Συγκριτική Απόδοση των Μεθόδων του Newton και Απότομης Κατάβασης.....	78
3.5.9 Ολικό Misadjustment.....	79
3.6 Ο Αλγόριθμος LMS	80
3.6.1 Εξαγωγή του Αλγόριθμου LMS.....	81
3.6.2 Σύγκλιση του Διανύσματος Βαρών (Convergence of the Weight Vector).....	82
3.6.3 Καμπύλη Μάθησης	83
3.6.4 Θόρυβος στην Επίλυση του Διανύσματος Βαρών (Noise in the Weight Vector Solution).....	84
3.6.5 Misadjustment.....	85
3.6.6 Απόδοση.....	87
3.7 Διερεύνηση της επιφάνειας με Χρήση Γενετικών Αλγορίθμων	88
3.7.1 Χρωμόσωμα - Κωδικοποίηση Παραμέτρων.....	90
3.7.1.1 Αρχικός Πληθυσμός.....	92
3.7.2 Φυσική Επίλογή	93
3.7.3 Επίλογή Γονέων	94
3.7.4 Δημιουργία Απογόνων - Διασταυρώσεις	96
3.7.5 Μετάλλαξη.....	97
3.7.6 Θεωρία «Σχημάτων»	98

3.7.6.1 Επίδραση των «Σχημάτων» στην Επιλογή	99
3.7.6.2 Επίδραση των «Σχημάτων» στη Διασταύρωση	100
3.7.6.3 Επίδραση των «Σχημάτων» στην Μετάλλαξη	100
3.7.7 Εξίσωση Ανάπτυξης των «Σχημάτων»	101
3.7.8 Υλοποίηση Γενετικών Αλγορίθμων	101
3.7.8.1 Περιγραφή του Γενετικού Αλγόριθμου	101
3.7.8.2 Επιλογή Γονέων	102
3.7.8.3 Διασταύρωση Γονέων	103
3.7.8.4 Μετάλλαξη	104
3.7.8.5 Μια Παραλλαγή του Γενετικού αλγόριθμου: Καιάδας.....	104
3.7.9 Αλγόριθμος	105
Κεφάλαιο 4.....	108
Μεθοδολογία Πρόβλεψης Χρονοσειρών με Χρήση ΤΝΔ	108
4.1 Εισαγωγή.....	108
4.2 Παρουσίαση Μεθόδων Πρόβλεψης	109
4.3 Μεθοδολογία Ανάπτυξης ΤΝΔ για Πρόβλεψη Χρονοσειρών	111
4.3.1 Ανάλυση Προβλήματος	111
4.3.2 Συλλογή Δεδομένων.....	112
4.3.3 Ανάλυση δεδομένων και επιλογή μεταβλητών εισοδου και εξοδου.....	112
4.3.3.1 Στατιστικές Μέθοδοι.....	113
4.3.3.2 Ανάλυση Ευαισθησίας - Εξομοίωση Εισοδου με Μέση Τιμή	113
4.3.4 Προ-επεξεργασία Δεδομένων	114
4.3.4.1 Αφαίρεση της Κλίσης	114
4.3.4.2 Κλιμάκωση (scaling) – Κανονικοποίηση (normalization).....	114
4.3.4.3 Διαχείριση Ακραίων Σημείων (Statistical outliers)	115
4.3.4.4 Λοιποί Μετασχηματισμοί Δεδομένων	116
4.3.4.5 Επιλογή Μοντέλου ΤΝΔ.....	117
4.3.5 Σχεδίαση – Διαμόρφωση Συνόλων Δεδομένων.....	117
4.3.6 Ανάπτυξη Μοντέλου και Βελτιστοποίηση.....	119
4.3.7 Συστηματική Βελτιστοποίηση Μοντέλου.....	120
4.3.8 Αξιολόγηση Μοντέλου	121
Βιβλιογραφία.....	123

Πρόλογος

Το αντικείμενο με το οποίο πραγματεύεται η παρακάτω εργασία είναι τα Τεχνητά Νευρωνικά Δίκτυα και η εφαρμογή τους σε προβλήματα πρόβλεψης Χρονοσειρών. Τα Τ.Ν.Δ ανήκουν στον τομέα της Τεχνητής Νοημοσύνης και γι αυτό ακριβώς τον λόγο η εργασία ξεκινά με κάποια εισαγωγικά στοιχεία της Τεχνητής Νοημοσύνης, τις κυριότερες εφαρμογές καθώς και τους βασικούς κλάδους της Τεχνητής Νοημοσύνης. Οι κλάδοι αυτοί είναι τα Έμπειρα Συστήματα, η Ασαφής Λογική και οι Γενετικοί Αλγόριθμοι. Στη συνέχεια στο Κεφάλαιο 3 παρουσιάζονται κάποια γενικά σημεία των Τ.Ν.Δ όπως τα δομικά του στοιχεία, αρχιτεκτονικές υλοποίησης, τρόποι εκπαίδευσης ενός Τ.Ν.Δ και κάποια βοηθητικά μέτρα όπως ο Ρυθμός Μάθησης και ο Συντελεστής Ορμής. Στο Κεφάλαιο 4 παρουσιάζεται ο Προσαρμοζόμενος Γραμμικός Συνδιαστής (adaptive linear combiner) ο οποίος θα βοηθήσει στην κατανόηση της λειτουργίας ενός Τ.Ν.Δ, στον τρόπο με τον οποίο δέχεται μια είσοδο, στον τρόπο με τον οποίο επεξεργάζεται την είσοδο και τον τρόπο με τον οποίο εκπαιδεύεται. Πιο συγκεκριμένα, αναλύονται οι ιδιότητες της συνάρτησης κόστους και της επιφάνειας σφάλματος που δημιουργείται, διάφορες μέθοδοι εξερεύνησης της επιφάνειας αυτής και εύρεσης της βέλτιστης τιμής. Τέλος στο Κεφάλαιο 5 η μεθοδολογία που θα πρέπει να ακολουθηθεί για την ανάπτυξη ενός Τ.Ν.Δ το οποίο θα έχει εφαρμογή στην πρόβλεψη Χρονοσειρών και πιο συγκεκριμένα στην πρόβλεψη οικονομικών χρονοσειρών

Κεφαλαίο 1

Τεχνητή Νοημοσύνη

1.1 Εισαγωγή

Η στροφή της έρευνας τα τελευταία χρόνια στην ανάπτυξη αλγορίθμων οι οποίοι είναι ικανοί να λύνουν σύνθετα φυσικά προβλήματα, να καταλαβαίνουν τη φυσική γλώσσα, να αντιλαμβάνονται τις οπτικές εικόνες, να διδάσκονται από την εμπειρία και γενικά να επιδεικνύουν μια νοήμονα συμπεριφορά, δημιούργησε ένα νέο τομέα στην επιστήμη της πληροφορικής. Ο τομέας αυτός ονομάζεται **Τεχνητή Νοημοσύνη** (Artificial Intelligence).

Ένα σημαντικό μέρος της προσπάθειας που γίνεται στον τομέα αυτό, στρέφεται στην ανάλυση και μοντελοποίηση των εκδηλώσεων της νοήμονος συμπεριφοράς του ανθρώπου. Αντικείμενο επομένως της Τεχνητής Νοημοσύνης είναι η προσπάθεια εμφύτευσης των ικανοτήτων της αντίληψης, του συλλογισμού και της κατανόησης στους ηλεκτρονικούς υπολογιστές ή πιο απλά είναι η μελέτη του πώς θα καταστήσουμε τους υπολογιστές ικανούς να κάνουν πράγματα στα οποία οι άνθρωποι προς το παρόν τα καταφέρνουν καλύτερα.

Τα προβλήματα που χειρίζεται η Τεχνητή Νοημοσύνη είναι προβλήματα των οποίων η επίλυση είτε δεν έχει επιχειρηθεί με τις κλασσικές μεθόδους προγραμματισμού, είτε έχει αποδειχθεί εξαιρετικά δύσκολη ή αδύνατη, εξαιτίας της πολυπλοκότητας και των ιδιομορφιών που παρουσιάζουν. Ένα σύνθημα γνώρισμα αυτών των προβλημάτων αποτελεί η “συνδυαστική έκρηξη” (φαινόμενο κατά το οποίο ξεκινώντας από μια αρχική κατάσταση ενός προβλήματος και οδεύοντας προς την τελική κατάσταση, αναφέεται ένα εκρηκτικά αυξανόμενο πλήθος δυνατών ενδιάμεσων καταστάσεων), η οποία καθιστά αδύνατη την επίλυσή τους μέσω της εξαντλητικής έρευνας όλων των δυνατών λύσεων, τουλάχιστον με τους υπολογιστές που διαθέτει σήμερα ο άνθρωπος.

Η ανθρώπινη νοημοσύνη χαρακτηρίζεται στη γενική της μορφή από δημιουργικότητα, φαντασία, κίνητρα, αναλογική σκέψη, κοινή λογική, παράλληλη

επεξεργασία διαφορετικών στοιχείων της γνώσης, κρίση, αυτοσχεδιασμό και εμπειρία. Όλα αυτά επιτρέπουν στον άνθρωπο να καταλήγει στην επιθυμητή λύση, χωρίς να ερευνά εξαντλητικά όλο το πεδίο της γνώσης του. Αντίθετα, επικαλείται την εμπειρία και την κρίση του για να περιορίσει την έρευνά του σε λίγες και συγκεκριμένες κατευθύνσεις, ανάμεσα στις οποίες τελικά θα βρει τη λύση. Αυτό εξηγεί και το γεγονός ότι η Τεχνητή Νοημοσύνη τείνει να αντιμετωπίζει κάθε πρόβλημα σαν πρόβλημα έρευνας σε ένα πεδίο δυνατών καταστάσεων. Η έρευνα αυτή διεξάγεται με τη βοήθεια ευρετικών μεθόδων, οι οποίες έχουν σαν σκοπό την καθοδήγησή της προς κατευθύνσεις, που εκτιμάται ότι θα οδηγήσουν ταχύτερα σε μια ικανοποιητική λύση.

1.2 Πεδία εφαρμογής της Τεχνητής Νοημοσύνης

Στη συνέχεια αναφέρονται μερικά από τα πεδία εφαρμογής της Τεχνητής Νοημοσύνης.

- **Ηλεκτρονικά Παιχνίδια (game playing)**

Μπορείτε να αγοράσετε υπολογιστές που μπορούν να παίξουν σκάκι υψηλού επιπέδου για μερικές εκατοντάδες Ευρώ .Δεν υπάρχει κάποια τεχνητή νοημοσύνη σε αυτό, αλλά παίζουν καλά ενάντια στους ανθρώπους κυρίως μέσω του υπολογισμού όλων των πιθανών καταστάσεων (Brute force computation), εξετάζοντας εκατοντάδες χιλιάδες διαφορετικών θέσεων.

- **Αναγνώριση Φωνής (speech recognition)**

Στη δεκαετία του '90, η αναγνώριση φωνής έφθασε σε ένα πρακτικό επίπεδο για περιορισμένους σκοπούς. Παρόλα αυτά ενώ είναι δυνατό να καθοδηγηθούν οι υπολογιστές με την χρήση της ομιλίας, οι περισσότεροι χρήστες βρίσκουν το πληκτρολόγιο και το ποντίκι πιο εύχρηστα.

- **Κατανόηση της φυσικής γλώσσας (understanding natural language)**

Το να δώσουμε μια ακολουθία λέξεων σε έναν υπολογιστή δεν είναι αρκετό. Ακόμα και η ανάλυση των προτάσεων δεν είναι αρκετή . Στον υπολογιστή θα πρέπει να παρᾶσχεθεί η δυνατότητα της κατανόησης ολόκληρου του κειμένου. Αυτό είναι προς το παρόν δυνατό μόνο σε πολύ περιορισμένες περιπτώσεις (domains).

- **Μηχανική όραση (machine vision)**

Ο κόσμος αποτελείται από τρισδιάστατα αντικείμενα, αλλά οι εισαγωγές στις φωτογραφικές μηχανές και στις βιντεοκάμερες των υπολογιστών είναι δύο διαστάσεων. Μερικά χρήσιμα προγράμματα μπορούν να λειτουργήσουν απλώς σε δύο διαστάσεις, αλλά η πλήρης όραση υπολογιστών επιθυμεί τις τρισδιάστατες πληροφορίες .

- **Έμπειρα συστήματα (expert systems)**

Ένας Μηχανικός "γνώσης" παίρνει συνέντευξη από εμπειρογνώμονες σε μια ορισμένη περιοχή και προσπαθεί να ενσωματώσει τη γνώση τους σε ένα πρόγραμμα υπολογιστή για κάποιο στόχο. Η απόδοση αυτής της εργασίας εξαρτάται από το εάν οι διανοητικοί μηχανισμοί που απαιτούνται για το στόχο είναι μέσα στα όρια της Τεχνητής Νοημοσύνης. Ένα από τα πρώτα έμπειρα συστήματα που υλοποιήθηκε ήταν το MYCIN το 1974, το οποίο εντόπιζε τις βακτηριακές μολύνσεις του αίματος και πρότεινε την ανάλογη θεραπεία. Το έμπειρο αυτό σύστημα τα κατάφερε καλύτερα από τους σπουδαστές ιατρικής ή τους εν ενεργεία γιατρούς, κάτω βέβαια από τους περιορισμούς του συστήματος.

- **Ευρετική ταξινόμηση (heuristic classification)**

Ένα από τα πιο εφικτά είδη έμπειρου συστήματος, δεδομένης της παρούσας γνώσης στην τεχνητή νοημοσύνη, είναι να τεθούν κάποιες πληροφορίες σε ένα χαρακτηριστικό από ένα σταθερό σύνολο κατηγοριών που χρησιμοποιούν διάφορες πηγές πληροφοριών. Ένα παράδειγμα ευρετικής ταξινόμησης συμβουλεύει εάν δεχτεί μια προτεινόμενη αγορά πιστωτικών καρτών. Οι πληροφορίες είναι διαθέσιμες για τον ιδιοκτήτη της πιστωτικής κάρτας, το αρχείο πληρωμής του και επίσης για το στοιχείο που αγοράζει και από πού το αγοράζει (π.χ., αν έχουν υπάρξει προηγούμενες απάτες πιστωτικών καρτών σε αυτό το κατάστημα).

- **Ασαφής Λογική**

Μια από τις πλέον διαδεδομένες μεθόδους Ευφυούς ή Έμπειρου Ελέγχου στηρίζεται στην προσέγγιση της διαδικασίας νόησης και ονομάζεται **Ασαφής Λογική**. Όπως γνωρίζουμε, η κλασική λογική είναι δίτιμη, χειρίζεται δηλαδή τη γνώση προσάπτοντας στα γεγονότα έναν από τους δύο χαρακτηρισμούς "αληθές" ή "ψευδές". Μ' αυτόν όμως τον τρόπο δεν γίνεται κατορθωτή μια ρεαλιστική περιγραφή των περισσότερων πραγματικών προβλημάτων, εφόσον ο κόσμος διέπεται από τυχαία

γεγονότα που ισχύουν με κάποιο βαθμό αβεβαιότητας. Η Ασαφής Λογική αποτελεί προσπάθεια αντιμετώπισης των δυσχερειών που έχει η δίτιμη λογική Boole στην έκφραση των αβέβαιων γεγονότων και εννοιών. Έτσι, ενώ η λογική Boole δέχεται μόνο δύο δυνατές τιμές για μία πρόταση (αληθής=1, μη αληθής=0), στην Ασαφή Λογική επιτρέπονται όλες οι τιμές μεταξύ 0 και 1.

1.3 Έμπειρα Συστήματα

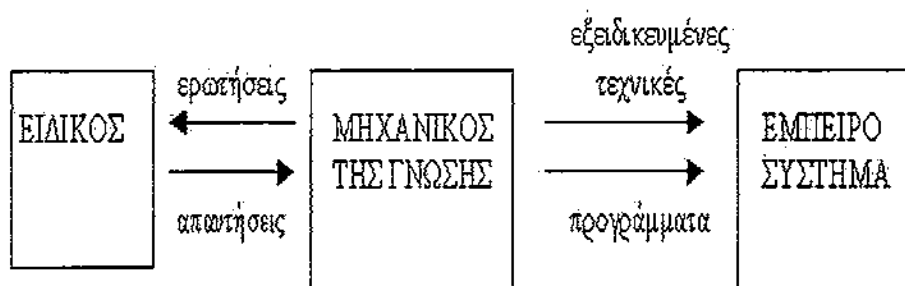
Οι ηλεκτρονικοί υπολογιστές έχουν καθιερωθεί ως ένα ισχυρό εργαλείο για την αντιμετώπιση πολύπλοκων και χρονοβόρων υπολογιστικών διαδικασιών. Σχετικά πρόσφατα όμως έχουν αρχίσει προσπάθειες με σκοπό την εκμετάλλευση των νέων δυνατοτήτων της επιστήμης της πληροφορικής και απώτερο στόχο την ανάπτυξη πληροφοριακών συστημάτων τα οποία θα έχουν τη δυνατότητα να σκέπτονται προσομοιώνοντας την ανθρώπινη λογική και σκέψη.

Η επιστήμη της Τεχνητής Νοημοσύνης, είναι ένας από τους νεότερους επιστημονικούς κλάδους, που σκοπό έχει την ανάπτυξη ευφυιών πληροφοριακών συστημάτων, μέσω των οποίων ο ηλεκτρονικός υπολογιστής θα μπορεί να εκτελέσει ορισμένες βασικές νοητικές λειτουργίες του ανθρώπινου εγκεφάλου, όπως η μάθηση, η πραγματοποίηση λογικών συλλογισμών για την επίλυση προβλημάτων, η κατανόηση της γλώσσας και της ομιλίας, κ.α..

Μέχρι σήμερα, παρόλη την έντονη και εκτενή επιστημονική μελέτη και έρευνα που γίνεται πάνω στον τομέα της τεχνητής νοημοσύνης, τα αποτελέσματα δεν έχουν φθάσει στα επιθυμητά επίπεδα. Αν και έχουν γίνει ορισμένα βήματα προόδου, παρόλα αυτά δεν έχει γίνει μέχρι σήμερα δυνατή μια ικανοποιητική προσομοίωση των ανθρώπινων νοητικών λειτουργιών μέσω ενός πληροφοριακού συστήματος. Μέσα όμως από τις προσπάθειες αυτές ανοίχτηκαν νέα πεδία έρευνας και αναπτύχθηκαν νέες προσεγγίσεις για την αξιοποίηση των δυνατοτήτων των υπολογιστών στην υποστήριξη της διαδικασίας λήψης αποφάσεων. Τα Έμπειρα Συστήματα, αποτέλεσμα των προσπαθειών που γίνονται στο χώρο της τεχνητής νοημοσύνης, παρέχουν μια νέα προσέγγιση στον τρόπο αντιμετώπισης της διαδικασίας λήψης αποφάσεων, εξομοιώνοντας με έναν απλό και κατανοητό τρόπο την ανθρώπινη σκέψη και λογική και έχουν ήδη αρχίσει να βρίσκουν εφαρμογή σε μια πληθώρα πρακτικών προβλημάτων λήψης αποφάσεων.

Τα Έμπειρα Συστήματα (Ε.Σ.) είναι προγράμματα, που συνδυάζουν τη γνώση των ειδικών και προσπαθούν να επιλύσουν ειδικά προβλήματα, μιμούμενα τη διαδικασία της λογικής τους. Ε.Σ. είναι εκείνο που αφενός μεν χειρίζεται πραγματικά σύνθετα προβλήματα που απαιτούν την παρουσία ειδικού, αφετέρου δε επιλύει τέτοια προβλήματα χρησιμοποιώντας ένα υπολογιστικό μοντέλο λογικής ειδικού, καταλήγοντας στα ίδια συμπεράσματα που ένας ειδικός θα έφτανε αν αντιμετώπιζε ένα ανάλογο πρόβλημα. Ένα Ε.Σ. προσπαθεί να συλλάβει αρκετή γνώση ενός ειδικού έτσι ώστε να επιλύει με έμπειρο τρόπο τα προβλήματα.

Τα Έμπειρα Συστήματα αποτελούν το κατ' εξοχήν εμπορικό προϊόν του τομέα της Τεχνητής Νοημοσύνης. Είναι προγράμματα τα οποία στην προσπάθειά τους να αντιμετωπίσουν πολύπλοκα προβλήματα του πραγματικού κόσμου, κωδικοποιούν τη γνώση και τη συλλογιστική ενός έμπειρου ειδικού σε έναν εξειδικευμένο τομέα της επιστήμης ή της τεχνολογίας. Είναι δηλαδή υπολογιστικά προγράμματα που αναφέρονται σε ένα εξειδικευμένο τομέα της ανθρώπινης γνώσης και επιδεικνύουν συμπεριφορά ανάλογη προς εκείνη του ανθρώπου-ειδικού στον εν λόγω τομέα. Ένα Έμπειρο Σύστημα διαθέτει μια σημαντική μάζα ειδικής γνώσης η οποία εκμαγιεύεται από ανθρώπους ειδικούς σε ένα τομέα και μεταφυτεύεται στο σύστημα με κατάλληλο τρόπο. Ο ειδικευμένος επιστήμονας που αποσπά τη γνώση από τους ειδικούς ονομάζεται "μηχανικός της γνώσης" (knowledge engineer).



Σχήμα 1.1: Έμπειρο Σύστημα

Στα πραγματικά προβλήματα η γνώση δεν είναι στατική, αλλά δυναμική υποκείμενη σε μεταβολές, προσθήκες, ανακατατάξεις και αναθεωρήσεις. Τα Έμπειρα Συστήματα χαρακτηρίζονται από σημαντική προσαρμοστικότητα και προσφέρουν

εξαιρετική ευελιξία, λόγω της ευκολίας που παρέχουν για προσθήκη νέας γνώσης και αναθεώρηση της ήδη υπάρχουσας και της ευχέρειας που προσφέρουν για σταδιακή ανάπτυξη. Αυτό ακριβώς συνιστά μια πρώτη ειδοποιό διαφορά των συστημάτων αυτών σε σχέση με τα παραδοσιακά προγράμματα. Η υφή του πραγματικού κόσμου εξάλλου, επιβάλλει την εξαγωγή συμπερασμάτων με χειρισμό ελλιπούς γνώσης, ασαφών και αβέβαιων γεγονότων, καθώς επίσης και μη αντικειμενικά ορισμένων εννοιών (όπως π.χ. “ψηλός”, που για τον καθένα εκφράζεται από διαφορετικό μέτρο). Η τεχνολογία των Έμπειρων Συστημάτων προσφέρει δυνατότητα έκφρασης τέτοιων χαρακτηριστικών του πραγματικού κόσμου με τη χρήση διάφορων μεθόδων και τεχνικών, οι οποίες έχουν ήδη αποδώσει ικανοποιητικά. Στην επόμενη παράγραφο περιγράφονται τα βασικά δομικά στοιχεία των Ε.Ε.

1.3.1 Δομή Έμπειρων Συστημάτων

- **Βάση Γνώσης (Knowledge Base)** – Η Βάση Γνώσης είναι μια συλλογή από γεγονότα (facts) και κανόνες (rules) σχετικά με το πεδίο εφαρμογής του Ε.Σ.
- **Μηχανή Συμπερασμού (Inference engine)** – Η Μηχανή Συμπερασμού αποτελείται από κωδικοποιημένα υπολογιστικά προγράμματα τα οποία εκτελούν την διαδικασία συλλογισμού του ΕΣ. Αυτά τα προγράμματα ταιριάζουν τις ερωτήσεις και απαντήσεις που δίνει ο χρήστης με τα γεγονότα και κανόνες που βρίσκονται στην Βάση Γνώσης .
- **Διεπαφή με τον Χρηστή (User interface)** – Διεπαφή με τον χρηστή είναι η επικοινωνία μεταξύ του χρηστή και της μηχανής συμπερασμού και παρουσιάζει τα αποτελέσματα από την μηχανή αιτιολόγησης
- **Μηχανή Αιτιολόγησης (Explanation facility)** – Η Μηχανή Αιτιολόγησης εξηγεί την λογική συμπερασμού του συστήματος και δικαιολογεί το πώς και το γιατί έφτασε στα συγκεκριμένα συμπεράσματα το Ε.Σ . Αυτό κάνει τους χρήστες να αισθάνονται πιο άνετα και πιο σίγουροι με το Ε.Σ.

1.4 Γενετικοί Αλγόριθμοι

Οι γενετικοί αλγόριθμοι είναι ευρεστικοί αλγόριθμοι που βασίζονται στους μηχανισμούς της φυσικής επιλογής και της γενετικής. Συνδυάζουν την αρχή της

επιβίωσης του ισχυρότερου και του καλύτερου μεταξύ δομών (αρχή του Δαρβίνου), με μια δομημένη αλλά τυχαία ανταλλαγή πληροφοριών, για να δημιουργήσουν έναν αλγόριθμο που προσεγγίζει τις καινοτομίες της ανθρώπινης αναζήτησης. Σε κάθε γενιά δημιουργείται ένα νέο σύνολο από τεχνητές οντότητες, χρησιμοποιώντας τα κομμάτια των ισχυρότερων της παλιάς γενιάς. Παρόλο που είναι πιθανοτικός αλγόριθμος, οι γενετικοί αλγόριθμοι δεν είναι απλοί τυχαίοι περίπατοι. Ερευνούν πληροφορίες του παρελθόντος για να μπορέσουν να κάνουν υποθέσεις σχετικά με νέα σημεία έρευνας, με βελτιωμένη απόδοση.

Τα πλεονεκτήματα των γενετικών τεχνικών έναντι των άλλων τεχνικών βελτιστοποίησης, είναι πολλά. Επιγραμματικά αναφέρονται τα εξής:

- Οι γενετικοί δεν ψάχνουν για μία πιθανή λύση σε κάποιο υποδιάστημα του χώρου των λύσεων, αλλά ερευνούν για μία πιθανή λύση σε όλο το χώρο των λύσεων του δοσμένου προβλήματος.

- Οι γενετικοί δεν ξεκινάνε από μία μόνο δοσμένη λύση-σημείο, αλλά αντίθετα, ξεκινάνε από ένα πλήθος λύσεων.

- Οι γενετικοί δε βασίζονται αποκλειστικά στην αρχή του “hill climbing” (που είναι εξαντλητική τεχνική), αλλά εντοπίζουν αυτά που είναι ανώτερα μέσα σε ένα σύνολο πιθανών λύσεων, με ένα πιθανοτικό τρόπο.

Οι Γενετικοί Αλγόριθμοι θα μελετηθούν εκτενέστερα σε επόμενη παράγραφο. [5]

Κεφάλαιο 2

Τεχνητά Νευρωνικά Δίκτυα

2.1 Εισαγωγή

Το πρώτο κύμα ενδιαφέροντος για τα Τεχνητά Νευρωνικά Δίκτυα (Τ.Ν.Δ) στην επιστημονική κοινότητα εκδηλώθηκε μετά την εισαγωγή ενός απλοποιημένου μοντέλου του νευρώνα (simplified neuron), από τους McCulloch και Pitts το 1943 [7]. Αυτοί οι νευρώνες παρουσιάστηκαν ως μοντέλα βιολογικών νευρώνων και ως μέρη κυκλωμάτων τα οποία θα μπορούσαν να εκτελέσουν υπολογισμούς. Όταν ο Minstky και ο Papert εξέδωσαν το βιβλίο το σύγγραμμα τους Perceptrons [8], στο οποίο παρέθεταν τα αρνητικά σημεία των μοντέλων Perceptron, η χρηματοδότηση των ερευνών για τα νευρωνικά δίκτυα σταμάτησε και στράφηκε σε άλλους τομείς. Ελάχιστοι ήταν οι ερευνητές που συνέχισαν να δραστηριοποιούνται στον τομέα αυτό, ανάμεσα τους οι Stephen Grossberg, James Anderson και Kunihiko Fukushima. Το ενδιαφέρον για τα Τ.Ν.Δ αναθερμάνθηκε στις αρχές της δεκαετίας του '80. Από τότε μέχρι και σήμερα σημειώθηκε σημαντική πρόοδος και σημαντικά ερευνητικά αποτελέσματα που σε συνδυασμό με την ραγδαία εξέλιξη του υλικού (hardware), οδήγησαν στην στροφή των χρηματοδοτήσεων ξανά προς αυτό τον τομέα. Στις μέρες μας η πλειονότητα των ακαδημαϊκών ιδρυμάτων διαθέτει μια ομάδα που ασχολείται ερευνητικά με τα Τεχνητά Νευρωνικά Δίκτυα.

Τα Τ.Ν.Δ μπορούν να χαρακτηριστούν ως υπολογιστικά μοντέλα με συγκεκριμένες ιδιότητες, όπως η ικανότητα να προσαρμόζονται και να μαθαίνουν, να γενικεύουν (generalize), να ομαδοποιούν και να οργανώνουν δεδομένα.

Το ερώτημα που τίθεται είναι το αν μπορεί να αποδειχθεί ότι τα Τ.Ν.Δ αποτελούν μια καλύτερη και αποτελεσματικότερη προσέγγιση σε αυτού του είδους τις εφαρμογές από τα υπάρχοντα μοντέλα. Το ερώτημα αυτό παραμένει μέχρι και σήμερα αναπάντητο.

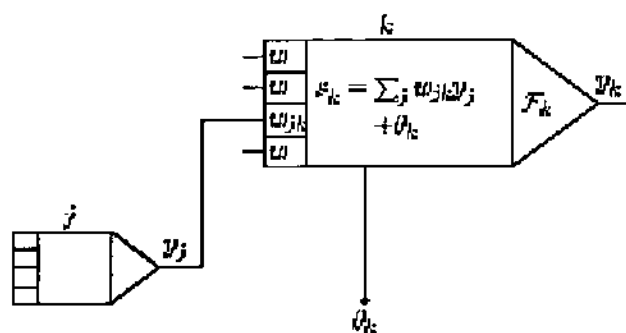
Συχνά γίνεται παραλληλισμός με τα βιολογικά νευρωνικά δίκτυα, όμως αυτά που ξέρουμε για αυτά είναι τόσο λίγα ώστε τα μοντέλα που χρησιμοποιούνται στα Τ.Ν.Δ να αποτελούν μια υπεραπλούστευση των βιολογικών μοντέλων.

Ο τρόπος με τον οποίο θα προσεγγιστούν παρακάτω τα Τ.Ν.Δ είναι καθαρά από την πλευρά της Επιστήμης των Υπολογιστών και όχι από την οπτική της Ψυχολογίας ή των βιολογικών συστημάτων.

2.2 Γενικά

Ένα Τεχνητό Νευρωνικό Δίκτυο αποτελείται από απλές υπολογιστικές μονάδες που επικοινωνούν μεταξύ τους στέλνοντας σήματα μέσω ενός μεγάλου αριθμού σταθμισμένων συνάψεων. Τα κύρια μέρη ενός καταμερισμένου παράλληλα μοντέλου είναι:

- Οι βασικές επεξεργαστικές μονάδες (Basic processing units).
- Η ενεργοποίηση Y_k κάθε μονάδας.
- Η διασύνδεση μεταξύ των μονάδων. Κάθε σύναψη χαρακτηρίζεται από ένα βάρος W_{jk} το οποίο προσδιορίζει την επίδραση που θα έχει το σήμα της μονάδας εξόδου j σε μια είσοδο της μονάδας k .
- Μια συνάρτηση ενεργοποίησης F_k , η οποία προσδιορίζει το νέο επίπεδο ενεργοποίησης με βάση την είσοδο $S_k(t)$ και την τρέχουσα ενεργοποίηση $Y_k(t)$.
- Μια εξωγενή είσοδο Θ_k για κάθε μονάδα, την τιμή κατωφλίου.
- Ένα κανόνα μάθησης (Learning Rule).
- Ένα περιβάλλον μέσα στο οποίο θα λειτουργήσει το σύστημα, που θα παρέχει τα δεδομένα εισόδου (input signals) και αν είναι αναγκαίο και σήματα σφαλμάτων (error signals).



Σχήμα 2.1: Τα συστατικά στοιχεία ενός Τ.Ν.Δ

2.2.1 Μονάδες Επεξεργασίας (Processing Units)

Κάθε μονάδα επιτελεί ένα σχετικά απλό έργο. Συγκεκριμένα δέχεται είσοδο από τις γειτονικές τις μονάδες ή από εξωγενείς πηγές εισόδου τις οποίες χρησιμοποιεί για να υπολογίσει την έξοδο της η οποία με τη σειρά της τροφοδοτεί τις εισόδους των επόμενων επεξεργαστικών μονάδων. Ένα επιπλέον έργο που επιτελούν η μονάδες αυτές είναι η προσαρμογή των βαρών (adjustment). Το σύστημα είναι εκ φύσεως παράλληλο, υπό την έννοια της πληθώρας των μονάδων που ταυτόχρονα μπορούν να κάνουν υπολογισμούς.

Στα νευρωνικά συστήματα θα πρέπει να αναφέρουμε ότι οι μονάδες επεξεργασίας διακρίνονται σε μονάδες εισόδου, μονάδες εξόδου και κρυφές μονάδες (input, output και hidden units) αντίστοιχα. Η λειτουργία κάθε μονάδας περιγράφεται αναλυτικότερα στη συνέχεια [5].

2.2.2 Συνάψεις Μονάδων

Στις περισσότερες των περιπτώσεων υποτίθεται πως κάθε μονάδα δίνει κάτι επιπρόσθετο στην είσοδο της μονάδας με την οποία είναι συνδεδεμένη, απο την δική της είσοδο. Η συνολική είσοδος της μονάδας k είναι το σταθμισμένο άθροισμα διάφορων εξόδων (outputs) οι οποίες καταλήγουν σε αυτή την μονάδα μαζί με την τιμή κατωφλίου Θ_k και δίνεται από τη σχέση:

$$s_k(t) = \sum_j w_{jk}(t) y_j(t) + \theta_k(t) \tag{2.1}$$

2.2.3 Κανόνες Ενεργοποίησης και Εξόδου (Activation and Output Rules)

Επίσης χρειάζεται ένας κανόνας που να δίνει την επίδραση της συνολικής εισόδου στην ενεργοποίηση της μονάδας. Χρειάζεται μια συνάρτηση F_k η οποία με βάση τη συνολική είσοδο $S_k(t)$ και την τρέχουσα ενεργοποίηση $y_k(t)$ να παράγει τη νέα τιμή ενεργοποίησης της μονάδας k

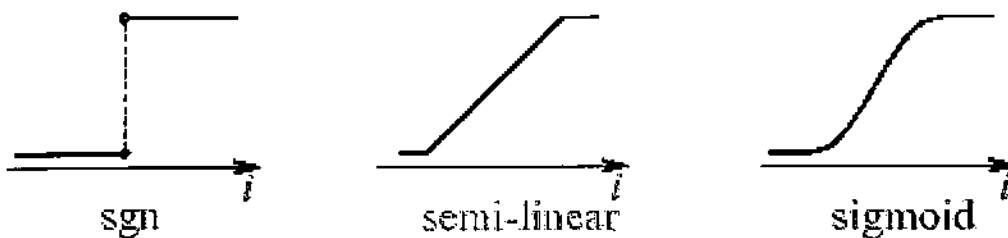
$$y_k(t+1) = F_k(y_k(t), s_k(t)) \quad (2.2)$$

Συχνά η συνάρτηση ενεργοποίησης είναι αύξουσα (non-decreasing) ώστε να μην μειώνει την συνολική είσοδο της μονάδας.

$$y_k(t+1) = F_k(s_k(t)) = F_k\left(\sum_j w_{jk}(t)y_j(t) + \theta_k(t)\right) \quad (2.3)$$

Στο Σχήμα 2.2 φαίνονται οι συναρτήσεις που χρησιμοποιούνται πιο συχνά όπως η συνάρτηση πρόσημου (sign function), η ημιγραμμική (semi-linear) και η σιγμοειδής (sigmoid). Η τελευταία ορίζεται από την ακόλουθη σχέση:

$$y_k = F(s_k) = \frac{1}{1 + e^{-s_k}} \quad (2.4)$$



Σχήμα 2.2: Συναρτήσεις ενεργοποίησης μονάδων

2.3 Τοπολογίες Δικτύων (Network Topologies)

Όσον αφορά τον τρόπο σύνδεσης τα Τ.Ν.Δ χωρίζονται σε 2 κατηγορίες :

- “Προς τα εμπρός” (Feed forward) Δίκτυα: Το χαρακτηριστικό αυτών των Νευρωνικών Δικτύων είναι ότι δεν υπάρχουν συνδέσεις ανάδρασης (feedback), δηλαδή συνδέσεις της εξόδου (output) μονάδας εξόδου σε είσοδο μονάδας εισόδου του ίδιου ή προηγούμενου στρώματος. Κλασικά παραδείγματα “Προς τα εμπρός” δικτύων είναι το Perceptron (Rosenblatt, 1959)[9] και το Adaline (Widrow & Hoff, 1960) [10].

Recurrent Δίκτυα: Αντίθετα εδώ υπάρχουν συνδέσεις ανατροφοδότησης οι οποίες προσδίδουν δυναμικές ιδιότητες στο σύστημα. Παραδείγματα recurrent δικτύων έχουν παρουσιαστεί από τον Anderson (Anderson,1977) [11], Kohonen (Kohonen,1977) [13] και τον Hopfield (Hopfield,1982) [12].

2.4 Εκπαίδευση Τ.Ν.Δ

Ένα Νευρωνικό Δίκτυο θα πρέπει να προσαρμόζει κατά τέτοιο τρόπο τα βάρη των συνάψεων του ώστε η εφαρμογή ενός συνόλου δεδομένων στις εισόδους των μονάδων εισόδου που διαθέτει να δίνει το επιθυμητό αποτέλεσμα στις εξόδους του Ν.Δ. Υπάρχουν πολλές μέθοδοι για να προσαρμοστούν κατά τον επιθυμητό τρόπο οι σύνδεσμοι ενός Ν.Δ. Ένας τρόπος είναι να τεθούν τα βάρη εκ των προτέρων εφόσον γνωρίζουμε ποια θα πρέπει να είναι η τιμή τους. Ένας άλλος τρόπος είναι να εκπαιδευτεί το Τ.Ν.Δ ενσωματώνοντας κάποιους αλγόριθμους μάθησης και αφήνοντας το να προσαρμόσει / ρυθμίσει τα βάρη του με βάση αυτούς τους κανόνες.

2.5 Τρόποι Μάθησης ενός Τ.Ν.Δ

Δυο είναι οι βασικοί τρόποι μάθησης ενός Τ.Ν.Δ:

- *Με επίβλεψη*, όπου το δίκτυο εκπαιδεύεται διαπερνώντας τις μονάδες εισόδου του με ένα σύνολο δεδομένων καθώς και τις επιθυμητές τιμές των μονάδων εξόδων του. Αυτά τα ζεύγη εισόδου-εξόδου δίνονται είτε από κάποιο εξωτερικό ‘δάσκαλο’ είτε από το σύστημα μέσα στο οποίο λειτουργεί το δίκτυο.
- *Χωρίς επίβλεψη*. Σε αυτή την περίπτωση το σύστημα ανακαλύπτει με στατιστικό τρόπο σημαντικά χαρακτηριστικά του πληθυσμού. Σε αντίθεση με την μάθηση με

επίβλεψη δεν υπάρχει έκ των προτέρων γνώση. Έτσι το σύστημα θα πρέπει να αναπτύξει τη δική του αναπαράσταση των 'ερεθισμάτων' που δέχεται μέσω των μονάδων εισόδου του.

2.6 Ρύθμιση των βαρών των συνάψεων

Οι δυο παραπάνω τύποι μάθησης οδηγούν στην προσαρμογή των βαρών που χρησιμοποιούνται σύμφωνα με κάποιο κανόνα προσαρμογής. Στην ουσία όλοι οι κανόνες αυτοί είναι παραλλαγές του κανόνα μάθησης Hebbian ο οποίος προτάθηκε από τον Hebb στην κλασική εργασία του «Οργάνωση της Συμπεριφοράς» (Organization of Behavior) [(1949)(Hebb,1949)] [14]. Η βασική ιδέα είναι ότι αν δύο (2) μονάδες j και k είναι ενεργές ταυτόχρονα τότε η μεταξύ τους διασύνδεση θα πρέπει να ενδυναμωθεί. Αν η j δέχεται είσοδο από την k η απλούστερη έκδοση του κανόνα αυτού συνιστά μετατροπή του βάρους W_{jk} με βάση τον ακόλουθο κανόνα:

$$\Delta w_{jk} = \gamma y_j y_k \quad (2.5)$$

όπου γ είναι μια θετική σταθερά που καθορίζει τον ρυθμό μάθησης (learning rate) του δικτύου. Ένας άλλος κλασικός κανόνας χρησιμοποιεί όχι αυτή καθαυτή την ενεργοποίηση της μονάδας k αλλά τη διαφορά της πραγματικής από την επιθυμητή ενεργοποίηση για την προσαρμογή των βαρών :

$$\Delta w_{jk} = \gamma y_j (d_k - y_k) \quad (2.6)$$

όπου d_k είναι η επιθυμητή ενεργοποίηση η οποία παρέχεται από κάποιο "δάσκαλο"

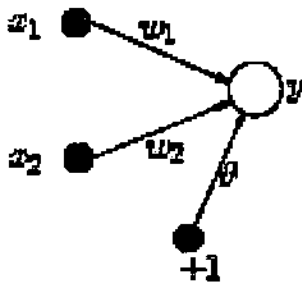
2.7 Perceptron και Adaline

Στη συνέχεια θα αναλυθούν δύο (2) μοντέλα Τ.Ν.Δ.. Πιο συγκεκριμένα το Perceptron το οποίο προτάθηκε από τον Rosenblatt (Rosenblatt,1959) [9] και το Adaline που προτάθηκε από τους Widrow και Hoff (Widrow and Hoff, 1960) [10].

2.7.1 Perceptron

Ένα “Προς τα εμπρός” δίκτυο με ένα επίπεδο. Έχει ένα ή περισσότερους νευρώνες εξόδου O . Καθένας από τους νευρώνες αυτούς είναι συνδεδεμένος με ένα βάρος W_{io} με όλες τις εισόδους. Το δομικό στοιχείο του μοντέλου αυτού αποτελείται από δύο (2) εισόδους και μια μόνο έξοδο (Σχήμα 2.3). Η είσοδος του νευρώνα είναι το σταθμισμένο άθροισμα των εισόδων συν την τιμή κατωφλίου (το bias). Όπως φαίνεται η έξοδος του δικτύου σχηματίζεται από την ενεργοποίηση του νευρώνα εξόδου και δίνεται από τη σχέση.

$$y = F\left(\sum_{i=1}^2 w_i x_i + \theta\right) \quad (2.7)$$



Σχήμα 2.3: Τ.Ν.Δ ενός επιπέδου με δύο εισόδους και μια έξοδο

Η συνάρτηση ενεργοποίησης F , όπως έχει ήδη αναφερθεί, μπορεί να είναι γραμμική, οπότε προκύπτει ένα γραμμικό δίκτυο, ή μη γραμμική. Υποθέτοντας, για παράδειγμα, την συνάρτηση πρόσημου

$$F(s) = \begin{cases} 1 & \alpha\nu s > 0 \\ -1 & \alpha\nu s \leq 0 \end{cases} \quad (2.8)$$

Η έξοδος του δικτύου θα είναι είτε 1 είτε -1 όπως φαίνεται από την παραπάνω σχέση. Με αυτό τον τρόπο το δίκτυο μπορεί να προχωρήσει στην κατηγοριοποίηση (classification) αποφασίζοντας αν ένα πρότυπο εισόδου ανήκει σε μια από τις δύο (2) κατηγορίες. Αν η συνολική είσοδος είναι θετική τότε το πρότυπο θα κατηγοριοποιηθεί ως +1 ενώ σε αντίθετη περίπτωση ως -1. Αυτό που χωρίζει αυτές τις δυο κατηγορίες είναι μια ευθεία η οποία ορίζεται από την εξίσωση :

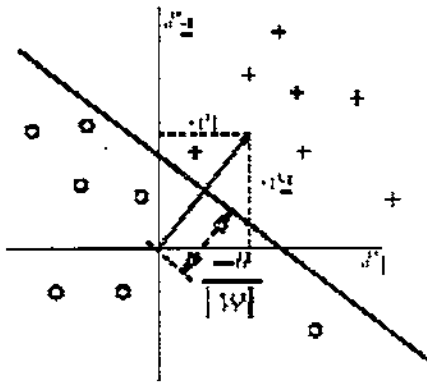
$$w_1 x_1 + w_2 x_2 + \theta = 0 \quad (2.9)$$

ή ισοδύναμα

$$x_2 = -\frac{w_1}{w_2} x_1 - \frac{\theta}{w_2} \quad (2.10)$$

Έτσι βλέπουμε πως ο λόγος των βαρών καθορίζει την κλίση της ευθείας ενώ ο λόγος του βάρους w_2 και της τιμής κατωφλίου την μετατόπιση.

Η γεωμετρική αναπαράσταση αυτού του δικτύου δίνεται στο σχήμα που ακολουθεί.



Σχήμα 2.4: Γεωμετρική αναπαράσταση της συνάρτησης προσήμου και των βαρών

Βλέποντας τις δυνατότητες αυτού του Δικτύου προκύπτει ένα νέο ερώτημα : πως ‘μαθαίνουμε’ τα βάρη και την τιμή κατωφλίου. Στη συνέχεια θα περιγράψουμε δυο μεθόδους μάθησης για τα δίκτυα αυτού του τύπου. Και οι δυο αυτές μέθοδοι είναι επαναληπτικές (iterative) διαδικασίες που προσαρμόζουν τα βάρη. Έτσι η νέα τιμή για κάθε βάρος και τιμή κατωφλίου υπολογίζεται προσθέτοντας μια σταθμισμένη διόρθωση στην παλιά τιμή σύμφωνα με τις παρακάτω σχέσεις :

$$w_i(t+1) = w_i(t) + \Delta w_i(t) \quad (2.11)$$

$$\theta(t+1) = \theta(t) + \Delta\theta(t) \quad (2.12)$$

Από τις παραπάνω σχέσεις είναι προφανές ότι το πρόβλημα μάθησης ανάγεται στο πρόβλημα υπολογισμού των $\Delta w_i(t)$ και $\Delta\theta(t)$ αντίστοιχα.

2.7.1.1 Ο Κανόνας Μάθησης που χρησιμοποιείται στο Perceptron και το Θεώρημα Σύγκλισης.

Έστω ότι έχουμε στη διάθεση μας ένα σύνολο δεδομένων μάθησης που αποτελούνται από το διάνυσμα εισόδου x και την επιθυμητή έξοδο d_x . Σε προβλήματα κατηγοριοποίησης το d_x είναι συνήθως $+1$ ή -1 . Ο αλγόριθμος μάθησης που χρησιμοποιείται στο Perceptron είναι απλός και μπορεί να διατυπωθεί με τα ακόλουθα βήματα :

B_1 : Ξεκίνα με τυχαία βάρη για τις συνδέσεις .

B_2 : Επέλεξε μια είσοδο x από το σύνολο εκπαίδευσης .

B_3 : Ενημέρωσε τα βάρη W_i των συνάψεων χρησιμοποιώντας την ακόλουθη σχέση:

$$\Delta w_i = \begin{cases} d_x \cdot x_i, & y \neq d_x \\ 0, & y = d_x \end{cases} \quad (2.13)$$

B_4 : Επέστρεψε στο B_2 ;

Παρατηρούμε ότι η διαδικασία είναι παρόμοια με τον κανόνα του Hebb με τη μόνη διαφορά είναι ότι όταν το δίκτυο ανταποκρίνεται σωστά, δεν αλλάζουν τα βάρη των συνδέσεων. Εκτός όμως από τα βάρη θα πρέπει να τροποποιηθεί και το θ . Ως θ θεωρείται η σύνδεση ανάμεσα στον νευρώνα εξόδου και μια dummy μονάδα η οποία είναι συνέχεια ανοιχτή ($x_0=1$). Έτσι με βάση τα παραπάνω το θ θα τροποποιείται ως εξής :

$$\Delta\theta = \begin{cases} 0, & \text{αν το perceptron ανταποκρίνεται σωστά} \\ d_x, & \text{διαφορετικά} \end{cases} \quad (2.14)$$

Θεώρημα Σύγκλισης (Convergence Theorem). Έστω ένα σει από βάρη w^* τα οποία είναι ικανά να συγκλίνουν στην λύση $y=d_x$, τότε ο κανόνας μάθησης για το Perceptron θα φτάσει σε κάποια λύση (η οποία μπορεί να είναι ή και να μην είναι ίδια με το w^*) ύστερα από ένα πεπερασμένο αριθμό βημάτων για κάθε αρχική επιλογή βαρών.

Απόδειξη: Υποθέτοντας ότι το μέγεθος του w^* δεν παίζει κανένα ρόλο (εξαιτίας του πρόσημου), παίρνουμε $|w^*|=1$. Επειδή έχουμε θεωρήσει ότι το w^* είναι η σωστή λύση, τότε το εσωτερικό γινόμενο $w^{*t}x$ θα είναι μεγαλύτερο του μηδενός ή αν υπάρχει ένα $\delta > 0$ τέτοιο ώστε $w^{*t}x > \delta$ για κάθε είσοδο x .

Ορίζουμε σαν $\cos a = w^t w^* / \|w\|$. Όταν τα βάρη, όπως γνωρίζουμε από τον κανόνα μάθησης για το Perceptron, τροποποιούνται για δεδομένη είσοδο x σύμφωνα με την Σχέση (2.13) και επομένως:

$$w = w + d_x x$$

Άρα

$$\begin{aligned} \|w^*\|^2 &= \|w + d_x x\|^2 \\ \|w^*\|^2 &= \|w\|^2 + 2d_x w^t x + \|x\|^2 \\ \|w^*\|^2 &< \|w\|^2 + \|x\|^2 \\ \|w^*\|^2 &= \|w\|^2 + M \end{aligned} \tag{2.15}$$

Όμως $w^t w^* = w^t w^* + \text{sgn}(x^t w^*) x^t w^* > w^t w^* + \delta$ και επομένως μετά από n επαναλήψεις θα έχουμε :

$$w_n^t w^* > w^t w^* + n\delta \tag{2.16}$$

Ορίζοντας την ποσότητα: $\cos(a_n) = \frac{w_n^t \cdot w^*}{\|w_n\|}$ (2.17)

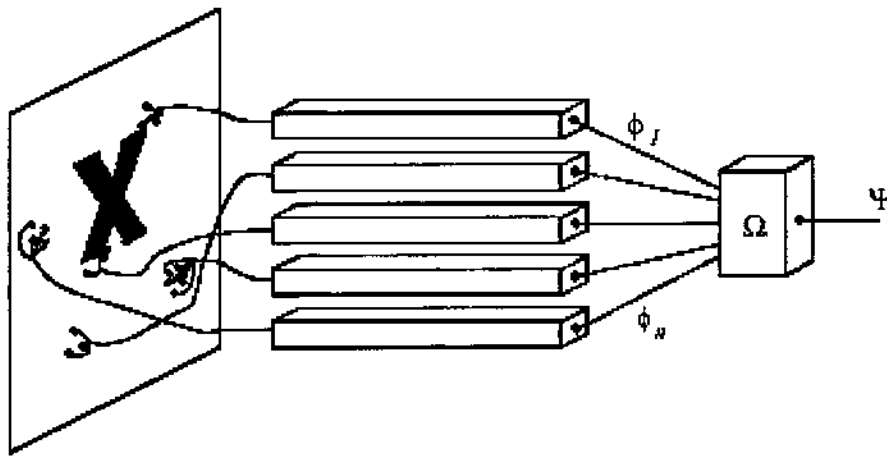
Έχουμε ότι $\cos(a_n) > \frac{w^t w^* + n\delta}{\sqrt{\|w\|^2 + nM}}$ (2.18)

και επομένως

$$\lim_{n \rightarrow \infty} \cos a_n = \lim_{n \rightarrow \infty} \frac{\delta}{\sqrt{M}} \sqrt{n} = \infty \tag{2.19}$$

το οποίο είναι άτοπο αφού εξορισμού $\cos(\alpha) \leq 1$.

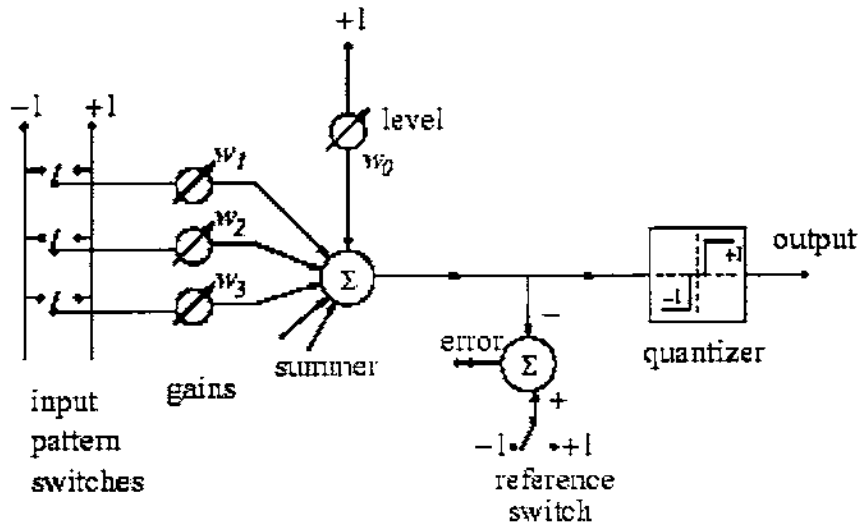
Συμπερασματικά θα πρέπει να υπάρχει ένα ανώτερο όριο n_{max} για το n . Το σύστημα τροποποιεί τις συνδέσεις του για ένα περιορισμένο αριθμό. Με αλλά λόγια, ύστερα από n_{max} ενημερώσεις το Perceptron λειτουργεί σωστά. Το n_{max} επιτυγχάνεται όταν $\cos(\alpha)=1$. Αν ξεκινήσουμε με $w=0$, $n_{max} = M/\delta^2$



Σχήμα 2.5: Το Perceptron

2.7.1.2 Το Adaline (Adaptive Linear Element)

Μια σημαντική γενίκευση του αλγόριθμου εκπαίδευσης του Perceptron παρουσιάστηκε από τους Widrow και Hoff, που είναι γνωστός σαν LMS (Least Mean Square). Η βασική λειτουργική διάφορα με το Perceptron είναι ο τρόπος με τον οποίο η έξοδος του συστήματος χρησιμοποιείται από τον αλγόριθμο μάθησης. Στο Perceptron χρησιμοποιείται η έξοδος της συνάρτησης πρόσημου (+1 ή -1) ενώ στη γενικευμένη του μορφή ο LMS χρησιμοποιεί την έξοδο του δικτύου χωρίς να το κατηγοριοποιεί σε τιμές +1 ή -1. Αυτός ο αλγόριθμος μάθησης χρησιμοποιήθηκε πρωταρχικά στο «Adaptive Linear Element» γνωστό και ως Adaline, που αναπτύχθηκε από τους Widrow και Hoff (Widrow and Hoff, 1960) [10].



Σχήμα 2.6: Το Adaline

Παρόλο που στην παραπάνω περίπτωση του σχήματος υπάρχει μόνο μια έξοδος, είναι σαφές πως ένα σύστημα με πολλαπλές παράλληλες εξόδους προέρχεται άμεσα από πολλαπλές μονάδες του παραπάνω τύπου. Αν $w_i, i=0, 1, 2, \dots, n$ τα βάρη των συνδέσεων του και y η έξοδος του, τότε αυτή ορίζεται ως

$$y = \sum_{i=1}^n w_i x_i + \theta, \quad (2.20)$$

όπου $\theta \equiv w_0$

Ο σκοπός αυτού του συστήματος είναι να δώσει σαν έξοδο ένα $y = d^p$ όταν οι εισοδοί είναι $x_i^p, i=1, 2, \dots, n$. Το πρόβλημα είναι να καθοριστούν με τέτοιο τρόπο τα $w_i, i=0, 1, 2, \dots, n$, ώστε το σύστημα να ανταποκρίνεται σωστά για ένα μεγάλο αριθμό τυχαία επιλεγμένων δειγμάτων εισόδου. Αν δεν είναι δυνατή μια ακριβής απεικόνιση της διαδικασίας από το σύστημα τότε θα πρέπει να ελαχιστοποιήσουμε το σφάλμα, για παράδειγμα με την έννοια των ελάχιστων τετράγωνων. Μια προσαρμοζόμενη διαδικασία (adaptive process) θα πρέπει να περιέχει ένα μηχανισμό τέτοιο ώστε τα w_i να ρυθμιστούν, με βάση κάποια επαναληπτική διαδικασία, έτσι ώστε να προσεγγίζουν τις επιθυμητές τιμές.

2.7.2 Δίκτυα με Γραμμικές Συναρτήσεις Ενεργοποίησης: Κανόνας Δέλτα

Για ένα σύστημα ενός επιπέδου (single layer) με μια μονάδα ενεργοποίησης η έξοδος δίνεται από την σχέση:

$$y = \sum_j w_j x_j + \theta \quad (2.21)$$

Ένα τόσο απλό δίκτυο έχει την δυνατότητα να αναπαραστήσει μία γραμμική σχέση ανάμεσα στην είσοδο και την έξοδο. Σε μεγάλο αριθμό διαστάσεων το δίκτυο αναπαριστάται από ένα υπερεπίπεδο και επομένως μπορούν να οριστούν και πολλαπλές μονάδες εξόδου.

Έστω ότι θέλουμε να εκπαιδεύσουμε το δίκτυο έτσι ώστε ένα υπερεπίπεδο να προσαρμοστεί όσο το δυνατό καλύτερα σε ένα σετ εκπαιδευτικών δειγμάτων (training samples) το οποίο αποτελείται από την είσοδο x και την επιθυμητή έξοδο d . Για κάθε δείγμα η έξοδος διαφέρει από το επιθυμητό κατά $d-y$, όπου y είναι η τιμή της εξόδου.

Ορίζοντας σαν συνάρτηση κόστους το άθροισμα των τετραγώνων του σφάλματος, το συνολικό σφάλμα θα δίνεται από τη σχέση

$$E = \sum_p (e^p)^2 = \frac{1}{2} \sum_p (d^p - y^p)^2 \quad (2.22)$$

όπου ο δείκτης p διατρέχει το πλήθος των τιμών της εισόδου του δικτύου και το e^p παριστά το σφάλμα στο πρότυπο p . Ο αλγόριθμος LMS βρίσκει τις τιμές εκείνες των βαρών οι οποίες ελαχιστοποιούν την συνάρτηση κόστους με μια μέθοδο η οποία ονομάζεται μέθοδος απότομης κατάβασης. Σύμφωνα με τη μέθοδο αυτή το βάρος αλλάζει με τρόπο ανάλογο προς το αρνητικό της μερικής παραγωγού του σφάλματος:

$$\Delta_p w_j = -\gamma \frac{\partial e^p}{\partial w_j} \quad (2.23)$$

όπου γ είναι ο συντελεστής της αναλογικότητας.

Όμως η μερική παράγωγος μπορεί να γραφεί ως εξής:

$$\frac{\partial e^p}{\partial w_j} = \frac{\partial e^p}{\partial y^p} \frac{\partial y^p}{\partial w_j} \quad (2.24)$$

και από την Σχέση (2.16) έχουμε

$$\frac{\partial y^p}{\partial w_j} = x_j \quad (2.25)$$

Παίρνοντας υπόψη μας τις παραπάνω σχέσεις και ότι

$$\frac{\partial e^p}{\partial y^p} = -(d^p - y^p) \quad (2.26)$$

Η Σχέση (2.18) γράφεται ισοδύναμα ως ακολούθως

$$\Delta_p w_j = \gamma e^p x_j \quad (2.27)$$

Ο κανόνας δέλτα τροποποιεί κατάλληλα τα βάρη έτσι ώστε να ελαχιστοποιηθεί η συνάρτηση κόστους της Σχέσης (2.17).

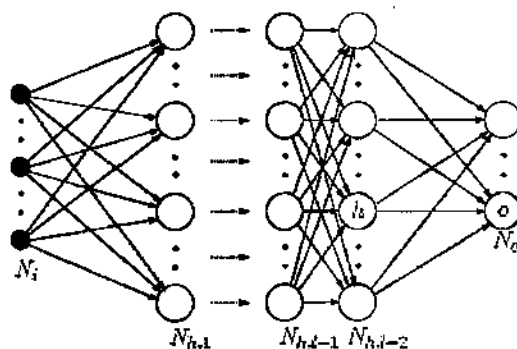
2.8 Ανάδρομη Διάδοση σφάλματος

Όπως είδαμε ένα δίκτυο ενός επιπέδου έχει πολλούς περιορισμούς όσον αφορά τις διεργασίες που μπορεί να πραγματοποιήσει. Γι' αυτό το λόγο παρακάτω θα εστιάσουμε σε "προς τα εμπρός" δίκτυα με περισσότερα στρώματα. Οι Minsky και Papert (Minsky and Papert, 1969) [8] έδειξαν ότι ένα "προς τα εμπρός" (feed forward) δίκτυο με δύο επίπεδα μπορεί να υπερβεί πολλούς από τους παραπάνω περιορισμούς αλλά δεν παρουσίασαν λύση για το πως θα προσαρμόζονταν τα βάρη από τις μονάδες εισόδου στις κρυφές μονάδες. Απάντηση σ' αυτό ήρθαν να δώσουν οι Rumelhart, Hinton και Williams το 1986 (Rumelhart, Hinton, & Williams, 1986) [15]. Το βασικό σκεπτικό αυτής της λύσης είναι ότι τα σφάλματα στο κρυφό επίπεδο καθορίζονται από την Ανάδρομη Διάδοση (back propagation) των σφαλμάτων της εξόδου. Γι' αυτό τον λόγο η μέθοδος αυτή αποκαλείται κανόνας μάθησης με Ανάδρομη Διάδοση Σφάλματος. Η Ανάδρομη Διάδοση Σφάλματος μπορεί να θεωρηθεί ως γενίκευση του Κανόνα Δέλτα για μη-γραμμικές συναρτήσεις και πολυεπίπεδα δίκτυα.

2.9 Πολυεπίπεδα "προς τα εμπρός" Δίκτυα

Τα "προς τα εμπρός" δίκτυα έχουν στρωματοποιημένη δομή. Κάθε στρώμα-επίπεδο αποτελείται από μονάδες οι οποίες λαμβάνουν είσοδο από μονάδες που βρίσκονται στο ακριβώς προηγούμενο επίπεδο και η έξοδος τους τροφοδοτεί τις εισόδους μονάδων που βρίσκονται στο ακριβώς επόμενο επίπεδο. Δεν υπάρχουν συνδέσεις μεταξύ

των μονάδων του ίδιου επιπέδου. Στις μονάδες εισόδου δεν λαμβάνει χώρα επεξεργασία. Η συνάρτηση ενεργοποίησης της κρυφής μονάδας είναι μια συνάρτηση F_i των σταθμισμένων εισόδων και του κατωφλίου. Η έξοδος της κρυφής μονάδας διανέμεται στο επόμενο επίπεδο έως ότου φτάσει στη N_o μονάδα εξόδου, όπως φαίνεται και στο Σχήμα 2.7.



Σχήμα 2.7: Ένα πολυεπίπεδο δίκτυο με l επίπεδα

Παρόλο που ο Back-propagation μπορεί να εφαρμοστεί σε δίκτυα με οποιονδήποτε αριθμό στρωμάτων, για τα δίκτυα με binary units έχει αποδειχθεί (Hornik, Stinchcombe & White 1989 [16], Funahashi, 1989 [17], Cybenko 1989 [18], Hartman, Keeler & Kowalski 1990 [19]) ότι αρκεί ένα επίπεδο κρυφών μονάδων για να προσεγγίσει οποιαδήποτε συνάρτηση αν χρησιμοποιήσουμε βέβαια σαν συναρτήσεις ενεργοποίησης των κρυφών μονάδων μη-γραμμικές συναρτήσεις (Universal Approximation Theorem). Η πιο συχνά χρησιμοποιούμενη συνάρτηση είναι η σιγμοειδής. Το θεώρημα αυτό θα αναπτυχθεί εκτενέστερα παρακάτω.

2.10 Πολυεπίπεδο Perceptron

Αντιπροσωπευτικό παράδειγμα πολυεπίπεδου δικτύου είναι το Πολυεπίπεδο Perceptron, που είναι ένα Νευρωνικό Δίκτυο το οποίο αποτελείται από:

- ένα επίπεδο εισόδου από κόμβους-πηγές
- ένα ή περισσότερα κρυφά επίπεδα (hidden layers) που αποτελούνται από υπολογιστικούς νευρώνες. Οι νευρώνες αυτοί λειτουργούν κατά κάποιο τρόπο ως ‘ανιχνευτές χαρακτηριστικών’ (‘feature detectors’) και
- ένα επίπεδο εξόδου το οποίο παρουσιάζει στο χρήστη τα αποτελέσματα.

Οι δύο βασικές ιδιότητες που χαρακτηρίζουν αυτές τις δομές είναι :

1. Τα Πολυεπίπεδα Perceptrons είναι “προς τα εμπρός” δίκτυα
2. Μπορούν να είναι πλήρως συνδεδεμένα (fully connected) με τέτοιο τρόπο ώστε κάθε κόμβος του δικτύου να είναι συνδεδεμένος με κάθε κόμβο του γειτονικού επιπέδου.

Ο αριθμός των κόμβων εισόδου καθορίζεται από την διάσταση (dimensionality) του πεδίου παρατήρησης που είναι υπεύθυνο για την δημιουργία των σημάτων εισόδου. Ο αριθμός των κόμβων εξόδου εξαρτάται από την διάσταση του επιθυμητού αποτελέσματος. Επομένως τρία είναι τα ζητήματα που ανακύπτουν κατά τον σχεδιασμό ενός τέτοιου δικτύου:

1. Ο προσδιορισμός του αριθμού των κρυφών επιπέδων
2. Ο προσδιορισμός του αριθμού των νευρώνων για κάθε κρυφό επίπεδο
3. Ο ακριβής προσδιορισμός των βαρών των συνάψεων που διασυνδέουν

τους νευρώνες σε κάθε επίπεδο του δικτύου.

Τα δυο πρώτα ζητήματα σχετίζονται με την πολυπλοκότητα (complexity) του δικτύου και αποτελούν το αδύνατο σημείο στον σχεδιασμό ενός πολυεπίπεδου perceptron. Όσον αφορά το τρίτο ζήτημα μπορεί να χρησιμοποιηθεί ο αλγόριθμος ανάδρομης διάδοσης σφάλματος, ο οποίος θα παρουσιαστεί σε επόμενη ενότητα της εργασίας.

2.11 Πολυπλοκότητα Δικτύου (Network Complexity)

Για την επίλυση πραγματικών προβλημάτων με πολυεπίπεδα perceptrons απαιτείται συνήθως η χρήση πολύπλοκων δικτύων. Έτσι το θέμα που ανακύπτει είναι η ελαχιστοποίηση του μεγέθους του δικτύου διατηρώντας όμως μια καλή απόδοση. Γενικότερα, ένα Νευρωνικό Δίκτυο με ελάχιστο μέγεθος είναι λιγότερο πιθανό να «μάθει» τις «ιδιαιτερότητες», τις «στατιστικές ιδιότητες» του θορύβου και των δεδομένων εκπαίδευσης (training data) και επομένως μπορεί πιο εύκολα να γενικεύσει σε νέα δεδομένα. Έτσι έχοντας πάρει μόνο τις απαραίτητες πληροφορίες από τα δεδομένα εκπαίδευσης, επιτυγχάνει μια «λογική » (reasonable) απόδοση και σε δεδομένα που προέρχονται από το ίδιο χώρο δεδομένων αλλά δεν έχουν εμφανιστεί κατά το παρελθόν.

Υπάρχουν δύο (2) μέθοδοι για τον σχεδιασμό του ελάχιστου δυνατού δικτύου:

- Ανάπτυξη Δικτύου (Network Growing), όπου ξεκινάμε με ένα δίκτυο μικρού μεγέθους για την εκπλήρωση του στόχου μας και προσθέτουμε συνεχώς νέους νευρώνες ή επίπεδα εκεί όπου το δίκτυο εμφανίζει αδυναμίες.
- Σμίκρυνση Δικτύου (Network Pruning), όπου ακολουθείται η αντίστροφη διαδικασία από αυτή της Ανάπτυξης Δικτύου. Δηλαδή ξεκινάμε με ένα υπερβολικά μεγάλο δίκτυο και κατά τη διαδικασία αφαιρούμε τα βάρη των συνάψεων που προσφέρουν λιγότερο στο δίκτυο.

Πρακτικά χρησιμοποιούνται και οι δυο μέθοδοι. Παρολαυτά πιο δημοφιλής είναι η Σμίκρυνση Δικτύου. Ο αντικειμενικός σκοπός της διαδικασίας αποκλεισμού βαρών (weight-eliminating procedure) είναι να βρεθεί το διάλυμα βάρους w το οποίο να ελαχιστοποιεί την ακόλουθη συνάρτηση κόστους:

$$R(w) = \varepsilon_s(w) + \lambda \varepsilon_c(w) \quad (2.28)$$

όπου $\varepsilon_s(w)$ είναι το μέτρο της συνηθισμένης επίδοσης του δικτύου, που εξαρτάται τόσο από το μοντέλο δικτύου όσο και από τα δεδομένα εισόδου και $\varepsilon_c(w)$ είναι το complexity penalty το οποίο εξαρτάται μόνο από το μοντέλο του δικτύου. Η εκτίμηση αυτού του μεγέθους εξαρτάται από τις συνάψεις του δικτύου. Στην διαδικασία αποκλεισμού βαρών, το complexity penalty ορίζεται από την ακόλουθη σχέση

$$\varepsilon_c(w) = \sum_{i \in I} \frac{(w_i / w_0)^2}{1 + (w_i / w_0)^2} \quad (2.29)$$

όπου w_0 είναι μια προκαθορισμένη ελεύθερη παράμετρος της διαδικασίας, και w_i αναφέρεται στο βάρος της σύναψης i . Το σύνολο I περιέχει όλες τις συνάψεις του δικτύου. Από την Σχέση (2.25), είναι εύκολο να δούμε ότι:

- αν $|w_i| \ll w_0$, το complexity penalty για το συγκεκριμένο βάρος τείνει στο μηδέν. Αυτός, ο περιορισμός εκφράζει πως όσον αφορά τη μάθηση από τα δεδομένα εκπαίδευσης, το i -οστό βάρος σύναψης δεν προσφέρει και επομένως μπορεί να απαλειφθεί.
- αν $|w_i| \gg w_0$, το complexity penalty τείνει στη μέγιστη τιμή του, γεγονός το οποίο σημαίνει πως w_i είναι σημαντικό.

Η παράμετρος λ παίζει το ρόλο του συντελεστή κανονικοποίησης. Όταν $\lambda=0$, το δίκτυο εξαρτάται αποκλειστικά από τα δεδομένα εκπαίδευσης. Όταν λ τείνει στο άπειρο, τα

δεδομένα εκπαίδευσης είναι αναξιόπιστα και θα πρέπει να αγνοηθούν. Σε πρακτικές εφαρμογές το $\lambda \in (0, \infty)$ [2].

2.12 Θεώρημα Καθολικής Προσέγγισης

Ένα πολυεπίπεδο Perceptron εκπαιδευμένο με ανάδρομη διάδοση σφάλματος αποτελεί ένα ισχυρό εργαλείο στην απεικόνιση εισόδου – εξόδου μιας μη γραμμικής διαδικασίας. Το ερώτημα εδώ είναι ποιος θα πρέπει να είναι ο αριθμός των κρυφών επιπέδων που χρειάζονται έτσι ώστε να απεικονισθεί καλύτερα η διαδικασία στο δίκτυο. Την απάντηση έρχεται να δώσει το Θεώρημα Καθολικής Προσέγγισης, στο οποίο έχουμε ήδη αναφερθεί και το οποίο αναπτύχθηκε από τους Cybenko, Funahashi και Hornik [16], [17], [18]. Το θεώρημα αυτό μπορεί να εκφραστεί ως ακολούθως:

Θεώρημα: Έστω f μια συνεχής, φραγμένη και γνησίως αύξουσα συνάρτηση. Αν I_{N_0} παριστά ένα υπέρ επίπεδο N_0 διαστάσεων και $C(I_{N_0})$ το πεδίο ορισμού των συνεχών συναρτήσεων στο I_{N_0} τότε, για οποιαδήποτε συνάρτηση $f \in C(I_{N_0})$ και $\varepsilon > 0$, υπάρχει ακέραιος N_1 και πραγματικές σταθερές a_i, b_i και w_{ip} , $i=1,2,\dots, N_1$ και $p=1,2,\dots, N_0$ τέτοια ώστε η συνάρτηση

$$F(x_1, x_2, \dots, x_{N_0}) = \sum_{i=1}^{N_1} a_i \phi \left(\sum_{p=1}^{N_0} w_{ip} x_p + b_i \right) \quad (2.30)$$

αποτελεί μια προσέγγιση της συνάρτησης $f(\cdot)$, και ικανοποιεί την ακόλουθη συνθήκη:

$$\left| F(x_1, x_2, \dots, x_{N_0}) - f(x_1, x_2, \dots, x_{N_0}) \right| < \varepsilon \quad (2.31)$$

για κάθε $\{x_1, x_2, \dots, x_{N_0}\} \in I_{N_0}$

Το παραπάνω θεώρημα εφαρμόζεται άμεσα στην περίπτωση ενός πολυεπίπεδου Perceptron ως εξής :

- ένα επίπεδο εισόδου με N_0 κόμβους των οποίων οι ανεξάρτητες εισοδοί συμβολίζονται με x_1, x_2, \dots, x_{N_0}

- ένα κρυφό επίπεδο με N_1 νευρώνες (neurons), $w_{i1}, w_{i2}, \dots, w_{iN_0}$ και

- ένα επίπεδο εξόδου που αποτελείται από ένα μόνο γραμμικό νευρώνα.

Αυτό που μας προσφέρει το θεώρημα είναι η μαθηματική απόδειξη για την προσέγγιση μιας αυθαίρετης συνεχούς συνάρτησης αντί για την ακριβή απεικόνιση της. Έτσι, το θεώρημα αυτό δηλώνει πως ένα μόνο κρυφό επίπεδο είναι αρκετό σε ένα πολυεπίπεδο Perceptron για να κάνει μια ενιαία προσέγγιση σε ένα δεδομένο σύνολο εκπαίδευσης με είσοδο x_1, x_2, \dots, x_{N_0} και επιθυμητή έξοδο $f(x_1, x_2, \dots, x_{N_0})$. Γι' αυτό το λόγο το Θεώρημα Καθολικής Προσέγγισης είναι εξαιρετικής σημασίας. Ωστόσο θα πρέπει να σημειώσουμε στο σημείο αυτό ότι δεν έχει αποδειχθεί πως ένα μόνο κρυφό επίπεδο είναι το βέλτιστο σε σχέση με το χρόνο μάθησης (learning time) και την ευκολία υλοποίησης (implementation) του δικτύου.

Πρακτικά, το πρόβλημα με τα πολυεπίπεδα perceptrons που έχουν μόνο ένα κρυφό επίπεδο είναι ότι οι νευρώνες του επιπέδου αυτού έχουν την τάση να αλληλεπιδρούν μεταξύ τους με αποτέλεσμα να οδηγείται το δίκτυο σε πολύπλοκες καταστάσεις. Η αλληλεπίδραση αυτή πολλές φορές δεν επιτρέπει την βελτίωση μιας προσέγγισης σε κάποιο σημείο χωρίς να προκαλεί το αντίθετο αποτέλεσμα σε κάποιο άλλο σημείο. Από την άλλη τα πολυεπίπεδα perceptrons με περισσότερα κρυφά επίπεδα έχουν καλύτερη συμπεριφορά στην διαδικασία προσέγγισης (curve fitting). Γι' αυτό το λόγο σε προβλήματα μεγάλης κλίμακας συνίσταται η χρήση περισσότερων κρυφών επιπέδων.

2.13 Ο Γενικευμένος Κανόνας Δέλτα

Απ' τη στιγμή που χρησιμοποιούμε μη-γραμμικές συναρτήσεις ενεργοποίησης, θα πρέπει να γενικεύσουμε τον Κανόνα Δέλτα έτσι ώστε να ισχύει και γι' αυτές. Η συνάρτηση ενεργοποίησης F είναι μια παραγωγίσιμη συνάρτηση της συνολικής εισόδου η οποία ορίζεται από τη σχέση:

$$y_k^p = F(s_k^p) \quad (2.32)$$

όπου

$$s_k^p = \sum_j w_j y_j^p + \theta_k \quad (2.33)$$

Για να εξαχθεί η σωστή γενίκευση του Κανόνα Δέλτα θα πρέπει να ορίσουμε τις ακόλουθες ποσότητες:

$$E^p = \frac{1}{2} \sum_{o=1}^{N_o} (d_o^p - y_o^p)^2 \quad (2.34)$$

$$\Delta_p w_{jk} = -\gamma \frac{\partial E^p}{\partial w_{jk}} \quad (2.35)$$

Το σφάλμα ορίζεται ως

$$(2.36)$$

$$E = \sum_p E^p$$

Έτσι μπορούμε να γράψουμε :

$$\frac{\partial E^p}{\partial w_{jk}} = \frac{\partial E^p}{\partial s_k^p} \frac{\partial s_k^p}{\partial w_{jk}} \quad (2.37)$$

από την Σχέση (2.29) παρατηρούμε ότι

$$\frac{\partial s_k^p}{\partial w_{jk}} = y_j^p \quad (2.38)$$

Ορίζοντας τώρα την ποσότητα

$$\delta_k^p = -\frac{\partial E^p}{\partial s_k^p} \quad (2.39)$$

Καταλήγουμε σ' ένα κανόνα που είναι όμοιος με τον Κανόνα Δέλτα που έχουμε περιγράψει σε προηγούμενη ενότητα (Σχέση (2.32)):

$$\Delta_p w_{jk} = \gamma \delta_k^p y_j^p \quad (2.40)$$

Θα πρέπει να κατανοήσουμε τι πρεσβεύει το δ_k^p για την κάθε μονάδα k του δικτύου. Το ενδιαφέρον συμπέρασμα το οποίο εξάγεται είναι ότι υπάρχει ένας απλός αναδρομικός υπολογισμός αυτών των ποσοτήτων ο οποίος εκτελείται με ανάδρομη μετάδοση (back propagation) των σφαλμάτων στο δίκτυο.

Για να υπολογίσουμε το δ_k^p υπολογίζουμε τις μερικές παραγωγούς ως προϊόν δύο παραγόντων. Ο ένας αντικατοπτρίζει την αλλαγή του σφάλματος ως συνάρτηση της εξόδου και ο άλλος την αλλαγή στην έξοδο ως συνάρτηση των αλλαγών στην είσοδο.

Έτσι έχουμε :

$$\delta_k^p = -\frac{\partial E^p}{\partial s_k^p} = -\frac{\partial E^p}{\partial y_k^p} \frac{\partial y_k^p}{\partial s_k^p} \quad (2.41)$$

Χρησιμοποιώντας την Σχέση (2.39) ο δεύτερος παράγοντας δίνεται από την σχέση:

$$\frac{\partial y_k^p}{\partial s_k^p} = F'(s_k^p) \quad (2.42)$$

όπου $F'(s_k^p)$ είναι η παράγωγος της $F(\cdot)$. Για να υπολογίσουμε τον πρώτο παράγοντα, υποθέτουμε πρώτα πως η μονάδα k είναι μονάδα εξόδου ($k=0$) του δικτύου. Έτσι έχουμε από τον ορισμό της Σχέσης (2.42) του σφάλματος:

$$\frac{\partial E^p}{\partial y_o^p} = -(d_o^p - y_o^p) \quad (2.43)$$

και χρησιμοποιώντας τις Σχέσεις (2.41) και (2.42) τελικά έχουμε:

$$\delta_o^p = (d_o^p - y_o^p) F'_o(s_o^p) \quad (2.44)$$

Αν η μονάδα k δεν είναι έξοδος αλλά μονάδα του κρυφού επιπέδου ($k=h$) μπορούμε να γράψουμε το σφάλμα ως συνάρτηση των εισόδων από το κρυφό στο επίπεδο εξόδου, δηλαδή

$$E^p = E^p(s_1^p, s_2^p, \dots, s_j^p, \dots) \quad (2.45)$$

και έτσι έχουμε ότι:

$$\frac{\partial E^p}{\partial y_h^p} = \sum_{o=1}^{N_o} \frac{\partial E^p}{\partial s_o^p} \frac{\partial s_o^p}{\partial y_h^p} = \sum_{o=1}^{N_o} \frac{\partial E^p}{\partial s_o^p} \frac{\partial}{\partial y_h^p} \sum_{j=1}^{N_h} w_{ko} y_j^p = \sum_{o=1}^{N_o} \frac{\partial E^p}{\partial s_o^p} w_{ho} = - \sum_{o=1}^{N_o} \delta_o^p w_{ho} \quad (2.46)$$

Συνδιάζοντας τις Σχέσεις (2.39) και (2.35) έχουμε:

$$\delta_h^p = F'(s_h^p) \sum_{o=1}^{N_o} \delta_o^p w_{ho} \quad (2.47)$$

Οι Εξισώσεις (2.43) και (2.47) δίνουν μια αναδρομική διαδικασία για τον υπολογισμό των αναγκαίων ποσοτήτων που χρειαζόμαστε για την ενημέρωση των βαρών σύμφωνα με την Σχέση (2.40). Αυτή η διαδικασία αποτελεί τον γενικευμένο Κανόνα Δέλτα για ένα "προς τα εμπρός" δίκτυο με μη γραμμικές μονάδες.

2.14 Κατανόηση της Ανάδρομης Διάδοσης Σφάλματος

Όλες οι παραπάνω εξισώσεις μπορεί να είναι μαθηματικά ορθές αλλά τι συμβολίζουν στην πραγματικότητα ; Η όλη διαδικασία της Ανάδρομης Διάδοσης Σφάλματος είναι αρκετά ξεκάθαρη. Αυτό που με απλό τρόπο εκφράζουν οι παραπάνω εξισώσεις είναι το εξής: Αφού επιλέξουμε ένα τρόπο μάθησης η είσοδος μέσω της συνάρτησης ενεργοποίησης οδηγείται στην έξοδο. Η έξοδος συγκρίνεται με την επιθυμητή έξοδο που έχουμε θέσει και καταλήγουμε σε μια διάφορα πραγματικής και επιθυμητής εξόδου που

αποτελεί το σφάλμα. Αν αποκαλέσουμε αυτό το σφάλμα μίας συγκεκριμένης μονάδας εξόδου e_o . Θα πρέπει αν όχι να το μηδενίσουμε να το ελαχιστοποιήσουμε. Αυτό επιτυγχάνεται με την αλλαγή των συνδέσεων στο δίκτυο με τέτοιο τρόπο ώστε να μηδενιστεί το σφάλμα. Γνωρίζουμε από τον Κανόνα Δέλτα πως για να μειώσουμε το σφάλμα θα πρέπει να προσαρμόσουμε τα βάρη σύμφωνα με την ακόλουθη σχέση:

$$\Delta w_{ho} = (d_o - y_o) y_h \quad (2.48)$$

Αυτό από μόνο του όμως δεν είναι αρκετό γιατί με αυτό τον τρόπο δεν αλλάζουν τα βάρη από τις μονάδες εισόδου στις κρυφές μονάδες και έτσι δεν επιτυγχάνεται η μέγιστη δυνατότητα απεικόνισης / αναπαράστασης που έχει ένα “προς τα εμπρός” δίκτυο, σύμφωνα με το Θεώρημα Καθολικής Προσέγγισης. Για να προσαρμόσουμε τα βάρη από τις μονάδες εισόδου στις κρυφές μονάδες πρέπει να εφαρμοστεί ξανά ο Κανόνας Δέλτα. Σ’ αυτή την περίπτωση όμως δεν έχουμε το δ για τις κρυφές μονάδες. Αυτό επιλύεται με διανομή του σφάλματος κάθε μονάδας εξόδου με την οποία είναι συνδεδεμένο, σταθμισμένο από τη συγκεκριμένη σύνδεση. Έτσι μια κρυφή μονάδα h λαμβάνει ένα δ από κάθε μονάδα εξόδου o ίσο με το δ της σύνδεσης αυτών των δυο μονάδων h και o . Τέλος θα πρέπει να συμπεριληφθεί και η συνάρτηση ενεργοποίησης $F(\cdot)$ πριν συνεχιστεί η διαδικασία ανάδρομης διάδοσης.

2.15 Δουλεύοντας με τον Back-Propagation

Η εφαρμογή του γενικευμένου Κανόνα Δέλτα περιλαμβάνει δύο (2) φάσεις. Κατά την πρώτη φάση η είσοδος x περνάει μέσα απ’ το δίκτυο για να υπολογιστεί η έξοδος. Η έξοδος αυτή συγκρίνεται με την επιθυμητή και καταλήγουμε σε μια διαφορά, το σφάλμα για κάθε μονάδα εξόδου. Η δεύτερη φάση περιλαμβάνει ένα πέρασμα προς τα πίσω από το δίκτυο για να υπολογιστούν οι αλλαγές των βαρών.

2.16 Προσαρμογή των βαρών με σιγμοειδή συνάρτηση ενεργοποίησης

Σύμφωνα με τα παραπάνω καταλήγουμε σε τρεις (3) εξισώσεις :

• Το βάρος μιας σύνδεσης προσαρμόζεται αναλογικά με το σφάλμα δ της μονάδας k το οποίο λαμβάνει την είσοδο και την έξοδο της μονάδας j στέλνοντας αυτό το σήμα στη σύνδεση :

$$\Delta_p w_{jk} = \gamma \delta_k^p y_j^p \quad (2.49)$$

• Αν η μονάδα είναι μονάδα εξόδου τότε το σφάλμα δίδεται από :

$$\delta_o^p = (d_o^p - y_o^p) F'(s_o^p) \quad (2.50)$$

Αν $F(\cdot)$ είναι η σιγμοειδής συνάρτηση :

$$y^p = F(s^p) = \frac{1}{1 + e^{-s^p}} \quad (2.51)$$

Τότε

$$\begin{aligned} F'(s^p) &= \frac{d}{ds^p} \frac{1}{1 + e^{-s^p}} \\ F'(s^p) &= \frac{1}{(1 + e^{-s^p})^2} (-e^{-s^p}) \\ F'(s^p) &= \frac{1}{(1 + e^{-s^p})} \frac{e^{-s^p}}{(1 + e^{-s^p})} \\ F'(s^p) &= y^p (1 - y^p) \end{aligned} \quad (2.52)$$

Άρα το σφάλμα να μπορεί να γραφτεί :

$$\delta_o^p = (d_o^p - y_o^p) y_o^p (1 - y_o^p) \quad (2.53)$$

• Το σφάλμα για μια κρυφή μονάδα (hidden unit) υπολογίζεται αναδρομικά. Για σιγμοειδή συνάρτηση ενεργοποίησης έχουμε :

$$\delta_o^p \delta_h^p = F'(s_h^p) \sum_{o=1}^{N_o} \delta_o^p w_{ho} = y_h^p (1 - y_h^p) \sum_{o=1}^{N_o} \delta_o^p w_{ho} = (d_o^p - y_o^p) y_o^p (1 - y_o^p) \quad (2.54)$$

2.17 Ρυθμός Μάθησης και Συντελεστής Ορμής

Η διαδικασία μάθησης απαιτεί η αλλαγή στο βάρος w να είναι ανάλογη της $\frac{\partial E^p}{\partial w}$ και τα “βήματα” να είναι πολύ μικρά. Η σταθερά της αναλογικότητας είναι αυτό που ονομάζεται ρυθμός μάθησης γ . Για πρακτικούς λόγους επιλέγουμε ένα ρυθμό

μάθησης ο οποίος να είναι όσο γίνεται μεγαλύτερος χωρίς όμως να καταλήγουμε σε ταλάντωση. Ένας τρόπος να αποφευχθεί αυτό είναι να κάνουμε την αλλαγή των βαρών να εξαρτάται από τις παρελθούσες αλλαγές των βαρών προσθέτοντας ένα συντελεστή ορμής:

$$\Delta w_{jk}(t+1) = \gamma \delta_k^j y_j^p + a \Delta w_{jk}(t) \quad (2.55)$$

όπου a ο Συντελεστής Ορμής.

2.17.1 Ρυθμός Μάθησης (Learning Rate)

Πιο συγκεκριμένα η παράμετρος του ρυθμού μάθησης γ είναι μια σταθερά περιορισμού ή εξομάλυνσης η οποία τίθεται από τον χρήστη έτσι ώστε οι αλλαγές στα βάρη Δw να μην είναι τόσο αργές ή απότομες. Η σταθερά αυτή παίρνει τιμές στο διάστημα $(0, 1)$. Είναι σαφές πως αυτή η διαδικασία προσαρμογής είναι ανάλογη της διαδικασίας της απλής εκθετικής εξομάλυνσης, όπου η τιμή του βάρους είναι ο σταθμισμένος κινητός μέσος των παρελθόντων βαρών. Αν το γ είναι πολύ μεγάλο τότε τα βάρη θα ταλαντώνονται γύρω από την βέλτιστη τιμή και η διαδικασία θα αποσταθεροποιηθεί. Όταν το γ είναι πολύ μικρό τότε θα πάρει πολύ περισσότερο χρόνο για να εκπαιδευτεί το δίκτυο.

2.17.2 Συντελεστής Ορμής

Ένα από τα προβλήματα που εμφανίζεται κατά τη διαδικασία εύρεσης του ελάχιστου σφάλματος είναι ότι τα τοπικά ελάχιστα μπορεί να αποσυντονίσουν την διαδικασία εύρεσης του ολικού ελάχιστου. Όταν πλησιάζουμε ένα ελάχιστο, η αλλαγή στα βάρη ελαττώνεται και μπορεί, λανθασμένα, να θεωρήσουμε ένα τοπικό ελάχιστο ως ολικό. Για να αποφευχθεί κάτι τέτοιο ενσωματώνεται ο Συντελεστής Ορμής στον Κανόνα Δέλτα. Το σκεπτικό της χρησιμοποίησης του Συντελεστή Ορμής είναι να συνεχιστούν οι σχετικά μεγάλες αλλαγές στα βάρη παρά την τάση μείωσης που μπορεί να υπάρχει, εξαιτίας ενός τοπικού ελάχιστου. Ωστόσο, αν ο Συντελεστής Ορμής είναι πολύ μεγάλος τότε υπάρχει η πιθανότητα να παραβλεφθεί το βέλτιστο σημείο. Επίσης ίσως να χρειαστούν κάποιες αλλαγές και προσαρμογές στο Συντελεστή Ορμής κατά τη διαδικασία της εκπαίδευσης. Δυστυχώς δεν υπάρχει κανένας καθολικός κανόνας που να καθορίζει βέλτιστα την τιμή του a . Ο μόνος τρόπος να βρούμε την κατάλληλη τιμή είναι μέσω δοκιμών.

2.18 Παράγοντες που επηρεάζουν την απόδοση ενός Πολυεπίπεδου “προς τα εμπρός” Δικτύου

1. Ο αλγόριθμος μάθησης και ο αριθμός των επαναλήψεων. Αυτός ο παράγοντας προσδιορίζει το πόσο σωστά έχει ελαχιστοποιηθεί το σφάλμα.

2. Ο αριθμός των δειγμάτων. Αυτός ο παράγοντας προσδιορίζει το κατά πόσο τα δείγματα εκπαίδευσης αναπαριστούν την πραγματική συνάρτηση. Όσο αυξάνεται ο αριθμός του δείγματος μάθησης (learning sample) τόσο το σύνολο δοκιμών (test set) όσο και το σύνολο μάθησης (learning set) προσεγγίζει το ολικό ελάχιστο.

3. Ο αριθμός των κρυφών μονάδων. Αυτό προσδιορίζει την “εκφραστική δύναμη” του συστήματος. Για ομαλές συναρτήσεις αρκούν ελάχιστες κρυφές μονάδες. Στην αντίθετη περίπτωση απαιτείται μεγάλος αριθμός κρυφών μονάδων. Εδώ όσο αυξάνεται ο αριθμός των κρυφών μονάδων το σύνολο δοκιμών (test set) προσεγγίζει ένα μικρό σφάλμα, αλλά αν συνεχίσει πέρα από κάποιο σημείο να αυξάνεται, τότε αυξάνεται και το σφάλμα. Αντίθετα όσο αυξάνεται ο αριθμός των κρυφών μονάδων τόσο μειώνεται το σφάλμα στο σύνολο μάθησης (learning set).

Κεφάλαιο 3

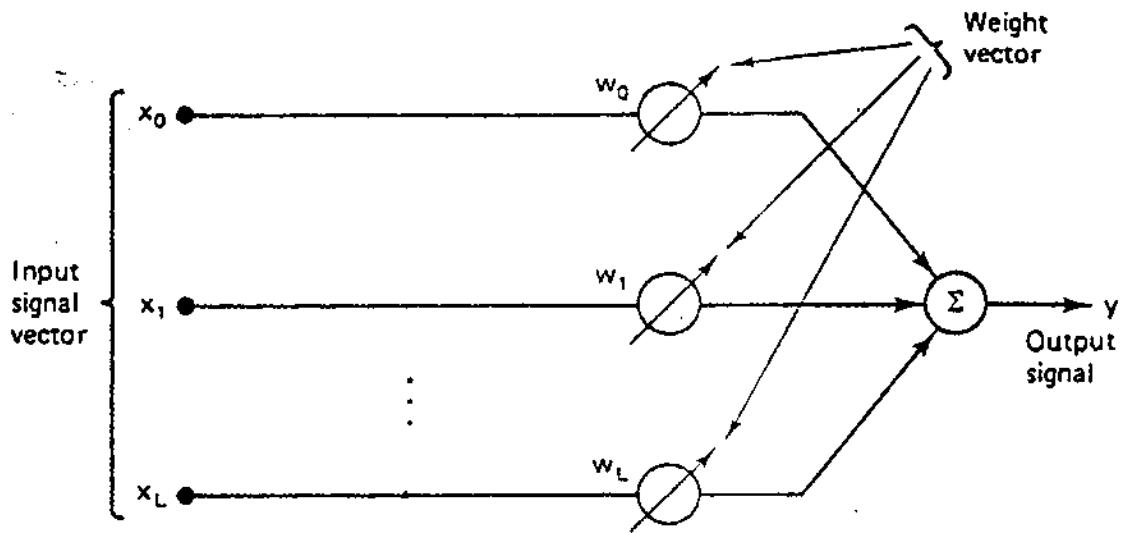
Αλγόριθμοι Διερεύνησης της Επιφάνειας Σφάλματος

3.1 Εισαγωγή

Σε αυτό το κεφάλαιο θα μελετηθούν κάποιοι αλγόριθμοι διερεύνησης της επιφάνειας σφάλματος. Πριν προχωρήσουμε όμως σ' αυτό θα πρέπει να κατανοηθεί τι είναι η επιφάνεια σφάλματος, ποιές είναι οι βασικές της ιδιότητες και ποιες προϋποθέσεις θα πρέπει να πληρούνται ώστε να είναι δυνατή η επεξεργασία μίας στοχαστικής διαδικασίας. Η επιφάνεια σφάλματος δεν είναι τίποτα παραπάνω από την απεικόνιση της συνάρτησης κόστους, που έχουμε ήδη δει, σε ένα σύστημα αξόνων. Βασική προϋπόθεση για να είναι δυνατή η επεξεργασία μίας στοχαστικής διαδικασίας είναι η στασιμότητα και η εργοδικότητα, οι οποίες όμως αναφέρονται σε επόμενη παράγραφο. Πριν από οτιδήποτε όμως θα πρέπει να αναφερθούμε στον Προσαρμοζόμενο Γραμμικό Συνδυαστή, ο οποίος είναι απαραίτητος για την ανάλυση στην οποία θα προχωρήσουμε [1].

3.2 Προσαρμοζόμενος Γραμμικός Συνδυαστής

Ο Προσαρμοζόμενος Γραμμικός Συνδυαστής (Adaptive linear combiner ή non-recursive adaptive filter) είναι θεμελιώδης στην προσαρμοστική επεξεργασία σημάτων (adaptive signal processing). Εμφανίζεται σε διάφορες μορφές στα περισσότερα προσαρμοζόμενα φίλτρα και συστήματα και είναι το πιο σημαντικό στοιχείο στα συστήματα που έχουν την δυνατότητα μάθησης και τις προσαρμοζόμενες διαδικασίες (adaptive processes) γενικότερα. Εξαιτίας της μη αναδρομικής δομής του ο Προσαρμοζόμενος Γραμμικός Συνδυαστής είναι πολύ απλός και εύκολος να κατανοηθεί και να αναλυθεί. Τόσο η συμπεριφορά του όσο και η τεκμηρίωση του στις διάφορες μορφές του μπορεί εύκολα να κατανοηθεί.



Σχήμα 3.1: Γενική μορφή του Adaptive Linear Combiner

Στο Σχήμα 3.1 φαίνεται η γενική μορφή του Προσαρμοζόμενου Γραμμικού Συνδυαστή. Όπως φαίνεται αποτελείται από ένα ένα σήμα εισόδου με στοιχεία X_0, X_1, \dots, X_L από μια ομάδα βαρών στάθμισης w_0, w_1, \dots, w_L , από μια αθροιστική μονάδα και ένα σήμα εξόδου y . Ο συνδυαστής αυτός αποκαλείται γραμμικός γιατί για μια δεδομένη ομάδα βαρών στάθμισης, το αποτέλεσμα (η έξοδος του) είναι ο γραμμικός συνδυασμός των στοιχείων της εισόδου.

3.2.1 Σήμα Εισόδου και Διάνυσμα Βαρών

Δυο είναι οι τρόποι με τους οποίους μπορεί να γίνει η φυσική απόδοση του σήματος εισόδου. Ο πρώτος τρόπος είναι να θεωρήσουμε ότι τα δείγματα της εισόδου προέρχονται από $L+1$ διαφορετικές πηγές σήματος. Αντίθετα, στο δεύτερο τρόπο τα δείγματα του σήματος μπαίνουν με ακολουθιακό τρόπο στο σύστημα, και προέρχονται από την ίδια πηγή σήματος. Η πρώτη περίπτωση αναφέρεται σαν περίπτωση πολλαπλών εισόδων, ενώ η δεύτερη σαν περίπτωση μοναδικής εισόδου. Και συμβολίζονται ως εξής:

$$\text{Περίπτωση Πολλαπλών Εισόδων: } X_k = [x_{0k} \quad x_{1k} \quad \dots \quad x_{Lk}]^T \quad (3.1)$$

$$\text{Περίπτωση Μοναδικής Εισόδου: } X_k = [x_k \quad x_{k-1} \quad \dots \quad x_{k-L}]^T \quad (3.2)$$

όπου το X' συμβολίζει τον ανάστροφο πίνακα του X και ο δείκτης k χρησιμοποιείται ως δείκτης χρόνου. Έτσι βλέπουμε ότι στην περίπτωση πολλαπλών εισόδων το k είναι σταθερό, δηλώνοντας ότι πρόκειται για δείγματα που φθάνουν ταυτόχρονα στην είσοδο του συστήματος. Αντίθετα στην περίπτωση μοναδικής εισόδου το διάνυσμα X_k περιέχει δείγματα εισόδου από την χρονική στιγμή $k-L$ έως και την χρονική στιγμή k . Στην περίπτωση της μιας εισόδου πρέπει να παρεμβάλουμε στοιχεία καθυστέρησης (unit delay elements). Η μορφή του συστήματος που προκύπτει είναι γνωστή σαν *adaptive transversal system*. Παρατηρούμε ότι στα βάρη έχει προστεθεί ένας δείκτης k έτσι ώστε να γίνεται σαφές ότι η διαδικασία είναι χρονομεταβαλλόμενη (time varying). Αυτή η μορφή το *adaptive transversal filter* είναι η *temporal* μορφή και έχει ένα μεγάλο εύρος εφαρμογών. Στα συστήματα πολλαπλών εισόδων, σε κάποιες περιπτώσεις μπορούμε να συμπεριλάβουμε μια επιπλέον είσοδο στην οποία συνήθως οδηγούμε μια τιμή κατωφλιού (bias). Χρησιμοποιώντας τα παραπάνω μπορούμε να εκφράσουμε τις εξόδους των δυο παραπάνω συστημάτων ως ακολούθως:

$$\text{Περίπτωση Πολλαπλών Εισόδων: } y_k = \sum_{l=0}^L w_{lk} x_{lk} \quad (3.3)$$

$$\text{Περίπτωση Μοναδικής Εισόδου: } y_k = \sum_{l=0}^L w_{lk} x_{k-l} \quad (3.4)$$

Ορίζοντας το διάνυσμα βαρών W_k με την ακόλουθη σχέση

$$W_k = [w_{0k} \quad w_{1k} \quad \dots \quad w_{Lk}]^T \quad (3.5)$$

Οι Σχέσεις (3.3) και (3.4) μπορούν να γραφτούν ισοδύναμα ως εξής:

$$y_k = x_k^T W_k = W_k^T x_k$$

Με βάση αυτή τη γενική περιγραφή της λειτουργίας του Προσαρμοζόμενου Γραμμικού Συνδυαστή μπορούμε να προχωρήσουμε στο πώς το σύστημα προσαρμόζει τα βάρη του καθώς το k αυξάνεται.

3.2.1.1 Στασιμότητα

Πριν προχωρήσουμε όμως θα πρέπει να ορίσουμε δύο βασικές έννοιες που αφορούν τις στοχαστικές διαδικασίες, την Στασιμότητα και την Εργοδικότητα. Πριν προχωρήσουμε όμως στην ανάλυση αυτή θα πρέπει να αναφερθεί πώς ο όρος στοχαστική διαδικασία

περιγράφει την εξέλιξη ενός στατιστικού φαινομένου στο χρόνο. Παραδείγματα τέτοιων διαδικασιών είναι σήματα ομιλίας, τηλεόρασης, ραντάρ, δεδομένα ψηφιακών υπολογιστών, σεισμολογικά δεδομένα, οικονομικά δεδομένα και θόρυβος. Η μορφή της στοχαστικής διαδικασίας που μας ενδιαφέρει είναι αυτή που ορίζεται σε διακριτές και ομοιόμορφα κατανεμημένες χρονικές στιγμές. Όμως μια στοχαστική διαδικασία δεν είναι μόνο μια απλή συνάρτηση του χρόνου αλλά αναπαριστά ένα μη πεπερασμένο αριθμό διαφορετικών υλοποιήσεων της συγκεκριμένης διαδικασίας. Μια συγκεκριμένη υλοποίηση ονομάζεται χρονοσειρά. Για παράδειγμα η ακολουθία $x(n), x(n-1), \dots, x(n-m)$ αναπαριστά μια χρονοσειρά η οποία αποτελείται από την παρούσα παρατήρηση $x(n)$ της στιγμής n και m παρελθούσες παρατηρήσεις των στιγμών $n-1, \dots, n-m$. Για στοχαστικές διαδικασίες διακριτού χρόνου της παραπάνω μορφής είναι αυστηρά στάσιμες όταν η συνάρτηση πυκνότητας πιθανότητας παραμένει η ίδια, ότι τιμές και αν δώσουμε στο n για σταθερό m .

3.2.1.2 Μερικός Χαρακτηρισμός μιας Στοχαστικής Διαδικασίας Διακριτού Χρόνου

Πρακτικά δεν είναι δυνατό να καθοριστεί η συνάρτηση πυκνότητας πιθανότητας για αυθαίρετες παρατηρήσεις στοχαστικής διαδικασίας. Έτσι θα πρέπει να περιοριστούμε σε μερικό χαρακτηρισμό στις πρώτης και δεύτερης τάξης στατιστικές της διαδικασίας.

Έστω η διαδικασία $x(n), x(n-1), \dots, x(n-m)$

Ορίζουμε τη μέση τιμή,

$$\mu_x(n) = E[x(n)] \quad (3.7)$$

τη συνάρτηση αυτοσυσχέτισης ως εξής :

$$r(n, n-k) = E[(x(n)x(n-k))^*] \quad , \quad k = 0, \pm 1, \pm 2, \dots \quad (3.8)$$

όπου το x^* υποδηλώνει τον συζυγή του x και την συνδυακόμενη ως

$$c(n, n-k) = E[(x(n) - \mu_x(n))(x(n-k) - \mu_x(n))^*] \quad (3.9)$$

Από τις παραπάνω σχέσεις παρατηρούμε ότι

$$c(n, n-k) = r(n, n-k) - \mu_x(n)\mu_x^*(n-k) \quad (3.10)$$

η συνδιακόμενη ισούται με την αυτοσυσχέτιση όταν

$$\mu_x(n) = 0 \quad (3.11)$$

για κάθε n .

Έτσι βλέπουμε πως για το μερικό χαρακτηρισμό μιας διαδικασίας πρέπει να προσδιοριστούν 3 μέτρα, ο μέσος, η αυτοσυσχέτιση και η συνδιακύμανση.

Μία στοχαστική διαδικασία διακριτού χρόνου είναι στάσιμη με την ευρεία έννοια όταν

- ο μέσος της στοχαστικής διαδικασίας είναι ο μέσος της ακολουθίας

$$\mu_x(n) = \mu_x \quad (3.12)$$

- τόσο η αυτοσυσχέτιση όσο και η συνδιακύμανση εξαρτώνται μόνο από τη διαφορά στις στιγμές των παρατηρήσεων n και $n-k$

$$r(n, n-k) = r(k) \quad (3.13)$$

$$c(n, n-k) = c(k) \quad (3.14)$$

3.2.1.3 Εργοδικότητα

Οι χρονικοί μέσοι μπορούν να χρησιμοποιηθούν στο χτίσιμο ενός στοχαστικού μοντέλου μιας φυσικής διαδικασίας εκτιμώντας άγνωστες παραμέτρους του μοντέλου.

Πιο συγκεκριμένα, έστω μια στοχαστική διαδικασία $x(n)$ που είναι στάσιμη με την ευρεία έννοια. Έστω

$$\hat{\mu}(N) = \frac{1}{N} \sum_{n=0}^{N-1} x(n) \quad (3.15)$$

ο αριθμητικός μέσος της χρονοσειράς.

Θα θέλαμε να ξέρουμε κάτω από ποιές προϋποθέσεις ο αριθμητικός μέσος $\hat{\mu}(N)$ είναι ένας αμερόληπτος εκτιμητής του μέσου της διαδικασίας. Θα λέμε πως η $x(n)$ είναι mean ergodic υπό την έννοια του MSE όταν το MSE ανάμεσα στο $\hat{\mu}(N)$ και στο μ όταν το $N \rightarrow \infty$, τείνει στο μηδέν, δηλαδή

$$\lim_{N \rightarrow \infty} (\mu - \hat{\mu}(N))^2 = 0 \quad (3.16)$$

Από την Σχέση (3.15) μπορούμε να γράψουμε:

$$\begin{aligned}
 E[(\mu - \hat{\mu}(N))^2] &= E\left[\left(\mu - \frac{1}{N} \sum_{n=0}^{N-1} x(n)\right)^2\right] \\
 E[(\mu - \hat{\mu}(N))^2] &= \frac{1}{N^2} E\left[\sum_{n=0}^{N-1} (x(n) - \mu)^2\right] \\
 E[(\mu - \hat{\mu}(N))^2] &= \frac{1}{N^2} E\left[\sum_{n=0}^{N-1} \sum_{k=0}^{N-1} (x(n) - \mu)(x(k) - \mu)^*\right] \\
 E[(\mu - \hat{\mu}(N))^2] &= \frac{1}{N^2} \sum_{n=0}^{N-1} \sum_{k=0}^{N-1} c(n-k)
 \end{aligned} \tag{3.18}$$

Αν $l=n-k$ τότε:

$$E[(\mu - \hat{\mu}(N))^2] = \frac{1}{N} \sum_{l=N+1}^{N-1} \left(1 - \frac{l}{N}\right) c(l) \tag{3.19}$$

Αναγκαία και ικανή συνθήκη για να είναι η $x(n)$ mean ergodic υπό την έννοια του MSE είναι

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{l=N+1}^{N-1} \left(1 - \frac{l}{N}\right) c(l) = 0 \tag{3.20}$$

Με αλλά λόγια αν η διαδικασία $x(n)$ είναι ασυμπτωτικά ασυσχέτιστη, υπό την έννοια της παραπάνω σχέσης, τότε συνάγεται ότι ο χρονικός μέσος $\hat{\mu}(N)$ συγκλίνει στο μ .

Η χρήση του θεωρήματος μέσης Εργοδικότητας μπορεί να επεκταθεί και σε άλλους χρονικούς μέσους της διαδικασίας, όπως τη συνάρτηση αυτοσυσχέτισης μιας στάσιμης διαδικασίας υπό την ευρεία έννοια.

$$\hat{r}(k, N) = \frac{1}{N} \sum_{n=0}^{N-1} x(n)x(n-k) \tag{3.21}$$

Η διαδικασία $x(n)$ είναι correlation ergodic όταν η μέση τετραγωνική τιμή της διαφοράς του $\hat{r}(k, N)$ και $r(k)$ τείνει στο μηδέν όταν $N \rightarrow \infty$. Η παραπάνω πρόταση μπορεί να αποδειχθεί εύκολα αν χρησιμοποιηθεί το θεώρημα εργοδικότητας του μέσου. Γι' αυτό το σκοπό έχουμε μια νέα διαδικασία $z(n, k)$ που ορίζεται από τη σχέση

$$z(n, k) = x(n)x(n - k) \quad (3.22)$$

Τότε η αναγκαία και ικανή συνθήκη για εργοδικότητα του μέσου της $z(n)$ οδηγεί στην εργοδικότητα της συσχέτισης.

3.2.2 Επιθυμητή Απόκριση και Ακολουθία Σφαλμάτων

Κατά την διαδικασία της προσαρμογής τα βάρη μεταβάλλονται κατά τέτοιο τρόπο ώστε η έξοδος του Προσαρμοζόμενου Γραμμικού Συνδυαστή να συμφωνεί όσο το δυνατόν περισσότερο με την επιθυμητή έξοδο. Για το σκοπό αυτό θα ορίσουμε την ακολουθία σφαλμάτων και θα μεταβάλλουμε τα βάρη του συστήματος έτσι ώστε να ελαχιστοποιήσουμε μια συνάρτηση κόστους της παραπάνω ακολουθίας.

Έτσι έχουμε

$$\begin{aligned} \varepsilon_k &= d_k - y_k \\ \varepsilon_k &= d_k - X_k^T W = d_k - W^T X_k \end{aligned} \quad (3.23)$$

Εδώ παραλείπεται ο δείκτης k από το W για ευκολία αφού σ' αυτό το σημείο δεν επιθυμούμε να προσαρμόσουμε τα βάρη.

Υποθέτοντας ότι τα ε_k , d_k και X_k είναι στάσιμα, μια κατάλληλη συνάρτηση κόστους είναι το μέσο τετραγωνικό σφάλμα που ορίζεται από τη σχέση:

$$E[\varepsilon_k^2] = E[d_k^2] + W^T E[X_k \cdot X_k^T] W - 2E[d_k X_k^T] \quad (3.24)$$

Η συνάρτηση του μέσου τετραγωνικού σφάλματος ίσως θα μπορούσε να αποτυπωθεί καλύτερα με την βοήθεια του πίνακα αυτοσυσχέτισης του σήματος εισόδου:

$$R = E[X_k \cdot X_k^T] = E \begin{bmatrix} x_{0k}^2 & x_{0k}x_{1k} & x_{0k}x_{2k} & \dots & x_{0k}x_{Lk} \\ x_{1k}x_{0k} & x_{1k}^2 & x_{1k}x_{2k} & \dots & x_{1k}x_{Lk} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{Lk}x_{0k} & x_{Lk}x_{1k} & x_{Lk}x_{2k} & \dots & x_{Lk}^2 \end{bmatrix} \quad (3.25)$$

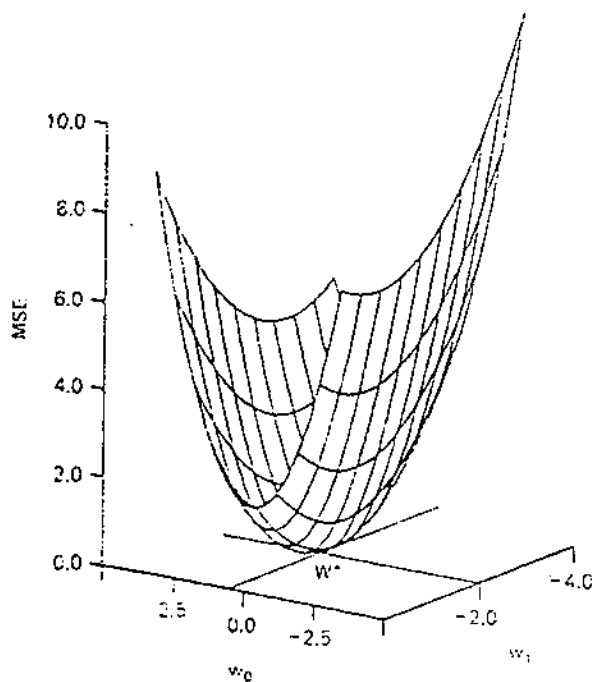
Έστω p είναι το διάνυσμα ετεροσυσχέτισης μεταξύ εισόδου και επιθυμητής απόκρισης του συστήματος, δηλαδή:

$$p = E[d_k X_k] = E[d_k x_{0k} \quad d_k x_{1k} \quad \dots \quad d_k x_{Lk}]^T \quad (3.26)$$

Λόγω στασιμότητας των X_k και d_k όλα τα στοιχεία των R και p έχουν σταθερές στατιστικές δεύτερης τάξης, και επομένως αν συμβολίσουμε με ξ το MSE τότε η Σχέση (3.24) γράφεται:

$$\xi = E[\varepsilon_k^2] = E[d_k^2] + W^T \cdot RW - 2p^T \cdot W \quad (3.27)$$

Είναι σαφές πως το ξ περιγράφει μια τετραγωνική συνάρτηση των βαρών W όταν τα X_k και d_k είναι στάσιμες στοχαστικές διαδικασίες. Η τυπική μορφή μιας τέτοιας συνάρτησης για την περίπτωση δυο βαρών $W = [w_1, w_2]$ του ελάχιστου τετραγωνικού σφάλματος απεικονίζεται στο Σχήμα 3.2. Ο κάθετος άξονας αναπαριστά το MSE και οι οριζόντιοι άξονες τις τιμές των δύο βαρών W . Το σημείο «στον πάτο του bowl» συμβολίζεται με W^* και είναι το σημείο στο οποίο ελαχιστοποιείται το σφάλμα. Είναι φανερό ότι με την μορφή της τετραγωνικής συνάρτησης κόστους εξασφαλίζεται ότι υπάρχει ένα ολικό ελάχιστο, δηλαδή δεν υπάρχουν τοπικά ελάχιστα.



Σχήμα 3.2: Γραφική αναπαράσταση της επιφάνειας σφάλματος στην περίπτωση των δύο βαρών

3.2.3 Ανάδελτα και Ελάχιστο MSE

Πολλές από τις προσαρμοζόμενες διαδικασίες που προσπαθούν να προσαρμόσουν τα βάρη αναζητώντας την ελάχιστη τιμή της επιφάνειας σφάλματος το πετυχαίνουν αυτό με ποσοτικές μεθόδους βασισμένες στον τελεστή ανάδελτα. Το ανάδελτα της επιφάνειας σφάλματος, που συμβολίζεται με $\nabla(\xi)$ δημιουργείται παραγωγίζοντας ως προς τα βάρη w_i τη (3.27). Έτσι έχουμε :

$$\nabla(\xi) = \left[\frac{\partial \xi}{\partial w_0} \quad \frac{\partial \xi}{\partial w_1} \quad \dots \quad \frac{\partial \xi}{\partial w_L} \right]^T = 2RW - 2p \quad (3.28)$$

Θέτοντας στην παραπάνω σχέση $\nabla(\xi) = 0$ καταλήγουμε στο:

$$2RW^* - 2p = 0 \quad (3.29)$$

$$W^* = R^{-1} \cdot p \quad (3.30)$$

Άρα το βέλτιστο W^* , που ονομάζεται Weiner διάνυσμα βαρών και το MSE δίνονται από τις Σχέσεις (3.30) και (3.31) και

$$\xi_{\min} = E[d_k^2] - P^T \cdot W^* \quad (3.31)$$

3.2.4 Αυτοσυσχέτιση του Σφάλματος και της Εισόδου

Είναι πολύ εύκολο να δούμε μια πολύ ενδιαφέρουσα και χρήσιμη σχέση που υπάρχει ανάμεσα στην ακολουθία σφαλμάτων ε_k και την είσοδο X_k όταν $W=W^*$, όταν δηλαδή έχουμε πετύχει το ελάχιστο σφάλμα. Πράγματι, πολλαπλασιάζοντας και τα δυο μέλη της Σχέσης (3.23) με το διάνυσμα X_k και παίρνοντας την αναμενόμενη τιμή έχουμε την ακόλουθη σχέση:

$$E[\varepsilon_k x_k] = p - RW \quad (3.32)$$

Και επομένως όταν $W = W^*$

$$E[\varepsilon_k x_k]_{W=W^*} = 0 \quad (3.33)$$

Αυτό το αποτέλεσμα είναι γνωστό αποτέλεσμα της θεωρίας των φίλτρων Wiener (Wiener filter theory) και το οποίο λέει ότι όταν τα βάρη λαμβάνουν την βέλτιστη τιμή τους τότε η ακολουθία σφαλμάτων είναι ασυσχέτιστη με το σήμα εισόδου.

Αφού αναλύθηκε η έννοια της Επιφάνειας Σφάλματος και πως αναπαριστάται σ' ένα προσαρμοζόμενο (adaptive) περιβάλλον, το επόμενο βήμα είναι να αναλυθεί το πώς μια προσαρμοζόμενη διαδικασία ξεκινάει από κάποιο σημείο στην επιφάνεια σφάλματος και προσπαθεί να μετακινηθεί προς το ολικό ελάχιστο όπου και θα πρέπει να μείνει. Βασική προϋπόθεση όπως έχει ήδη αναφερθεί είναι η διαδικασία εισόδου να είναι στάσιμη. Αν αυτό δεν συμβαίνει δηλαδή αν τα δεδομένα είναι μη στάσιμα και οι στατιστικές τους ιδιότητες αλλάζουν, τότε η επιφάνεια του σφάλματος γίνεται ασαφής και ασταθής. Σ' αυτή την περίπτωση η προσαρμοζόμενη διαδικασία θα πρέπει εκτός από το να βρεί το ολικό ελάχιστο στην επιφάνεια, να εντοπίζει και την καινούργια θέση του ελαχίστου καθώς κινείται στο σύστημα αξόνων.

3.3 Ιδιότητες της Τετραγωνικής Επιφάνειας Σφάλματος

Εφόσον έχει ήδη ορισθεί η συνάρτηση κόστους για τα προσαρμοζόμενα συστήματα, θα πρέπει να προχωρήσουμε στην αναφορά και ανάλυση αλγορίθμων οι οποίοι μεταβάλλουν τα βάρη κατά τέτοιο τρόπο ώστε να κινηθούμε στην επιφάνεια σφάλματος προς το ελάχιστο σημείο. Προτού, όμως προχωρήσουμε σ' αυτό θα δούμε πρώτα κάποιες βασικές ιδιότητες της τετραγωνικής επιφάνειας σφάλματος. Αυτές πηγάζουν πρωταρχικά από τον πίνακα συσχέτισης των στοιχείων εισόδου (input correlation matrix).

Χρησιμοποιώντας τις σχέσεις / ορισμούς του ξ και ξ_{\min} το MSE μπορεί να γραφεί στην ακόλουθη μορφή

$$\xi = \xi_{\min} + (W - W^*)^T R (W - W^*) \quad (3.34)$$

όπου W^* το διανυσμα βαρών Weiner που ορίστηκε στην Σχέση (3.30).

Ορίζοντας το διανυσμα:

$$V = W - W^* = [v_0 \quad v_1 \quad \dots \quad v_L]^T \quad (3.35)$$

και παίρνοντας υπόψη μας ότι το ανάδελτα είναι το ίδιο με αυτό της (3.28) γιατί το W διαφέρει από το V κατά μια σταθερά, παίρνουμε:

$$\nabla(\xi) = 2RV = 2(RW - p) \quad (3.36)$$

Αυτή η, τελευταία, έκφραση χρησιμοποιείται στην ανάπτυξη και ανάλυση πολλών προσαρμοστικών (adaptive) αλγορίθμων

3.3.1 Κανονική Μορφή του Πίνακα Αυτοσυσχέτισης

Είναι γνωστό ότι οι ιδιοτιμές λ_n και τα ιδιοδιανύσματα q_n του πίνακα R αποτελούν τις λύσεις του ομογενούς γραμμικού συστήματος $[R - \lambda_n I]q_n = 0$. Αυτή η εξίσωση έχει μη μηδενικές λύσεις για λ_n και q_n , αν και μόνο αν η παρακάτω ορίζουσα μηδενίζεται.

$$|R - \lambda I| = 0 \quad (3.39)$$

Η εξίσωση (3.39) ονομάζεται «χαρακτηριστική εξίσωση» του R και είναι μια αλγεβρική εξίσωση του λ βαθμού $L+1$. Οι $L+1$ λύσεις της είναι οι $\lambda_0, \lambda_1, \dots, \lambda_L$. Αυτές είναι οι ιδιοτιμές του R , οι οποίες στη γενική περίπτωση μπορεί να μην είναι η μια διαφορετική από την άλλη.

$$R q_n = \lambda_n q_n \quad (3.40)$$

Το διάνυσμα q_n είναι το n -οστό ιδιοδιάνυσμα του R και είναι συσχετισμένο με την ιδιοτιμή λ_n . Εξαιτίας του ότι ο πίνακας R είναι συμμετρικός και θετικά ορισμένος* υπάρχει ένας ορθογώνιος* πίνακας Q , του οποίου οι στήλες είναι τα ιδιοδιανύσματα q_n και ένας διαγώνιος πίνακας Λ που περιέχει τις ιδιοτιμές λ_n , όπου $n=1, 2, \dots, L$

$$R[q_0 \ q_1 \ \dots \ q_L] = [q_0 \ q_1 \ \dots \ q_L] \begin{bmatrix} \lambda_0 & \dots & 0 \\ \vdots & \lambda_1 & \vdots \\ 0 & \dots & \lambda_L \end{bmatrix} \quad (3.41)$$

ή ισοδύναμα

$$R = Q\Lambda Q^{-1} \quad (3.42)$$

Η εξίσωση (3.42) μας δίνει την κανονική μορφή του R . Οι πίνακες Λ και Q είναι διαστάσεων $(L+1) \times (L+1)$ όπως και η διάσταση του πίνακα R .

3.3.2 Ιδιοτιμές και Ιδιοδιανύσματα του Πίνακα Συσχέτισης Εισοδου

Από τον ορισμό του R είναι σαφές ότι είναι ένας συμμετρικός πίνακας ($R = R^T$) και επομένως τα ιδιοδιανύσματά του θα πρέπει να είναι ορθογώνια $q_m^T q_n = \begin{cases} 1, & n = m \\ 0, & n \neq m \end{cases}$

για κάθε ζευγάρι n, m . Αυτό μπορεί εύκολα να αποδειχθεί ως ακολούθως:

Έστω λ_1 και λ_2 δυο ξεχωριστές ιδιοτιμές του πίνακα. Τότε

$$R q_1 = \lambda_1 q_1 \quad (3.43)$$

και

$$R q_2 = \lambda_2 q_2 \quad (3.44)$$

Αν εφαρμόσουμε τον τελεστή αναστροφής στην (3.43) και πολλαπλασιάσουμε από δεξιά με q_2 τότε:

$$q_1^T R q_2 = \lambda_1 q_1^T q_2 \quad (3.45)$$

Αν πολλαπλασιάσουμε από τα αριστερά την (3.44) με q_1^T τότε $q_1^T R q_2 = \lambda_2 q_1^T q_2$ ή ισοδύναμα:

$$\lambda_1 q_1^T q_2 = \lambda_2 q_1^T q_2 \quad (3.46)$$

Όμως $\lambda_1 \neq \lambda_2$ και επομένως

$$q_1^T \cdot q_2 = 0 \quad (3.47)$$

Άρα τα ιδιοδιανύσματα λ_1 και λ_2 είναι ορθογώνια.

Υποθέτουμε ότι ο Q είναι ορθοκανονικός πίνακας τότε:

$$\begin{aligned} Q \cdot Q^T &= I \\ \text{και } Q^{-1} &= Q^T \end{aligned} \quad (3.48)$$

Τελος μπορούμε να δείξουμε ότι οι ιδιοτιμές του R είναι θετικές. Πράγματι,

$$\begin{aligned} V^T R V &\geq 0 \\ \text{Επίσης } R &= Q \Lambda \Lambda^{-1} = Q \Lambda \Lambda^T \\ Q^T R Q &\geq 0 \\ \text{ή } Q^T Q \Lambda \Lambda^T Q &\geq 0 \\ \text{ισοδύναμα } \Lambda &\geq 0 \end{aligned} \quad (3.49)$$

3.3.3 Γεωμετρική Σημασία των Ιδιοτιμών και Ιδιοδιανυσμάτων

Τα ιδιοδιανύσματα και οι ιδιοτιμές σχετίζονται άμεσα με κάποιες ιδιότητες της επιφάνειας σφάλματος ξ . Αυτή η επιφάνεια περιγράφεται από μια υπερπαραβολή με $L+2$ διαστάσεις όπου $L+1$ είναι ο αριθμός των βαρών w που χρησιμοποιούνται w_0, w_1, \dots, w_L . Καταρχήν θα αναλύσουμε την επιφάνεια με δύο (2) βάρη w_0 και w_1 έτσι

ώστε να μπορεί να απεικονισθεί στις τρεις (3) διαστάσεις και στην συνέχεια θα γενικεύσουμε τα αποτελέσματά μας. Στην περίπτωση αυτή έχουμε ένα παραβολοειδές. Αν “κόψουμε” αυτό το παραβολοειδές σε φέτες παράλληλες, στο επίπεδο $w_0 w_1$ έτσι ώστε να έχουμε την κάτοψη, παίρνουμε ομόκεντρες ελλείψεις σταθερού MSE. Δηλαδή:

$$W^T R W - 2P^T W = \text{σταθερά} \quad (3.50\alpha)$$

ή χρησιμοποιώντας τον Γραμμικό μετασηματισμό $V = W - R^{-1}P = W - W^*$, μπορούμε να γράψουμε:

$$V^T R V = \text{σταθερά} \quad (3.50\beta)$$

το οποίο αναπαριστά επίσης μια έλλειψη με κέντρο το κέντρο του επίπεδου $v_0 v_1$. Σ' αυτό το νέο σύστημα αξόνων υπάρχουν δυο γραμμές κάθετες στις ελλείψεις, γνωστές ως βασικοί άξονες των ελλείψεων και της επιφάνειας σφάλματος. Οι άξονες αυτοί ονομάζονται v'_0 και v'_1 . Κάθε διάνυσμα το οποίο είναι κάθετο στις ελλείψεις μπορεί να εκφραστεί θεωρώντας τις ελλείψεις ως περιγραφή της $F(v) = V^T R V$ και παίρνοντας την παράγωγο της $F(\cdot)$, που είναι και παράγωγος της ξ , αφού F και ξ διαφέρουν μόνο κατά μια σταθερά

$$\nabla(F) = \left[\frac{\partial F}{\partial v_0} \quad \frac{\partial F}{\partial v_1} \quad \dots \quad \frac{\partial F}{\partial v_L} \right]^T = 2RV \quad (3.51)$$

Από την άλλη, κάθε διάνυσμα που περνά από την αρχή των αξόνων $(v_0, v_1) = (0,0)$ θα είναι της μορφής μV . Και εφόσον περνά από το κέντρο και είναι κάθετο στο $F(v)$ τότε μπορούμε να γράψουμε:

$$2RV' = \mu V' \quad (3.52)$$

ή ισοδύναμα:

$$\left[R - \frac{\mu}{2} I \right] V' = 0 \quad (3.53)$$

όπου το V' τώρα είναι ο βασικός άξονας και επομένως θα πρέπει να είναι και ιδιοδιάνυσμα του πίνακα R . Έτσι μπορούμε να πούμε πως :

Τα ιδιοδιανύσματα του πίνακα συσχέτισης των στοιχείων εισόδου προσδιορίζουν τους βασικούς άξονες της επιφάνειας σφάλματος.

Συνοψίζοντας μπορούμε να καταλήξουμε στα παρακάτω όσον αφορά την επιφάνεια σφάλματος στις τρεις (3) διαστάσεις.

$$\begin{aligned}
 \xi &= \xi_{\min} + (W - W^*)^T R (W - W^*) \\
 \xi &= \xi_{\min} + V^T R V \\
 \xi &= \xi_{\min} + V^T (Q \Lambda Q^T) V \\
 \xi &= \xi_{\min} + (Q^T V) \Lambda (Q^T V) \\
 \xi &= \xi_{\min} + V'^T \Lambda V'
 \end{aligned} \tag{3.53}$$

Η τελευταία εξίσωση μπορεί να επαληθευθεί παραγωγίζοντας. Πράγματι,

$$\nabla \xi = 2 \Lambda V' = 2 \begin{bmatrix} \lambda_0 v_0' & \lambda_1 v_1' & \dots & \lambda_L v_L' \end{bmatrix} \tag{3.54}$$

όπου $V = W - W^*$ και $V' = Q^T V = Q^{-1} V$.

Ακόμη, οι ιδιοτιμές του R έχουν και αυτές γεωμετρική σημασία. Όπως φαίνεται στην (3.34) η μερική παράγωγος του ξ μπορεί να γραφτεί

$$\frac{\partial \xi}{\partial v_n'} = 2 \lambda_n v_n' \tag{3.55a}$$

και

$$\frac{\partial^2 \xi}{\partial v_n'^2} = 2 \lambda_n, \quad n = 0, 1, \dots, L \tag{3.55b}$$

Παρατηρούμε ότι η δεύτερη μερική παράγωγος του ξ είναι διπλάσια από την ιδιοτιμή σε κάθε κύριο άξονα, έτσι:

Οι ιδιοτιμές του πίνακα συσχέτισης των στοιχείων εισόδου R είναι ίσες με τις δεύτερες μερικές παραγώγους της επιφάνειας σφάλματος ξ .

3.4 Ερευνώντας την Επιφάνεια Σφάλματος

Όπως έχει ήδη αναλυθεί η επιφάνεια του μέσου τετραγωνικού σφάλματος για τον Προσαρμοζόμενο Γραμμικό Συνδυαστή είναι μια τετραγωνική συνάρτηση των βαρών όταν η χρονοσειρά X_k και η επιθυμητή απόκριση d_k είναι στάσιμες διαδικασίες. Το ζητούμενο είναι να αναπτυχθούν συστηματοποιημένες διαδικασίες και αλγόριθμοι ικανοί να εντοπίζουν τα βέλτιστα βάρη όταν είναι διαθέσιμα μόνο μετρήσιμα ή εκτιμώμενα

δεδομένα. Οι περισσότερες απ' αυτές τις διαδικασίες δεν ερευνούν διεξοδικά την επιφάνεια αλλά βρίσκουν το βέλτιστο ή το «σχεδόν» βέλτιστο κάνοντας προσαρμογές γύρω απ' αυτό.

3.4.1 Μέθοδοι Εξερεύνησης της Επιφάνειας Σφάλματος

Δύο είναι οι βασικές μέθοδοι εξερεύνησης της επιφάνειας σφάλματος η μέθοδος του Newton και η μέθοδος απότομης κατάβασης (steepest descent). Και οι δύο κάνουν χρήση της παραγώγου ώστε να εντοπίσουν την κατεύθυνση στην οποία βρίσκεται το ελάχιστο. Αυτές οι μέθοδοι μπορούν να χρησιμοποιηθούν τόσο στις τετραγωνικές επιφάνειες σφάλματος όσο και σε άλλες γενικότερης μορφής επιφάνειες.

Η μέθοδος του Newton έχει θεμελιώδη μαθηματική σημασία, αν και δύσκολα εφαρμόζεται στην πράξη. Είναι μια μέθοδος αναζήτησης βασισμένη στον τελεστή ανάδελτα η οποία προκαλεί τα βάρη και μεταβάλλονται σε κάθε βήμα. Οι μεταβολές αυτές είναι πάντα προς την κατεύθυνση του ελάχιστου, με την προϋπόθεση ότι η επιφάνεια είναι τετραγωνική.

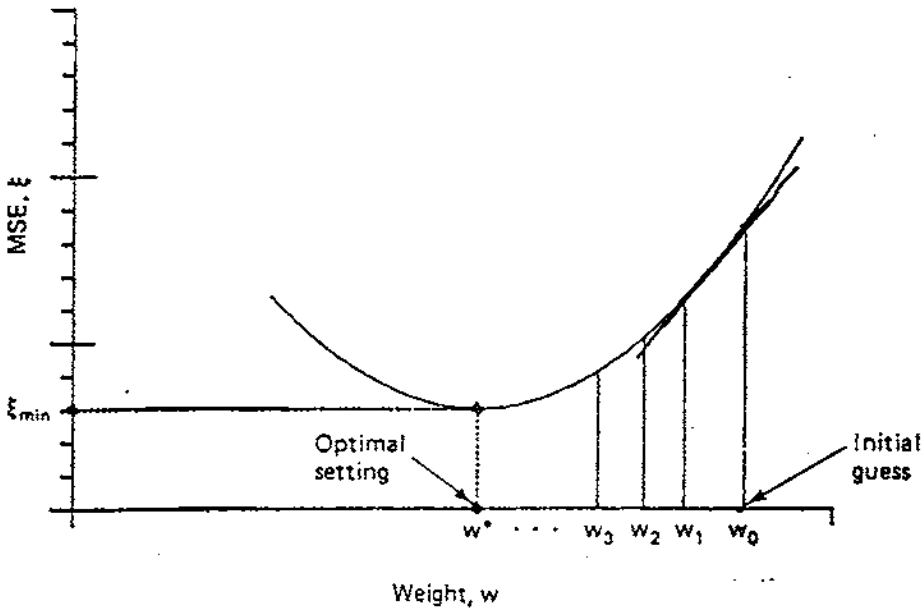
Η μέθοδος απότομης κατάβασης είναι αναγνωρισμένης αξίας και βρίσκει πρακτική εφαρμογή σε πολλές περιπτώσεις. Είναι κι αυτή μια μέθοδος αναζήτησης βασισμένη στον τελεστή ανάδελτα που αναγκάζει όλα τα στοιχεία του παράγοντα βάρους να μεταβάλλονται σε κάθε βήμα. Σ' αυτή την περίπτωση οι μεταβολές γίνονται προς την κατεύθυνση της αρνητικής παραγώγου της επιφάνειας, γεγονός το οποίο δεν σημαίνει αναγκαστικά ότι οδηγούμαστε και προς το ελάχιστο.

3.4.2 Βασικές Αρχές της Μεθόδου Αναζήτησης Βασισμένης στο Ανάδελτα

Για να κατανοηθούν καλύτερα οι μέθοδοι αναζήτησης με χρήση της παραγώγου θεωρούμε την περίπτωση όπου υπάρχει μόνο ένα βάρος. Σ' αυτή την περίπτωση, που δεν έχει πρακτική σημασία, όλες αυτές οι μέθοδοι ταυτίζονται.

Η επιφάνεια του σφάλματος είναι μια παραβολή (Σχήμα 3.3). Η επιφάνεια μπορεί να εκφραστεί μαθηματικά ως εξής:

$$\xi = \xi_{\min} + \lambda(W - W^*)^2 \quad (3.56)$$



Σχήμα 3.3: Εξερεύνηση της επιφάνειας σφάλματος με βάση το Ανάδελτα στην περίπτωση της μίας μεταβλητής

Παρατηρούμε πως η ιδιοτιμή λ , ισούται με r_{00} . Εύκολα προκύπτει ότι η πρώτη και δεύτερη παράγωγος δίνονται απο τις σχέσεις:

$$\frac{d\xi}{dw} = 2\lambda(w - w^*) \tag{3.57}$$

$$\frac{d^2\xi}{dw^2} = 2\lambda \tag{3.58}$$

Το ζητούμενο είναι να βρεθεί το W^* το οποίο ελαχιστοποιεί το MSE. Μη γνωρίζοντας την μορφή της συνάρτησης σφάλματος, ξεκινάμε με μια τυχαία τιμή W_0 και υπολογίζουμε την κλίση της καμπύλης στο αντίστοιχο σημείο. Στη συνέχεια επιλέγουμε μια νέα τιμή η οποία ισούται με W_0 συν μία αύξηση ανάλογη προς το αρνητικό της κλίσης. Με τον ίδιο τρόπο παίρνουμε μια νέα τιμή W_2 . Αυτή η διαδικασία επαναλαμβάνεται ώσπου να βρεθεί το W^* . Η τιμή η οποία παίρνουμε από την μέτρηση της κλίσης της καμπύλης στα διαφορετικά W_0, W_1, W_2, \dots καλείται εκτίμηση του Ανάδελτα.

3.4.3 Ένας Απλός Αλγόριθμος Αναζήτησης Βασισμένος στο Ανάδελτα

Η διαδικασία που περιγράψαμε στην προηγούμενη παράγραφο μπορεί να εκφραστεί από τις παρακάτω εξισώσεις:

$$W_{k+1} = W_k + \mu(-\nabla_k \xi) \quad (3.59)$$

όπου W_k είναι η τρέχουσα τιμή του W και W_k είναι η νέα τιμή που θέλουμε να βρούμε. Η παράμετρος μ είναι μια σταθερά με την οποία ελέγχεται η σταθερότητα και ο ρυθμός σύγκλισης:

$$\nabla_k \xi = \left. \frac{d\xi}{dw} \right|_{w=w_k} = 2\lambda(w_k - w^*)$$

Άρα

$$W_{k+1} = (1 - 2\mu\lambda)W_k + 2\mu\lambda W^* \quad (3.60)$$

Αυτή η εξίσωση εφόσον είναι γραμμική πρώτου βαθμού μπορεί να λυθεί με μαθηματική επαγωγή:

$$\begin{aligned} W_1 &= (1 - 2\mu\lambda)W_0 + 2\mu\lambda W^* \\ W_2 &= (1 - 2\mu\lambda)^2 W_0 + 2\mu\lambda W^* [(1 - 2\mu\lambda) + 1] \\ W_3 &= (1 - 2\mu\lambda)^3 W_0 + 2\mu\lambda W^* [(1 - 2\mu\lambda)^2 + (1 - 2\mu\lambda) + 1] \\ &\vdots \end{aligned} \quad (3.61)$$

και επομένως:

$$\begin{aligned} W_k &= (1 - 2\mu\lambda)^k W_0 + 2\mu\lambda W^* \sum_{n=0}^{k-1} (1 - 2\mu\lambda)^n \\ W_k &= (1 - 2\mu\lambda)^k W_0 + 2\mu\lambda W^* \frac{1 - (1 - 2\mu\lambda)^k}{1 - (1 - 2\mu\lambda)} \\ W_k &= W^* + (1 - 2\mu\lambda)^k (W_0 - W^*) \end{aligned} \quad (3.62)$$

3.4.4 Σταθερότητα και Ρυθμός Σύγκλισης

Στη Σχέση (3.62) το $r=1-2\mu\lambda$ είναι γνωστό ως γεωμετρικός λόγος (geometric ratio) γιατί είναι η αναλογία των συνεχόμενων όρων στο γεωμετρικό άθροισμα της (3.62). Η εξίσωση θεωρείται σταθερή αν και μόνο αν ο γεωμετρικός λόγος ικανοποιεί την συνθήκη:

$$|r| = |1 - 2\mu\lambda| < 1 \quad (3.63)$$

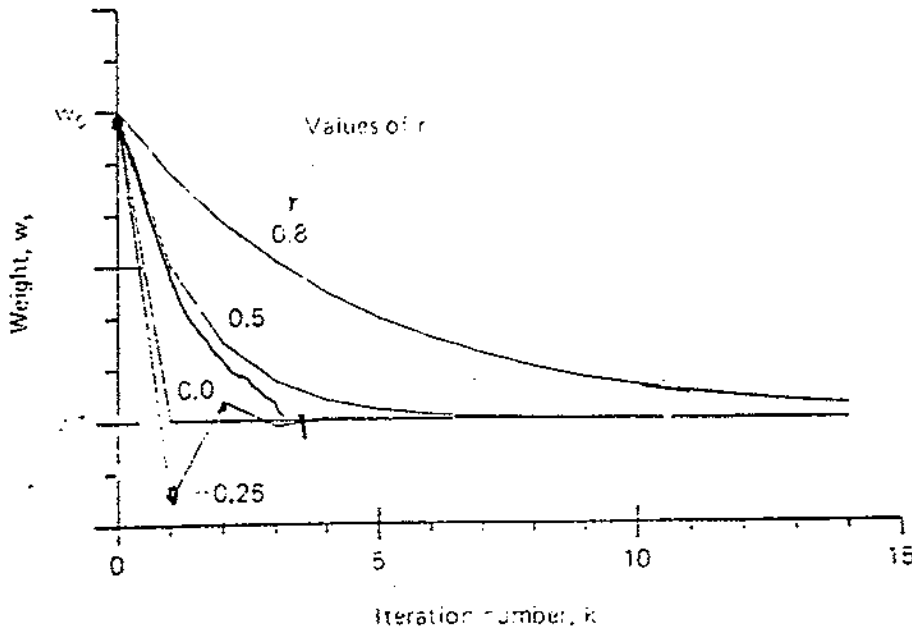
ή ισοδύναμα

$$\frac{1}{\lambda} > \mu > 0$$

στην περίπτωση που οι τιμές της παραμέτρου μ ικανοποιούν την παραπάνω σχέση, ο αλγόριθμος τείνει προς την βέλτιστη λύση:

$$\lim_{k \rightarrow \infty} (W_k) = W^* \quad (3.64)$$

Ο ρυθμός σύγκλισης (rate of Convergence) εξαρτάται και αυτός από το r . Στο Σχήμα 3.4 είναι ολοφάνερο το τι συμβαίνει για διαφορετικές τιμές του γεωμετρικού λόγου r .



Σχήμα 3.4: Ρύθμιση του βάρους για διαφορετικές τιμές του r

Παρατηρούμε πως η απόλυτη τιμή του r είναι <1 και ο ρυθμός σύγκλισης αυξάνεται καθώς το r πλησιάζει το μηδέν. Φαίνεται επίσης ότι για θετικές τιμές του r , μικρότερες από την μονάδα δεν παρατηρείται ταλάντωση στις τιμές των βαρών, ενώ για αρνητικές τιμές του r τα βάρη ξεπερνούν τις βέλτιστες τιμές τους και στη συνέχεια ταλαντώνονται γύρω από αυτές. Στην πρώτη περίπτωση η διαδικασία ονομάζεται 'overdamped', ενώ στη δεύτερη 'underdamped'. Για $r=0$ η διαδικασία είναι όμοια με την μέθοδο Newton. Συνοψίζοντας καταλήγουμε στον πίνακα που ακολουθεί:

$0 < \mu < \frac{1}{\lambda}$,	$ r < 1 \rightarrow \text{stable}$
$0 < \mu < \frac{1}{2\lambda}$,	$1 > r > 0 \rightarrow \text{Overdamped}$
$\mu = \frac{1}{2\lambda}$,	$r = 0 \rightarrow \text{Critically damped}$
$\frac{1}{2\lambda} < \mu < \frac{1}{\lambda}$,	$0 > r > -1 \rightarrow \text{Underdamped}$
$\mu \geq \frac{1}{\lambda}$ κ' $\mu \leq 0$,	$ r > 1 \rightarrow \text{Unstable}$

Πίνακας 3.1: Η επίδραση του μ στη σύγκλιση στην περίπτωση της διαδικασίας εξερεύνησης με ένα βάρους.

3.4.5 Η Καμπύλη Μάθησης

Αν ξ_k είναι το MSE που αντιστοιχεί στο διάνυσμα συντελεστών W_k τότε μπορεί να γραφτεί σαν:

$$\xi_k = \xi_{\min} + \lambda(W_k - W^*)^2$$

ή χρησιμοποιώντας τη Σχέση (3.62) έχουμε ισοδύναμα

$$\xi_k = \xi_{\min} + \lambda(W_0 - W^*)^2 (1 - 2\mu\lambda)^{2k} \quad (3.65)$$

Είναι φανερό πως καθώς το W_k κατευθύνεται προς το W^* , το MSE κατευθύνεται και αυτό με τη σειρά του προς το ξ_{\min} . Ο γεωμετρικός λόγος του MSE είναι

$$r_{mse} = r^2 = (1 - 2\mu\lambda)^2 \quad (3.66)$$

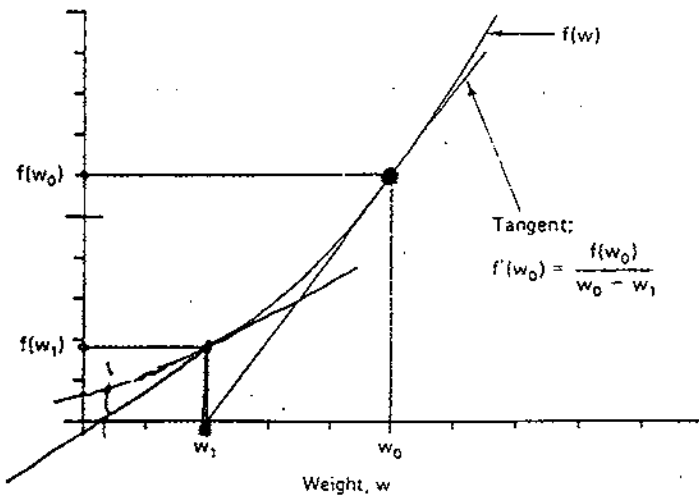
και αφού το r είναι μη αρνητικό το MSE δε θα ταλαντωθεί ποτέ και επομένως διασφαλίζεται η σταθερότητα. Η καμπύλη που δημιουργείται εδώ ονομάζεται 'καμπύλη μάθησης' και αναπαριστά την μείωση του MSE κατά την διαδικασία των επαναλήψεων.

3.4.6 Η Μέθοδος του Newton

Έχουμε ήδη δει πως η διαδικασία αναζήτησης του W^* με χρήση της παραγωγού απορρίπτεται οριακά όταν $r=0$

$$r = 1 - 2\mu\lambda = 0 \Rightarrow \mu = \frac{1}{2\lambda} \quad (3.67)$$

Σ' αυτή την περίπτωση η διαδικασία μετασχηματίζεται σε διαδικασία ενός βήματος για τετραγωνικές MSE συναρτήσεις και αποκαλείται μέθοδος του Newton εξαιτίας της σχέσης της με μια απλή αριθμητική μέθοδο για την εύρεση των ριζών ενός πολυώνυμου. Θα ερευνηθεί πρώτα η εφαρμογή της μεθόδου αυτής σε συνάρτηση με ένα βάρος και έπειτα θα επεκταθούμε και σε συναρτήσεις με πολλά βάρη. Η μέθοδος αυτή αρχικά βρίσκει τις ρίζες μιας συνάρτησης θέτοντας $f(w)=0$. Ξεκινάμε αρχικά με ένα τυχαίο W_0 , και βρίσκοντας την παραγωγό υπολογίζουμε το επόμενο W_1 . Όπως φαίνεται στο Σχήμα 3.5 το W_1 βρίσκεται εκεί που η εφαπτόμενη στο $f(w_0)$ τέμνει τον άξονα W .



Σχήμα 3.5: Η μέθοδος του Newton

$$f'(w_0) = \frac{f'(w_0)}{w_0 - w_1} \Rightarrow w_1 = w_0 - \frac{f(w_0)}{f'(w_0)} \quad (3.68)$$

Το επόμενο σημείο εντοπίζεται χρησιμοποιώντας το w_1 ως το αρχικό τυχαίο βάρος και προχωρούμε ομοίως. Γενικά,

$$w_{k+1} = w_k - \frac{f(w_k)}{f'(w_k)} \quad (3.69)$$

Η μέθοδος αυτή βασίζεται αρχικά στη τυχαία τιμή w_0 , και στη φύση της $f(w)$. Αν η $f(w)$ δίνεται αναλυτικά και είναι συνεχής τότε μπορούμε να υπολογίσουμε την παράγωγό της

και να χρησιμοποιήσουμε την αναδρομική Σχέση (3.68). Αν όμως η $f'(w)$ πρέπει να υπολογισθεί προσεγγιστικά τότε αυτό γίνεται χρησιμοποιώντας τον παρακάτω τύπο:

$$f'_{(k)} = \frac{f(w_k) - f(w_{k-1})}{w_k - w_{k-1}}$$

και επομένως

$$w_{k+1} = w_k - \frac{f(w_k)(w_k - w_{k-1})}{f(w_k) - f(w_{k-1})} \quad (3.70)$$

Για να εφαρμοσθεί αυτή η μέθοδος στην ‘ψηλάφηση’ της επιφάνειας σφάλματος, πρέπει να ξεκινήσουμε με μια εξίσωση της μορφής $f(w)=0$. Στην περίπτωση μας έχουμε $\xi(w)=0$ ή $\nabla \xi = 0$ γιατί θέλουμε να ελαχιστοποιήσουμε το $\xi(w)$:

$$f(w) = \xi'(w)$$

$$w_{k+1} = w_k - \frac{\xi'(w_k)}{\xi''(w_k)} \quad (3.71)$$

Αν η συνάρτηση είναι τετραγωνική τότε οδηγούμεθα σε λύση ενός βήματος.

$$w_1 = w_0 - \frac{2\lambda(w_0 - w^*)}{2\lambda} = w^*$$

Η μέθοδος του Newton μπορεί να γίνει περίπλοκη σε δύο περιπτώσεις. Πρώτον αν δεν γνωρίσουμε επακριβώς το $\xi(w)$ και έτσι θα πρέπει να υπολογίσουμε διαφορετικά τα ξ' και ξ'' ή στην περίπτωση που η επιφάνεια δεν είναι τετραγωνική.

3.4.7 Η Μέθοδος του Newton στην Πολυδιάστατη Περίπτωση

Γενικεύοντας την μέθοδο του Newton στην περίπτωση πολλών μεταβλητών, παρατηρούμε ότι:

$$W^* = R^{-1}P$$

$$\nabla(\xi) = 2RW - 2P \quad (3.72)$$

Πολλαπλασιάζοντας από τα αριστερά με $\frac{1}{2}R^{-1}$ και συνδυάζοντας τις δυο εξισώσεις

παίρνουμε:

$$W_{k+1} = W_k - \frac{1}{2}R^{-1}\nabla_k(\xi) \quad (3.73)$$

Η εξίσωση (3.73) είναι η μέθοδος Newton για την περίπτωση που έχουμε πολλές μεταβλητές – βάρη. Όταν η επιφάνεια σφάλματος είναι τετραγωνική τότε προχωρεί στην

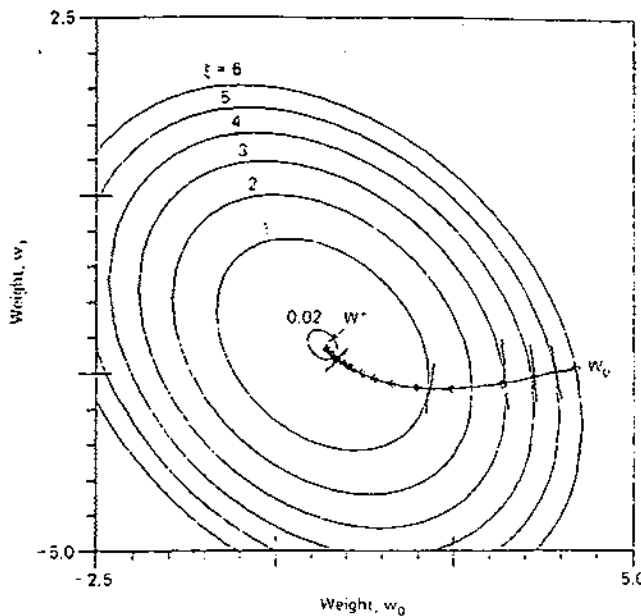
εύρεση του βέλτιστου σε ένα βήμα. Η γενική μορφή του αλγορίθμου της μεθόδου Newton περιέχει και τον συντελεστή μ για να ομαδοποιηθεί ο ρυθμός σύγκλισης (convergence rate).

Θέτοντας $\mu=1/2$ έχουμε μια διαδικασία ενός βήματος. Σε διαφορετική περίπτωση θα πρέπει να επιλεγεί μια τιμή $0 < \mu < 1$ έτσι ώστε η διαδικασία να είναι σταθερή. Επίσης, μπορούμε να γράψουμε: $W_{k+1} = (1 - 2\mu)W_k - 2\mu W^*$. Τώρα έχουμε μια εξίσωση της μορφής της (3.62) και μπορούμε με τον ίδιο τρόπο να καταλήξουμε στο :

$$W_k = W^* + (1 - 2\mu)^k (W_0 - W^*) \quad (3.74)$$

3.4.8 Αναζήτηση με τη Μέθοδο Απότομης Κατάβασης

Η δεύτερη βασική μέθοδος αναζήτησης ονομάζεται απότομη κατάβαση. Σε αυτή, σε αντίθεση με τη μέθοδο Newton, τα βάρη προσαρμόζονται προς την κατεύθυνση της παραγώγου, σε κάθε βήμα.



Σχήμα 3.6: Η μέθοδος Steepest Descent στην περίπτωση δύο βαρών

Εδώ επίσης δεν προχωράμε σε ένα βήμα αλλά σταδιακά σε πολλά μικρά βήματα. Αυτό είναι πλεονέκτημα καθώς σε ένα περιβάλλον 'θορυβώδες' και στοχαστικό, όπου η συνάρτηση είναι άγνωστη, η αργή προσαρμογή προσφέρει μια διαδικασία φιλτραρίσματος η οποία εξασθενίζει τις επιρροές του θορύβου που προκαλείται από τη

μέτρηση της παραγωγού. Τελος, η μέθοδος της απότομης κατάβασης είναι αναγνωρισμένης αξίας και βρίσκει εφαρμογή σε πολλές περιπτώσεις.

Εξορισμού η μέθοδος της απότομης κατάβασης εκφράζεται από τον παρακάτω αναδρομή:

$$W_{k+1} = W_k + \mu(-\nabla_k \xi) \quad (3.75)$$

όπου μ είναι μια σταθερά εξομάλυνσης του μεγέθους των βημάτων και έχει διαστάσεις του αντιστρόφου της ισχύος του σφάλματος.

$$\begin{aligned} W_{k+1} &= W_k - 2\mu R V_k \\ W_{k+1} &= W_k + 2\mu R(W^* - W_k) \\ W_{k+1} &= (I - 2\mu R)W_k + 2\mu R W^* \end{aligned} \quad (3.76)$$

Η λύση αυτής της εξίσωσης είναι πολύπλοκη εξαιτίας του γεγονότος ότι τα διάφορα W_k διασταυρώνονται ανά ζευγάρια. Ο πίνακας $(I - 2\mu R)$ δεν είναι διαγώνιος γιατί περιέχει τον όρο $2\mu R$ και ο πίνακας R δεν είναι διαγώνιος. Παρολαυτα μπορεί να λυθεί η τελευταία σχέση μετατοπιζόμενοι στο κύριο σύστημα αξόνων.

$$\begin{aligned} V &= W - W^* \\ V_{k+1} &= (I - 2\mu R)V_k \end{aligned} \quad (3.77)$$

Στην συνέχεια περιστρέφουμε την επιφάνεια σφάλματος στους κυρίως άξονες.

$$\begin{aligned} V &= QV' \\ QV'_{k+1} &= (I - 2\mu R)QV'_k \end{aligned} \quad (3.78)$$

Πολλαπλασιάζοντας από τα αριστερά με Q^{-1} έχουμε

$$\begin{aligned} V'_{k+1} &= Q^{-1}(I - 2\mu R)QV'_k \\ V'_{k+1} &= (Q^{-1}IQ - 2\mu Q^{-1}RQ)V'_k \\ V'_{k+1} &= (I - 2\mu \Lambda)V'_k \end{aligned} \quad (3.79)$$

Τώρα ο πίνακας των Ιδιοτιμών Λ είναι διαγώνιος. Στην περίπτωση αυτή δεν υπάρχει cross coupling. Έτσι:

$$V' = (I - 2\mu \Lambda)^k V'_0 \quad (3.80)$$

Η παραπάνω σχέση μας δείχνει πως ο αλγόριθμος είναι σταθερός και έχει την δυνατότητα να συγκλίνει όταν ισχύει η παρακάτω συνθήκη

$$\lim_{k \rightarrow \infty} (I - 2\mu\Lambda)^k = 0 \quad (3.81)$$

ή παίρνοντας υπόψη μας ότι ο πίνακας $(I - 2\mu\Lambda)$ είναι διαγώνιος και γράφεται ισοδύναμα

$$\begin{bmatrix} \lim_{k \rightarrow \infty} (I - 2\mu\lambda_0)^k & & & \\ & \lim_{k \rightarrow \infty} (I - 2\mu\lambda_1)^k & & \\ & & \dots & \\ & & & \lim_{k \rightarrow \infty} (I - 2\mu\lambda_n)^k \end{bmatrix} = \mathbf{0} \quad (3.82)$$

Ο αλγόριθμος είναι σταθερός αν και μόνο αν

$$0 < \mu < \frac{1}{\lambda_{\max}} \quad (3.83)$$

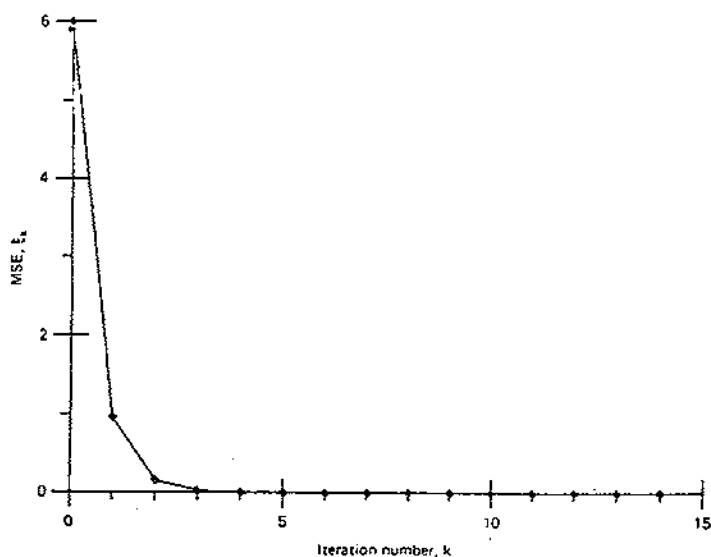
3.4.9 Σύγκριση των Καμπυλών Μάθησης

Η σύγκριση των μεθόδων του Newton και της απότομης κατάβασης μέσα από τη σύγκριση των καμπυλών μάθησης τους, παρουσιάζει εξαιρετικό ενδιαφέρον.

Πρώτον, πρέπει να βρεθεί η MSE συνάρτηση

$$\begin{aligned} \xi &= \xi_{\min} + V^T k V \\ \xi_k &= \xi_{\min} + (1 - 2\mu)^{2k} V_0^T R V_0 \end{aligned} \quad (3.84)$$

Αυτή είναι μια απλή γεωμετρική πρόοδος με $r = (1 - 2\mu)^2$. Η καμπύλη μάθησης για τη μέθοδο Newton είναι η εξής

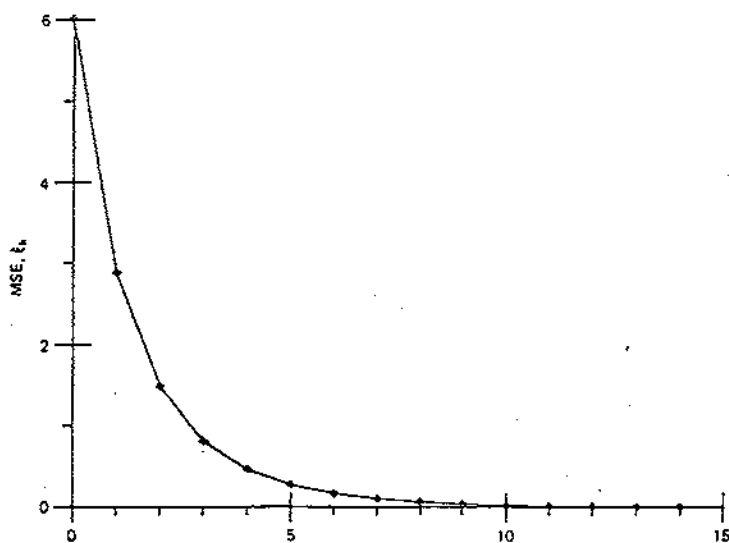


Σχήμα 3.7: Η καμπύλη μάθησης της μεθόδου του Newton

Η καμπύλη μάθησης για τη μέθοδο της απότομης κατάβασης εξάγεται ομοίως

$$\xi = \xi_{\min} + V'^T \Lambda V'$$

$$\xi_k = \xi_{\min} + V_0'^T (I - 2\mu\Lambda)^{2k} \Lambda V_0' = \xi_{\min} + \sum_{n=0}^L v_0'^2 \lambda_n (1 - 2\mu\lambda_n)^{2k} \quad (3.85)$$



Σχήμα 3.8: Η καμπύλη μάθησης για τη μέθοδο Steepest Descent

Ετσι φαίνεται πως η καμπύλη μάθησης για τη μεθοδο Απότομης Κατάβασης (steepest descent) είναι το άθροισμα γεωμετρικών προόδων με r της μορφής :

$$(r_{mse})_n = r_n^2 = (1 - 2\mu\lambda_n)^2 \quad (3.86)$$

Συγκρίνοντας τις δυο καμπύλες, παρατηρούμε πως η μέθοδος του Newton εξελίσσεται ραγδαία σε αντίθεση με τον τη μεθοδο Απότομης Κατάβασης. Αυτό συμβαίνει επειδή η πρώτη χρησιμοποιεί τον πίνακα R για να βρει τον 'απευθείας δρόμο ' στην επιφάνεια σφάλματος για το ξ_{\min} .

3.5 Υπολογισμός του Ανάδελτα και η Επίδρασή του στη Διαδικασία Προσαρμογής

Έως τώρα γινόταν η υπόθεση πως είναι διαθέσιμη μια ακριβής μέτρηση του ανάδελτα (gradient vector). Αυτό όμως στην πλειονότητα των εφαρμογών δεν ισχύει. Αυτό που συμβαίνει είναι ότι η εκτίμηση της παραγωγού βασίζεται σε ένα περιορισμένο στατιστικό δείγμα. Μια τέτοια εκτίμηση είναι 'θορυβώδης ' (noisy) και θεωρείται ως η παραγωγός συν κάποιο προσθετικό θόρυβο.

Η μέθοδος που χρησιμοποιείται για την μέτρηση του ανάδελτα αποκαλείται μέτρηση της παραγωγού (derivative measurement). Εκτός από αυτή υπάρχει και άλλη μια μέθοδος, η οποία ονομάζεται στιγμιαία (instantaneous method). Αυτή η μέθοδος είναι η βάση για τον LMS. Είναι λιγότερο γενική από την 'μέτρηση της παραγωγού γιατί απαιτεί συγκεκριμένη γνώση για την φύση της επιφάνειας σφάλματος. Αντίθετα η στιγμιαία μέθοδος απαιτεί μόνο πολύ γενικές γνώσεις της επιφάνειας και γι' αυτό το λόγο θα αναλυθεί και πρώτη.

3.5.1 Εκτίμηση του Ανάδελτα με Μέτρηση της Παραγωγού στην Περίπτωση ενός Βάρους

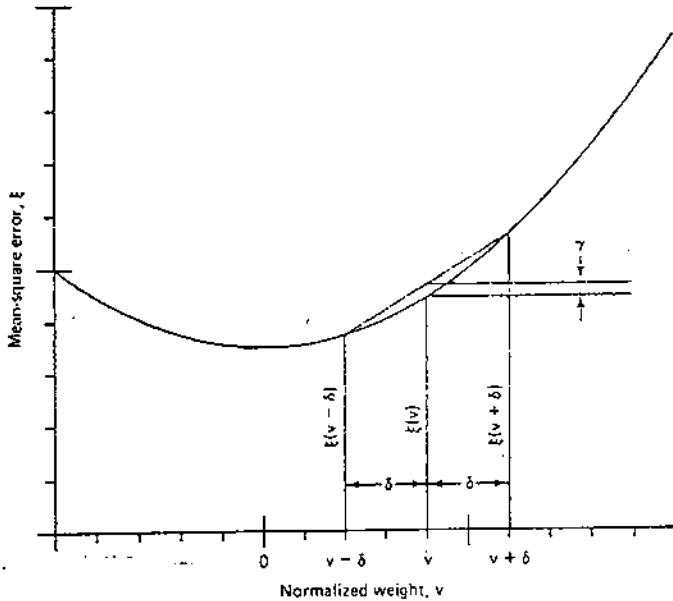
Η παραγωγός μιας τετραγωνικής συνάρτησης με ένα βάρος μπορεί να μετρηθεί ως εξής :

$$\xi(v) = \xi_{\min} + \lambda v^2 \quad (3.87)$$

και

$$\begin{aligned} \frac{d\xi}{dv} &= 2\lambda v \\ \frac{d^2\xi}{dv^2} &= 2\lambda \end{aligned} \quad (3.88)$$

Θα προσπαθήσουμε να εκτιμήσουμε τη παραγωγή χρησιμοποιώντας τις κεντρικές διαφορές



Σχήμα 3.9: Μέτρηση της Παραγωγού

Πράγματι έστω

$$\frac{d\xi}{dv} = \frac{\xi(v+\delta) - \xi(v-\delta)}{2\delta}$$

όμως

$$\frac{\xi(v+\delta) - \xi(v-\delta)}{2\delta} = \frac{\lambda(v+\delta)^2 - \lambda(v-\delta)^2}{2\delta}$$

και επομένως

$$\frac{\xi(v+\delta) - \xi(v-\delta)}{2\delta} = \frac{\lambda}{2\delta} (v^2 + \delta^2 + 2v\delta - v^2 - \delta^2 + 2v\delta)$$

και ύστερα από πράξεις καταλήγουμε στην σχέση

$$\frac{\xi(v+\delta) - \xi(v-\delta)}{2\delta} = 2\lambda v = \frac{d\xi}{dv} \quad (3.89)$$

Όμως μπορεί εύκολα να αποδειχθεί ότι αν χρησιμοποιηθούν δεύτερης τάξης διαφορές η εκτιμήσή μας ταυτίζεται με την δεύτερη παράγωγο της τετραγωνικής συνάρτησης, ήτοι:

$$\frac{d^2\xi}{dv^2} = \frac{\xi(v+\delta) - 2\xi(v) + \xi(v-\delta)}{\delta^2}$$

ή ισοδύναμα

$$\frac{d^2\xi}{d^2\nu} = 2\lambda \quad (3.90)$$

3.5.2 Performance Penalty

Η διαδικασία αυτή προϋποθέτει ότι καθώς γίνεται η εκτίμηση της παραγωγού η ρύθμιση των βαρών (weight adjustment) αλλάζει. Για να βρεθεί η πρώτη παραγωγός το προσαρμοστικό σύστημα ξοδεύει χρόνο για να ρυθμιστεί στα $\nu-\delta$ και $\nu+\delta$, όπως φαίνεται και από το Σχήμα 3.9, ενώ δεν ξοδεύεται χρόνος για το ν . Το performance penalty ορίζεται ως η αύξηση γ , που προκαλείται από την απόκλιση δ που παίρνουμε στα βάρη. Έτσι :

$$\begin{aligned} \gamma &= \frac{1}{2} [2\xi_{\min} + \lambda(\nu + \delta)^2 + \lambda(\nu - \delta)^2] - (\xi_{\min} + \lambda) \\ \gamma &= \lambda\delta^2 \end{aligned} \quad (3.91)$$

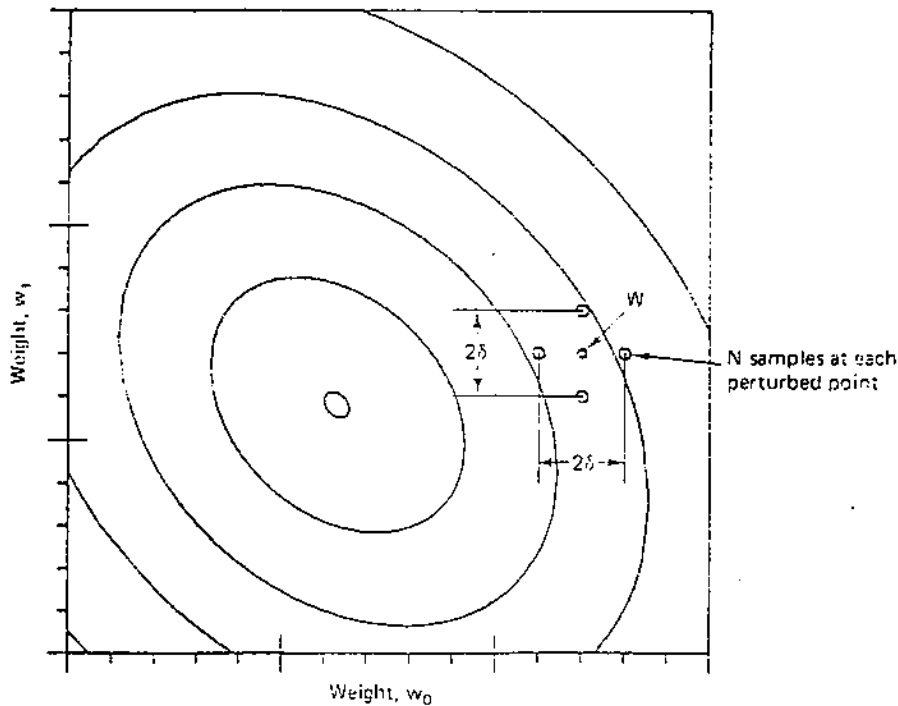
Παρατηρούμε πως το γ δεν είναι συνάρτηση του ν . Η επίδραση αυτή της μέτρησης της παραγωγού στην ρύθμιση της προσαρμογής (adaptive adjustment) καλείται διαταραχή (perturbation) P . Το P μπορεί να εκφραστεί σε όρους του γ ως εξής:

$$P = \frac{\gamma}{\xi_{\min}} = \frac{\lambda\delta^2}{\xi_{\min}} \quad (3.92)$$

Αυτή η εξίσωση μας δίνει την μέση αύξηση στο MSE με βάση το ελάχιστο εφικτό.

Στην περίπτωση που είναι αναγκαίο να εκτιμηθεί και η δεύτερη παραγωγός (όπως στην μέθοδο του Newton), θεωρούμε ότι δεν έχει καμία επίδραση στο perturbation και μπορεί να αγνοηθεί. Αυτό συμβαίνει γιατί η δεύτερη παράγωγος στις τετραγωνικές επιφάνειες είναι σταθερά και μετράται σπάνια.

3.5.3 Μέτρηση της Παραγώγου και Διαταραχή στην Περίπτωση πολλών Βαρών



Σχήμα 3.10: Μέτρηση της Παραγώγου στην περίπτωση 2 βαρών

Στην περίπτωση των δύο (2) βαρών η συνάρτηση είναι

$$\begin{aligned} \xi &= \xi_{\min} + V^T R V \\ \xi &= \xi_{\min} + [v_0 \ v_1] \begin{bmatrix} r_{00} & r_{01} \\ r_{10} & r_{11} \end{bmatrix} \begin{bmatrix} v_0 \\ v_1 \end{bmatrix} \\ \xi &= \xi_{\min} + r_{00} v_0^2 + r_{11} v_1^2 + 2r_{01} v_0 v_1 \end{aligned} \quad (3.93)$$

Για τις μερικές παραγώγους της συνάρτησης ως προς v_0 , v_1 έχουμε

$$\begin{aligned} P_0 &= \frac{r_{00} \delta^2}{\xi_{\min}} \\ P_1 &= \frac{r_{11} \delta^2}{\xi_{\min}} \end{aligned} \quad (3.94)$$

και έστω ότι ο χρόνος που απαιτείται για το υπολογισμό ίσος. Τότε το μέσο P είναι :

$$P = \frac{1}{2}(P_0 + P_1) = \frac{\delta^2}{\xi_{\min}} \frac{(r_{00} + r_{11})}{2} \quad (3.95)$$

Έτσι μπορούμε να γενικεύσουμε και να πούμε ότι το P για L+1 βάρη είναι :

$$P = \frac{\delta^2}{\xi_{\min}} \frac{\text{trace}[R]}{L+1} \quad (3.96)$$

Εφόσον το trace[R] είναι ίσο με το άθροισμα των ιδιοτιμών του πίνακα R όπως και με το άθροισμα των στοιχείων της κύριας διαγωνίου μπορούμε να γράψουμε :

$$P = \frac{\delta^2}{\xi_{\min}} \frac{\sum_{n=0}^L \lambda_n}{L+1} = \frac{\delta^2}{\xi_{\min}} \lambda_{av} \quad (3.97)$$

όπου

$$\lambda_{av} = \frac{1}{L+1} \sum_{n=0}^L \lambda_n \quad (3.98)$$

η μέση τιμή των ιδιοτιμών του πίνακα R.

3.5.4 Διακύμανση του Εκτιμώμενου Ανάδελτα

Οι παράγωγοι που εκτιμώνται κατ' αυτό τον τρόπο είναι θορυβώδεις εξαιτίας τις παρουσίας θορύβου στις μετρήσεις του ξ . Το πρώτο βήμα στον υπολογισμό της διακύμανσης της εκτιμώμενης παραγώγου είναι να βρεθεί η διακυμανση του εκτιμώμενου ξ , όπου $\xi = [\varepsilon_k^2] = \hat{\xi}$. Ξεκινώντας από τον υπολογισμό της διακύμανσης του $\hat{\xi}$ και της παραγώγου, θα πρέπει πρώτα να ορίσουμε την ροπή τάξης r της τ.μ. ε_k :

$$\begin{aligned} \hat{\alpha}_r &= \frac{1}{N} \sum_{k=1}^N (\varepsilon_k)^r \\ \alpha_r &= E[\hat{\alpha}_r] = E[\varepsilon_k^r] \end{aligned} \quad (3.99)$$

Έστω $r=4$, ε_k είναι Gaussian με μέσο ίσο με το μηδέν, δηλαδή η πυκνότητα της πιθανότητας της ε , $p(\varepsilon)$, έχει μέσο ίσο με το μηδέν και τυπική απόκλιση ίση με σ . Τότε:

$$\alpha_4 = \int_{-\infty}^{+\infty} \varepsilon^4 p(\varepsilon) d\varepsilon = \int_{-\infty}^{+\infty} \frac{\varepsilon^4}{\sigma\sqrt{2\Pi}} e^{-\varepsilon^2/2\sigma^2} d\varepsilon = 3\sigma^4 \quad (3.100)$$

Δοθέντος του $\rho(\varepsilon)$ μπορούμε βέβαια να καθορίσουμε κάθε τάξης ροπές της τ.μ. Για παράδειγμα, για $r=2$ μπορούμε εύκολα να δούμε ότι $\alpha_2 = E[\varepsilon_k^2] = \xi$

Τώρα μπορούμε να υπολογίσουμε τη διακύμανση ως την ελάχιστη τετραγωνική απόκλιση από τον μέσο, ήτοι:

$$\begin{aligned} \text{Var}[\hat{\alpha}_r] &= E[\alpha_r^2 + \hat{\alpha}_r^2 - 2\alpha_r \hat{\alpha}_r] \\ \text{Var}[\hat{\alpha}_r] &= E[\hat{\alpha}_r^2] - \alpha_r^2 \\ \text{Var}[\hat{\alpha}_r] &= \frac{1}{N^2} \sum_{k=1}^N \sum_{l=1}^N E[(\varepsilon_k \varepsilon_l)^r] - \alpha_r^2 \end{aligned} \quad (3.101)$$

Η τελευταία σχέση μπορεί να απλοποιηθεί διαχωρίζοντας την περίπτωση $k=l$. Έτσι έχουμε:

$$E[(\varepsilon_k \varepsilon_l)^r] = \begin{cases} E[\varepsilon_k^{2r}] = \alpha_{2r} & k=l \\ E[\varepsilon_k^r]E[\varepsilon_l^r] = \alpha_r^2 & k \neq l \end{cases} \quad (3.102)$$

$$\Rightarrow \text{Var}[\hat{\alpha}_r] = \frac{1}{N^2} [N\alpha_{2r} + (N^2 - N)\alpha_r^2] - \alpha_r^2 = \frac{\alpha_{2r} - \alpha_r^2}{N} \quad (3.103)$$

$$\text{Αφού } \alpha_2 = \xi \text{ τότε } \text{Var}[\hat{\xi}] = \text{Var}[\hat{\alpha}_2] = \frac{\alpha_4 - \alpha_2^2}{N} \quad (3.104)$$

Οι τιμές του α , εδώ, εξαρτώνται από την κατανομή του ε_k . Αν υποθέσουμε ότι το ε_k έχει κανονική κατανομή, με μηδενικό μέσο και διακύμανση σ_ε^2 [3,5]. Τότε το α_4 που βρέθηκε να είναι $3\sigma_\varepsilon^4$ και $\alpha_2 = \sigma_\varepsilon^2$, τότε έχουμε:

$$\text{Var}[\hat{\xi}] = \frac{3\sigma_\varepsilon^4 - \sigma_\varepsilon^4}{N} = \frac{2\sigma_\varepsilon^4}{N} = \frac{2\xi^2}{N} \quad (3.105)$$

Επομένως όταν ε_k έχει κανονική κατανομή με μηδενικό μέσο, η διακύμανση του ξ^2 είναι ανάλογη με το ξ^2 και αντιστρόφως ανάλογη με το N . Με βάση τα παραπάνω μπορούμε να γενικεύσουμε και να γράψουμε:

$$\text{Var}[\hat{\xi}] = \frac{\kappa \xi^2}{N} \quad (3.106)$$

Όταν η κατανομή είναι κανονική αλλά ο μέσος μη μηδενικός, το κ είναι μικρότερο του 2. Όταν η κατανομή δεν είναι Gaussian, το κ και πάλι είναι μικρότερο του 2. Όταν ούτε η κατανομή είναι Gaussian ούτε ο μέσος μηδέν, θα προκαλέσει αύξηση ή μείωση του κ , ανάλογα με την κατανομή.

Για να παρουσιάσουμε τη non- Gaussian περίπτωση, κάνουμε την υπόθεση ότι το ε_k είναι uniformly distributed, με μηδενικό μέσο και τυπική απόκλιση σ . Έτσι έχουμε:

$$\alpha_r = \int_{-\infty}^{\infty} \varepsilon^r p(\varepsilon) d\varepsilon$$

$$\alpha_r = \int_{-\sigma\sqrt{3}}^{\sigma\sqrt{3}} \varepsilon^r \left(\frac{1}{2\sigma\sqrt{3}} \right) d\varepsilon = \frac{\sigma^r 3^{r/2}}{r+1} \quad (3.107)$$

$$Var[\xi] = \frac{\alpha_4 - \alpha_2^2}{N} = \frac{4\sigma^4}{5N} \quad (3.108)$$

Αφού ο μέσος του ε_k είναι μηδέν έχουμε $\xi = \sigma^2$.

Αφού βρήκαμε τη διακύμανση του ξ , τώρα θα προχωρήσουμε στον προσδιορισμό της διακύμανσης της εκτιμώμενης παραγώγου. Για να το καταφέρουμε αυτό κάνουμε και πάλι την υπόθεση ότι τα δείγματα του σφάλματος είναι στατιστικά ανεξάρτητα μεταξύ τους, που σημαίνει και πως οι τιμές του ξ είναι ανεξάρτητες.

$$\frac{d\hat{\xi}}{dv} = \frac{1}{2\delta} \xi(v+\delta) - \frac{1}{2\delta} \xi(v-\delta) \quad (3.109)$$

Σύμφωνα με την υπόθεση που έγινε (ανεξαρτησίας), η διακύμανση είναι ίση με το άθροισμα των διακυμάνσεων των δυο όρων που εμφανίζονται στη Σχέση (3.109). Ακόμη η διακύμανση του $c\xi$, όπου c^2 φορές η διακύμανση του ξ . Έτσι από (3.106) και (3.109) με $\kappa=2$, έχουμε

$$Var\left[\frac{d\hat{\xi}}{dv}\right] = \frac{1}{4\delta^2} Var\left[\xi(v+\delta)\right] + \frac{1}{4\delta^2} Var\left[\xi(v-\delta)\right] = \frac{1}{2N\delta^2} [\xi^2(v+\delta) + \xi^2(v-\delta)] \quad (3.110)$$

Έστω ότι η διαταραχή (perturbation) δ είναι τόσο μικρή ώστε η προσαρμοστική διαδικασία έχει φτάσει σ' ένα αποτέλεσμα κοντά στο w^* , τότε οι δυο τιμές του ξ στην (3.110) είναι σχεδόν ίσες με το ξ_{min} . Έτσι

$$\text{Var} \left[\frac{\hat{d\xi}}{dv} \right] = \frac{\xi_{\min}^2}{N\delta^2} \quad (3.111)$$

Εφόσον τα N κ δ είναι για την εκτίμηση όλων των στοιχείων της παραγώγου και εφόσον τα ε_k είναι ανεξάρτητα, τα σφάλματα ε σ' όλες τις περιπτώσεις που είναι ανεξάρτητα έχουν την ίδια διακύμανση.

Ο πίνακας συνδιακύμανσης της εκτιμώμενης παραγώγου στο k -στο βήμα δίνεται από τον τύπο.

$$\text{Cov}[\hat{\nabla} \xi] \Rightarrow E \left[(\hat{\nabla}_k \xi - \nabla_k \xi)(\hat{\nabla}_k \xi - \nabla_k \xi)^T \right] = \frac{\xi_{\min}^2}{N\delta^2} \mathbf{I} \quad (3.112)$$

Ακόμη πρέπει:

$$0 < \mu < \frac{1}{\lambda_{\max}} \quad (3.113)$$

όπου λ_{\max} είναι η μεγαλύτερη ιδιοτιμή του R . Αυτή η συνθήκη είναι αναγκαία και ικανή για τη σύγκλιση του αλγορίθμου απότομης κατάβασης σε τετραγωνική επιφάνεια. Αν ικανοποιείται αυτή η συνθήκη τότε:

$$\lim_{k \rightarrow \infty} V'_k = 0 \quad (3.114)$$

και επομένως:

$$V' = Q^{-1}V = Q^{-1}(W - W^*) \Rightarrow \lim_{k \rightarrow \infty} W_k = W^* \quad (3.115)$$

Επομένως η μέθοδος απότομης κατάβασης είναι σταθερή (stable) και συγκλίνουσα αν και μόνο αν $0 < \mu < 1/\lambda_{\max}$.

Μια άλλη οπτική στον απότομης κατάβασης είναι η εξής: με βάση την (3.86), ο ρυθμός σύγκλισης σε κάθε άξονα εξαρτάται από μια μοναδική γεωμετρική αναλογία.

$$\begin{aligned} r_0 &= 1 - 2\mu\lambda_0 \\ r_1 &= 1 - 2\mu\lambda_1 \\ &\vdots \\ &\vdots \\ r_L &= 1 - 2\mu\lambda_L \end{aligned} \quad (3.116)$$

Αυτές είναι οι προβολές του w_k στον κάθε κύριο άξονα, η οποία είναι γεωμετρική πρόοδος με βάση κάποιο ποσοστό που καθορίζεται από την ιδιοτιμή. Η συχνότητα των προβολών του w_k στις αρχικές συντεταγμένες είναι στο άθροισμα γεωμετρικών προόδων και είναι πολύ πιο περίπλοκη.

Η δυναμική αυτή συμπεριφορά της σταδιακής διαδικασίας, επανεμφανίζοντας την σε όρους του w_k .

3.5.5 Επιδράσεις στο Διάνυσμα Βαρών κατά τη Διαδικασία Προσαρμογής

Έχοντας εξάγει τους τύπους για τη διακύμανση της εκτίμησης του ανάδελτα, μπορούμε να προχωρήσουμε στη μελέτη της επίδρασης μιας “θορυβώδους” εκτίμησης της παραγώγου στα βάρη κατά τη διάρκεια της διαδικασίας προσαρμογής (adaptation process). Θα δούμε πως η διαδικασία αυτή, η οποία βασίζεται στην “θορυβώδη” εκτίμηση του ανάδελτα οδηγεί σε θόρυβο στη λύση (εύρεση ελαχίστων) των βαρών και σε απώλειες. Η ακριβής φάση αυτών των επιδράσεων εξαρτάται από τη μέθοδο που χρησιμοποιείται. Την επίδραση στα βάρη θα τη δούμε τόσο στη μέθοδο του Newton όσο και στη μέθοδο της απότομης κατάβασης. Για το σκοπό αυτό ορίζουμε τον θόρυβο (διάνυσμα) της εκτίμησης της παραγώγου στο k -στο βήμα όταν $w = w_k, N_k$. Έτσι, μπορούμε να γράψουμε:

$$\hat{\nabla}_k \xi = \nabla_k \xi + N_k \quad (3.117)$$

Όσον αφορά την μέθοδο του Newton μπορούμε να γράψουμε:

$$W_{k+1} = W_k - \mu R^{-1} \nabla_k \quad (3.118)$$

και επομένως

$$\begin{aligned} W_{k+1} &= W_k - \mu R^{-1} (\hat{\nabla}_k - N_k) \\ W_{k+1} &= W_k - \mu R^{-1} \nabla_k - \mu R^{-1} N_k \end{aligned}$$

ή

$$\begin{aligned} V_{k+1} &= V_k - \mu R^{-1} \nabla_k - \mu R^{-1} N_k \\ V_{k+1} &= V_k - 2\mu V_k - \mu R^{-1} N_k \\ V_{k+1} &= (1 - 2\mu)V_k - \mu R^{-1} N_k \end{aligned}$$

Το σύστημα αυτό είναι διασταυρωμένο ανά ζεύγη (cross-coupled), επειδή το N_k πολλαπλασιάζεται με το R . Αυτό μπορεί να εξαλειφθεί περιστρέφοντας το σύστημα

Ξέρουμε ότι : $V = QV'$ και $R^{-1} = Q\Lambda Q^{-1}$

$$\begin{aligned} QV'_{k+1} &= (1-2\mu)QV'_k - \mu Q\Lambda^{-1}Q^{-1}N_k \\ V'_{k+1} &= (1-2\mu)V'_k - \mu\Lambda^{-1}(Q^{-1}N_k) \end{aligned} \quad (3.119)$$

Αν ορίσουμε $Q^{-1}N = N'$ τότε :

$$V'_{k+1} = (1-2\mu)V'_k - \mu\Lambda^{-1}N_k \quad (3.120)$$

Εδώ, αφού ο Λ^{-1} είναι διαγώνιος λύνεται το παραπάνω πρόβλημα. Το V'_k μπορεί να εξαχθεί με μαθηματική επαγωγή όπως έχουμε ήδη δει και θα δίνεται από τις σχέσεις

$$\begin{aligned} V'_1 &= (1-2\mu)V'_0 - \mu\Lambda^{-1}N'_0 \\ V'_2 &= (1-2\mu)^2V'_0 - \mu\Lambda^{-1}[(1-2\mu)N'_0 + N'_1] \\ V'_3 &= (1-2\mu)^3V'_0 - \mu\Lambda^{-1}[(1-2\mu)^2N'_0 + (4-2\mu)N'_1 + N'_2] \\ &\vdots \\ &\vdots \\ V'_k &= (1-2\mu)^kV'_0 - \mu\Lambda^{-1}\sum_{n=0}^{k-1}(1-2\mu)^nN'_{k-n-1} \end{aligned} \quad (3.121)$$

Υποθέτοντας ότι το k τείνει στο άπειρο και το σύστημα είναι σταθερό ($0 < \mu < 1/2$), το $(1-2\mu)^k$ γίνεται αμελητέο και έτσι έχουμε:

$$V'_k = -\mu\Lambda^{-1}\sum_{n=0}^{k-1}(1-2\mu)^nN'_{k-n-1} \quad (3.122)$$

Αυτή η εξίσωση μας δίνει το steady state σφάλμα για τη μέθοδο του Newton σε όρους των ιδιοτιμών του Λ^{-1} και $N'=Q^{-1}N$. Παρατηρούμε πως τα στοιχεία του διαγωνίου πίνακα Λ^{-1} είναι τα $1/\lambda_0, 1/\lambda_1, 1/\lambda_2, \dots, 1/\lambda_L$

Αφού είδαμε τη μέθοδο του Newton μπορούμε να προχωρήσουμε στη μέθοδο Απότομης Κατάβασης. Πράττουμε ομοίως και έτσι έχουμε:

$$W_{k+1} = W_k - \mu\nabla_k$$

χρησιμοποιώντας τις σχέσεις

$$\begin{aligned} V &= W - W^* \\ \hat{\nabla} &= 2RV + N \end{aligned}$$

έχουμε $V_{k+1} = V_k - \mu(2RV_k + N_k)$

$$\text{ή } V_{k+1} = (I - 2\mu R)V_k - \mu N_k \quad (3.123)$$

Περιστρέφουμε θέτωντας $V = QV'$

$$\begin{aligned} QV'_{k+1} &= (I - 2\mu R)QV'_k - \mu N_k \\ V'_{k+1} &= (I - 2\mu Q^{-1}RQ)V'_k - \mu Q^{-1}N_k \\ V'_{k+1} &= (I - 2\mu\Lambda)V'_k - \mu N'_k \end{aligned} \quad (3.124)$$

Για να βρούμε το V'_k προχωράμε ομοίως με επαγωγή και έτσι έχουμε:

$$V'_k = (I - 2\mu\Lambda)^k V'_0 - \mu \sum_{n=0}^{k-1} (I - 2\mu\Lambda)^n N'_{k-n-1} \quad (3.125)$$

Η steady state λύση εδώ είναι:

$$V'_k = -\mu \sum_{n=0}^{k-1} (I - 2\mu\Lambda)^n N'_{k-n-1} \quad (3.126)$$

Επίσης μπορούμε να εξάγουμε τύπους για τη συνδιακύμανση (covariance) αν έχουμε τον πίνακα του "θορύβου" της παραγωγού.

Ο πίνακας συνδιακύμανσης για τα βάρη δίνεται από τον τύπο:

$$\text{Cov}[V'_k] = E[V'_k V'_k{}^T] \quad (3.127)$$

Για τη μέθοδο του Newton έχουμε:

$$\begin{aligned} V'_k &= -\mu \sum_{n=0}^{k-1} (I - 2\mu\Lambda)^n N'_{k-n-1} V'_k V'_k{}^T = \\ &= (1 - 2\mu)V'_{k-1} V'_{k-1}{}^T + \mu^2 \Lambda^{-1} N'_{k-1} N'_{k-1}{}^T (\Lambda^{-1})^T - \mu(1 - 2\mu) [V'_{k-1} N'_{k-1}{}^T (\Lambda^{-1})^T + \Lambda^{-1} N'_{k-1} V'_{k-1}{}^T] \end{aligned} \quad (3.128)$$

Ο Λ^{-1} είναι διαγώνιος και επομένως $\Lambda^{-1} = (\Lambda^{-1})'$ και V'_k και N'_k , θεωρούμε ότι είναι ανεξάρτητα με μηδενικό μέσο

$$\begin{aligned} \text{Cov}[V'_k] &= (1 - 2\mu)^2 \text{cov}[V'_k] + \mu^2 (\Lambda^{-1})^2 \text{cov}[N'_k] = \\ &= \frac{\mu(\Lambda^{-1})^2}{4(1 - \mu)} \text{cov}[N'_k] \end{aligned} \quad (3.129)$$

Για την μέθοδο Ατότομης Κατάβασης έχουμε:

$$\begin{aligned} V'_k V'_k{}^T &= (1 - 2\mu\Lambda) V'_{k-1} V'_{k-1}{}^T (1 - 2\mu\Lambda)^T + \mu^2 N'_{k-1} N'_{k-1}{}^T - \\ &- \mu [(1 - 2\mu\Lambda) V'_{k-1} N'_{k-1}{}^T + N'_{k-1} V'_{k-1}{}^T (1 - 2\mu\Lambda)^T] \end{aligned} \quad (3.130)$$

Ομοίως

$$\begin{aligned} \text{Cov}[V'_k] &= (1 - 2\mu\Lambda)^2 \text{cov}[V'_{k-1}] + \mu^2 \text{cov}[N'_k] \\ \text{Cov}[V'_k] &= \frac{\mu}{4} (\Lambda - \mu\Lambda^2)^{-1} \text{cov}[N'_k] \end{aligned} \quad (3.131)$$

Χρησιμοποιώντας τις σχέσεις (3.112) και (3.117) μπορούμε να γράψουμε

$$\begin{aligned} \text{Cov}[N'_k] &= E[N'_k N'_k{}^T] = Q^{-1} E[N_k N_k{}^T] Q \\ \text{Cov}[N'_k] &= Q^{-1} E[(\hat{\nabla}_k - \nabla_k)(\hat{\nabla}_k - \nabla_k)^T] Q \\ \text{Cov}[N'_k] &= Q^{-1} \text{cov}[\hat{\nabla}_k] Q \\ \text{Cov}[N'_k] &= \frac{\xi_{\min}^2}{N\delta^2} Q^{-1} I Q \\ \text{Cov}[N'_k] &= \frac{\xi_{\min}^2}{N\delta^2} I \end{aligned} \quad (3.132)$$

Συγκεντρωτικά έχουμε:

$$\text{Newton's Method:} \quad \text{Cov}[V'_k] = \frac{\mu(\Lambda^{-1})^2 \xi_{\min}^2}{4N\delta^2(1-\mu)} \quad (3.133)$$

$$\text{Steepest descent:} \quad \text{Cov}[V'_k] = \frac{\mu(\Lambda - \mu\Lambda^2)^{-1} \xi_{\min}^2}{4N\delta^2} \quad (3.134)$$

Για να περιστρέψουμε ξανά την επιφάνεια πρέπει να λάβουμε υπόψη ότι:

$$\text{Cov}[V_k] = E[V_k V_k{}^T] = Q E[V'_k V'_k{}^T] Q^{-1} = Q \text{Cov}[V'_k] Q^{-1} \quad (3.135)$$

Επομένως μπορούμε γράψουμε:

$$\text{Newtons Method:} \quad \text{Cov}[V'_k] = \frac{\mu \xi_{\min}^2 (R^{-1})^2}{4N\delta^2(1-\mu)} \quad (3.136)$$

$$\text{Steepest Descent:} \quad \text{Cov}[V'_k] = \frac{\mu \xi_{\min}^2 (R - \mu R^2)^{-1}}{4N\delta^2} \quad (3.137)$$

3.5.6 Πλεονάζον Μέσο Τετραγωνικό Σφάλμα και η Επίδραση του στη Διαδικασία Προσαρμογής

Αν δεν υπήρχε θόρυβος η διαδικασία προσαρμογής και με τις 2 μεθόδους (Newton και απότομη κατάβασης), θα προκαλούσε τα βάρη να καταλήξουν σε μια steady state λύση, στο ελάχιστο σημείο της επιφάνειας του MSE θα ήταν ίσο με ξ_{\min} . Παρόλα αυτά ο

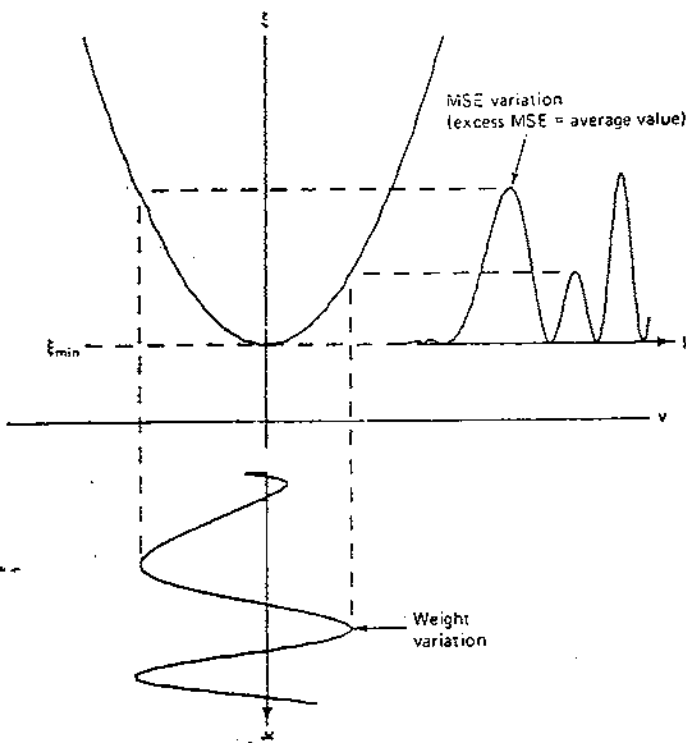
“θόρυβος” προκαλεί τον παράγοντα βάρους να κυμαίνεται με τυχαίο τρόπο γύρω από το ελάχιστο. Το αποτέλεσμα είναι το “ πλεονάζον ” MSE και η τιμή του ξ να είναι μεγαλύτερη από το ξ_{min} .

Αν ο παράγοντας βάρους W δεν έχει καταλήξει κάπου, τότε το στιγμιαίο MSE, ορίζεται ως η αναμενόμενη τιμή του σφάλματος $W=W_k$ για τη δεδομένη στιγμή. Για να καταλήξουμε σε κάποιο μαθηματικό τύπο για το “πλεονάζον” MSE, θα πρέπει να οριστεί κάποια μέση τιμή για ένα τυχαίο και μεγάλο αριθμό βημάτων (iterations) για το ξ_k . Αυτό σημαίνει πως θα πρέπει να επιτραπεί στο N να είναι “θορυβώδες” και να κυμαίνεται γύρω από το ελάχιστο, έτσι ώστε να δούμε την επίδραση που θα έχει στο ξ_k . Χρησιμοποιώντας τη σχέση μεταξύ ξ και V μπορεί το “ πλεονάζον ” MSE, να οριστεί ως

$$\text{excess MSE} = E[\xi_k - \xi_{min}] = E[V_k^T R V_k] \quad (3.138)$$

Αυτός ο ορισμός έχει εφαρμογή μόνο όταν το V_k περνάει από μια στάσιμη διαδικασία, δηλαδή θα πρέπει ο θόρυβος N_k , το input signal x_k και το επιθυμητό αποτέλεσμα d_k να είναι στάσιμα. Ακόμη έχει εφαρμογή μόνο σε steady state κατάσταση, δηλαδή όταν η προσαρμοστική (adaptive) διαδικασία δεν προχωράει άλλο.

Η φυσική σημασία αυτών απεικονίζεται στο Σχήμα 3.11



Σχήμα 3.11: Αναπαράσταση του excess MSE

Ο τύπος για το “πλεονάζον” MSE (3.138) είναι παρόμοιος με συνδιακύμανση V'_k . Έτσι θα πρέπει ξανά να διαχωρίσουμε σε 2 περιπτώσεις στη μέθοδο του Newton και στη μέθοδο της απότομης κατάβασης. Πριν προχωρήσουμε όμως σ’ αυτό μπορούμε να αλλάξουμε την (3.138) με βάση V'_k , έτσι έχουμε:

$$\text{excessMSE} = E[V'^T \Lambda V'] \quad (3.139)$$

$$V^T R V = V'^T \Lambda V' \quad (3.140)$$

Μεθοδος Newton : $r = 1 - 2\mu$

$$\text{excessMSE} = \mu^2 E \left[\sum_{n=0}^{\infty} r^n N'_{k-n-1}{}^T \Lambda^{-1} \Lambda \Lambda^{-1} \sum_{m=0}^{\infty} r^m N'_{k-m-1} \right]$$

$$\text{ή ισοδύναμα } \text{excessMSE} = \mu^2 \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} r^{n+m} E[N'_{k-n-1}{}^T \Lambda^{-1} N'_{k-m-1}] \quad (3.141)$$

Έχει ήδη ειπωθεί πως ο θόρυβος του ανάδελτα, N , που προέρχεται από ανεξάρτητα σφάλματα, είναι ανεξάρτητος από τη μία επανάληψη στην άλλη. Έτσι όλοι οι όροι όπου $m \neq n$ θα πρέπει να εξαφανιστούν.

$$\text{excessMSE} = \mu^2 \sum_{n=0}^{\infty} r^{2n} E[N'_{k-n-1}{}^T \Lambda^{-1} N'_{k-n-1}] \quad (3.142)$$

Ακόμη ο θόρυβος του ανάδελτα θεωρείται ότι είναι στάσιμος έτσι η παραπάνω σχέση γίνεται:

$$\begin{aligned} \text{excessMSE} &= \mu^2 E[N'_k{}^T \Lambda^{-1} N'_k] \sum_{n=0}^{\infty} r^{2n} \\ \text{excessMSE} &= \frac{\mu^2}{1-r^2} E[N'_k{}^T \Lambda^{-1} N'_k] \end{aligned} \quad (3.143)$$

Επίσης έχει ήδη υποτεθεί ότι η διαδικασία είναι σταθερή με $r < 1$. Έτσι για $N'_k{}^T \Lambda N'_k$ έχουμε:

$$\begin{aligned} N'_k{}^T \Lambda^{-1} N'_k &= \begin{bmatrix} n_{0k} & \dots & n_{Lk} \end{bmatrix} \begin{bmatrix} \lambda_0^{-1} & & 0 \\ & \ddots & \\ 0 & & \lambda_L^{-1} \end{bmatrix} \begin{bmatrix} n_{0k} \\ \vdots \\ n_{Lk} \end{bmatrix} \\ N'_k{}^T \Lambda^{-1} N'_k &= \sum_{m=0}^L \lambda_m^{-1} n_{mk}^2 \end{aligned} \quad (3.144)$$

από (3.147) έχουμε

$$\text{excessMSE} = \frac{\mu^2}{1-r^2} \sum_{m=0}^L \lambda_m^{-1} E[n_{mk}^{\prime 2}] \quad (3.145)$$

Όμως το $E[n_{mk}^{\prime 2}]$ είναι στοιχείο της διαγωνίου του πίνακα N' . Είναι γνωστό πως $\xi_{\min}^2 / N\delta^2$

$$\text{excessMSE} = \frac{\xi_{\min}^2 \mu^2}{N\delta^2(1-r^2)} \sum_{m=0}^L \frac{1}{\lambda_m} \quad (3.146)$$

χρησιμοποιώντας τη σχέση $r=1-2\mu$ έχουμε

$$\text{excessMSE} = \frac{\xi_{\min}^2 \mu}{4N\delta^2(1-\mu)} \sum_{m=0}^L \frac{1}{\lambda_m} \quad (3.147)$$

Θα ήταν όμως καλύτερο για πρακτικούς λόγους, να εκφραστεί αυτό το αποτέλεσμα με βάση κάποιους συντελεστές χρόνου. Για να καθοριστεί ο αριθμός αυτών των συντελεστών πρέπει να κατασκευαστεί ένας εκθετικός χώρος για την γεωμετρική συχνότητα των δειγμάτων.

Ο χώρος αυτός μπορεί να εκφραστεί από $\exp(-t/\tau)$, όπου t είναι ο χρόνος και τ ο συντελεστής χρόνου. Αν μια μονάδα χρόνου αναλογεί σε μία επανάληψη, τότε:

$$\exp(-1/\tau) = r$$

ο οποίο μπορεί να αναπτυχθεί σε σειριακές ακολουθίες:

$$r = \exp(-1/\tau) = 1 - \frac{1}{\tau} + \frac{1}{2!\tau^2} - \frac{1}{3!\tau^3} + \dots$$

Εφόσον στις περισσότερες εφαρμογές το τ είναι μεγαλύτερο από 10 και το r είναι μικρό (μικρότερο αλλά κοντά στο 1), τότε μπορούμε να γράψουμε:

$$r \cong 1 - \frac{1}{\tau}$$

χρησιμοποιώντας τη σχέση $r=1-2\mu$ έχουμε

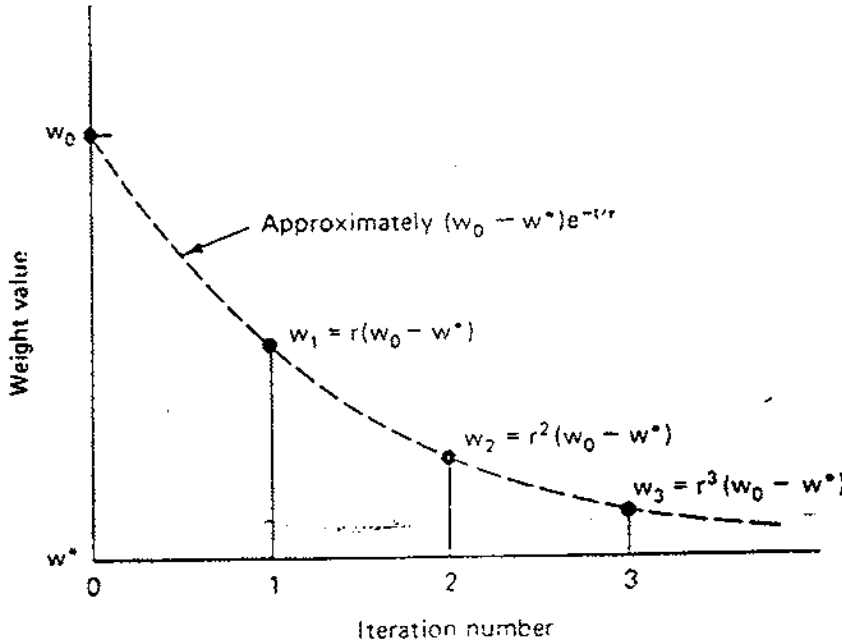
$$\mu \approx \frac{1}{2\tau}$$

ισοδύναμα

$$\text{excessMSE} \approx \frac{\xi_{\min}^2}{4N\delta^2(2\tau-1)} \sum_{m=0}^L \frac{1}{\lambda_m}$$

$$\text{excessMSE} \approx \frac{\xi_{\min}^2}{8N\delta^2} \sum_{m=0}^L \frac{1}{\lambda_m} \quad (\text{για μεγάλο } \tau) \quad (3.148)$$

Η παραπάνω σχέση μπορεί να απλοποιηθεί αν εκφράσουμε το άθροισμα με κάποιο μέσο.



Σχήμα 3.12: Εκθετική προσέγγιση

$$\sum_{m=0}^L \frac{1}{\lambda^m} = (L+1) \left(\frac{1}{\lambda}\right)_{av}. \quad (3.149)$$

Χρησιμοποιώντας τη διαταραχή (perturbation), δηλαδή τη μέση αύξηση στο MSE εξομαλυνμένο με βάρη το ξ_{min} .

$$\text{excess MSE} \approx \frac{(L+1)\xi_{min}(1/\lambda)_{av}}{8NP\tau} \quad (3.150)$$

Μέθοδος Απότομης Κατάβασης: Για να βρούμε το "υπερβάλον» MSE για τη μέθοδο Steepest Decent, ακολουθούμε παρόμοια διαδικασία:

$$\begin{aligned} \text{excess MSE} &= E[V_k^T \Lambda V_k] \\ \text{excess MSE} &= \mu^2 \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} E[N'_{k-n-1}^T \Lambda (1-2\mu\Lambda)^{n+m} N'_{k-m-1}] \end{aligned} \quad (3.151)$$

Ομοίως με τη μέθοδο του Newton έχουμε:

$$\begin{aligned} \text{excessMSE} &= \mu^2 \sum_{n=0}^{\infty} E[N_k'^T \Lambda (I - 2\Lambda)^{2n} N_k'] \\ \text{excessMSE} &= \mu^2 E \left[N_k'^T \Lambda \left(\sum_{n=0}^{\infty} (I - 2\Lambda)^{2n} \right) N_k' \right] \\ \text{ή } \sum_{n=0}^{\infty} (I - 2\Lambda)^{2n} &= \frac{1}{4\mu} (\Lambda - \mu\Lambda^2)^{-1} \end{aligned} \quad (3.152)$$

$$\begin{aligned} \text{ισοδύναμα } \text{excessMSE} &= \mu^2 E \left[N_k'^T \frac{1}{4\mu} (\Lambda - \mu\Lambda^2)^{-1} N_k' \right] \\ \text{excessMSE} &= \frac{\mu}{4} E \left[N_k'^T \frac{1}{4\mu} (\Lambda - \mu\Lambda^2)^{-1} N_k' \right] \end{aligned}$$

Έχουμε καταλήξει σ' ένα αποτέλεσμα παρόμοιο με το (3.143) αν χρησιμοποιήσουμε $\mu/4$ αντί για $\mu^2/(\mu-r^2)$ και $(I-\mu\Lambda)$ αντί για Λ .

$$\text{excessMSE} = \frac{\mu \xi_{\min}^2}{4N\delta^2} \sum_{m=0}^L \frac{1}{1 - \mu\lambda_m} \quad (3.153)$$

$$\text{excessMSE} = \frac{\mu \xi_{\min} \lambda_{av}}{4NP} \sum_{m=0}^L \frac{1}{1 - \mu\lambda_m} \quad (3.154)$$

Όπου P είναι η διαταραχή από τη μέτρηση της παραγώγου ίση με $\delta^2 \lambda_{av} / \xi_{\min}$

Όπως και παραπάνω, θα ήταν χρήσιμο να ορίζαμε δυο άλλους συντελεστές χρόνου, που θα περιγράφουν την αποτελεσματικότητα και το ρυθμό σύγκλισης (rate of convergence) της προσαρμοστικής διαδικασίας. Αυτοί οι συντελεστές του χρόνου για την καμπύλη μάθησης, τ_{MSE} και ο συντελεστής χρόνου της προσαρμογής (adaptation), T_{MSE} .

Αυτοί οι συντελεστές επίσης σχετίζονται με το γεωμετρικό λόγο r των βαρών. Το r για τη καμπύλη μάθησης είναι

$$\begin{aligned} r_{MSE} &= r^2 \\ \exp\left(-\frac{1}{\tau_{MSE}}\right) &= \exp\left(-\frac{2}{\tau}\right) = r^2 = r_{MSE} \\ \tau_{MSE} &= \frac{\tau}{2} \end{aligned} \quad (3.155)$$

Η βασική μονάδα του T_{MSE} είναι ο αριθμός των επαναλήψεων. Από την άλλη η βασική μονάδα για το T_{MSE} είναι το δείγμα των δεδομένων. Έτσι εφόσον χρειάζονται $2N$ δείγματα για κάθε παράγοντα της μέτρησης της παραγώγου, για κάθε επανάληψη. Χρειάζονται $2(L+1)*N$. έτσι:

$$T_{MSE} = 2(L+1)N\tau_{mse} = N(L+1)\tau \quad (3.156)$$

$$\tau_n = 1 - 2\mu\lambda_n \quad (3.157)$$

$$\tau_n \approx 1 - \frac{1}{\tau_n} \quad (\text{για μεγάλο } \tau_n) \quad (3.158)$$

από τις (3.161) και (3.162) έχουμε

$$\mu\lambda_n \approx \frac{1}{2\tau_n} \quad (3.159)$$

$$\text{ισοδύναμα } (\tau_{MSE})_n = \frac{\tau_n}{2} \quad (3.160)$$

ομοίως $T_{MSE} = N(L+1)\tau_n = 2N(L+1)(\tau_{mse})_n$

$$\text{ισοδύναμα } \mu\lambda_n \approx \frac{N(L+1)}{2(T_{MSE})_n} \quad (3.161)$$

$$\text{ή } \mu\lambda_{n,av} \approx \frac{N(L+1)}{2} \left(\frac{1}{T_{MSE}} \right)_{av} \quad (3.162)$$

Έτσι το πλεονάζον MSE για τη μέθοδο Απτόμης Κατάβασης μπορεί να βρεθεί αντικαθιστώντας αυτά τα αποτελέσματα με την (3.154), στην (3.150)

$$\text{excessMSE} \approx \frac{(L+1)\xi_{\min}}{8P} \left(\frac{1}{T_{MSE}} \right)_{av} \sum_{m=0}^L \frac{1}{1-1/2\tau_m} \quad (3.163)$$

Έχει γίνει η υπόθεση πως η steady state λύση του W είναι κοντά στο βέλτιστο w^* . Σύμφωνα μ' αυτό η διαδικασία εξελίσσεται αργά, με μικρό μ και μεγάλο T_n . Έτσι μπορούμε να γράψουμε:

$$\frac{1}{1-1/2\tau_n} = 1, \quad n = 0,1,\dots,L$$

ή (3.164)

$$\text{excessMSE} \approx \frac{(L+1)\xi_{\min}}{8P} \left(\frac{1}{T_{MSE}} \right)_{av}$$

Συνοπτικά γι' αυτήν την ενότητα μπορούμε να γράψουμε:

$$\begin{aligned} \text{excess MSE} &= \frac{(L+1)\xi_{\min}\lambda_{av}(1/\lambda)_{av}}{8NP\tau} \\ &= \frac{(L+1)^2\xi_{\min}\lambda_{av}(1/\lambda)_{av}}{8NP T_{MSE}} \quad (\alpha) \end{aligned}$$

$$\text{excess MSE} = \frac{(L+1)^2\xi_{\min}}{8NP} \left(\frac{1}{T_{MSE}} \right)_{av} \quad (\beta)$$

όπου $L+1$ = ο αριθμός των βαρών
 ξ_{\min} = το ελάχιστο MSE
 N = ο αριθμός των παρατηρήσεων για το σφάλμα
 P = Perturbation
 τ = ο συντελεστής χρόνου για τη διαδικασία σύγκλισης
 T_{MSE} = ο συντελεστής χρόνου για τη διαδικασία προσαρμογής
 λ = ιδιοτιμή του πίνακα συσχέτισης εισόδου (input correlation matrix)

Πίνακας 3.2: Συνοπτικός πίνακας του πλεονάζοντος MSE

3.5.7 Misadjustment

Το πλεονάζον MSE είναι το μέσο MSE μείον το ελάχιστο MSE. Είναι ένα μέτρο της διαφοράς ανάμεσα στην πραγματική και στη βέλτιστη επίδοση. Ένα άλλο μέτρο της διαφοράς αυτής, εξαιρετικά χρήσιμο είναι το misadjustment M . Αυτό ορίζεται από το "υπερβάλον" MSE προς το ελάχιστο MSE:

$$M = \frac{\text{excess MSE}}{\xi_{\min}} \quad (3.165)$$

Το misadjustment είναι ένας καθαρός αριθμός της διαφοράς ανάμεσα στην προσαρμοστική και στη βέλτιστη επίδοση ως αποτέλεσμα του θορύβου της εκτίμησης

της παραγώγου. Παρατηρούμε πως το M δεν περιλαμβάνει την διαταραχή, η οποία οφείλεται στη διακύμανση των βαρών παρά εξαιτίας του θορύβου.

Το M στην περίπτωση του Newton θα είναι:

$$M \approx \frac{(L+1)\lambda_{av.}(1/\lambda)_{av.}}{8NP\tau}$$

ή

$$M \approx \frac{(L+1)^2 \lambda_{av.}(1/\lambda)_{av.}}{8NPT_{MSE}} \quad (3.166)$$

Η παραπάνω σχέση προσφέρει ένα τρόπο εκτίμησης της μεθόδου Newton. Παρατηρούμε πως το M μειώνεται καθώς η διαταραχή και ο συντελεστής χρόνου αυξάνεται. Μεγαλύτερη διαταραχή επιτρέπει πιο ακριβείς μετρήσεις της παραγώγου με δεδομένη την ποσότητα των δεδομένων, ενώ μεγαλύτερος συντελεστής χρόνου επιτρέπει σε περισσότερα δεδομένα να υπολογίσουν τη μέση τιμή. Ακόμη το M μαζί το τετράγωνο του αριθμού των βαρών $((L+1)^2)$.

Όσον αφορά το πλεονάζον MSE από την σχέση (α) του πίνακα 3.3 μπορούμε να δούμε πως για τη μέθοδο του Newton παραμένει σταθερό αν και ο συντελεστής χρόνου διπλασιαστεί ενώ ταυτόχρονα το N γίνει μισό. Το τ μπορεί να διπλασιαστεί διαιρώντας με το μ δια 2 και το N μπορεί να μειωθεί παίρνοντας λιγότερα δείγματα σφάλματος σε κάθε επανάληψη. Αντίθετα στο Misadjustment ο σημαντικός παράγοντας, φαίνεται και από (3.161) είναι η ποσότητα των δεδομένων που επεξεργάζεται τη μονάδα συντελεστή χρόνου προσαρμογής.

Ο αλγόριθμος που βασίζεται στη μέθοδο Newton θα ήταν πιο αποδοτικός αν άλλαζε με κάποιο τρόπο η διαδικασία μέτρησης της παραγώγου. Αυτή η αλλαγή έγκειται στο να γίνει η μέτρηση του ξ στο σημείο $W = W_k$, να βρεθεί η διαφορά ανάμεσα σ' αυτή τη μέτρηση και σε άλλες μετρήσεις του ξ και να ρυθμιστεί ένα βάρος σε κάποια χρονική στιγμή με βάση το δ . Ο αριθμός των δεδομένων για κάθε διάνυμα του εκτιμητή του ανάδελτα θα πρέπει να αλλάξει με βάση το $(L+2)/2(L+1)=1/2$. η ανάλυση αυτής της διαδικασίας όμως θα ήταν περίπλοκη επειδή οι παράγοντες του θορύβου της παραγώγου E_k συσχετίζονται μεταξύ τους. Σε κάθε περίπτωση όμως το M θα συνέχιζε να αυξάνεται μαζί με το $(L+1)^2$

Για τη μέθοδο απότομης κατάβασης έχουμε

$$M \approx \frac{(L+1)^2}{8P} \left(\frac{1}{T_{MSE}} \right)_{av}. \quad (3.167)$$

Η παραπάνω σχέση είναι παρόμοια με την (3.166). Η τελευταία παράγει τον όρο $\lambda_{av} (1/\lambda)_{av}$ ο οποίος δεν εμφανίζεται στην (3.167), αντίθετα στην (3.167) υπάρχει ο όρος $(1/T_{MSE})_{av}$ ο οποίος στην (3.166) εμφανίζεται ως $1/T_{MSE}$. Έτσι βλέπουμε πως στη μέθοδο του Newton έχουμε ένα συντελεστή χρόνου ενώ στη μέθοδο απότομης κατάβασης βρίσκουμε $L+1$ διαφορετικά T_{MSE} . Ακόμη είναι προφανές πως και οι δύο τύποι θα ήταν ίδιοι αν όλες οι ιδιοτιμές του R ήταν ίσες. Υπό αυτή την προϋπόθεση η επιφάνεια σφάλματος θα είχε κυκλική συμμετρία και το αρνητικό της παραγώγου θα έδειχνε πάντα προς την κατεύθυνση του ελαχίστου. Με λίγα λόγια οι δύο μέθοδοι θα ήταν ταυτόσημες.

3.5.8 Συγκριτική Απόδοση των Μεθόδων του Newton και Απότομης Κατάβασης

Για δεδομένη διαταραχή και αριθμό βαρών και οι δύο μέθοδοι παρουσιάζουν μια αύξηση του misadjustment όταν αυξάνεται η ταχύτητα προσαρμογής. Παρόλα αυτά οι δύο μέθοδοι δεν έχουν αναγκαστικά το ίδιο misadjustment για την ίδια ταχύτητα. Σε μια δεδομένη κατάσταση καλύτερος αλγόριθμος θα είναι αυτός που θα έχει συγκεκριμένο επίπεδο misadjustment παρουσιάζει τη μεγαλύτερη ταχύτητα προσαρμογής ή για συγκεκριμένη ταχύτητα παρουσιάζει τη μικρότερη misadjustment.

Για να συγκριθούν οι δύο αυτές μέθοδοι ως προς αυτά τα χαρακτηριστικά θα πρέπει η (3.167) να εκφραστεί με διαφορετικό τρόπο. Η προσαρμογή με τη μέθοδο απότομης κατάβασης περιλαμβάνει πολλαπλούς συντελεστές χρόνου. Ενώ ο συντελεστής χρόνου είναι (T_{MSE}) .

$$\left(\frac{1}{T_{MSE}} \right)_{max} = \frac{(L+1)N}{2\mu\lambda_{min}} \quad (3.168)$$

Η μέση ιδιοτιμή λ_{av} μπορεί να γραφτεί:

$$\lambda_{av} = \frac{(L+1)N}{2\mu} \left(\frac{1}{T_{MSE}} \right)_{av}. \quad (3.169)$$

$$\text{ισοδύναμα} \left(\frac{1}{T_{\text{MSE}}} \right)_{\text{av.}} = \frac{\lambda_{\text{av}}}{\lambda_{\text{min}} (T_{\text{MSE}})_{\text{max}}} \quad (3.170)$$

Και το M μπορεί να επανεκφραστεί ως εξής:

$$M \approx \frac{(L+1)^2 \lambda_{\text{av.}}}{8P \lambda_{\text{min}} (T_{\text{MSE}})_{\text{max}}} \quad (3.171)$$

Αν το $(T_{\text{MSE}})_{\text{max}}$ της (3.170) είναι ίσο με το T_{MSE} και τα L και P είναι ίδια και στις δύο εξισώσεις τότε το misadjustment τόσο για τη μέθοδο του Newton όσο και για την μέθοδο απότομης κατάβασης είναι το ίδιο εκτός από τον όρο $(1/\lambda)$ για τη πρώτη μέθοδο και το $1/\lambda_{\text{min}}$ για τη δεύτερη.

Επομένως το misadjustment της μεθόδου Newton θα είναι γενικά μικρότερο, εφόσον.

$$\frac{1}{\lambda_{\text{min}}} \geq \left(\frac{1}{\lambda} \right)_{\text{av.}} \quad (3.172)$$

Όταν υπάρχει μεγάλος αριθμός ιδιοτιμών είναι δυνατό σε κάποιες περιπτώσεις να παρουσιάζονται ως όμοια κατανεμημένα ανάμεσα στα λ_{min} και στα λ_{max} . Έτσι το λ/λ θα είναι

$$\left(\frac{1}{\lambda} \right)_{\text{av.}} = \frac{1}{(\lambda_{\text{max}}) - \lambda_{\text{min}}} \int_{\lambda_{\text{min}}}^{\lambda_{\text{max}}} \frac{d\lambda}{\lambda} = \frac{\ln(\lambda_{\text{max}}/\lambda_{\text{min}})}{\lambda_{\text{max}} - \lambda_{\text{min}}} \quad (3.173)$$

Αν $\lambda_{\text{max}} = 10\lambda_{\text{min}}$ βλέπουμε πως το $(1/\lambda)$ αν θα είναι περίπου 0,26 φορές μεγαλύτερο από το $(1/\lambda_{\text{min}})$ και επομένως το M της απότομης κατάβασης θα είναι περίπου 4 φορές το M για τη μέθοδο του Newton.

3.5.9 Ολικό Misadjustment

Οι (3.166) και (3.167) δείχνουν πως το misadjustment είναι αντιστρόφως ανάλογο με το P . Επομένως, είναι προφανές και misadjustment μπορεί να γίνει όσο μικρό επιθυμούμε αυξάνοντας το perturbation. Όμως μια απεριόριστη αύξηση στο P δεν είναι δυνατή.

Η διαταραχή έχει οριστεί ως ένας καθαρός αριθμός του βαθμού του οποίου το MSE επηρεάζεται από την μέτρηση της παραγώγου. Το μέγεθος αυτό είναι παρόμοιο με το

misadjustment που προκαλείται από τη μέτρηση της παραγώγου σ' ένα " on-line" σύστημα. Το «ολικό» misadjustment ενός τέτοιου συστήματος είναι το άθροισμα των δυο αυτών μεγεθών. Έτσι

$$M_{tot} = M + P \quad (3.174)$$

$$\text{Newtons Method: } M_{tot} \approx \frac{(L+1)^2 \lambda_{av.}(1/\lambda)_{av.}}{8PT_{MSE}} + P \quad (3.175)$$

$$\text{Steepest Descent: } M_{tot} \approx \frac{(L+1)^2}{8P} \left(\frac{1}{T_{MSE}} \right)_{av.} + P \quad (3.176)$$

Βλέπουμε ότι αυτοί οι δύο τύποι μας δίνουν το M_{tot} με την μορφή $P+A/P$, όπου το A είναι συνάρτηση του P . Αν θέσουμε την παράγωγο αυτού ίσαμε το μηδέν ως n τότε θα βρούμε το βέλτιστο P , P_{opt} , το οποίο ελαχιστοποιεί το M_{tot}

$$P_{opt} \approx \frac{A}{P_{opt}} \approx M_{tot} - P_{opt} = 1/2 M_{tot} \quad (3.177)$$

Έτσι το ολικό misadjustment ελαχιστοποιείται όταν το perturbation είναι περίπου ίσο με το μισό M_{tot}

3.6 Ο Αλγόριθμος LMS

Στις προηγούμενες ενότητες είδαμε δυο μεθόδους για την διερεύνηση της επιφάνειας σφάλματος και την εύρεση του ελάχιστου σημείου. Όπως είδαμε σ' αυτές ήταν απαραίτητος ο υπολογισμός /εκτίμηση της παραγώγου η οποία έγινε με κάποιες μεθόδους οι οποίες έχουν ήδη αναλυθεί.

Σ' αυτή την ενότητα θα παρουσιαστεί ένας άλλος αλγόριθμος ανίχνευσης της επιφάνειας σφάλματος, γνωστός και σαν αλγόριθμος LMS (Least Mean Square). Ο αλγόριθμος αυτός χρησιμοποιεί ένα ειδικό μέτρο για την εκτίμηση της παραγώγου το οποίο ισχύει για τον Προσαρμοζόμενο Γραμμικό Συνδυαστή. Επομένως η χρήση του είναι περισσότερο περιορισμένη απο τους άλλους αλγόριθμους. Απο την άλλη, ο LMS είναι εξαιρετικής σημασίας εξαιτίας της απλότητας του, της ευκολίας υπολογισμού του και επειδή δεν απαιτεί off-line εκτίμηση της παραγώγου και επανάληψη δεδομένων. Αν το

σύστημα είναι ένα προσαρμοστικό γραμμικό σύστημα και αν τα x και d είναι διαθέσιμα σε κάθε επανάληψη, τότε ο LMS είναι ο καλύτερος αλγόριθμος.

3.6.1 Εξαγωγή του Αλγόριθμου LMS

Όπως έχουμε ήδη δει, το σφάλμα στον Προσαρμοζόμενο Γραμμικό Συνδυαστή είναι:

$$E_k = d_k - X_k^T W_k \quad (3.178)$$

Σύμφωνα με τις προηγούμενες μεθόδους θα πρέπει να προχωρήσουμε στη μέτρηση της παραγώγου $\xi = E[\varepsilon_k^2]$ παίρνοντας διαφορές ανάμεσα σε μέσους του ε_k^2 . Αντίθετα στον LMS παίρνουμε το ε_k^2 αυτούσιο ως εκτιμητή του ξ_k . Έτσι σε κάθε επανάληψη έχουμε μια παράγωγο της μορφής

$$\hat{\nabla}_k = \begin{bmatrix} \frac{\partial \varepsilon_k^2}{\partial w_0} \\ \vdots \\ \frac{\partial \varepsilon_k^2}{\partial w_L} \end{bmatrix} = 2\varepsilon_k \begin{bmatrix} \frac{\partial \varepsilon_k}{\partial w_0} \\ \vdots \\ \frac{\partial \varepsilon_k}{\partial w_L} \end{bmatrix} = -2\varepsilon_k X_k \quad (3.179)$$

Με αυτή την απλή μέτρηση της παραγώγου, μπορούμε να αναπτύξουμε έναν αλγόριθμο τύπου Απότομης Κατάβασης.

$$W_{k+1} = W_k - \mu \hat{\nabla}_k \quad (3.180)$$

$$W_{k+1} = W_k + 2\mu \varepsilon_k X_k \quad (3.181)$$

Αυτός είναι ο αλγόριθμος LMS. Όπως και προηγουμένως το μ είναι ο συντελεστής ο οποίος καθορίζει την ταχύτητα και τη σταθερότητα προσαρμογής. Και σ' αυτή την περίπτωση περιμένουμε η διαδικασία αυτή να είναι 'θορυβώδης' αφού βασίζεται σε προσεγγιστικό υπολογισμό της παραγώγου.

Απο την μορφή που έχει ο LMS στην τελευταία σχέση βλέπουμε πως μπορεί να υπολογισθεί χωρίς τετραγωνισμό, μέση τιμή και παραγωγή. Κάθε στοιχείο της παραγώγου εξάγεται από ένα και μόνο δείγμα χωρίς διαταραχή (perturbing) στο W . Έτσι αφού δεν υπάρχουν μέσες τιμές οι παράγωγοι δεν περιέχουν τόσο μεγάλο θόρυβο.

3.6.2 Σύγκλιση του Διανύσματος Βαρών (Convergence of the Weight Vector)

Όπως σε κάθε προσαρμοστικό αλγόριθμο έτσι και στον LMS πρωταρχικός στόχος είναι η μετάβαση προς το βέλτιστο διάνυσμα συντελεστών στο οποίο το $E[\varepsilon_k^2]$ ελαχιστοποιείται. Η αναμενόμενη τιμή της παραγώγου όταν $W_k=W$ είναι:

$$\begin{aligned} E[\hat{\nabla}] &= -2E[\varepsilon_k X_k] \\ E[\hat{\nabla}] &= -2E[d_k X_k - X_k X_k^T W] \\ E[\hat{\nabla}] &= 2(RW - P) = \nabla \end{aligned} \quad (3.182)$$

Εφόσον $\hat{\nabla}_k$ είναι ίσο με το ∇_k τότε το $\hat{\nabla}_k$ θα πρέπει να είναι αμερόληπτος εκτιμητής. Βλέποντας πως η εκτίμηση της παραγώγου είναι αμερόληπτη μπορούμε να υπολογίσουμε το ∇_k σε κάθε βήμα χωρίς να προσαρμόζουμε τα βάρη για αρκετά βήματα. Έτσι το $\hat{\nabla}_k$ μπορεί να προσεγγίσει το ∇_k .

Με το διάνυσμα των συντελεστών να αλλάζει σε κάθε επανάληψη πρέπει η μετάβαση να εξετασθεί με διαφορετικό τρόπο. Από την Σχέση (3.176) φαίνεται πως το W_k είναι συνάρτηση μόνο των παρελθόντων διανυσμάτων εισόδου. Υποθέτοντας πως συνεχόμενα διανύσματα εισόδου είναι ανεξάρτητα στη διάρκεια του χρόνου, το W_k είναι ανεξάρτητο από το X_k . Για στάσιμες διαδικασίες, η αναμενόμενη τιμή του W_k ύστερα από έναν αριθμό επαναλήψεων είναι:

$$\begin{aligned} E[W_{k+1}] &= E[W_k] + 2\mu E[\varepsilon_k X_k] \\ E[W_{k+1}] &= E[W_k] + 2\mu (E[d_k X_k] - E[X_k X_k^T W_k]) \end{aligned} \quad (3.183)$$

Ξέρουμε επίσης ότι $W^* = R^{-1}P$ και επομένως

$$\begin{aligned} E[W_{k+1}] &= E[W_k] + 2\mu(P - RE[W_k]) \\ E[W_{k+1}] &= (I - 2\mu R)E[W_k] + 2\mu RW^* \end{aligned} \quad (3.184)$$

Όπως έχουμε ήδη αποδείξει στις Σχέσεις (3.76) και (3.80) μπορεί να γράφει:

$$E[V_k'] = (I - 2\mu\Lambda)^k V_0' \quad (3.185)$$

Απο την παραπάνω εξίσωση συμπεραίνουμε πως καθώς το k αυξάνει, το διάνυσμα των συντελεστών τείνει στη βέλτιστη λύση, μόνο όταν το δεξιό μέλος τείνει στο μηδέν και όπως είδαμε στη Σχέση (3.83) αυτό συμβαίνει αν και μόνο αν:

$$\frac{1}{\lambda_{\max}} > \mu > 0 \quad (3.186)$$

Μέσα σε αυτά τα όρια, η ταχύτητα προσαρμογής και ο θόρυβος στην επίλυση του W καθορίζεται απο το μ . Επίσης παρατηρούμε πως το λ_{\max} δεν μπορεί να είναι μεγαλύτερο απο το ίχνος (trace) του R , που είναι το άθροισμα των διαγωνίων στοιχείων του R .

$$\lambda_{\max} \leq \text{tr}[R] \quad (3.187)$$

Η οριοθέτηση του μ εδώ είναι πιο περιοριστική, αλλά είναι πολύ πιο εύκολη. Οι υποθέσεις που έγιναν (π.χ. στασιμότητας) για το διάνυσμα εισόδου δεν είναι αναγκαστικές, απλά έγιναν για ευκολότερη ανάλυση.

3.6.3 Καμπύλη Μάθησης

Έχουμε ήδη δει την καμπύλη μάθησης για τη μέθοδο απότομης κατάβασης, η οποία ήταν θεωρητική γιατί δεν ήταν δυνατό να έχουμε την τιμή της παραγωγού σε κάθε επανάληψη. Αυτή η καμπύλη εξελίσσεται με ρυθμό γεωμετρικής προόδου, δηλαδή:

$$r_n = 1 - 2\mu\lambda_n, \quad n = 0, 1, \dots, L \quad (3.188)$$

Αποτέλεσμα αυτού είναι ο συντελεστής χρόνου να γίνει :

$$\tau_n \approx \frac{1}{2\mu\lambda_n} \quad (3.189)$$

ή ισοδύναμα

$$(\tau_{MSE})_n \approx \frac{1}{4\mu\lambda_n} \quad (3.190)$$

Επιπλέον, εφόσον ο LMS βασίζεται σε μια μόνο παρατήρηση τότε το τ_{MSE} είναι ίσο με T_{MSE} . Επομένως για τον LMS έχουμε:

$$(T_{MSE})_n = (\tau_{MSE})_n \approx \frac{1}{4\mu\lambda_n} \quad (3.191)$$

Στις περιπτώσεις εκείνες που το ε_k^2 είναι μια καλή προσέγγιση του $E[\varepsilon_k^2]$ σε κάθε επανάληψη, για τις τρέχουσες τιμές βάρους, ο LMS δίνει μια καλή προσέγγιση στους συντελεστές χρόνου για την καμπύλη μάθησης.

Γενικότερα, όμως, με τον LMS η προηγούμενη σχέση είναι μια αισιόδοξη εκτίμηση, αφού στη γενική περίπτωση το ε_k^2 δεν αποτελεί μια καλή προσέγγιση του $E[\varepsilon_k^2]$ και η διαδικασία μετάβασης στο ελάχιστο σημείο είναι ακανόνιστη.

3.6.4 Θόρυβος στην Επίλυση του Διανύσματος Βαρών (Noise in the Weight Vector Solution)

Έχει ήδη αναλυθεί η διακύμανση της εκτίμησης της παραγώγου και ο θόρυβος στο διάνυσμα βαρών υπό την προϋπόθεση ότι η εκτίμηση της παραγώγου βασίστηκε στην διαταραχή (perturbation) των βαρών. Στον LMS όμως, η εκτίμηση της παραγώγου δεν βασίζεται σ' αυτή (perturbation), έτσι θα πρέπει να επανεξετασθεί η διακύμανση του ανάδελα.

$$\hat{\nabla}_k = \nabla_k + N_k \quad (3.192)$$

Έστω ότι ο LMS χρησιμοποιώντας μικρό μ , έχει φτάσει σε μια steady state λύση κοντά στο W^* , έτσι στην τελευταία σχέση το ανάδελα θα είναι κοντά στο μηδέν και επομένως:

$$N_k = \hat{\nabla}_k = -2\varepsilon_k X_k \quad (3.193)$$

Άρα, ο πίνακας των συνδιασπορών δίνεται από την ακόλουθη σχέση:

$$\text{Cov}[N_k] = E[N_k N_k^T] = 4E[\varepsilon_k^2 X_k X_k^T] \quad (3.194)$$

Αν υποθέσουμε τώρα πως το W_k παραμένει κοντά στο W^* , τότε το σφάλμα είναι σχεδόν ασυσχέτιστο με το διάνυσμα εισόδου.

$$\text{Cov}[N_k] \approx 4E[\varepsilon_k^2]E[X_k X_k^T] = 4\xi_{\min} R \quad (3.195)$$

Μεταφερόμενοι τώρα στο κύριο σύστημα αξόνων, ο πίνακας συνδιασπορών γράφεται ισοδύναμα

$$\begin{aligned}
 \text{Cov}[N_k] &= \text{Cov}[Q^{-1}N_k] \\
 \text{Cov}[N_k] &= E\left[Q^{-1}N_k(Q^{-1}N_k)^T\right] \\
 \text{Cov}[N_k] &= Q^{-1}E[N_k N_k^T]Q \\
 \text{Cov}[N_k] &= Q^{-1}\text{Cov}[N_k]Q \\
 \text{Cov}[N_k] &= 4\xi_{\min}\Lambda
 \end{aligned}
 \tag{3.196}$$

Στην πραγματικότητα, σε πρακτικές εφαρμογές ο όρος $\mu\Lambda$ είναι μικρότερος της μονάδας.

Έτσι μπορούμε να απλοποιήσουμε την παραπάνω σχέση αγνοώντας τον όρο $\mu\Lambda^2$

$$\text{Cov}[V'_k] = \mu\xi_{\min}\Lambda^{-1}\Lambda = \mu\xi_{\min}I \tag{3.197}$$

Περιστρέφοντας ξανά το σύστημα συντεταγμένων έχουμε:

$$\begin{aligned}
 \text{Cov}[V'_k] &= Q\text{Cov}[V'_k]Q^{-1} \\
 \text{Cov}[V'_k] &= \mu\xi_{\min}QIQ^{-1} \\
 \text{Cov}[V'_k] &= \mu\xi_{\min}I
 \end{aligned}
 \tag{3.198}$$

3.6.5 Misadjustment

Ξέρουμε πως το misadjustment σε μια προσαρμοστική διαδικασία είναι ο λόγος του πλεονάζον MSE προς το ελάχιστο τετραγωνικό σφάλμα, και είναι επομένως ένα μέτρο του πόσο κοντά προσεγγίζει η διαδικασία τις βέλτιστες λύσεις, δηλαδή το «κόστος προσαρμοστικότητας» (cost of adaptability).

$$\text{excessMSE} = E[V'_k \Lambda V_k'^T]$$

Έστω $L+1$ τα στοιχεία του V'_k και Λ είναι διαγώνιος πίνακας τότε:

$$\text{Steepest Descent: } T_{\text{MSE}} = \frac{(L+1)^2}{8MP} = 50(L+1)^2 \tag{3.199}$$

$$\text{LMS: } T_{\text{MSE}} = \frac{L+1}{4M} = 2,5(L+1) \tag{3.200}$$

Αν υποθέσουμε πως η διαδικασία έχει 'λιμνάσει' που σημαίνει πως βρισκόμαστε κοντά στο ελάχιστο της επιφάνειας σφάλματος, μπορούμε να πούμε πως $E[V_{nk}'^2]$ είναι στοιχείο του $\text{Cov}[V'_k]$.

$$excessMSE = \mu \xi_{\min} \sum_{n=0}^L \lambda_n \quad (3.201)$$

$$excessMSE = \mu \xi_{\min} tr[R]$$

Χρησιμοποιώντας τις παραπάνω σχέσεις και τον ορισμό του misadjustment έχουμε:

$$M = \frac{excessMSE}{\xi_{\min}} = \mu tr[R] \quad (3.202)$$

Βλέπουμε πως το M είναι ευθέως ανάλογο με το μ . Επομένως υπάρχει μια σχέση ανάμεσα στο misadjustment και τον ρυθμό προσαρμογής (rate of adaption). Για να δούμε αυτή τη σχέση καλύτερα, παίρνουμε τον συντελεστή χρόνου της καμπύλης μάθησης:

$$(\tau_{MSE})_n = \frac{1}{4\mu\lambda_n} \quad (3.203)$$

$$tr[R] = \sum_{n=0}^L \lambda_n$$

ή

$$\frac{1}{4\mu} \sum \frac{1}{(\tau_{MSE})_n} = \frac{L+1}{4\mu} \left(\frac{1}{\tau_{MSE}_{av}} \right) \quad (3.204)$$

$$\text{ισοδύναμα } M \approx \frac{L+1}{4\mu} \left(\frac{1}{\tau_{MSE}_{av}} \right) \quad (3.205)$$

Στην ειδική περίπτωση που οι ιδιοτιμές είναι ίσες έχουμε:

$$M \approx \frac{L+1}{4\tau_{MSE}} \quad (3.206)$$

Η εμπειρία έχει δείξει πως η παραπάνω σχέση είναι μια καλή προσέγγιση της σχέσης ανάμεσα στο misadjustment, τον συντελεστή χρόνου και το πλήθος των συντελεστών. Μια τέτοια σχέση είναι απαραίτητη στο σχεδιασμό ενός προσαρμοστικού συστήματος όταν οι ιδιοτιμές είναι άγνωστες.

Εφόσον το $tr[R]$ είναι, ως γνωστόν, η συνολική ισχύς των εισόδων στα βάρη, η (3.198) μπορεί να χρησιμοποιηθεί για τον προσδιορισμό του μ . Χρησιμοποιώντας τις (3.202) και (3.198) εξάγουμε τον εξής τύπο:

$$(\tau_{MSE}) = \frac{L+1}{4\mu tr[R]} \quad (3.207)$$

Ετσι εφόσον η διαδικασία έχει 'λιμνάσει', καταλήγουμε στο ακόλουθο συμπέρασμα:

Το misadjustment ισούται με τον αριθμό των βαρών προς τον χρόνο, όταν οι ιδιοτιμες είναι ίσες.

3.6.6 Απόδοση

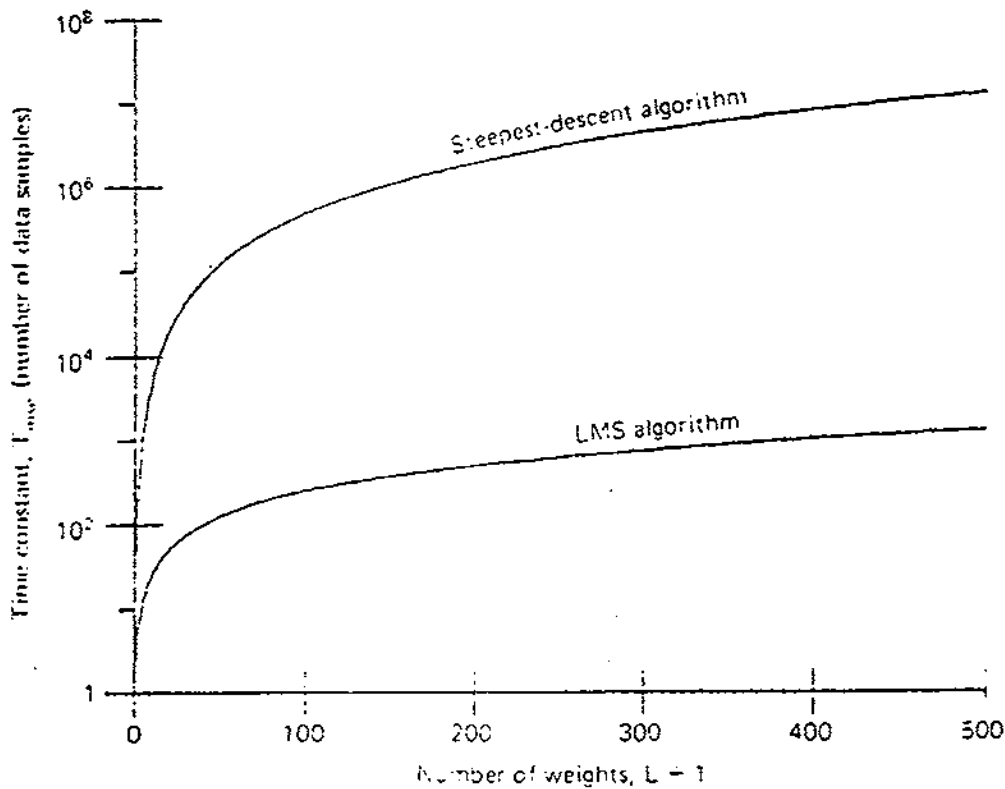
Η κύρια διάφορα του LMS από τους υπόλοιπους αλγόριθμους που αναλύθηκαν έγκειται στον τρόπο με τον οποίο εκτιμάται η παραγωγός σε κάθε βήμα. Ο LMS εκμεταλλεύεται επιπρόσθετες πληροφορίες που είναι γνωστές έκτων προτέρων, δηλαδή ότι η επιφάνεια σφάλματος είναι τετραγωνική. Ετσι περιμένουμε πως ο LMS θα έχει πλεονεκτήματα σε σχέση με τους υπόλοιπους αλγόριθμους που χρησιμοποιούν την μέθοδο των διαφορών για των προσδιορισμό του ∇_k .

Το πρακτικό πλεονέκτημα του LMS έναντι της μεθόδου απότομης κατάβασης μπορεί εύκολα να γίνει κατανοητό συγκρίνοντας τους τύπους για το misadjustment και το συντελεστή χρόνου. Είναι προφανές πως και στις δυο περιπτώσεις το misadjustment μειώνεται με αργή προσαρμογή, δηλαδή κάνοντας τον συντελεστή χρόνου πολύ μεγάλο. Για δεδομένο, όμως, συντελεστή χρόνου το M για τον LMS αυξάνεται ανάλογα με τον αριθμό των βαρών και όχι με το τετράγωνο και επομένως, υπό κανονικές συνθήκες, είναι δυνατή πιο γρήγορη προσαρμογή.

Επιπλέον, η διαφορά των δυο αυτών αλγορίθμων φαίνεται από τον συντελεστή χρόνου συνάρτηση του πλήθους των βαρών. Για να συγκρίνουμε, έστω το M του LMS είναι 10% και το M_{SD} για την μέθοδο απότομης κατάβασης είναι 10% με P_{opt} σύμφωνα με την (3.172). Ετσι:

$$\text{Steepest Descent: } T_{\text{MSE}} = \frac{(L+1)^2}{8MP} = 50(L+1)^2 \quad (3.208)$$

$$\text{LMS: } T_{\text{MSE}} = \frac{L+1}{4M} = 2,5(L+1) \quad (3.209)$$



Σχήμα 3.13: Σχέση της σταθεράς του χρόνου προσαρμογής και του αριθμού των βαρών για τη Steepest Descent και τον LMS

3.7 Διερεύνηση της επιφάνειας με Χρήση Γενετικών Αλγορίθμων

Ένας άλλος τρόπος διερεύνησης της επιφάνειας σφάλματος είναι οι Γενετικοί Αλγόριθμοι, στους οποίους αναφερθήκαμε σε προηγούμενη παραγραφο.

Το κυριότερο πρόβλημα που ένας γενετικός καλείται να λύσει είναι η έρευνα μέσα σε ένα σύνολο από υποψήφιες υποθέσεις - λύσεις, για να μπορέσει να εντοπίσει την καλύτερη υπόθεση. Αυτές οι υποθέσεις στους γενετικούς αλγορίθμους αναπαριστώνται με τη μορφή δυαδικών συμβολοσειρών που ονομάζονται και **χρωμοσώματα**. Βασικό δομικό στοιχείο των χρωμοσωμάτων είναι τα **γονίδια** τα οποία συνήθως αναπαριστούν σε δυαδική κωδικοποιημένη μορφή τις παραμέτρους του προβλήματος που πρέπει να βελτιστοποιηθούν. Το σύνολο των γονιδίων στην αποκωδικοποιημένη τους μορφή αποτελούν το **φαινότυπο** του χρωμοσώματος.

Για να μπορέσουν οι γενετικοί να θεωρήσουν μία λύση ως καλύτερη έναντι μιας άλλης, πρέπει να έχει οριστεί ένα μέτρο σύγκρισης μεταξύ των χρωμοσωμάτων. Τον ρόλο αυτό αναλαμβάνει η **συνάρτηση κόστους**, η οποία και αποδίδει αριθμητικές τιμές στα χρωμοσώματα. Για παράδειγμα, στην περίπτωση που έπρεπε να προβλεφτεί μια συνάρτηση βάση εισόδων και εξόδων της συνάρτησης, η συνάρτηση κόστους θα έδινε μικρότερη τιμή σε εκείνα τα χρωμοσώματα που προσέγγιζαν καλύτερα τα δεδομένα. Σαν ιδανικό χρωμόσωμα, θεωρείται εκείνο που δίνει την ελάχιστη δυνατή τιμή στην συνάρτηση κόστους.

Αν και έχουν προταθεί πολλές παραλλαγές γενετικών αλγορίθμων, γενικά όλοι ακολουθούν μια βασική δομή. Ξεκινώντας από ένα αρχικό σύνολο χρωμοσωμάτων, που καλείται **αρχικός πληθυσμός**, ο αλγόριθμος με μία επαναληπτική διαδικασία αλλάζει την μορφή των χρωμοσωμάτων. Όλα τα χρωμοσώματα αξιολογούνται ξανά και βαθμολογούνται βάση της συνάρτησης κόστους, έτσι ώστε να υπάρχει πάντα μια βαθμολογία επιλογής. Στη συνέχεια, για το πέρασμα σε ένα νέο πληθυσμό, ο αλγόριθμος διαλέγει με πιθανοτικό τρόπο τις οντότητες που θα περάσουν στην επόμενη γενιά χωρίς καμία αλλαγή, τις οντότητες που θα μεταλλαχθούν και τις οντότητες που θα συνδυαστούν για να δημιουργήσουν νέες οντότητες.

Ο ρόλος της συνάρτησης κόστους είναι σημαντικός καθώς, όπως θα δούμε και παρακάτω, η πιθανότητα για να επιλεγθεί κάποιο χρωμόσωμα και να περάσει χωρίς καμία αλλαγή στην επόμενη γενιά, εξαρτάται άμεσα από την συνάρτηση αυτή.

Μόλις επιλεγούν τα μέλη του πληθυσμού που θα επιβιώσουν ως έχουν, τα υπόλοιπα μέλη του πληθυσμού παράγονται με την **γενετική διασταύρωση** (crossover). Η γενετική διασταύρωση παίρνει δύο αρχικά χρωμοσώματα από την αρχική γενιά και τα συνδυάζει – μετασχηματίζει – έτσι ώστε να δημιουργήσει δύο νέα. Αξίζει να αναφερθεί πως για την επιλογή των δύο αρχικών χρωμοσωμάτων ακολουθούνται οι ίδιες πιθανότητες με αυτές της επιβίωσης, όπως θα δούμε αναλυτικά σε επόμενες παραγράφους.

Μόλις εκτελεστούν όλες οι γενετικές διασταυρώσεις και οι απευθείας μεταβάσεις από την παλιά στη νέα γενιά, τότε συμπληρώνεται ο επιθυμητός νέος πληθυσμός. Στην νέα γενιά, ένα συγκεκριμένο ποσοστό από τα μέλη της νέας γενιάς διαλέγονται τυχαία και γίνονται τυχαίες μεταλλάξεις που αλλάζουν την μορφή των μελών.

Στον παρακάτω πίνακα παρουσιάζεται συνοπτικά τα βήματα ενός γενετικού αλγορίθμου.

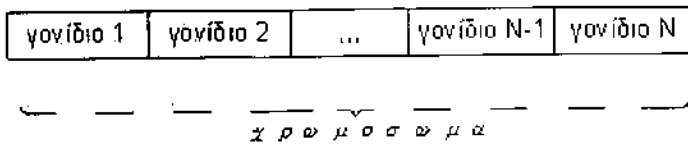
1.	Τυχαία δημιουργήσε ένα αρχικό πληθυσμό $X(0) = \{x_1, x_2, \dots, x_N\}$
2.	Υπολόγισε την fitness τιμή $F(x_i)$ για κάθε χρωμόσωμα x_i στον τρέχοντα πληθυσμό $X(t)$
3.	Δημιούργησε νέα χρωμοσώματα $X_i(t)$ ζευγαρώνοντας ορισμένα από τα τρέχοντα χρωμοσώματα και εφαρμόζοντας ύστερα τον τελεστή μετάλλαξης
4.	Διέγραψε ορισμένα χρωμοσώματα ώστε να δημιουργηθεί χώρος για τα νέα.
5.	Υπολόγισε την συνάρτηση fitness του $X_i(t)$ και εισήγαγέ τον στον πληθυσμό
6.	$t := t + 1$ αν δεν τελείωσε πήγαινε στο βήμα 3, διαφορετικά σταμάτα και επέστρεψε το καλύτερο χρωμόσωμα.

Πίνακας 3.4: Περιγραφή ενός απλού γενετικού αλγόριθμου

Στην συνέχεια παρουσιάζονται αναλυτικότερα οι βασικές έννοιες των γενετικών αλγορίθμων.

3.7.1 Χρωμόσωμα - Κωδικοποίηση Παραμέτρων

Ένα από τα βασικά δομικά στοιχεία ενός γενετικού αλγόριθμου είναι το χρωμόσωμα. Το χρωμόσωμα έχει συγκεκριμένη δομή που εξαρτάται κάθε φορά από το πρόβλημα που καλείται να λύσει ο γενετικός αλγόριθμος. Κάθε χρωμόσωμα αποτελείται από ένα σταθερό αριθμό γονιδίων τα οποία αντιπροσωπεύουν την κωδικοποιημένη, με ένα συγκεκριμένο τρόπο, μορφή των παραμέτρων. Η θέση που θα έχει μια συγκεκριμένη παράμετρος μέσα στο χρωμόσωμα παίζει σημαντικό ρόλο στην αποτελεσματικότητα του γενετικού αλγορίθμου. Η γενική μορφή που έχει το χρωμόσωμα φαίνεται στο σχήμα που ακολουθεί.



Σχήμα 3.14: Γενική μορφή χρωμοσώματος

Ο δυαδικός γενετικός αλγόριθμος δουλεύει με ένα πεπερασμένο (συνήθως αρκετά μεγάλο) πλήθος παραμέτρων. Αυτό το χαρακτηριστικό κάνει τον γενετικό αλγόριθμο ιδανικό για την βελτιστοποίηση κόστους το οποίο εξαρτάται από παραμέτρους που έχουν ένα πεπερασμένο πεδίο τιμών.

Εάν μια παράμετρος είναι συνεχής, τότε πρέπει να κβαντιστεί. Ο κβαντισμός μιας παραμέτρου ή ενός σήματος είναι μια αρκετά συνηθισμένη διαδικασία και μπορεί να γίνει ως ακολούθως. Πρώτα, το πεδίο τιμών χωρίζεται σε ίσα επίπεδα κβάντισης (*quantization levels*). Κάθε τιμή που πέφτει μέσα σε κάποιο από τα επίπεδα τίθεται ίση με την μέση, την μέγιστη ή την ελάχιστη τιμή του επιπέδου αυτού. Γενικά, το να τίθεται η τιμή ίση με την μέση τιμή του επιπέδου κβάντισης είναι το καλύτερο διότι το μέγιστο λάθος που μπορεί να υπάρξει είναι στο μέσο του επιπέδου. Στρογγυλοποιώντας μια τιμή στην ελάχιστη ή την μέγιστη τιμή του επιπέδου επιτρέπει ένα μέγιστο λάθος ίσο με το επίπεδο κβάντισης. Οι μαθηματικές εκφράσεις για την δυαδική κωδικοποίηση και αποκωδικοποίηση της n -οστής παραμέτρου, p_n , δίνονται από τις σχέσεις.

Κωδικοποίηση:

$$p_{norm} = \frac{p_n - p_{lo}}{p_{hi} - p_{lo}} \tag{3.210}$$

$$gene[m] = \text{round} \left\{ p_{norm} \cdot 2^{-m} - \sum_{p=1}^{m-1} gene[p] \cdot 2^{-p} \right\} \tag{3.211}$$

Αποκωδικοποίηση:

$$p_{quant} = \sum_{m=1}^{N_{gene}} gene[m] \cdot 2^{-m} + 2^{-(N_{gene}+1)} \tag{3.212}$$

$$q_n = p_{quant} (p_{hi} - p_{lo}) + p_{lo} \tag{3.213}$$

όπου:

p_{norm} = η κανονικοποιημένη παράμετρος, $0 \leq p_{norm} \leq 1$

p_{lo} = η ελάχιστη τιμή της παραμέτρου

p_{hi} = η μέγιστη τιμή της παραμέτρου

$gene[m]$ = η δυαδική τιμή της p_n

$round\{\}$ = στρογγυλοποίηση στον κοντινότερο ακέραιο

N_{gene} = αριθμός των bits στο γονίδιο

p_{quant} = κβαντισμένη έκδοση της p_{norm}

q_n = κβαντισμένη έκδοση της p_n

Ο γενετικός αλγόριθμος δουλεύει με τις δυαδικές κωδικοποιήσεις αλλά η συνάρτηση κόστους συνήθως χρησιμοποιεί συνεχείς παραμέτρους. Όποτε υπολογίζεται η συνάρτηση κόστους, το χρωμόσωμα πρέπει πρώτα να αποκωδικοποιηθεί χρησιμοποιώντας την σχέση (3.212).

Ένα παράδειγμα ενός δυαδικά κωδικοποιημένου χρωμοσώματος το οποίο έχει N_{par} παραμέτρους κάθε μία από τις οποίες είναι κωδικοποιημένη με $N_{gene} = 10\text{bits}$, είναι το ακόλουθο:

$$chromosome = \left[\underbrace{1111001001}_{gene_1} \underbrace{0011011111}_{gene_2} \dots \underbrace{0000101001}_{gene_{N_{par}}} \right]$$

Αντικαθιστώντας κάθε γονίδιο από το χρωμόσωμα στην σχέση (3.212) προκύπτει ένα διάνυσμα από τις κβαντισμένες εκδόσεις των παραμέτρων. Το παραπάνω χρωμόσωμα είναι μια δυαδική συμβολοσειρά (string) μήκους $N_{bits} = N_{gene} \times N_{par} = 10 \times N_{par}$ bits.

3.7.1.1 Αρχικός Πληθυσμός

Ο γενετικός αλγόριθμος ξεκινά με μια αρκετά μεγάλη κοινότητα χρωμοσωμάτων γνωστή ως *αρχικός πληθυσμός*. Ο αρχικός πληθυσμός έχει N_{pop} χρωμοσώματα και ο οποίος στην ανοιγμένη του μορφή είναι ένας $N_{pop} \times N_{bits}$ πίνακας ο οποίος περιέχει μονάδες και μηδενικά και κάθε γραμμή του αντιστοιχεί σ' ένα χρωμόσωμα. Στην συνέχεια οι παράμετροι που περιέχονται στα χρωμοσώματα περνούν στην συνάρτηση κόστους. Ένας μεγάλος αρχικός πληθυσμός παρέχει στον γενετικό αλγόριθμο ένα αρκετά καλό δείγμα του χώρου ανίχνευσης. Συνήθως, δεν χρησιμοποιείται όλος ο αρχικός

πληθυσμός στην μετέπειτα επαναληπτική διαδικασία του γενετικού αλγορίθμου αλλά ένα μέρος του συνήθως από τα καλύτερα χρωμοσώματα.

3.7.2 Φυσική Επιλογή

Ο αρχικός πληθυσμός είναι αρκετά μεγάλος ώστε να χρησιμοποιηθεί ολόκληρος στην επαναληπτική διαδικασία του γενετικού αλγορίθμου. Έτσι, ένα μεγάλο μέρος του από τα χρωμοσώματα με το μεγαλύτερο κόστος απορρίπτονται μέσω της φυσικής επιλογής ή της επιβίωσης του ισχυρότερου. Πρώτα, τα N_{pop} κόστη και τα σχετιζόμενα μ' αυτά χρωμοσώματα κατατάσσονται από το χαμηλότερο κόστος στο υψηλότερο. Τότε μόνο τα καλύτερα $N_{pop} \leq N_{ipop}$ μέλη του πληθυσμού κρατούνται για κάθε επανάληψη του γενετικού αλγορίθμου, ενώ τα υπόλοιπα απορρίπτονται. Αφήνοντας τον αρχικό πληθυσμό να είναι μεγαλύτερος από το πληθυσμό που χρησιμοποιείται στην συνέχεια από την επαναληπτική διαδικασία, δίνεται στον γενετικό αλγόριθμο μια καλή αρχή με το να παρέχεται σ' αυτόν ένα πολύ καλό δείγμα της επιφάνειας κόστους. Η φυσική επιλογή συμβαίνει σε κάθε γενιά ή επανάληψη του αλγορίθμου. Από τα N_{pop} χρωμοσώματα σε μια γενιά, μόνο τα N_{good} πρώτα επιζούν για το ζευγάρι και τα N_{bad} τελευταία απορρίπτονται για να κάνουν χώρο στους νέους απόγονους.

Η απόφαση για το πόσα χρωμοσώματα θα επιλεγθούν είναι λίγο πολύ αυθαίρετη. Αφήνοντας μόνο λίγα χρωμοσώματα να επιζήσουν στην επόμενη γενιά περιορίζουμε τα διαθέσιμα γονίδια των απογόνων. Αν κρατηθούν πάρα πολλά χρωμοσώματα τότε δίνεται η ευκαιρία σε υψηλού κόστους (κακά) χρωμοσώματα να περάσουν τα γνωρίσματά τους στην επόμενη γενιά. Συνήθως το ποσοστό επιλογής των χρωμοσωμάτων στην διαδικασία της φυσικής επιλογής είναι 50%.

Μια διαφορετική προσέγγιση που ακολουθείται κατά την φάση της φυσικής επιλογής είναι η μέθοδος του κατωφλίου (*thresholding*). Σ' αυτή την προσέγγιση, όλα τα χρωμοσώματα που έχουν ένα κόστος μικρότερο από μια τιμή κατωφλίου επιζούν. Η τιμή του κατωφλίου θα πρέπει να αφήνει ορισμένα χρωμοσώματα να συνεχίζουν ώστε να υπάρχουν γονείς για την παραγωγή νέων απογόνων, διαφορετικά θα πρέπει να δημιουργηθεί εκ νέου μια γενιά ώστε να βρεθούν χρωμοσώματα που να περνάνε το τεστ του κατωφλίου. Στην αρχή μόνο λίγα χρωμοσώματα θα επιζούν. Στις επόμενες γενιές, ωστόσο, τα περισσότερα από τα χρωμοσώματα θα επιζήσουν εκτός και αν η τιμή του

κατωφλίου αλλάζει. Ένα αρκετά ελκυστικό χαρακτηριστικό της μεθόδου είναι ότι ο πληθυσμός δεν χρειάζεται να ταξινομηθεί βάσει της τιμής της συνάρτησης κόστους και επομένως αποφεύγεται το υπολογιστικό κόστος της ταξινόμησης.

3.7.3 Επιλογή Γονέων

Για την δημιουργία των απογόνων επιλέγονται δύο χρωμοσώματα από τα N_{good} που αποτελούν το σύνολο των γονέων. Το ζευγάριμα συνεχίζεται μέχρι να δημιουργηθούν από τους επιλεγμένους γονείς, N_{bad} απόγονοι που θα αντικαταστήσουν τα χρωμοσώματα που απορρίφθηκαν στην φάση της φυσικής επιλογής. Υπάρχουν διάφορες μέθοδοι που μπορούν να χρησιμοποιηθούν για την επιλογή των γονέων οι οποίες παρουσιάζονται στην συνέχεια ξεκινώντας από την απλούστερη.

1. *Ζευγάριμα από την κορυφή μέχρι το τέλος (top to bottom)*. Ξεκινώντας από την κορυφή της λίστας συνδυάζονται ανά δύο τα χρωμοσώματα μέχρι να επιλεγθούν τα πρώτα N_{good} . Συνεπώς, ο αλγόριθμος αυτός ζευγαρώνει το $chromosome_{2i-1}$ με το $chromosome_{2i}$ για $i=1,2,\dots$. Η προσέγγιση αυτή αν και δεν μοντελοποιεί τον τρόπο που λειτουργεί η φύση έχει το πλεονέκτημα ότι μπορεί να προγραμματιστεί πολύ εύκολα σε έναν υπολογιστή.

2. *Τυχαίο ζευγάριμα*. Με την προσέγγιση αυτή χρησιμοποιείται μια γεννήτρια τυχαίων αριθμών με κανονική κατανομή για να επιλέξει τα ζευγάρια. Στα χρωμοσώματα αντιστοιχείται ένας αριθμός από το 1 έως το N_{good} και στην συνέχεια δύο τυχαίοι αριθμοί παράγονται για να επιλέξουν το ζευγάρι.

3. *Ζυγισμένο τυχαίο ζευγάριμα*. Αυτή η προσέγγιση αναθέτει στα χρωμοσώματα που ανήκουν στο σύνολο των γονέων πιθανότητες σύμφωνα με την τιμή της συνάρτησης κόστους. Ένα χρωμόσωμα με την μικρότερη τιμή κόστους έχει την μεγαλύτερη πιθανότητα να επιλεγθεί ως ένας από τους γονείς ενός ζευγαριού, αντίστοιχα το χρωμόσωμα με την μεγαλύτερη τιμή κόστους έχει την μικρότερη πιθανότητα να ζευγαρωθεί. Ένας τυχαίος αριθμός καθορίζει ποιο χρωμόσωμα θα επιλεγθεί. Αυτός ο τύπος του ζυγίσματος συχνά αναφέρεται και ως ζύγισμα ρουλέτας. Υπάρχουν δύο τεχνικές: το ζύγισμα με βάση τη θέση και το ζύγισμα με βάση το κόστος. Στην πρώτη τεχνική η πιθανότητα που θα αντιστοιχηθεί στο χρωμόσωμα βρίσκεται με βάση την θέση του χρωμοσώματος στην λίστα και δίνεται από τον παρακάτω τύπο

$$P_n = \frac{N_{good} - n + 1}{\sum_{n=1}^{N_{good}} n} \quad (3.214)$$

Στο ζύγισμα με βάση το κόστος η πιθανότητα υπολογίζεται χρησιμοποιώντας το κόστος του χρωμοσώματος. Ένα κανονικοποιημένο κόστος υπολογίζεται για κάθε χρωμόσωμα αφαιρώντας το μικρότερο κόστος από τα απορριφθέντα χρωμοσώματα ($cost_{N_{good}+1}$) από το κόστος του κάθε χρωμοσώματος από το σύνολο των γονέων:

$$C_n = cost_n - cost_{N_{good}+1} \quad (3.215)$$

Αφαιρώντας το $cost_{N_{good}+1}$ εξασφαλίζουμε ότι τα κανονικοποιημένα κόστη είναι αρνητικά.

Η απόλυτη τιμή $f_n = |C_n|$ του κανονικοποιημένου κόστους ονομάζεται και *τιμή fitness*. Η πιθανότητα που θα αντιστοιχηθεί σε κάθε χρωμόσωμα δίνεται από την παρακάτω σχέση

$$P_n = \left| \frac{C_n}{\sum_{p=1}^{N_{good}} C_p} \right| = \frac{f_n}{\sum_{p=1}^{N_{good}} f_p} \quad (3.216)$$

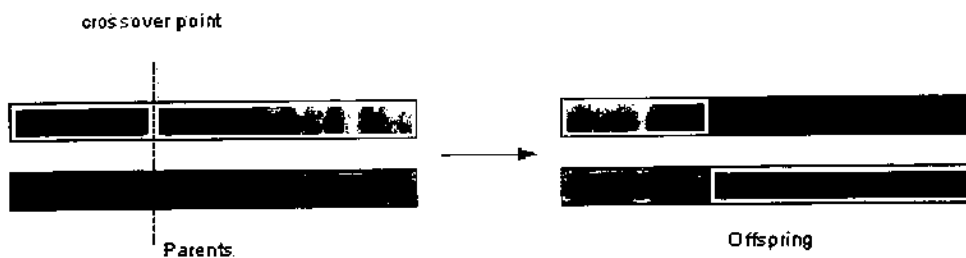
Αυτή η τεχνική τείνει να δίνει περισσότερη βαρύτητα στα χρωμοσώματα στην κορυφή όταν υπάρχει μεγάλη διαφορά μεταξύ του πρώτου και του τελευταίου χρωμοσώματος. Αντίθετα, τείνει να δίνει την ίδια βαρύτητα όταν τα χρωμοσώματα έχουν σχεδόν τα ίδια κόστη.

Για την επιλογή των γονέων και με τις δύο τεχνικές δημιουργείται ένας τυχαίος αριθμός μεταξύ του μηδέν και του ένα. Ξεκινώντας από την κορυφή της λίστας, το πρώτο χρωμόσωμα που η αθροιστική πιθανότητα είναι μεγαλύτερη από αυτόν τον αριθμό επιλέγεται ως γονέας. Αυτές οι τεχνικές προσεγγίζουν περισσότερο τον τρόπο με τον οποίο δουλεύει η φύση.

4. *Επιλογή πρωταθλήματος (tournament selection)*. Μια άλλη προσέγγιση η οποία μιμείται αρκετά τον ανταγωνισμό που γίνεται στην φύση είναι η τυχαία επιλογή ενός μικρού υποσυνόλου χρωμοσωμάτων από το σύνολο των γονέων και το χρωμόσωμα που έχει το μικρότερο κόστος γίνεται γονέας. Αυτή η διαδικασία επαναλαμβάνεται για κάθε γονέα. Ένας συνδυασμός της τεχνικής αυτής μαζί με την κατωφλίωση είναι αρκετά χρήσιμος γιατί δεν χρειάζεται έτσι να ταξινομείται ο πληθυσμός.

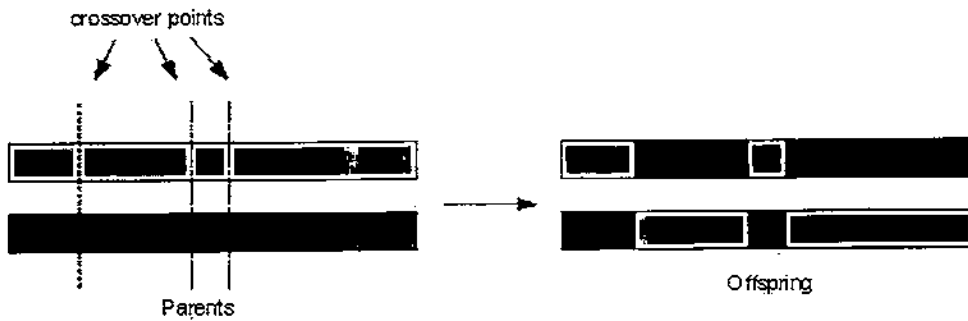
3.7.4 Δημιουργία Απογόνων - Διασταυρώσεις

Μετά την επιλογή των γονέων η επόμενη διαδικασία που ακολουθείται σε έναν γενετικό αλγόριθμο είναι αυτή της δημιουργίας των απογόνων από τους γονείς που επιλέχθηκαν. Είναι ο πρώτος τρόπος με τον οποίο ο γενετικός αλγόριθμος εξερευνά την επιφάνεια κόστους. Για την εξερεύνηση ο γενετικός αλγόριθμος χρησιμοποιεί τους συνδυασμούς των bit που υπάρχουν μέσα στα χρωμοσώματα. Η γενετική σύνθεση του πληθυσμού περιορίζεται από τα τρέχοντα μέλη του και βοηθά στην σύγκλιση του γενετικού αλγορίθμου. Η πιο κοινή μορφή δημιουργίας απογόνων περιλαμβάνει δύο γονείς που παράγουν δύο παιδιά σύμφωνα με την διαδικασία που φαίνεται στο Σχήμα 3.15. Ένα σημείο τομής ή διασταύρωσης (crossover point) επιλέγεται ανάμεσα στα πρώτα και τα τελευταία bits των χρωμοσωμάτων των γονέων. Αρχικά, ο πρώτος γονέας ($parent_1$) περνά τα bit αριστερά του σημείου διασταύρωσης στο πρώτο παιδί ($offspring_1$). Με τον ίδιο τρόπο ο δεύτερος γονέας ($parent_2$) περνά τα bit του αριστερά του σημείου διασταύρωσης στο δεύτερο παιδί ($offspring_2$). Στην συνέχεια, ο $parent_1$ περνά τα υπόλοιπα bits δεξιά του σημείου διασταύρωσης στο δεύτερο παιδί και ομοίως κάνει και ο δεύτερος γονέας. Οι γονείς παράγουν συνολικά N_{bad} απογόνους και έτσι ο συνολικός πληθυσμός φτάνει πάλι στους N_{pop} . Η παραπάνω διαδικασία ονομάζεται απλή διασταύρωση ενός σημείου.



Σχήμα 3.15: Διασταύρωση ενός σημείου

Ένα άλλο είδος διασταύρωσης είναι η διασταύρωση πολλαπλών σημείων. Αυτό το είδος της διασταύρωσης όπως φαίνεται και στο Σχήμα 3.16 είναι παρόμοιο με την απλή διασταύρωση με την διαφορά ότι επιλέγονται με τυχαίο τρόπο αντί ενός, m σημεία διασταύρωσης.



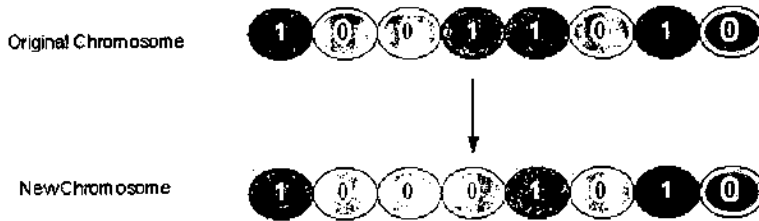
Σχήμα 3.16: Διασταύρωση πολλών σημείων ($m=3$)

Οι απλές και πολλαπλές διασταυρώσεις ορίζουν σημεία διασταύρωσης όπου το χρωμόσωμα μπορεί να χωριστεί, η *ομοιόμορφη διασταύρωση* γενικεύει αυτές τις μορφές. Σύμφωνα μ' αυτή, κάθε σημείο μπορεί να είναι ένα πιθανό σημείο διασταύρωσης. Μια τυχαία δυαδική συμβολοσειρά ίδιου μήκους με αυτό του χρωμοσώματος, δείχνει ποιος από τους γονείς θα προμηθεύσει στο παιδί το αντίστοιχο του bit. Σε κάθε θέση, τα αντίστοιχα bits των γονέων ανταλλάσσονται εάν η τυχαία δυαδική συμβολοσειρά περιέχει 1 σ' αυτές τις θέσεις. Εάν συμβαίνει το αντίθετο τότε δεν γίνεται καμία ανταλλαγή σ' αυτές τις θέσεις.

3.7.5 Μετάλλαξη

Ένας δεύτερος μηχανισμός που παρέχει την δυνατότητα στον γενετικό αλγόριθμο να εξερευνά την επιφάνεια κόστους είναι αυτός της *μετάλλαξης (mutation)*. Οι τυχαίες μεταλλάξεις αλλάζουν ένα μικρό ποσοστό των bits στην λίστα των χρωμοσωμάτων. Ο μηχανισμός αυτός μπορεί να εισαγάγει στα χρωμοσώματα γνωρίσματα τα οποία δεν υπάρχουν στον υπάρχοντα πληθυσμό και αποτρέπει την γρήγορη σύγκλιση του αλγορίθμου. Μια μετάλλαξη ενός σημείου αλλάζει την τιμή ενός bit του χρωμοσώματος από μηδέν σε ένα και αντίστροφα, όπως φαίνεται στο Σχήμα 3.17. Τα σημεία μετάλλαξης επιλέγονται τυχαία από τα $N_{pop} \times N_{bits}$ συνολικά bits του πληθυσμού. Αυξάνοντας τον αριθμό των μεταλλάξεων αυξάνεται η ελευθερία του αλγορίθμου να ψάχνει και έξω από την περιοχή του πεδίου τιμών των παραμέτρων. Επίσης τείνει να αποτρέπει τον αλγόριθμο από το να συγκλίνει σε μια λύση. Τυπικά, το ποσοστό των bits που υπόκεινται σε μεταλλάξεις κυμαίνεται από 1% σε 5% του συνόλου των bits του πληθυσμού σε κάθε επανάληψη του αλγορίθμου. Στο σημείο αυτό θα πρέπει να τονίσουμε ότι οι μεταλλάξεις δεν εφαρμόζονται στην τελευταία επανάληψη του αλγορίθμου, καθώς επίσης και σε

όλες τις επαναλήψεις του αλγορίθμου στο χρωμόσωμα του πληθυσμού που έχει το μικρότερο κόστος.



Σχήμα 3.17: Μετάλλαξη του 4^{ου} bit του αρχικού χρωμοσώματος

Οι περισσότερες μεταλλάξεις αυξάνουν το κόστος του χρωμοσώματος. Η σποραδική μείωση του κόστους των χρωμοσωμάτων προσθέτει ποικιλία και ενδυναμώνει τον πληθυσμό. Οι μεταλλάξεις παρέχουν καλύτερη εξερεύνηση των επιφανειών κόστους, επειδή συνήθως απομακρύνουν τον αλγόριθμο από τον μονοπάτι της σύγκλισης σε μια νέα περιοχή της επιφάνειας κόστους.

3.7.6 Θεωρία «Σχημάτων»

Λόγω της τυχαιότητας και της πολυπλοκότητας των γενετικών αλγορίθμων δεν υπάρχει κάποια μαθηματική απόδειξη ότι οι γενετικοί αλγόριθμοι συγκλίνουν σε ένα ολικό βέλτιστο [22]. Ένας προσεγγιστικός τρόπος με τον οποίο μπορεί να αποδειχθεί το παραπάνω είναι η χρήση της Θεωρίας «Σχημάτων» (Schema Theory) [20],[21].

Η μεθοδολογία του σχεδιασμού των γενετικών αλγορίθμων βασίζεται κύρια στην έννοια των «Σχημάτων» του Holland [23]. Τα «Σχήματα» είναι σύνολα από συμβολοσειρές (strings) τα οποία έχουν ένα ή περισσότερα κοινά χαρακτηριστικά. Ένα «Σχήμα» σχηματίζεται με την εισαγωγή ενός συμβόλου “don’t care”, “#”, μέσα στο αλφάβητο των γονιδίων π.χ. #11011#0. Ένα σχήμα αναπαριστά όλες τις συμβολοσειρές οι οποίες ταιριάζουν σε όλες τις θέσεις των συμβόλων εκτός απ’ αυτές που έχουν το #. Είναι προφανές ότι κάθε «Σχήμα» ταιριάζει με ακριβώς 2^r συμβολοσειρές, όπου r είναι ο αριθμός των “don’t care” συμβόλων. Για παράδειγμα, το σύνολο των συμβολοσειρών που προκύπτουν από το «Σχήμα» #1101#0 είναι {1110110, 1110100, 0110110, 0110100}.

3.7.6.1 Επίδραση των «Σχημάτων» στην Επιλογή

Εφ' όσον ένα «Σχήμα» αναπαριστά ένα σύνολο συμβολοσειρών, μπορεί να συνδεθεί μια fitness value $f(S_i, t)$ για κάθε συμβολοσειρά S_i του «Σχήματος» S καθώς και την μέση τιμή του «Σχήματος» $f(S, t)$ και η οποία θα καθορίζεται από όλες τις συμβολοσειρές που ανήκουν στο συγκεκριμένο «Σχήμα» και υπάρχουν στον πληθυσμό την χρονική στιγμή t .

Επομένως αν $\zeta(S, t)$ είναι ο αριθμός των συμβολοσειρών που υπάρχουν στην τρέχουσα γενιά και ανήκουν στο «Σχήμα» S , τότε η πιθανότητα για την επιλογή μίας συμβολοσειράς του «Σχήματος» (ζυγισμένο τυχαίο ζευγάρι) είναι ίση με $f(S, t)/F(t)$, όπου $F(t)$ είναι η μέση τιμή fitness του τρέχοντος πληθυσμού και κατά συνέπεια ο αναμενόμενος αριθμός εμφανίσεων συμβολοσειρών του S στην επόμενη γενιά είναι:

$$\zeta(S, t+1) = \zeta(S, t) \times \frac{f(S, t)}{F(t)} \quad (3.217)$$

Έστω:

$$\varepsilon = \frac{(f(S, t) - F(t))}{F(t)} \quad (3.218)$$

Εάν το $\varepsilon > 0$, σημαίνει ότι το «Σχήμα» έχει fitness τιμή πάνω από το μέσο όρο και το αντίστροφο.

Αντικαθιστώντας την εξίσωση (3.215) στην (3.214) μπορεί εύκολα να αποδειχτεί ότι ένα «Σχήμα» με τιμή fitness πάνω από τη μέση τιμή fitness του «Σχήματος», έχει μια εκθετική αύξηση των αριθμών συμβολοσειρών του στις επόμενες γενεές, δηλαδή:

$$\zeta(S, t) = \zeta(S, 0)(1 + \varepsilon)^t \quad (3.219)$$

3.7.6.2 Επίδραση των «Σχημάτων» στη Διασταύρωση

Κατά την διάρκεια της εξέλιξης ενός γενετικού αλγορίθμου, οι γενετικοί τελεστές δρουν διασπαστικά στα τρέχοντα «Σχήματα». Υποθέτουμε ότι το μήκος ενός χρωμοσώματος είναι L και ότι για την δημιουργία των απογόνων εφαρμόζεται η διασταύρωση ενός σημείου το οποίο επιλέγεται τυχαία από $L-1$ δυνατές θέσεις.

Αυτό συνεπάγεται ότι η πιθανότητα καταστροφής ενός «Σχήματος» S είναι:

$$p_d(S) = \frac{\sigma(S)}{L-1} \quad (3.220)$$

ή η πιθανότητα ένα «Σχήμα» να επιβιώσει είναι:

$$p_s(S) = 1 - \frac{\sigma(S)}{L-1} \quad (3.221)$$

όπου το σ είναι το “defining length” του «Σχήματος» S που ορίζεται ως η απόσταση μεταξύ των δύο εξώτατων σταθερών θέσεων του «Σχήματος». Αυτό ορίζει την περιεκτικότητα της πληροφορίας που έχει το «Σχήμα». Για παράδειγμα, στο «Σχήμα» #000# οι εξώτατες σταθερές θέσεις είναι η 2 και η 4 άρα το defining length είναι το $4-2=2$ ενώ στο 1#00# είναι η 1 και η 4 οπότε εδώ το defining length είναι $4-1=3$.

Θεωρώντας ότι η συχνότητα λειτουργίας του ζευγαρώματος είναι p_c τότε η πιθανότητα επιβίωσης του «Σχήματος» είναι:

$$p_s(S) \geq 1 - p_c \frac{\sigma(S)}{L-1} \quad (3.222)$$

3.7.6.3 Επίδραση των «Σχημάτων» στην Μετάλλαξη

Εάν η πιθανότητα μετάλλαξης ενός bit είναι p_m τότε η πιθανότητα ένα μόνο bit να επιβιώσει είναι $1 - p_m$. Ορίζοντας την τάξη του «Σχήματος» S $o(S)$ ως τον αριθμό των σταθερών θέσεων (π.χ. οι θέσεις με 0 ή 1) που υπάρχουν στο «Σχήμα», η πιθανότητα επιβίωσης ενός «Σχήματος» μετά από μετάλλαξη είναι:

$$p_s(S) = (1 - p_m)^{o(S)} \quad (3.223)$$

η πιθανότητα μπορεί να προσεγγιστεί με:

$$p_c(S) = 1 - o(S) \cdot p_m \quad (3.224)$$

3.7.7 Εξίσωση Ανάπτυξης των «Σχημάτων»

Αν συνδυαστούν τα αποτελέσματα της επιλογής, του ζευγαρώματος και της μετάλλαξης, σχηματίζεται μια νέα εξίσωση για την αναπαραγωγική αύξηση των «Σχημάτων»:

$$\zeta(S, t+1) \geq \zeta(S, t) \frac{f(S, t)}{F(t)} \left[1 - p_c \frac{\sigma(S)}{L-1} - o(S) \cdot p_m \right] \quad (3.225)$$

Από την εξίσωση (3.221) είναι εύκολο να συμπεράνουμε ότι μια υψηλή μέσης τιμής συνάρτηση fitness δεν αρκεί για να υπάρχει υψηλός βαθμός ανάπτυξης του «Σχήματος». Στην πραγματικότητα, μικρά, χαμηλής τάξης και πάνω από το μέσο όρο «Σχήματα» λαμβάνουν μέρος με εκθετικό και αυξητικό τρόπο σε διαδοχικές γενεές ενός γενετικού αλγορίθμου.

3.7.8 Υλοποίηση Γενετικών Αλγορίθμων

Στην ενότητα αυτή θα παρουσιάσουμε ένα παράδειγμα εφαρμογής του γενετικού αλγόριθμου στην ταυτοποίηση ενός άγνωστου συστήματος ARMA. Συγκεκριμένα θα παρουσιαστούν ένας απλός γενετικός αλγόριθμος με τα χαρακτηριστικά που περιγράφηκαν στις προηγούμενες ενότητες και μια παραλλαγή του. Στο τέλος της ενότητας υπάρχουν τα αποτελέσματα της εφαρμογής των γενετικών αλγορίθμων στην ταυτοποίηση του άγνωστου συστήματος ARMA.

3.7.8.1 Περιγραφή του Γενετικού Αλγόριθμου

Ο γενετικός αλγόριθμος ξεκινά με τον αρχικό πληθυσμό ο οποίος δημιουργείται με τυχαίο τρόπο και εφαρμόζει πάνω σ' αυτόν τους γενετικούς τελεστές της επιλογής, διασταύρωσης και μετάλλαξης. Η διαδικασία αυτή συνεχίζεται για κάθε γενιά που δημιουργείται μέχρι είτε την δημιουργία ενός συγκεκριμένου αριθμού γενεών είτε να φτάσουμε σε μία "ικανοποιητική" τιμή της συνάρτησης κόστους.

Άλλο σημαντικό στοιχείο του γενετικού αλγορίθμου είναι η συνάρτηση κόστους. Με αυτή την συνάρτηση καθορίζεται η ποιότητα του κάθε χρωμοσώματος και κατά πόσο πλησιάζει την βέλτιστη λύση. Αυτή η συνάρτηση εξαρτάται από το πρόβλημα που

καλείται να λύση ο γενετικός αλγόριθμος και επαφίεται στην ευχέρεια του χρήστη να την επιλέξει. Στο συγκεκριμένο πρόβλημα της ταυτοποίησης του συστήματος η συνάρτηση κόστους που έχει επιλεγεί ορίζεται από την σχέση:

$$e_{ji} = \frac{1}{d} \sum_{k=1}^d (y(k) - \hat{y}_{ji}(k))^2 \quad (3.226)$$

Η σχέση αυτή ορίζει το σφάλμα του μοντέλου e_{ji} που αναπαριστάται από το j -χρωμόσωμα της i -οστής γενιάς από το πραγματικό σύστημα και $\hat{y}_{ji}(k)$, $y(k)$ είναι η έξοδος του μοντέλου του συστήματος αντίστοιχα την χρονική στιγμή k .

Για να υπολογιστεί το κόστος του κάθε χρωμοσώματος θα πρέπει το χρωμόσωμα να αποκωδικοποιηθεί στην lattice μορφή του συστήματος και στην συνέχεια να υπολογιστεί η έξοδος του συστήματος στο παράθυρο των δεδομένων εισόδου.

Αρχικά ο γενετικός αλγόριθμος επιλέγει τα καλύτερα χρωμοσώματα από τον αρχικό πληθυσμό. Αυτό γίνεται με το να υπολογίσει σε όλα τα χρωμοσώματα την συνάρτηση κόστους και στη συνέχεια ταξινομεί τα χρωμοσώματα με βάση το κόστος από το μικρότερο στο μεγαλύτερο. Από τον ταξινομημένο αρχικό πληθυσμό επιλέγει τα $K < M$ καλύτερα, δηλαδή αυτά που έχουν μικρότερο κόστος. Ο γενετικός αλγόριθμος απ' εδώ και πέρα θα χρησιμοποιεί K χρωμοσώματα. Στην συγκεκριμένη υλοποίηση ισχύει $K=M/2$.

Οι επόμενες παράγραφοι θα περιγράψουν τις διαδικασίες της επιλογής, διασταύρωσης και μετάλλαξης που χρησιμοποιήθηκαν κατά την διεξαγωγή των πειραμάτων.

3.7.8.2 Επιλογή Γονέων

Όπως αναφέρθηκε στις προηγούμενες ενότητες του κεφαλαίου, ένας βασικός τελεστής των γενετικών αλγορίθμων είναι ο τελεστής της επιλογής των γονέων που θα παράγουν τα χρωμοσώματα των επόμενων γενεών. Για την επιλογή των γονέων έχουν προταθεί διάφοροι τρόποι. Στην μέθοδο που περιγράφεται εδώ ο τρόπος επιλογής των γονέων γίνεται με την ανάθεση σε κάθε χρωμόσωμα μιας πιθανότητας επιλογής η οποία καθορίζει, όπως λέει και το όνομά της, την πιθανότητα να επιλεγεί ένα χρωμόσωμα για γονέας. Η πιθανότητα αυτή δίνεται από την σχέση

$$P_j^i = \frac{e_{\max}^i - e_{ji}}{\sum_{k=1}^K (e_{\max}^i - e_{ki})} \quad (3.227)$$

Όπου P_j^i είναι η πιθανότητα επιλογής του j -οστού χρωμοσώματος της i -οστής γενιάς, e_{\max}^i είναι το μέγιστος κόστος της i -οστής γενιάς και e_{ji} είναι το κόστος του j -οστού χρωμοσώματος της ίδιας γενιάς. K είναι ο αριθμός των χρωμοσωμάτων στην παρούσα γενιά.

Μετά την απόδοση των πιθανοτήτων σε κάθε χρωμόσωμα επιλέγονται τυχαία, με βάση τις πιθανότητες επιλογής, $D < K$ χρωμοσώματα τα οποία θα αποτελέσουν τους γονείς.

Ένα ποσοστό από τα χρωμοσώματα που επιλέχθηκαν για γονείς θα επιβιώσει στην επόμενη γενιά και τα υπόλοιπα χρωμοσώματα που δεν θα επιβιώσουν θα αντικατασταθούν από τα παιδιά. Οι γονείς που θα επιλεγθούν για να επιβιώσουν αποτελούν αυτούς που έχουν την μικρότερη τιμή κόστους.

3.7.8.3 Διασταύρωση Γονέων

Οι γονείς οι οποίοι έχουν επιλεγθεί σχηματίζουν ζευγάρια και διασταυρώνουν τα κομμάτια τους για να παραχθούν οι απόγονοι. Η διασταύρωση των κομματιών γίνεται με τυχαία επιλογή διασταύρωσης ενός σημείου. Το δεξί κομμάτι του ενός χρωμοσώματος από το ζευγάρι γίνεται δεξί κομμάτι του άλλου, και η ίδια διαδικασία γίνεται για το αριστερό. Αν και όπως αναφέρθηκε στην περιγραφή των γενετικών αλγορίθμων υπάρχει η δυνατότητα να επιλεγούν περισσότερα από ένα σημεία διασταύρωσης στην παρούσα υλοποίηση έχει επιλεγεί η διασταύρωση ενός σημείου.

Από την απλή διασταύρωση ενός σημείου οι δύο γονείς σχηματίζουν δύο απόγονους. Ο αριθμός των απόγονων που θα παραχθούν εξαρτάται από το ποσοστό των γονέων που θα επιβιώσουν στην επόμενη γενιά. Αν D είναι ο αριθμός των χρωμοσωμάτων που έχουν επιλεγεί για γονείς και pD ($0 < p < 1$) είναι ο αριθμός των γονέων που θα επιβιώσουν στην επόμενη γενιά τότε ο αριθμός των απογόνων που θα πρέπει να παραχθούν είναι $K - pD$, ώστε να διατηρηθεί σταθερός ο αριθμός του πληθυσμού, κάτι που στην περίπτωσή μας ισχύει για την τιμή του $p = 1/2$.

3.7.8.4 Μετάλλαξη

Ο τελεστής της μετάλλαξης ενός γενετικού αλγορίθμου μπορεί να εφαρμοστεί είτε μόνο στα παιδιά είτε σε ολόκληρο τον πληθυσμό με εξαίρεση το χρωμόσωμα με το μικρότερο κόστος. Η μετάλλαξη αλλάζει τα bits του χρωμοσώματος από 1 σε 0 ή το αντίστροφο με πιθανότητα P_m . Στην συγκεκριμένη μεθοδολογία για την μετάλλαξη του πληθυσμού (παιδιά ή ολόκληρος), δημιουργείται ένας πίνακας από τυχαίους άσσους και μηδενικά τέτοιος ώστε το ποσοστό των άσσων επί του συνόλου του πίνακα να είναι ίσος με το P_m . Στον πίνακα που δημιουργήθηκε και στα χρωμοσώματα που είναι για μετάλλαξη εφαρμόζεται η πράξη xor και το αποτέλεσμα είναι τα μεταλλαγμένα χρωμοσώματα.

Στην συνέχεια ο μεταλλαγμένος πληθυσμός προστίθεται στον υπόλοιπο που δεν έχει μεταλλαχθεί και υπολογίζεται το κόστος των νέων χρωμοσωμάτων και ταξινομείται με βάση το μικρότερο ώστε να επαναληφθεί η όλη διαδικασία της εφαρμογής των τελεστών του γενετικού αλγορίθμου.

3.7.8.5 Μια Παραλλαγή του Γενετικού αλγόριθμου: Καιιάδας

Οι γενετικοί αλγόριθμοι έχουν την ιδιαιτερότητα να υφίστανται πολλές αλλαγές και έτσι να υπάρχουν αρκετές παραλλαγές τους. Στην παρούσα μέθοδο εκτός από τον γενετικό αλγόριθμο που περιγράφηκε παραπάνω υλοποιήθηκε και μια παραλλαγή του και έγιναν πειράματα χρησιμοποιώντας και αυτή.

Η διαφορά με τον προηγούμενο αλγόριθμο έγκειται στον τρόπο με τον οποίο εισέρχονται οι απόγονοι στον νέο πληθυσμό. Ειδικότερα η παραλλαγή αυτή του γενετικού αλγορίθμου εισαγάγει ένα νέο τρόπο με τον οποίο ανανεώνεται ο πληθυσμός των χρωμοσωμάτων. Στον προηγούμενο γενετικό αλγόριθμο επιβιώνει ένα ποσοστό των γονέων και στην συνέχεια δημιουργούνται τόσοι απόγονοι ώστε ο πληθυσμός να παραμένει σταθερός. Στην παραλλαγή του γενετικού αλγορίθμου παράγεται σταθερός αριθμός απογόνων ο οποίος εισάγεται στον τρέχον πληθυσμό. Μετά την πράξη της μετάλλαξης ο επαυξημένος πληθυσμός ταξινομείται με βάση το κόστος του και κρατούνται τα K καλύτερα χρωμοσώματα. Επειδή μ' αυτό τον τρόπο κάθε φορά τα χειρότερα χρωμοσώματα απορρίπτεται η παραλλαγή αυτή ονομάστηκε «καιιάδας» από

τον γκρεμό που πετούσαν οι αρχαίοι Σπαρτιάτες τα μωρά που παρουσίαζαν πρόβλημα. Η παραλλαγή αυτή είναι πιο απαιτητική σε υπολογιστική ισχύ και μνήμη εφ' όσον επεξεργάζεται μεγαλύτερο πληθυσμό από την προηγούμενη. Τα αποτελέσματα της παραλλαγής αυτής ως προς το πρόβλημα ταυτοποίησης φαίνονται στην παρακάτω ενότητα με τις πειραματικές μετρήσεις.

3.7.9 Αλγόριθμος

Όπως προαναφέρθηκαν προηγουμένως υπάρχουν αρκετές παράμετροι ρύθμισης του αλγορίθμου. Αυτές οι παράμετροι είναι οι ακόλουθες:

N	Η τάξη του ARMA συστήματος
I_{pop}	Το πλήθος του αρχικού πληθυσμού του γενετικού αλγορίθμου
$bits$	Ο αριθμός των bits που θα κωδικοποιείται κάθε γονίδιο του χρωμοσώματος
d	Το μήκος του παραθύρου των δεδομένων για τον υπολογισμό της συνάρτησης κόστους
K_{init}	Το πλήθος του αρχικού πληθυσμού των χρωμοσωμάτων του γενετικού αλγορίθμου
K	Το πλήθος του πληθυσμού των χρωμοσωμάτων του γενετικού αλγορίθμου
D	Ο αριθμός των επιλεγμένων γονέων
p	Το ποσοστό των γονέων που θα επιβιώσουν
M	Ο αριθμός των απογόνων που θα δημιουργηθούν (παραλλαγή καιάδα)
P_m	Η πιθανότητα μετάλλαξης του

	χρωμοσώματος
$GA_{iterations}$	Ο αριθμός των γενεών που θα παραχθούν μέχρι να τερματίσει ο γενετικός αλγόριθμος

Πίνακας 3.5: Παράμετροι του νέου αλγορίθμου.

Παρακάτω φαίνεται η περιγραφή της μεθόδου σε βήματα

1. Δημιούργησε με τυχαίο τρόπο τον αρχικό πληθυσμό.
2. Υπολόγισε το κόστος του κάθε χρωμοσώματος του αρχικού πληθυσμού και ταξινόμησέ τα με βάση το κόστος από το μικρότερο στο μεγαλύτερο.
3. Κράτησε τα K καλύτερα χρωμοσώματα και ξεκίνησε τον γενετικό αλγόριθμο.
4. $k:=1$
5. Υπολόγισε την πιθανότητα επιλογής και επέλεξε D υποψήφιους γονείς τυχαία με βάση την πιθανότητα αυτή.
6. Σχημάτισε τα ζευγάρια των γονέων και με απλή διασταύρωση δημιούργησε $K \cdot pD$ απογόνους.
7. Σχημάτισε τα ζευγάρια των γονέων και με απλή διασταύρωση δημιούργησε M απογόνους. (παραλλαγή καιάδα)
8.
 - a. Εφάρμοσε μετάλλαξη με πιθανότητα P_m στους απογόνους και πρόσθεσέ τα μεταλλαγμένα χρωμοσώματα μαζί με τους pD γονείς που επιβίωσαν σχηματίζοντας την νέα γενιά
 - b. Υπολόγισε το κόστος του κάθε χρωμοσώματος της νέας γενιάς και ταξινόμησέ του όπως στο βήμα 6.
8. «παραλλαγή καιάδα»
9. Εφάρμοσε μετάλλαξη με πιθανότητα P_m στους απογόνους και πρόσθεσέ τα μεταλλαγμένα χρωμοσώματα μαζί με τον υπόλοιπο πληθυσμό.
 - a. Υπολόγισε το κόστος του κάθε χρωμοσώματος της νέας γενιάς και ταξινόμησέ του όπως στο βήμα 6.
 - b. Αφαίρεσε τα M χειρότερα χρωμοσώματα από τον τρέχον πληθυσμό.
10. $k:=k+1$
11. αν $k < GA_{iterations}$ πήγαινε στο βήμα 9.

12. Επέλεξε το καλύτερο χρωμόσωμα της τελευταίας γενιάς, αποκωδικοποίησέ το και μετέτρεψε στην κανονική μορφή του μοντέλου.

Κεφάλαιο 4

Μεθοδολογία Πρόβλεψης Χρονοσειρών με Χρήση ΤΝΔ

4.1 Εισαγωγή

Οι κλασικές μέθοδοι της στατιστικής επεξεργασίας σημάτων στηρίζονται στις ακόλουθες βασικές παραδοχές σχετικά με τη φύση των σημάτων: γραμμικότητα (linearity), στασιμότητα (stationarity) και κανονική κατανομή.

Οι παραδοχές αυτές έχουν τεθεί ώστε να γίνει δυνατή η μαθηματική μοντελοποίηση των συστημάτων. Στην πράξη όμως οι περισσότερες παραδοχές, εάν όχι όλες, δεν ισχύουν σε πραγματικές εφαρμογές, αφού τα σήματα εισόδου προέρχονται από δυναμικά συστήματα. Το τελικό αποτέλεσμα είναι η σχεδίαση υπο-βέλτιστων συστημάτων, τα οποία όμως, διαθέτουν καλά τεκμηριωμένα μαθηματικά μοντέλα. Μια από τις νέες τεχνολογίες η οποία μπορεί να χρησιμοποιηθεί για τη βελτίωση πολλών υφιστάμενων συστημάτων επεξεργασίας σημάτων καθώς και την υλοποίηση νέων είναι τα Τεχνητά Νευρωνικά Δίκτυα.

Η πρόβλεψη σημάτων με χρονική εξάρτηση (χρονοσειρών) αποτελεί μια από τις σημαντικότερες περιοχές της επεξεργασίας σήματος και έχει ως στόχο την πρόβλεψη των μελλοντικών τιμών ενός σήματος, με χρήση μιας σειράς μετρήσεων του παρελθόντος του, καθώς και άλλων βοηθητικών σημάτων.

Ειδικά στην πρόβλεψη χρονοσειρών τα ΤΝΔ εμφανίζουν σημαντικά πλεονεκτήματα σε σχέση με άλλες συμβατικές μεθόδους πρόβλεψης, διότι έχουν τη δυνατότητα μοντελοποίησης μη-γραμμικών, μη-στάσιμων και μη-γκαουσιανών διεργασιών. Δεδομένης της έλλειψης καλά τεκμηριωμένης θεωρητικής βάσης της τεχνολογίας των νευρωνικών δικτύων η επιτυχής υλοποίηση συστημάτων απαιτεί

εξαιρετικά συστηματική προσέγγιση. Παρακάτω θα παρουσιαστεί η μεθοδολογία σχεδίασης και υλοποίησης αποδοτικών συστημάτων με χρήση ΤΝΔ.

Υπάρχουν πολλά παραδείγματα χρονοσειρών τα οποία είναι εξαιρετικής σημασίας στην επίλυση πραγματικών προβλημάτων. Ένα τέτοιο πρόβλημα είναι η πρόβλεψη της ζήτησης ηλεκτρικής ενέργειας. Επίσης η επιστήμη της μετεωρολογίας διαθέτει ένα πλήθος τεχνικών προβλέψεων. Ακόμη, η ικανότητα πρόβλεψης της χρονικής εξέλιξης οικονομικών μεγεθών. [24]

4.2 Παρουσίαση Μεθόδων Πρόβλεψης

Οι περισσότερες μέθοδοι είναι ποσοτικές από τη φύση τους αλλά υπάρχουν και χρήσιμες ποιοτικές προσεγγίσεις. Οι κυριότερες μέθοδοι πρόβλεψης παρουσιάζονται στον παρακάτω πίνακα

Μέθοδος	Περιγραφή	Εφαρμογές
Κινητός μέσος όρος (Moving Average)	Τεχνική πρόβλεψης που βασίζεται στη μέση τιμή των παρατηρήσεων γειτονικών χρονικών στιγμών. Προϋποθέτει στασιμότητα των χρονοσειρών αφού σταθμίζονται όλα τα δεδομένα με τον ίδιο συντελεστή. Δεν είναι δυνατή η πρόβλεψη μεταβολής της διεύθυνσης της χρονοσειράς	Προβλέψεις μικρού χρονικού εύρους για δραστηριότητες που χρειάζονται μεγάλο αριθμό μεγεθών.
Εκθετική Εξομάλυνση (Exponential Smoothing)	Οι προβλέψεις είναι σταθμισμένοι συνδυασμοί των παρατηρήσεων και προβλέψεων του παρελθόντος. Δίδεται περισσότερο βάρος στα πιο πρόσφατα γεγονότα. Δεν είναι δυνατή πρόβλεψη μεταβολής της διεύθυνσης της χρονοσειράς	Εφαρμογές παρόμοιες με αυτές του Κινητού μέσου όρου με μικρό χρονικό εύρος.
Μέθοδοι Διαχωρισμού (Decomposition)	Η μέθοδος προϋποθέτει την ύπαρξη σχέσης μεταξύ του χρόνου και της μεταβλητής που θα προβλεφθεί. Η χρονοσειρά χωρίζεται σε συστηματικές και μη-συστηματικές συνιστώσες.	Εφαρμογή σε μεγέθη που μπορούν να αναλυθούν σε συνιστώσες τάσης, εποχικότητας και κυκλικότητας.

Μέθοδος	Περιγραφή	Εφαρμογές
Ανάλυση Παλινδρόμησης (Regression Analysis)	Ερμηνευτική πρόβλεψη που υποθέτει την ύπαρξη σχέσης μεταξύ της εξαρτημένης και μιας ή περισσότερων ερμηνευτικών μεταβλητών.	Πρόβλεψη μικρού και μέσου χρονικού ορίζοντα. Αποτελεί την τυπική προσέγγιση για την ανάπτυξη ενός ερμηνευτικού μοντέλου.
Οικονομικά Μοντέλα (Econometric Models)	Συστήματα ερμηνευτικών εξισώσεων που περιλαμβάνει εξωγενείς και ενδογενείς μεταβλητές.	Βραχυχρόνια και μεσοπρόθεσμη πρόβλεψη για τη ζήτηση προϊόντων, το κόστος, τις τιμές, τις δαπάνες, και τα έσοδα.
Χρήση Προσαρμοζόμενων Φίλτρων (Adaptive Filtering)	Αποτελεί επέκταση της εκθετικής εξομάλυνσης και χρησιμοποιεί μια επαναληπτική προσέγγιση για τον προσδιορισμό των καλύτερων συντελεστών στάθμισης.	Βραχυχρόνια και μεσοπρόθεσμη πρόβλεψη για τα κέρδη, τις τιμές τα αποθέματα και τον έλεγχο της παραγωγής.
Τεχνικές Box-Jenkins	Χρησιμοποιείται μια επαναληπτική προσέγγιση για την ανεύρεση του αποδοτικότερου μοντέλου μεταξύ μιας γενικότερης τάξης μοντέλων. Δεν υποθέτει ύπαρξη πρότυπου στα δεδομένα.	Βραχυχρόνια και μεσοπρόθεσμη πρόβλεψη για τα κέρδη, τις τιμές τα αποθέματα και τον έλεγχο της παραγωγής.
Τεχνητά Νευρωνικά Δίκτυα	Χρησιμοποιείται για την ανεύρεση μιας απεικόνισης μεταξύ των δεδομένων εισόδου και των επιθυμητών μεγεθών	Βραχυχρόνια και μεσοπρόθεσμη πρόβλεψη για τα κέρδη, τις τιμές τα αποθέματα και τον έλεγχο της παραγωγής.

Πίνακας 4.3: Παρουσίαση Μεθόδων Πρόβλεψης

4.3 Μεθοδολογία Ανάπτυξης ΤΝΔ για Πρόβλεψη Χρονοσειρών

Η μεθοδολογία Ανάπτυξης ΤΝΔ για Πρόβλεψη Χρονοσειρών αποτελείται από τα ακόλουθα βήματα:

1. Ανάλυση προβλήματος
2. Συλλογή δεδομένων
3. Ανάλυση δεδομένων και επιλογή μεταβλητών εισόδου και εξόδου
4. Προ-επεξεργασία δεδομένων
5. Κατασκευή Συνόλων Δεδομένων
6. Ανάπτυξη μοντέλου και βελτιστοποίηση
7. Αξιολόγηση μοντέλου

Θα πρέπει να τονισθεί πως η ανωτέρω διαδικασία δεν εξαντλείται από τα βήματα αυτά αλλά σχεδόν πάντοτε είναι επαναληπτική, δεδομένου ότι, τα αποτελέσματα που προκύπτουν από τα επόμενα βήματα οδηγούν συχνά σε αναθεωρήσεις των προηγούμενων επιλογών.

4.3.1 Ανάλυση Προβλήματος

Στο στάδιο αυτό διεξάγονται οι ακόλουθες εργασίες:

- Καθορίζεται επακριβώς το ζητούμενο προς μοντελοποίηση πρόβλημα.
- Συγκεντρώνονται όλα τα δεδομένα γύρω από το πρόβλημα και οργανώνονται σε μια βάση γνώσης.
- Αναλύεται το πρόβλημα με βάση τα διαθέσιμα στοιχεία σε επιμέρους προβλήματα.
- Διερευνώνται και αξιολογούνται εναλλακτικές μέθοδοι επίλυσης του προβλήματος (εκτός των ΤΝΔ) και εξετάζεται εάν θα υλοποιηθούν είτε ως συμπληρωματικές μέθοδοι είτε ως συστήματα αναφοράς για την απόδοση του νέου μοντέλου (benchmark).

4.3.2 Συλλογή Δεδομένων

Κατά τη συλλογή των δεδομένων για την ανάπτυξη ενός μοντέλου ΤΝΔ υπάρχουν τρία κύρια θέματα που πρέπει να εξετάζονται:

- Η διαθεσιμότητα τους υπό πραγματικές συνθήκες
- Η διαχείριση των ελλείπων δεδομένων και
- Η εκκαθάριση των ασύμβατων δεδομένων ή αυτών με υψηλό ποσοστό θορύβου

Κατά την ανάπτυξη μοντέλων για χρήματο-οικονομικές εφαρμογές, ειδικά για αυτές που θα λειτουργούν σε περιβάλλον πραγματικού χρόνου, απαιτείται εξαιρετικά προσεκτική επιλογή των δεδομένων ώστε αυτά να είναι διαθέσιμα σε πραγματικό ή σχεδόν πραγματικό χρόνο. Τυπικό παράδειγμα αποτελεί ο όγκος συναλλαγών.

Η διαχείριση των ελλείπων δεδομένων αποτελεί ένα από τα δυσκολότερα σημεία στην ανάπτυξη μοντέλων ΤΝΔ για χρήματο-οικονομικές εφαρμογές διότι το φαινόμενο αυτό παρατηρείται σχεδόν σε μόνιμη βάση. Η έλλειψη αυτή οφείλεται σε αυτές συστηματικές και μη. Για τις συστηματικές αιτίες θα πρέπει να εντοπισθούν οι συγκεκριμένοι περιορισμοί και να ληφθούν υπόψη κατά τη φάση της σχεδίασης ώστε να αντιμετωπισθούν κατάλληλα. Όταν η έλλειψη οφείλεται σε μη συστηματικές αιτίες υπάρχουν οι παρακάτω επιλογές:

- Αφαίρεση των πλειάδων στις οποίες υπάρχουν ελλιπή δεδομένα
- Αντικατάσταση των δεδομένων από προσεγγίσεις με βάση τα γειτονικά δείγματα

4.3.3 Ανάλυση δεδομένων και επιλογή μεταβλητών εισόδου και εξόδου

Η φάση αυτή έχει ως σκοπό να επιλέξει τις καταλληλότερες μεταβλητές από τα διαθέσιμα δεδομένα κατά τρόπο ώστε το προκύπτον δίκτυο να διαθέτει τους ελάχιστους βαθμούς ελευθέριας. Στη φάση προκύπτει ένα από τα σημαντικότερα σχεδιαστικά ζητήματα των ΤΝΔ που είναι γνωστό ως δίλημμα της γενίκευσης-απομνημόνευσης. Αυτό συμβαίνει διότι όσο περισσότεροι είναι οι βαθμοί ελευθέριας ενός ΤΝΔ (βάρος) τόσο ακριβέστερη η προσέγγιση της ζητούμενης απεικόνισης στα συγκεκριμένα σημεία του συνόλου μάθησης, αλλά ταυτόχρονα τόσο περισσότερο το δίκτυο τείνει να απομνημονεύει τα δεδομένα παρά να ανακαλύπτει τις σχέσεις που τα συνδέουν.

Σε όλες τις περιπτώσεις θα πρέπει να αναζητείται το πιο φειδωλό μοντέλο (parsimonious model). Δεδομένου ότι δεν υπάρχει κάποιος ποσοτικός κανόνας που να

προσδιορίζει την ελάχιστη τιμή του αριθμού των βαρών στις ενδιάμεσες βαθμίδες (hidden layers) η ανεύρεση του φειδωλού μοντέλου πρέπει να ξεκινά από την ελαχιστοποίηση των μεταβλητών εισόδου.

Δύο είναι οι εναλλακτικές προσεγγίσεις στην επιλογή των προσφορότερων μεταβλητών και το συγγενές πρόβλημα της συχνότητας δειγματοληψίας. Οι προσεγγίσεις αυτές είναι:

- Η χρήση αναλυτικών τεχνικών από την θεωρία της στατιστικής και της ανάλυσης μη-γραμμικών δυναμικών συστημάτων
- Η χρήση ανάλυσης ευαισθησίας στις εισόδους – η εξομοίωση μιας μεταβλητής με τη μέση τιμή της.

4.3.3.1 Στατιστικές Μέθοδοι

Οι στατιστικές μέθοδοι που μπορούν να χρησιμοποιηθούν κατά τη διαδικασία εξέτασης των δεδομένων και της ανεύρεση των σχέσεων μεταξύ τους είναι:

- Υπολογισμός συντελεστών συσχέτισης (Correlation Analysis)
- Ανάλυση παλινδρόμησης ελάχιστων τετράγωνων (Least Squares Regression Analysis)
- Έλεγχος t-statistic ή adjusted R^2 με βάση την ανάλυση παλινδρόμησης ελάχιστων τετράγωνων
- Ανάλυση πρωτευουσών συνιστωσών (Principal Components Analysis)

4.3.3.2 Ανάλυση Ευαισθησίας - Εξομοίωση Εισόδου με Μέση Τιμή

Υπάρχουν δύο εναλλακτικοί τρόποι διεξαγωγής της ανάλυσης ευαισθησίας:

- Διαδοχική αφαίρεση κάθε μιας από τις υπό εξέταση εισόδους και παρατήρηση της μεταβολής των κριτηρίων αξιολόγησης ως συνάρτηση των εξαιρούμενων εισόδων
- Αντικατάσταση των τιμών όλων των δειγμάτων μιας εισόδου με μια σταθερή τιμή

Εάν το σφάλμα που προκύπτει αφαιρώντας μια μεταβλητή παραμένει στα ίδια επίπεδα το συμπέρασμα είναι ότι η υπόψη μεταβλητή δεν είναι σημαντική με βάση το συγκεκριμένο κριτήριο αξιολόγησης. Μια ασφαλέστερη μέθοδος είναι η εξέταση της μεταβλητότητας της εξόδου. Στην περίπτωση της σταθεροποίησης μιας εισόδου θα πρέπει να επιλέγει η συγκεκριμένη σταθερή τιμή που θα χρησιμοποιηθεί. Στην

πλειονότητα των περιπτώσεων χρησιμοποιείται η μέση τιμή των δειγμάτων μιας μεταβλητής.

4.3.4 Προ-επεξεργασία Δεδομένων

Η ανάλυση των δεδομένων που έχουν συλλεχθεί και η προ-επεξεργασία τους επιμερίζεται στις ακόλουθες ενέργειες:

- Αφαίρεση της κλίσης (detrending)
- Κλιμάκωση (scaling) – κανονικοποίησης (normalization)
- Διαχείριση των ακραίων σημείων
- Μετασχηματισμοί δεδομένων

4.3.4.1 Αφαίρεση της Κλίσης

Η αφαίρεση της κλίσης τόσο από τις εξαρτημένες μεταβλητές όσο και από τις ανεξάρτητες είναι απαραίτητη, διότι η ύπαρξη τους μπορεί να οδηγήσει σε ψευδοσυσχετίσεις. Σε κάθε περίπτωση όμως θα πρέπει να ληφθεί υπόψη ότι η αφαίρεση της κλίσης ουσιαστικά αφαιρεί πληροφορία από τη χρονοσειρά. Η πληροφορία αυτή σε πολλές περιπτώσεις μπορεί να είναι χρήσιμη. Εάν από τα δεδομένα έχει αφαιρεθεί η κλίση το προκύπτον μοντέλο δε θα διαθέτει αυτή την πληροφορία και κατά συνέπεια δε θα μπορέσει να μοντελοποιήσει τα διαφορετικά δυναμικά χαρακτηριστικά που χαρακτηρίζουν κάθε φάση της αγοράς. Μια δυνατή λύση στο πρόβλημα είναι η εισαγωγή της πληροφορίας της κλίσης μέσω μιας πληροφοριακής μεταβλητής.

4.3.4.2 Κλιμάκωση (scaling) – Κανονικοποίηση (normalization)

Η κλιμάκωση των μεταβλητών έχει ως στόχο την κανονικοποίηση τους και είναι απαραίτητη στα ΤΝΔ για τους ακόλουθους λόγους:

- Αποφεύγεται το ενδεχόμενο το δίκτυο να συντονισθεί σε μια συγκεκριμένη περιοχή των δεδομένων εισοδου – εξοδου
- Κλιμακώνονται οι τιμές εισοδου ώστε να εμπίπτουν εντός της περιοχής λειτουργίας
- Η απ' ευθείας τροφοδότηση του δικτύου με τις πρωτογενείς τιμές των μεταβλητών εισοδου θα είχε ως αποτέλεσμα μεγάλες διαφορές στις περιοχές τιμών

μεταξύ των αντίστοιχων εισόδων. Τις διαφορές αυτές το δίκτυο μπορεί να τις εκλάβει ως σημαίνουσες στη βαρύτητα των αντιστοιχών εισόδων. Στην περίπτωση αυτή το μοντέλο θα εξαρτάται σημαντικά από τις εισόδους με υψηλές τιμές και θα ελαχιστοποιηθεί η επιρροή των εισόδων με μικρότερες απόλυτες τιμές, χωρίς αυτό να είναι πληροφοριακά ορθό.

Η συνήθης τακτική για την κανονικοποίηση των μεταβλητών εισόδου είναι ο μετασχηματισμός με κέντρο το μηδέν (μηδενική μέση τιμή) και μοναδιαία τυπική απόκλιση. Οι σχέσεις που χρησιμοποιούνται για το σκοπό αυτό είναι οι ακόλουθες:

$$\tilde{x}_i(t) = \frac{x_i(t) - \bar{x}}{\sigma_x} \quad (4.1)$$

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^N (x_i(t) - \bar{x})^2}{N-1}} \quad (4.2)$$

οπου \tilde{x} είναι η κανονικοποιημένη τιμή της μεταβλητής x_i και \bar{x} η μέση τιμή της.

4.3.4.3 Διαχείριση Ακραίων Σημείων (Statistical outliers)

Τα ακραία σημεία στις μεταβλητές εισόδου θέλουν ειδική μεταχείριση γιατί δημιουργούν τα ακόλουθα προβλήματα:

- Ωθούν την πλειοψηφία του συνόλου των τιμών της μεταβλητής σε πολύ περιορισμένη δυναμική περιοχή η οποία απαιτεί εξαιρετικά υψηλή αριθμητική ακρίβεια κατά την υλοποίηση του αλγόριθμου εκπαίδευσης
- Η συντριπτική πλειοψηφία των παρατηρήσεων θα εμπίπτει στη γραμμική περιοχή της σιγμοειδούς συνάρτησης με αποτέλεσμα να απολεσθούν οι ιδιότητες μη-γραμμικότητας του μοντέλου

Ο εντοπισμός των ακραίων σημείων γίνεται συνήθως με τη γραφική αναπαράσταση των μεταβλητών. Σε άλλες περιπτώσεις η παραπάνω μέθοδος δεν είναι δυνατή σε περισσότερες από δυο ή τρεις διαστάσεις ή όταν τα δεδομένα είναι πολυάριθμα.

Μια άλλη μέθοδος εντοπισμού των ακραίων σημείων είναι η σχεδίαση ιστογραμμάτων συχνότητας τιμών. Τα ιστογράμματα καταδεικνύουν τα ακραία σημεία και ταυτόχρονα παρέχουν πληροφορίες σχετικά με την στατιστική κατανομή των αντίστοιχων διανυσμάτων η οποία είναι χρήσιμη.

4.3.4.4 Λοιποί Μετασχηματισμοί Δεδομένων

Οι μετασχηματισμοί αυτοί μπορούν να κατηγοριοποιηθούν στις ακόλουθες κατηγορίες:

- Αφαίρεση θορύβου από τα δεδομένα.
- Μετασχηματισμός πληροφοριακών εισόδων σε προβλεπτικές.
- Μεταβολή της στατιστικής κατανομής.

Στο σύνολο των σημάτων που προέρχονται από μετρήσεις, ειδικά σε οικονομικά μεγέθη, υπάρχει ένα υψηλό ποσοστό θορύβου το οποίο καθιστά δύσκολη την ανακάλυψη των σχέσεων που διέπουν τις μεταβολές μεγεθών.

Η τεχνική που χρησιμοποιείται εκτενώς για την αφαίρεση του θορύβου και την ανεύρεση των τάσεων (trends) σε οικονομικά μεγέθη είναι ο κινητός μέσος όρος (moving average). Στην πράξη χρησιμοποιούνται τρεις παραλλαγές της τεχνικής αυτής:

- Ο απλός κινητός μέσος όρος (simple moving average)
- Ο σταθμισμένος κινητός μέσος (weighted moving average)
- Ο εκθετικός κινητός μέσος (exponential moving average)

Ο μετασχηματισμός πληροφοριακών δεδομένων σε προβλεπτικά αποτελεί έναν από τους σημαντικότερους στόχους της προ-επεξεργασίας των δεδομένων ενός ΤΝΔ. Οι κλασικότερες μέθοδοι μετασχηματισμού που χρησιμοποιούνται είναι η διαφόριση μιας μεταβλητής, και ο υπολογισμός της διαφοράς και του λόγου δυο μεταβλητών.

Η διαφόριση μιας μεταβλητής παρέχει τα εξής πλεονεκτήματα:

- Επιτυγχάνεται η αφαίρεση της κλίσης από τη χρονοσειρά. Η κλίση μετατρέπεται σε μια σταθερή απόκλιση με αποτέλεσμα να βελτιώνεται σημαντικά η δυναμική περιοχή των δεδομένων

- Οι κανονικοποιημένες διαφορές της μορφής $\frac{\Delta\psi}{\psi}$, χρησιμοποιούνται πολύ συχνά στην μοντελοποίηση διότι επιλύουν το πρόβλημα της απόλυτης τιμής με αποτελεσματικό τρόπο

Ένα πιθανό μειονέκτημα της διαφόρισης το οποίο εντοπίζεται στην υπερβατική συμπεριφορά της συνάρτησης μεταφοράς που υλοποιεί (high pass filter) είναι ότι στις περιπτώσεις που η πληροφορία χαμηλών συχνοτήτων είναι χρήσιμη θα πρέπει να ληφθούν μέτρα ώστε να συμπεριληφθεί στο δίκτυο. Ειδικά σε χρηματοοικονομικές εφαρμογές η πληροφορία αυτή είναι συνήθως χρήσιμη.

Ο λόγος δυο μεταβλητών μετασχηματίζει με απλό τρόπο πληροφοριακές μεταβλητές σε προβλεπτικές, δεδομένου ότι λειτουργεί ως μια λογική σύγκριση. Τυπικό παράδειγμα αποτελεί ο τεχνικός δείκτης Ανόδου / Καθόδου ο οποίος προκύπτει από τον λόγο του αριθμού των μετοχών των οποίων η τιμή τους αυξήθηκε σε δεδομένη χρονική περίοδο ως προς τον αριθμό των μετοχών που μειώθηκε.

Μια άλλη οικογένεια μετασχηματισμών έχει ως στόχο τον μετασχηματισμό της στατιστικής κατανομής των δεδομένων σε κανονική. Αυτοί χρησιμοποιούνται ώστε να εξασφαλισθεί η απαίτηση που τίθεται από πολλά στατιστικά μοντέλα για δεδομένα που ακολουθούν κανονική κατανομή. Δεδομένου ότι τα νευρωνικά δίκτυα αποτελούν μια επέκταση των κλασικών στατιστικών μεθόδων, αναμένεται να αποδίδουν καλύτερα όταν τα δεδομένα ακολουθούν την κανονική κατανομή.

Για τον έλεγχο της συμμετρίας και της διασποράς μιας κατανομής χρησιμοποιούνται δυο εργαλεία της στατιστικής ανάλυσης, ο συντελεστής ασυμμετρίας και ο συντελεστής κύρτωσης.

Ο συντελεστής ασυμμετρίας χαρακτηρίζει ποσοτικά τον βαθμό της συμμετρίας μιας κατανομής περί τη μέση τιμή της.

Ο συντελεστής κύρτωσης χαρακτηρίζει ποσοτικά τον βαθμό της κύρτωσης μιας κατανομής σε σχέση με την κανονική κατανομή.

4.3.4.5 Επιλογή Μοντέλου ΤΝΔ

Σε αρκετές περιπτώσεις προτείνεται η χρήση ΤΝΔ με ανατροφοδότηση (recurrent neural networks) για τη μοντελοποίηση χρονοσειρών. Στην πράξη παρά τις προσπάθειες πολλών ερευνητών για την ανάπτυξη επιτυχών και αποδοτικών μοντέλων τα αποτελέσματα ήταν πάντοτε κατώτερα αυτών με αρχιτεκτονική Ανάδρομης Διάδοσης Σφάλματος (Back-propagation). Ταυτόχρονα στην πλειονότητα των αναφορών που έχουν ως θέμα την πρόβλεψη χρονοσειρών χρησιμοποιείται ως αλγόριθμος εκπαίδευσης του δικτύου ο γενικευμένος κανόνας Δέλτα με συχνότερη χρήση του LMS.

4.3.5 Σχεδίαση – Διαμόρφωση Συνόλων Δεδομένων

Ένα από τα πιο χρονοβόρα βήματα στην ανάπτυξη ενός μοντέλου είναι η σχεδίαση και διαμόρφωση των συνόλων δεδομένων για τις διαδικασίες μάθησης (training set), δοκίμων (test set), και αξιολόγησης γενίκευσης (validation set). Στην πράξη

αποδεικνύεται ότι αποτελεί το σημαντικότερο βήμα της φάσης υλοποίησης για την επιτυχία ενός μοντέλου.

Οι εργασίες που εντάσσονται στην κατασκευή των συνόλων δεδομένων είναι:

- Επιλογή των δεδομένων που θα χρησιμοποιηθούν στη φάση εκμάθησης
- Διαμέριση σε σύνολα εκμάθησης, δοκίμων και αξιολόγησης

Γενικά το σύνολο εκμάθησης θα πρέπει να παρέχει όσο το δυνατόν αντιπροσωπευτικότερο δείγμα των δεδομένων τα οποία καλείτε να επεξεργάζεται το δίκτυο.

Μεγάλος αριθμός δεδομένων ελαττώνει την πιθανότητα υποδειγματοληψίας. Αντίθετα, μικρά σύνολα με υψηλό θόρυβο εμπεριέχουν αυξημένο κίνδυνο να απομνημονευθούν από το δίκτυο με αποτέλεσμα να εμφανίζονται ικανοποιητικά αποτελέσματα στη φάση της εκπαίδευσης αλλά το δίκτυο να αποτυγχάνει να ανταποκριθεί σωστά σε πραγματικά δεδομένα.

Δεδομένου ότι έχουν συλλέγει τα δεδομένα, το πλήθος που θα χρησιμοποιηθεί για τη διαμόρφωση των απαιτούμενων συνόλων εξαρτάται από πολλούς παράγοντες:

Το μέγεθος του δικτύου, τις απαιτήσεις για δοκιμές και αξιολόγηση της γενίκευσης, τις κατανομές των εισόδων και των εξόδων, κ.α. συνήθως η σημαντικότερη παράμετρος είναι το μέγεθος του δικτύου.

Περιορισμό στη διαμέριση των δεδομένων αποτελεί το γεγονός ότι η αξιολόγηση απαιτεί ένα ανεξάρτητο σύνολο δεδομένων. Ο απλούστερος τρόπος για τη δημιουργία του συνόλου αξιολόγησης είναι η διαμέριση του αρχικού συνόλου σε δυο τμήματα, κρατώντας ένα ποσοστό της τάξης του 10%-30% για τις ανάγκες της αξιολόγησης. Η τυχαία επιλογή των πρότυπων που θα συμπεριληφθούν στο σύνολο αξιολόγησης ελαχιστοποιεί την πιθανότητα συστηματικών συσχετίσεων μεταξύ των συνόλων εκπαίδευσης και αξιολόγησης.

Οι παραπάνω παρατηρήσεις είναι γενικές και ισχύουν για κάθε μοντέλο ΤΝΔ. Ειδικά για τα μοντέλα πρόβλεψης χρονοσειρών είναι αποδοτικότερη η λεπτομερής εξέταση και επιλογή των δειγμάτων μάθησης και η επιβλεπόμενη διαμεριση τους στα σύνολα εκμάθησης, έλεγχου και αξιολόγησης γενίκευσης.

4.3.6 Ανάπτυξη Μοντέλου και Βελτιστοποίηση

Στο βήμα αυτό εφαρμόζονται στην πράξη όσα αναλύθηκαν παραπάνω. Πιο συγκεκριμένα το βήμα της ανάπτυξης ενός μοντέλου μπορεί να αναλυθεί στις ακόλουθες εργασίες:

Εργασία 1: Αρχική επιλογή των παραμέτρων του δικτύου

Εργασία 2: Εκπαίδευση

Εργασία 3: Επισκόπηση και ανάλυση αποτελεσμάτων εκπαίδευσης

Εργασία 4: Αναθεώρηση παραμέτρων

Εργασία 5: Βελτιστοποίηση παραμέτρων

Η αρχική επιλογή των παραμέτρων του δικτύου αποτελεί μια εργασία η οποία στηρίζεται, σε μεγάλο βαθμό, στην εμπειρία από αντιστοιχα η ανάλογα σε μέγεθος προβλήματα. Μια αποδοτική μέθοδος είναι η εκκίνηση με πολλές εκδόσεις του ίδιου μοντέλου στις οποίες υπάρχει σημαντική διαφοροποίηση των παραμέτρων. Η σύγκριση της απόδοσης των εκδόσεων γίνεται αφού διεξαχθεί η εκπαίδευση για ένα συγκεκριμένο, αρχικό, αριθμό εποχών.

Κατά τη φάση της επισκόπησης και ανάλυσης των αποτελεσμάτων πρέπει να εξετάζονται δυο μέτρα απόδοσης:

- Τα κριτήρια αξιολόγησης ή ακρίβειας προσεγγίσης τα οποία είναι συνήθως μετρικές του σφάλματος
- Η πορεία της σύγκλισης, κυρίως στο σύνολο έλεγχου

Με την εργασία αυτή επιλέγονται τα δίκτυα που πέτυχαν την υψηλότερη απόδοση καθώς και αυτά στα οποία η κλίση της καμπύλης σύγκλισης στο σύνολο έλεγχου είναι πτωτική και κατά συνέπεια έχουν δυνατότητα επίτευξης ακόμη καλύτερων αποτελεσμάτων.

Στη συνέχεια βελτιστοποιούνται οι παράμετροι των δικτύων που επιλέγονται κατά την ανάλυση για συνέχιση της εκπαίδευσης. Γενικά η μεθοδολογία της διαδικασίας είναι παρόμοια με αυτή των Γενετικών Αλγορίθμων. Συγκεκριμένα στόχος είναι η διάχυση των επιτυχών παραμέτρων σε άλλα δίκτυα με μικρότερη απόδοση ώστε να υλοποιηθεί ένα είδος τελεστών μετάλλαξης και διασταύρωσης. Η ανωτέρω διαδικασία είναι συριστική και οδηγεί σε πολύ καλά αποτελέσματα εάν γίνει προσεκτική εφαρμογή της.

Ειδικά για την περίπτωση των δικτύων Ανάδρομης Διάδοσης Σφάλματος, τυπικές παράμετροι βελτιστοποίησης είναι:

- Ο αριθμός των νευρώνων σε κάθε ενδιάμεση βαθμίδα
- Οι αρχικές τιμές των βαρών
- Ο ρυθμός μάθησης και ο συντελεστής ορμής
- Η διατεταγμένη ή τυχαία παρουσίαση των πρότυπων στο δίκτυο
- Η περίοδος εφαρμογής του συνόλου δοκίμων

Δεδομένου ότι δεν υπάρχει κανένας κανόνας προσδιορισμού των βέλτιστων τιμών των παραμέτρων, η υλοποίηση της ανωτέρω διαδικασίας γίνεται με τη μέθοδο του πειραματισμού και σφάλματος.

4.3.7 Συστηματική Βελτιστοποίηση Μοντέλου

Θα πρέπει να τονισθεί πως η αξία των νευρωνικών δικτύων έγκειται στη ικανότητα τους να παρέχουν ικανοποιητικά αποτελέσματα όταν τροφοδοτηθούν με δεδομένα εισόδου στα οποία δεν έχουν εκπαιδευτεί. Η ικανότητα αυτή των δικτύων ονομάζεται γενίκευση (generalization).

Για τη βελτιστοποίηση της γενίκευσης του δικτύου προτείνεται η χρήση μιας τεχνικής εκπαίδευσης στην οποία χρησιμοποιείται ένα τρίτο σύνολο το οποίο ονομάζεται σύνολο δοκίμων (test set). Στη φάση της εκπαίδευσης σε περιοδικά διαστήματα εξετάζεται το σφάλμα του δικτύου στο σύνολο δοκίμων (test set). Το σφάλμα αυτό δε χρησιμοποιείται για την αλλαγή των βαρών, αλλά παρέχει ένδειξη για πιθανή υπερεκμάθηση του δικτύου. Τα βάρη που θα χρησιμοποιηθούν στη φάση της αξιολόγησης του δικτύου πρέπει να αντιστοιχούν στο ελάχιστο σφάλμα του συνόλου δοκίμων (Min Test Set Error), καθώς από το σημείο που αυτό θα αρχίσει να αυξάνεται έχουμε το φαινόμενο της υπερεκμάθησης. Επίσης στη πραγματική λειτουργία του συστήματος χρησιμοποιείται το σύνολο βαρών που καταγράφηκε στην ελάχιστη τιμή του σφάλματος στο σύνολο δοκίμων (Best Test Set Weights File). Έτσι ελαχιστοποιείται το φαινόμενο της υπέρ-εκπαίδευσης (over-training) και αντιστοίχα μεγιστοποιείται η ικανότητα γενίκευσης.

4.3.8 Αξιολόγηση Μοντέλου

Τα κριτήρια με τα οποία αξιολογείται η εγκυρότητα της υλοποίησης ενός μοντέλου ΤΝΔ είναι τα ακόλουθα:

- Απόδοση (Performance)

Η απόδοση του συστήματος αξιολογείται με βάση μετρικές της ικανότητας γενίκευσης

- Γενικότητα (Generality)

Το κριτήριο αυτό εξετάζει εάν η μέθοδος που υλοποιεί το μοντέλο μπορεί να εφαρμοσθεί επιτυχώς σε πολλούς διαφορετικούς τομείς.

- Ευστάθεια (Stability)

Το κριτήριο της ευστάθειας αναφέρεται στην ικανότητα του μοντέλου να διατηρεί τη γνώση που έχει συσσωρεύσει. Ένα σύστημα που έχει τη δυνατότητα μάθησης θα πρέπει να μπορεί να ξεχνά παλιότερη γνώση όταν τα όρια της μνήμης του ξεπεραστούν, είτε όταν νέα δεδομένα δε συμφωνούν με παλαιά γνώση. Για να χαρακτηριστεί ένα μοντέλο ως ευσταθές θα πρέπει η συσσώρευση της γνώσης και η απόδοσή του να είναι μονοτονικά αύξουσες στην πάροδο του χρόνου.

- Αποδοτικότητα (Efficiency)

Η ιδιότητα της αποδοτικότητας εκτιμά την υπολογιστική πολυπλοκότητα του μοντέλου και την ταχύτητα εκμάθησης. Η αξιολόγηση γίνεται μεταξύ των μοντέλων τα οποία διαθέτουν ορισμένες σταθερές παραμέτρους απόδοσης.

- Συνέπεια (Consistency)

Η συνέπεια εξετάζει αν τα συστήματα που παράγονται από το μοντέλο διαθέτουν συνεπή απόδοση σε παρόμοια σύνολα δεδομένων. Το κριτήριο αυτό είναι ιδιαίτερα σημαντικό για μη – επιβλεπόμενα συστήματα.

- Δομική Ευστάθεια – Σταδιακή Μάθηση (Incremental Learning)

Ένα σύστημα το οποίο διαθέτει την ιδιότητα της σταδιακής μάθησης μπορεί να προσαρμόζεται σε νέα δεδομένα και κατά συνέπεια διαθέτει δομική ευστάθεια. Η ιδιότητα αυτή είναι ιδιαίτερα σημαντική για συστήματα που λειτουργούν σε πραγματικό χρόνο (real time).

- Ευαισθησία (Sensitivity)

Η απόδοση του συστήματος πρέπει να παραμένει σταθερή για μικρές διαταραχές των παραμέτρων του συστήματος ή των δεδομένων εκπαίδευσης. Υψηλή ευαισθησία σε

μικρές αλλαγές στο ίδιο σύστημα ή το περιβάλλον του συνιστούν μικρή ανεκτικότητα στο θόρυβο και τα σφάλματα.

Βιβλιογραφία

- [1] Bernard Widrow and Samuel Stearns, "Adaptive Signal Processing," Prentice Hall inc. Signal Processing Series, U.S.A, 1985.
- [2] S. Haykin, "Adaptive Filter Theory", Prentice Hall, U.S.A, 2002.
- [3] Braspenning P.J., Thuijman, F., Weijter, Ed., "Artificial Neural Networks: An Introduction to ANN Theory and Practice," Springer, Hiedelberg, Germany, 1991.
- [4] Ben Krose and Patrick van der Smagt, "Artificial Neural Networks: An Introduction to Neural Networks," University of Amsterdam, 1996.
- [5] Μαλαβάζος Κωνσταντίνος, "Γενετικοί Αλγόριθμοι και το Πρόβλημα Ταυτοποίησης Συστημάτων ARMA," Master Thesis, Τμήμα Η/Υ και Πληροφορικής Πανεπιστημιού Πατρών, Πάτρα, 2002
- [6] Azoff, E.M., "Neural Network Time Series Forecasting of Financial Markets," Chichester, John Wiley & Sons – (Wiley finance editions), 1994
- [7] McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics, 5, 115-133.
- [8] Minsky, M., & Papert, S. (1969). Perceptrons: An Introduction to Computational Geometry. The MIT Press.
- [9] Rosenblatt, F. (1959). Principles of Neurodynamics. New York: Spartan Books.
- [10] Widrow, B., & Hoff, M. E. (1960). Adaptive switching circuits. In 1960 IRE WESCON Convention Record (pp. 96-104). DUNNO.
- [11] Anderson, J. A. (1977). Neural models with cognitive implications. In D. LaBerge & S. J. Samuels (Eds.), Basic Processes in Reading Perception and Comprehension Models (p.27-90). Hillsdale, NJ: Erlbaum
- [12] Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. Proceedings of the National Academy of Sciences, 79, 2554-2558.
- [13] Kohonen, T. (1977). Associative Memory: A System-Theoretical Approach. Springer-Verlag.
- [14] Hebb, D. O. (1949). The Organization of Behaviour. New York: Wiley.
- [15] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by backpropagating errors. Nature, 323, 533-536.

- [16] Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5), 359-366.
- [17] Funahashi, K.-I. (1989). On the approximate realization of continuous mappings by neural networks. *Neural Networks*, 2(3), 193-192.
- [18] Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems*, 2(4), 303-314.
- [19] Hartman, E. J., Keeler, J. D., & Kowalski, J. M. (1990). Layered neural networks with Gaussian hidden units as universal approximations. *Neural Computation*, 2(2), 210-215.
- [20] K.S. Tang, K.F. Man, S. Kwong and Q. He, "Genetic Algorithms and their Applications", *IEEE Signal Processing Magazine*, November 1996, pp 22-37.
- [21] "How Do Genetic Algorithms Work?", http://library.thinkquest.org/18242/ga_math.shtml
- [22] R.L. Haupt and S. Haupt, "Practical Genetic Algorithms", John Wiley & Sons, inc, 1997.
- [23] H. Holland, "Adaptation on Natural and Artificial Systems", Ann Arbor: The University of Michigan Press, 1975.
- [24] Ψωμάς Παναγιώτης, "Πολλαπλά Νευρωνικά Δίκτυα και Εφαρμογές σε Πρόβλεψη Χρονοσειρών," Διδακτορική Διατριβή, Εθνικό Μετσόβιο Πολυτεχνείο. Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών, 1996.

