Αναγνώριση Προτύπων ΙΙ

Ενότητα 1: Εκπαίδευση Συστημάτων Αναγνώρισης Προτύπων

Αν. Καθηγητής Δερματάς Ευάγγελος

Τμήμα Ηλεκτρολόγων Μηχανικών και Τεχνολογίας Υπολογιστών

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Άδειες Χρήσης

Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Πατρών**» έχει χρηματοδοτήσει μόνο τη αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Περιεχόμενα

1. Εκπαίδευση Συστημάτων Αναγνώρισης Προτύπων

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ

ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ
ΤΕΧΝΟΛΟΓΙΑΣ ΥΠΟΛΟΓΙΣΤΩΝ

ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ ΙΙ

Βαγγέλης Δερματάς

Πάτρα 1998

Πίνακας περιεχομένων

ΚΕΦΑΛΑΙΟ

1

Εκπαίδευση συστημάτων αναγνώρισης προτύπων

Εισαγωγή

Η γενική της μορφή μιας συνάρτησης που χρησιμοποιείται για να υλοποιήσει ένα αιτιοκρατικό σύστημα ταξινόμησης για πρότυπα που δεν μεταβάλλονται στο χρόνο, δίνεται από την σχέση R($\mathbf{w}$, $\mathbf{x}$), όπου $\mathbf{w} \in \Re^K$ είναι η "μνήμη" του συστήματος, και $\mathbf{x} \in \Re^N$ είναι η διανυσματική παράσταση του άγνωστου πρότυπου.

Η συνάρτηση $\mathbf{R}$: $\Re^N \rightarrow \Re^M$ έχει συνήθως αριθμό εξόδων ίδιο με τον αριθμό των κατηγοριών που το σύστημα αναγνωρίζει. Κάθε έξοδος της συνάρτησης ταξινόμησης αντιστοιχεί σε μία κατηγορία προτύπων. Η ταξινόμηση πραγματοποιείται με την εύρεση της συνιστώσας που έχει την μεγαλύτερη αριθμητική τιμή όταν στην είσοδο τοποθετηθεί το άγνωστο πρότυπο $\mathbf{x}$.

Η κατασκευή συστήματος ταξινόμησης προτύπων απαιτεί την λύση τριών προβλημάτων:

Επιλογή της συνάρτησης ταξινόμησης. Δυστυχώς δεν υπάρχει γενική μέθοδος η οποία να μας βοηθά στην επιλογή της μορφής της συνάρτησης ταξινόμησης βελτιστοποιώντας κάποιο αντικειμενικό κριτήριο.

Εύρεση παραδειγμάτων εκπαίδευσης. Ο πλήρης ορισμός της συνάρτησης ταξινόμησης απαιτεί τον προσδιορισμό της παραμέτρου $\mathbf{w} \in \Re^K$. Ένα καλό κριτήριο εκτίμησης της παραμέτρου (χρησιμοποιείται σχεδόν πάντα κατά την σχεδίαση συστημάτων ταξινόμησης προτύπων), είναι η εύρεση της αριθμητικής τιμής που ελαχιστοποιεί το μέσο σφάλμα της εξόδου της συνάρτησης ταξινόμησης από επιθυμητές τιμές ($\mathbf{d}_i \in \Re^M$) σε διακεκριμένα σημεία του πεδίου ορισμού της ($\mathbf{x}_i \in \Re^N$).

Στην αναγνώριση προτύπων, τα παραδείγματα εκπαίδευσης κατασκευάζονται σαν ένα σύνολο διατεταγμένων ζευγών τα οποία αποτελούνται από: την διανυσματική παράσταση του πρότυπου, και την πληροφορία που περιέχει την αντίστοιχη κατηγορία του πρότυπου. Η πληροφορία αυτή μετασχηματίζεται σε επιθυμητή τιμή της συνάρτησης ταξινόμησης ορίζοντας το διάνυσμα $\mathbf{d}_i = (v_l, v_l, \ldots v_l, v_m, v_l, \ldots, v_l)$, $v_l, v_m \in \Re^N$ και $v_l < v_m$ στο οποίο, η αριθμητική τιμή v_l τοποθετείται σε όλες τις συνιστώσες του εκτός από την συνιστώσα εκείνη στην οποία αναγνωρίζουμε την κατηγορία του πρότυπου $\mathbf{x}_i$ στην οποία τοποθετείται μεγαλύτερη αριθμητική τιμή η v_m. Το ιδεατό διάνυσμα εξόδου θα θέλαμε να προκύπτει όταν τοποθετήσουμε στην είσοδο της συνάρτησης ταξινόμησης το διάνυσμα $\mathbf{x}_i$. Στην έξοδο θα ταξινομήσουμε το πρότυπο στην έξοδο που τοποθετήσαμε την μεγαλύτερη αριθμητική τιμή, την τιμή v_m.

Στην πράξη πολλές φορές παρουσιάζονται μερικές διαφοροποιήσεις από την ιδανική προσέγγιση που δόθηκε:

- Σε έναν αριθμό παραδειγμάτων εκπαίδευσης δεν είναι γνωστή η αντίστοιχη κατηγορία του πρότυπου. Το πρόβλημα που έχουμε να αντιμετωπίσουμε σε αυτή την περίπτωση είναι να χρησιμοποιήσουμε και τις δύο κατηγορίες παραδειγμάτων, τα παραδείγματα στα οποία γνωρίζουμε την αντίστοιχη κατηγορία προτύπων και εκείνα για τα οποία γνωρίζουμε μόνο την διανυσματική περιγραφή του πρότυπου. Στόχος μας είναι να μεγιστοποιήσουμε την ακρίβεια με την οποία υπολογίζουμε τις σταθερές παραμέτρους της συνάρτησης ταξινόμησης, το **x.**

- Τα πρότυπα εκπαίδευσης είναι ημιτελή, δηλαδή δεν γνωρίζουμε όλες τις συνιστώσες των διανυσμάτων εκπαίδευσης. Το πρόβλημα των ημιτελών προτύπων παρουσιάζεται συχνά σε συστήματα ιατρικών διαγνώσεων. Το πρόβλημα της εκπαίδευσης απαιτεί τον προσδιορισμό των σταθερών συντελεστών της συνάρτησης ταξινόμησης από αρχειοθετημένες περιπτώσεις ασθενών στις οποίες δεν είναι πλήρως καταχωρημένες οι φυσικές και εργαστηριακές εξετάσεις. Ημιτελή πρότυπα εμφανίζονται επίσης και κατά την διαδικασία ταξινόμησης, δηλαδή ζητείται η πραγματοποίηση ιατρικής διάγνωσης παρόλο που πολλές φορές δεν είναι διαθέσιμα όλες οι πληροφορίες που απαιτούνται για την πλήρη συμπλήρωση του πρότυπου. Πρόσφατα έχουν αντιμετωπισθεί με σχετική επιτυχία προβλήματα που έχουν να κάνουν με ημιτελή γνώση παραδειγμάτων εκπαίδευσης.

Εκπαίδευση συστημάτων ταξινόμησης προτύπων ή ισοδύναμα υπολογισμός των σταθερών συντελεστών της συνάρτησης ταξινόμησης.

Η ακρίβεια υπολογισμού των σταθερών παραμέτρων της συνάρτησης ταξινόμησης συνεισφέρει σημαντικά στην τελική αξιοπιστία του συστήματος ταξινόμησης. Υποθέτοντας ότι τα παραδείγματα εκπαίδευσης είναι αρκετά σε αριθμό και αντιπροσωπευτικά (έχουν ληφθεί με τυχαία δειγματοληψία) το πρόβλημα της εκπαίδευσης μπορεί να περιγραφεί από το ακόλουθο πρόβλημα βελτιστοποίησης:

$$w_o = \arg\min_w \Sigma\phi(w) = \arg\min_w \sum_{i=1}^{I} D(R(w, x_i), d_i)$$

όπου $D(\mathbf{x}_1, \mathbf{x}_2)$ είναι η συνάρτηση που αποδίδει την "απόσταση" ανάμεσα στα διανύσματα $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_i$, i=1,I είναι τα I διανύσματα εκπαίδευσης και $\mathbf{d}_i$, i=1,I είναι οι αντίστοιχες επιθυμητές τιμές της εξόδου της συνάρτησης ταξινόμησης.

Δυστυχώς η επίλυση της εξίσωσης στην γενική της μορφή δεν είναι δυνατή. Εξαρτάται από την μορφή της συνάρτησης απόστασης $D(\mathbf{y},\mathbf{x})$, η οποία είναι μία μονοσήμαντη απεικόνιση $D: (\Re^M, \Re^M) \to \Re$, και από την μορφή της συνάρτησης ταξινόμησης, η οποία περιγράφει ένα αιτιοκρατικό (συνήθως μη γραμμικό) δίκτυο.

Στο παρόν κεφάλαιο περιγράφονται μερικές γενικές μέθοδοι προσδιορισμού των σταθερών παραμέτρων της συνάρτησης ταξινόμησης. Σαν συνάρτηση απόστασης θα χρησιμοποιηθεί κυρίως το τετράγωνο της Ευκλείδειας απόστασης της εξόδου της συνάρτησης από την επιθυμητή (ιδεατή) έξοδο. Το συνολικό αθροιστικό σφάλμα υπολογίζεται στα σημεία που ορίζονται από τα παραδείγματα εκπαίδευσης. Θα δοθούν επίσης και άλλες εναλλακτικοί τρόποι υπολογισμού της απόστασης δύο διανυσμάτων. Το τετράγωνο της Ευκλείδειας απόστασης είναι η πλέον διαδεδομένη συνάρτηση απόστασης και χρησιμοποιείται διότι έχει πολλές χρήσιμες θεωρητικές και πρακτικές ιδιότητες.

Γενικά το πρόβλημα της εκπαίδευσης συστήματος ταξινόμησης προτύπων από παραδείγματα έχει ως εξής. Έστω ότι είναι διαθέσιμα I παραδείγματα,

$$\Omega=\{(\mathbf{x}_1, \omega_1), (\mathbf{x}_2, \omega_2),..., (\mathbf{x}_I, \omega_I)\}.$$

Ζητείται να κατασκευαστεί αλγόριθμος ο οποίος να υπολογίζει τους σταθερούς συντελεστές συνάρτησης ταξινόμησης έτσι ώστε η συνάρτηση ταξινόμησης να ελαχιστοποιεί το σφάλμα των πραγματικών από τις ιδεατές τιμές για το σύνολο των παραδειγμάτων εκπαίδευσης.

Μερικές από τις λύσεις στο πρόβλημα εύρεσης του βέλτιστου αλγόριθμου δίνονται στο παρόν κεφάλαιο. Η πλέον διαδεδομένη ακολουθία ενεργειών που εκτελούμε για την επιτυχημένη εκπαίδευση συστήματος ταξινόμησης προτύπων είναι η ακόλουθη:

Ο πρώτος μετασχηματισμός που λαμβάνει χώρα είναι η μετατροπή του προβλήματος εκπαίδευσης σε πρόβλημα βελτιστοποίησης. Συγκεκριμένα το σύνολο των παραδειγμάτων εκπαίδευσης μετατρέπεται σε διατεταγμένα ζεύγη τιμών εισόδου-ιδεατής εξόδου για την συνάρτηση ταξινόμησης:

$$\Omega=\{(\mathbf{x}_1, \omega_1), (\mathbf{x}_2, \omega_2),..., (\mathbf{x}_I, \omega_I)\} \to \Omega_o=\{(\mathbf{x}_1, d_1), (\mathbf{x}_2, d_2),..., (\mathbf{x}_I, d_I)\}$$

Η δεύτερη ενέργεια που πρέπει να γίνει είναι να οριστεί η συνάρτηση σφάλματος η οποία θα ποσοτικοποιεί την έννοια της απόστασης του διανύσματος της ιδεατής εξόδου από την πραγματική έξοδο που μας δίνει η συνάρτηση ταξινόμησης.

Η τρίτη και δυσκολότερη ενέργεια αναφέρεται στην εύρεση του αναλυτικού η επαναληπτικού αλγόριθμου εκπαίδευσης που οδηγεί στην επίλυση του προβλήματος βελτιστοποίησης που περιγράφεται από την εξίσωση.

Στην συνέχεια αυτού του κεφαλαίου θα προχωρήσουμε στην αναλυτική περιγραφή κάθε μίας των σημαντικότερων ενεργειών που πρέπει να εκτελέσουμε για την κατασκευή ενός αιτιοκρατικού συστήματος ταξινόμησης προτύπων.

Η συνάρτηση σφάλματος

Η συνάρτηση που αποδίδει την διαφορά που υπάρχει ανάμεσα στις επιθυμητές (ιδεατές) τιμές της συνάρτησης ταξινόμησης από τις πραγματικές της τιμές (για δεδομένη τιμή της μνήμης της) δεν είναι γνωστή εκ των προτέρων. Το πρόβλημα ορισμού της συνάρτησης απόστασης δεν έχει μοναδική λύση διότι εξαρτάται από την στοχαστική συμπεριφορά των προτύπων κάθε κατηγορίας η οποία στην πλειονότητα των προβλημάτων που αντιμετωπίζουμε δεν είναι γνωστή εκ των προτέρων. Σε αυτή την παράγραφο αποδεικνύουμε ότι μπορούμε να ορίσουμε μία βέλτιστη συνάρτηση απόστασης όταν γνωρίζουμε την πυκνότητα πιθανότητας των προτύπων και ικανοποιούνται κάποιες προϋποθέσεις.

Ξεκινάμε με την υπόθεση ότι *δεν υπάρχει αριθμητική τιμή για την μνήμη της συνάρτησης ταξινόμησης τέτοια ώστε να επιτευχθεί ταυτότητα αριθμητικών τιμών μεταξύ της πραγματικής εξόδου και της αντίστοιχης ιδεατής της τιμής για όλα τα παραδείγματα εκπαίδευσης.*

Η υπόθεση αυτή γίνεται βεβαιότητα όταν ο αριθμός των παραδειγμάτων εκπαίδευσης είναι μεγαλύτερος από τον αριθμό των ελεύθερων παραμέτρων της μνήμης του συστήματος. Η συνθήκη αυτή καλύπτει την πλειονότητα των περιπτώσεων εκπαίδευσης που συναντώνται σε πρακτικές εφαρμογές.

Η δεύτερη υπόθεση που κάνουμε είναι η ακόλουθη: *κάθε ιδεατή τιμή της εξόδου μπορεί να αναλυθεί σαν το άθροισμα της πραγματικής εξόδου της συνάρτησης ταξινόμησης και της αριθμητικής τιμής μίας στοχαστικής διεργασίας που παράγει πολυδιάστατα διανύσματα που ακολουθούν μία συνάρτηση πυκνότητας πιθανότητας με στατιστικά χαρακτηριστικά μέσης τιμής μηδέν και διασποράς Σ.*

Η υπόθεση αυτή εκφράζει και την επιθυμία μας να εκτιμήσουμε την συνάρτηση ταξινόμησης με βέλτιστο τρόπο, δίχως δηλαδή την παρουσία κάποιου συστηματικού σφάλματος, γιαυτό το λόγο θέλουμε το λάθος εκτίμησης να έχει μέσο στατιστικό σφάλμα μηδέν. Συνεπώς και η στοχαστική διεργασία που θα περιγράφει το σφάλμα εκτίμησης πρέπει και αυτό να έχει μέση τιμή μηδέν.

Η τυχαία μεταβλητή του σφάλματος δίνεται από την ακόλουθη σχέση:

$$\boldsymbol{d}_i = \boldsymbol{y}_i + \boldsymbol{e}_i \rightarrow \boldsymbol{e}_i = \boldsymbol{y}_i - d_i$$

Λαμβάνοντας υπόψη τα παραπάνω, η εκπαίδευση του συστήματος ταξινόμησης μπορεί να περιγραφεί με το ακόλουθο πρόβλημα στοχαστικής βελτιστοποίησης. Αν είναι γνωστό το σύνολο των παραδειγμάτων εκπαίδευσης, ποιά είναι η μνήμη της συνάρτησης ταξινόμησης η οποία μεγιστοποιεί την πιθανότητα παραγωγής των ιδεατών τιμών στην εξοδό της;

Αν εκφράσουμε την ίδια πρόταση αλγεβρικά πρέπει να γράψουμε:

$$w_o = \arg\max_w p((d_1 \mid x_1),(d_2 \mid x_2),...,(d_I \mid x_I))$$

Υποθέτοντας επίσης ότι *η στοχαστική συμπεριφορά των πρότυπων είναι αμετάβλητη στο χρόνο και δεν υπάρχει στοχαστική συσχέτιση μεταξύ των προτύπων* $\boldsymbol{x}_i$, το κριτήριο υπολογισμού της μνήμης της συνάρτησης ταξινόμησης το οποίο ονομάζεται κριτήριο μέγιστης πιθανότητας (ΜΠ), (στην διεθνή βιβλιογραφία θα το βρείτε σαν maximum likelihood - ML), μπορεί να αναλυθεί ως εξής:

$$w_o = \arg\max_w \prod_{i=1}^{I} p(d_i \mid x_i)$$

Επειδή η συνάρτηση του λογάριθμου είναι μονότονα αύξουσα, μετατρέπουμε το γινόμενο της σχέσης σε ένα ισοδύναμο άθροισμα:

$$w_o = \arg\max_w \sum_{i=1}^{I} \log(p(d_i \mid x_i))$$

Η συνάρτηση πυκνότητας πιθανότητας της μεταβλητής $\mathbf{d}_i$ ακολουθεί την ίδια πυκνότητα πιθανότητα με την μεταβλητή $\mathbf{e}_i$. Συνεπώς αναλύουμε την συνάρτηση πυκνότητας πιθανότητας $p(d_i|\mathbf{x}_i)$ σαν συνάρτηση της γνωστής πυκνότητας πιθανότητας του σφάλματος και λύνουμε το πρόβλημα βελτιστοποίησης. Για να δώσουμε αναλυτική λύση πρέπει να προχωρήσουμε σε μία υπόθεση που έχει να κάνει με την μορφή της συνάρτησης πυκνότητας πιθανότητας.

Η Ευκλείδεια συνάρτηση σφάλματος

Αν υποθέσουμε ότι η συνάρτηση πυκνότητας πιθανότητας του σφάλματος d_i είναι κανονική κατανομή με μέση τιμή μηδέν και πίνακα διασπορά Σ, τότε αν αντικαταστήσουμε στην συνάρτηση του κριτηρίου έχουμε:

$$w_o = \arg\max_w \sum_{i=1}^{I} \log(N(y_i - d_i, 0, \Sigma)) = \arg\max_w \sum_{i=1}^{I} \log(N(R(w, x_i) - d_i, 0, \Sigma))$$

Αντικαθιστώντας την αναλυτική εξίσωση της κανονικής κατανομής για πολυδιάστατη μεταβλητή και απλοποιώντας τις εκφράσεις που προκύπτουν, έχουμε τελικά:

$$w_o = \arg\min_w \sum_{i=1}^{I} (R(w,x_i) - d_i)^T \Sigma^{-1} (R(w,x_i) - d_i)$$

Η τελευταία εξίσωση μας λεει ότι: αν θέλουμε να εκτιμήσουμε την μνήμη της συνάρτησης ταξινόμησης από πεπερασμένο αριθμό παραδειγμάτων (βάσει του κριτηρίου εκπαίδευσης της μέγιστης πιθανότητας) και ικανοποιούνται οι στοχαστικές υποθέσεις που κάναμε, τότε η βέλτιστη μνήμη είναι εκείνη η οποία ελαχιστοποιεί το αθροιστικό σφάλμα του τετραγώνου της γενικευμένης Ευκλείδειας συνάρτησης από τις αντίστοιχες επιθυμητές τιμές.

Αν προχωρήσουμε σε μία ακόμα πιο αυστηρή υπόθεση, ότι ο πίνακας της διασποράς Σ είναι διαγώνιος, τότε η συνάρτηση βελτιστοποίησης απλοποιείται ως εξής:

$$w_o = \arg\min_w \sum_{i=1}^{I} \sum_{k=1}^{K} (R_k(w,x_i) - d_{ik})^2$$

όπου Μ είναι ο συνολικός αριθμός των παραμέτρων της μνήμης της συνάρτησης ταξινόμησης και s_k η διασπορά της k παραμέτρου του διανύσματος της εξόδου.

Είναι εύκολο να αποδειχθεί ότι, αν υποθέσουμε ότι όλες οι μεταβλητές έχουν ίδια διασπορά τότε η βέλτιστη μνήμη του συστήματος ταξινόμησης προκύπτει σαν την μνήμη που ελαχιστοποιεί το τετράγωνο της Ευκλείδειας απόστασης των ιδεατών από τις πραγματικές τιμές της συνάρτησης:

$$w_o = \arg\min_w \sum_{i=1}^{I} \sum_{k=1}^{K} (R\ (w,x_i) - d_i)^T (R\ (w,x_i) - d_i)$$

Η εξίσωση περιγράφει το πλέον δημοφιλές κριτήριο βελτιστοποίησης. Χρησιμοποιείται ευρύτατα λόγω της υπολογιστικής του απλότητας.

Η συνάρτηση σφάλματος Minkowski

Η μορφή της απόδειξης που ακολουθήσαμε χρησιμοποιείται ευρύτατα στην αναγνώριση προτύπων και στην ψηφιακή επεξεργασία σήματος. Η τεχνική αυτή συνδέει την στοχαστική συμπεριφορά φαινομένων με αντίστοιχους αιτιοκρατικούς αλγόριθμους βελτιστοποίησης και χρησιμοποιείται ευρύτατα σε μη-γραμμικά συστήματα προσομοίωσης.

Γενικεύοντας την προσέγγιση που κάναμε στην προηγούμενη παράγραφο αναφέρουμε την συνάρτηση σφάλματος Minkowski η οποία προκύπτει όταν χρησιμοποιήσουμε σαν συνάρτηση πυκνότητας πιθανότητας μία ομάδα συναρτήσεων που έχει την μορφή:

$$p(x) = \frac{rK^{1/r}}{2\Gamma(1/r)} e^{-K|x|^r}$$

όπου Γ(.) είναι η συνάρτηση γάμμα, και η παράμετρος Κ τίθεται σε τιμή τέτοια έτσι ώστε το ολοκλήρωμα της συνάρτησης να δίνει τιμή 1, και $r \in \Re^+$. Εύκολα μπορούμε να αποδείξουμε ότι και η κανονική πυκνότητα πιθανότητας ανήκει σε αυτή την κατηγορία συναρτήσεων, για την τιμή r=2.

Αν επαναλάβουμε την ίδια ανάλυση θα βρούμε ότι το κριτήριο μέγιστης πιθανότητας ισοδυναμεί με την μνήμη της συνάρτησης ταξινόμησης που ελαχιστοποιεί την ακόλουθη ποσότητα:

$$w_o = \arg\min_w \sum_{i=1}^{I} | R(w, x_i) - d_i |^r$$

Αν επαναλάβουμε την ίδια ανάλυση χρησιμοποιώντας διαφορετικές συναρτήσεις πυκνότητας πιθανότητας είμαστε σε θέση να βρούμε εναλλακτικές μεθόδους υπολογισμού της μνήμης της συνάρτησης ταξινόμησης.

Εκπαίδευση αποκρατικών συστημάτων

Παρατηρώντας την μορφή της συνάρτησης εκπαίδευσης για όλα τα κριτήρια που έχουν αναφερθεί παρατηρούμε ότι σε όλες τις περιπτώσεις απαιτείται η παραμετρική εύρεση του ελάχιστου μίας συνάρτησης η οποία εκφράζεται σαν το άθροισμα μίας γενικά μη γραμμικής συνάρτησης απόστασης σε διακεκριμένα σημεία του πεδίου ορισμού της. Τα σημεία αυτά είναι τα πρότυπα των παραδειγμάτων.

Αναλυτική λύση στο πρόβλημα εύρεσης της μνήμης του συστήματος ταξινόμησης το οποίο ελαχιστοποιεί το σφάλμα προσέγγισης από τις ιδεατές τιμές που τίθενται στα

$$\Sigma\phi(x) = \sum_{i=1}^{I} | w^T x_i - d_i |^2$$

παραδείγματα εκπαίδευσης δεν υπάρχει. Σαν μία εξαίρεση σε αυτόν τον κανόνα μπορούμε να αναφέρουμε την περίπτωση του φίλτρου Wiener, το οποίο αναφέρεται στο πρόβλημα εύρεσης των σταθερών παραμέτρων της γραμμικής συνάρτησης ταξινόμησης όταν χρησιμοποιήσουμε σαν συνάρτηση σφάλματος το τετράγωνο της Ευκλείδειας συνάρτησης:

Η αναλυτική λύση προκύπτει εύκολα όταν λύσουμε την εξίσωση που προκύπτει από τον μηδενισμό της πρώτης παραγώγου του συνολικού σφάλματος Σφ(**w**) ως προς **w**. Η εξίσωση του συνολικού σφάλματος είναι δευτεροβάθμια ως προς **w**. Η πρώτη παράγωγος του σφάλματος είναι μία πρωτοβάθμια εξίσωση, εξασφαλίζοντας την ύπαρξη μοναδικής ρίζας.

Στα περισσότερα συστήματα αναγνώρισης προτύπων η συνάρτηση ταξινόμησης είναι μία μη γραμμική εξίσωση των σταθερών της συντελεστών, γεγονός το οποίο μας αναγκάζει κατά την διάρκεια της εκπαίδευσης να χρησιμοποιήσουμε επαναληπτικούς αλγόριθμους εύρεσης ελάχιστων πολυδιάστατων συναρτήσεων. Στην πλειονότητα αυτών των προσεγγίσεων δεν μπορούμε να πετύχουμε αναλυτική λύση.

Οι μέθοδοι ελαχιστοποίησης της συνάρτησης του σφάλματος είναι πολλοί σε αριθμό γιαυτό και σε αυτό το κεφάλαιο θα αναπτύξουμε τους πλέον σημαντικούς, απλούς αλγοριθμικά και εύκολους στην υλοποίηση με λογισμικό.

Τυχαία αναζήτηση

Όταν έχουμε στην διάθεσή μας, ένα ισχυρό υπολογιστικό σύστημα, αρκετό διαθέσιμο χρόνο και θέλουμε να φτιάξουμε τον πλέον απλό αλγόριθμο για την επίλυση του προβλήματος βελτιστοποίησης που περιγράφεται από το γενικό πρόβλημα εκπαίδευσης ενός αιτιοκρατικού συστήματος ταξινόμησης, όπως αυτό δίνεται αλγεβρικά από την εξίσωση του σφάλματος, τότε μπορούμε να επιλέξουμε τους συντελεστές μνήμης οι οποίοι για μεγάλο δείγμα τυχαίας δειγματοληψίας αποδίδουν το μικρότερο σφάλμα.

Αναλυτικά ο αλγόριθμος εκπαίδευσης με τυχαία αναζήτηση έχει ως εξής:

Αλγόριθμος

- **Τοποθέτηση αρχικών τιμών**. ΕΣ = 10^{30}. Τα επόμενα τρία βήματα επαναλαμβάνονται Κ φορές.

- **Επιλογή τυχαίου σημείου**. Μέσα στο πεδίο ορισμού της μνήμης της συνάρτησης ταξινόμησης επιλέγουμε με τυχαίο τρόπο μία τιμή, έστω **w**.

- **Υπολογισμός του σφάλματος**. Για την τρέχουσα τιμή της μνήμης της συνάρτησης ταξινόμησης υπολογίζω την αριθμητική τιμή του σφάλματος:

$$\Sigma\phi(w) = \sum_{i=1}^{I} D(R(w, x_i), d_i)^2$$

- **Επανεκτίμηση της μνήμης**. Αν η τρέχουσα τιμή του σφάλματος βρεθεί να είναι μικρότερη από το ελάχιστο σφάλμα που έχει βρεθεί από τις τυχαίες αναζητήσεις, το ΕΣ, τότε η καλύτερη τιμή της μνήμης φυλάσσεται $\mathbf{w}_o = \mathbf{w}$ και ταυτόχρονα επαναπροσδιορίζεται το ελάχιστο σφάλμα ΕΣ = Σφ($\mathbf{w}$) .

- **Τερματισμός**. Η καλύτερη τιμή της μνήμης του συστήματος ταξινόμησης που έχει βρεθεί με τυχαία αναζήτηση είναι η $\mathbf{w}_o$.

Η μέθοδος τυχαίας αναζήτησης είναι ιδιαίτερα χρήσιμη στις περιπτώσεις κατά τις οποίες:

1. **Η συνάρτηση ταξινόμησης έχει πολλά σημεία ασυνέχειας**. Σε αντίθετη περίπτωση είναι προτιμότερο να χρησιμοποιήσουμε τις μεθόδους που βασίζονται σε πληροφορίες που μας παρέχουν οι πρώτες παράγωγοι της συνάρτησης του σφάλματος, οι σημαντικότερες των οποίων περιγράφονται στις παραγράφους που ακολουθούν.

2. **Η συνάρτηση σφάλματος έχει πολλά τοπικά ακρότατα**. Η μέθοδος τυχαίας αναζήτησης δεν επηρεάζεται από το είδος και τον αριθμό των τοπικών ακρότατων σε αντίθεση τις μεθόδους που βασίζονται στην πληροφορία που μας παρέχουν οι παράγωγοι της συνάρτησης του σφάλματος. Οι μέθοδοι αυτοί "εγκλωβίζονται" συχνά σε "κοιλάδες" της συνάρτησης σφάλματος που ανήκουν σε τοπικά ακρότατα. Η αποτελεσματική αντιμετώπιση του προβλήματος συνίσταται στην χρήση υβριδικών μεθόδων, δηλαδή με τυχαίο τρόπο επιλέγουμε την αρχική εκτίμηση της μνήμης του συστήματος ταξινόμησης και στην συνέχεια εφαρμόζουμε κάποια από τις μεθόδους εύρεσης του ελάχιστου της συνάρτησης σφάλματος.

3. **Ο αριθμός των ελεύθερων παραμέτρων της μνήμης του συστήματος είναι μικρός**. Η συνθήκη αυτή αυξάνει την πιθανότητα να βρούμε με τυχαίο τρόπο την βέλτιστη λύση διότι το πεδίο αναζήτησης είναι μικρό. Όταν ο αριθμός των ελεύθερων παραμέτρων είναι μεγάλος, όπως για παράδειγμα στα νευρωνικά δίκτυα, ο χώρος αναζήτησης είναι πολύ μεγάλος με συνέπεια να απαιτούνται πάρα πολλές επαναλήψεις του αλγόριθμου έτσι ώστε να αυξηθεί σε ικανοποιητικό βαθμό η πιθανότητα να βρούμε σημείο στο πεδίο αναζήτησης το οποίο να βρίσκεται στην γειτονιά της βέλτιστης λύσης.

Σε περίπτωση που ο αριθμός των ελεύθερων παραμέτρων είναι πολύ μεγάλος η προσπάθειά μας επικεντρώνεται στην χρήση τεχνικών προσανατολισμένης αναζήτησης με τυχαία δειγματοληψία. Δύο είδη τέτοιων τεχνικών, που χρησιμοποιούνται αρκετά συχνά σε προβλήματα βελτιστοποίησης, η τυχαία γραμμική εκπαίδευση και οι γενετικοί αλγόριθμοι περιγράφονται συνοπτικά στα πλαίσια του παρόντος μαθήματος.

Τυχαία γραμμική εκπαίδευση

Την μέθοδο της τυχαίας γραμμικής εκπαίδευσης θα την βρούμε στην βιβλιογραφία σαν line search method. Αποτελεί περισσότερο μία στρατηγική αναζήτησης της βέλτιστης λύσης και όχι μία συγκεκριμένη αλγοριθμική μέθοδο. Η στρατηγική αυτή αποτελεί ένα ενδιάμεσο στάδιο για την μετάβαση από τις τεχνικές της τυχαίας αναζήτησης στις τεχνικές που χρησιμοποιούν πληροφορία από τις πρώτες παραγώγους της συνάρτησης του σφάλματος.

πιο συγκεκριμένα η στρατηγική της γραμμικής αναζήτησης περιορίζει αυθαίρετα τον χώρο αναζήτησης σε πρόβλημα βελτιστοποίησης συνάρτησης μίας μεταβλητής, εισάγοντας K-1 τυχαίες πεπλεγμένες γραμμικές εξισώσεις των μεταβλητών της συνάρτησης του σφάλματος:

$$\begin{gathered} a_{1,1}w_1 + \ldots + a_{1,K}w_K = b_1 \\ a_{2,1}w_1 + \ldots + a_{2,K}w_K = b_1 \\ \ldots = \ldots \\ a_{K-11,1}w_1 + \ldots + a_{K-1,K}w_K = b_{K-1} \end{gathered}$$

Επιλέγοντας με τυχαίο τρόπο τις σταθερές a_{ij} και b_i, έτσι ώστε να εξασφαλίσουμε επίσης ότι τα αντίστοιχα διανύσματα τους δεν είναι παράλληλα, μπορούμε να εκφράσουμε την συνάρτηση σφάλματος σαν συνάρτηση μόνο μίας μεταβλητής. Στην συνέχεια με την βοήθεια μεθόδου βελτιστοποίησης βρίσκουμε το ελάχιστο της συνάρτησης του σφάλματος ως προς την μοναδική μεταβλητή. Σε αυτό το σημείο έχουμε την ελευθερία να χρησιμοποιήσουμε οποιαδήποτε μέθοδο θέλουμε για να βρούμε το ελάχιστο πάνω στην ευθεία που ορίζεται από τις K-1 γραμμικές εξισώσεις.

Μετά το τέλος και της διαδικασίας βελτιστοποίησης έχουμε ορίσει πλήρως ένα σημείο στο πεδίο ορισμού της μεταβλητής **w**. Ισοδύναμα μπορούμε να πούμε ότι έχουμε βρει και άλλη μία εξίσωση η οποία δεν είναι γραμμικά εξαρτημένη από τις προηγούμενες.

Η νέα αναζήτηση επαναλαμβάνεται σε τυχαία ευθεία η οποία όμως διέρχεται από το σημείο που ορίστηκε μετά το τέλος της διαδικασίας βελτιστοποίησης. Η συνθήκη αυτή ικανοποιείται εύκολα όταν αφαιρέσουμε από τις K γραμμικές εξισώσεις μία τυχαία, η οποία δεν είναι ίδια με αυτή που χρησιμοποιήθηκε στον αλγόριθμο βελτιστοποίησης του προηγούμενου βήματος του αλγόριθμου.

Η διαδικασία αυτή επαναλαμβάνεται αρκετές φορές μέχρι να επιτευχθεί ένα ικανοποιητικό ελάχιστο.

Μία υπολογιστικά απλοποιημένη έκδοση της μεθόδου τυχαίας γραμμικής αναζήτησης, προκύπτει όταν την γραμμική αναζήτηση την υλοποιήσουμε σε τυχαίες ευθείες που είναι πάντα παράλληλες στο ορθοκανονικό σύστημα συντεταγμένων που ορίζεται από τις παραμέτρους. Με απλά λόγια, σε κάθε επανάληψη του αλγόριθμου θεωρούμε σαν μεταβλητή μονάχα μία παράμετρο, για την οποία εφαρμόζουμε έναν αλγόριθμο βελτιστοποίησης. Στην επόμενη επανάληψη, απλά επιλέγουμε μία άλλη παράμετρο την οποία προσπαθούμε να βελτιστοποιήσουμε. Ο αλγόριθμος τερματίζει μετά από έναν σταθερό αριθμό επαναλήψεων.

Αλγόριθμος

- **Τοποθέτηση αρχικών τιμών.** Τοποθετούμε τυχαίες τιμές στις παραμέτρους **w**. ΕΣ = 10^{30}. Τα επόμενα τρία βήματα επαναλαμβάνονται Κ φορές.

- **Επιλογή της ελεύθερης μεταβλητής**. Με τυχαίο τρόπο επιλέγουμε μία συνιστώσα του διανύσματος **w** σαν ελεύθερη μεταβλητή, έστω την w_i.

- **Εύρεση του ελάχιστου**. Υλοποιούμε μία μέθοδο βελτιστοποίησης με την οποία βρίσκουμε το ελάχιστο της συνάρτησης του σφάλματος θεωρώντας σαν μοναδική μεταβλητή την w_i.

- **Επιλογή**. Αν το σφάλμα είναι μικρότερο του μικρότερου σφάλματος το οποίο έχει βρεθεί κατά την διάρκεια του επαναληπτικού τμήματος του αλγόριθμου, τότε η νέα αυτή λύση αντικαθιστά την προηγούμενη εκτίμηση της βέλτιστης λύσης.

Παράδειγμα. Έστω ότι η συνάρτηση ταξινόμησης δίνεται από την σχέση:

$$R(w,x) = \left(\frac{1}{(w^T x)^2 + 1)^2}, e^{(w-x)^T (w-x)} \right)^T$$

Θέλουμε να κατασκευάσουμε σύστημα ταξινόμησης προτύπων δύο κατηγοριών διαθέτοντας τα εξής παραδείγματα:

$$\Omega_1 = \{ (1,1),(3,5),(-1,3),(-3,1)\}$$

$$\Omega_2 = \{ (-1,-1),(-2,1),(1,-1),(2,0),(0,0)\}$$

για τις δύο κατηγορίες αντίστοιχα.

Υπολογίστε την μνήμη της συνάρτησης ταξινόμησης με την βοήθεια της στρατηγικής τυχαίας γραμμικής αναζήτησης.

Εκπαίδευση κατά την διεύθυνση της αρνητικής παραγώγου

Η απλούστερη επαναληπτική μέθοδος εύρεσης του ελάχιστου συνάρτησης η οποία χρησιμοποιεί πληροφορία τοπικών παραγώγων είναι η αναζήτηση του ελάχιστου μετακινώντας την βέλτιστη λύση κατά την αρνητική παράγωγο της συνάρτησης του σφάλματος. Η μέθοδος αυτή χρησιμοποιείται πολύ συχνά διότι είναι αλγοριθμικά απλή, συγκλίνει γρήγορα και στις περισσότερες εφαρμογές δίνει ικανοποιητικές λύσεις. Στην βιβλιογραφία θα την βρείτε αν gradient descent ή steepest descent method.

Η αρχή αναζήτησης του ελάχιστου κατά την διεύθυνση της αρνητικής παραγώγου βασίζεται στο γεγονός ότι όταν αναζητούμε το ελάχιστο μίας **συνεχούς** συνάρτησης και γνωρίζουμε την τιμή της πρώτης παραγώγου σε κάποιο σημείο της μπορούμε να αποφανθούμε για την **κατεύθυνση στην οποία βρίσκεται το κοντινότερο τοπικό ελάχιστο**. Βρίσκεται προς την κατεύθυνση της αρνητικής τιμής της πρώτης παραγώγου της συνάρτησης. Η πλέον απλή μορφή του αλγόριθμου έχει ως εξής:

Αλγόριθμος

Τοποθέτηση αρχικών τιμών. Τοποθετούμε τυχαίες τιμές στις παραμέτρους $\mathbf{w}^{(0)}$,

$t \rightarrow 0$.

Νέα εκτίμηση της μνήμης της συνάρτησης ταξινόμησης. $t \rightarrow t+1$. Ακολουθεί επαναπροσδιορισμός της βέλτιστης λύσης:

$$\mathbf{w}^{(t)} = \mathbf{w}^{(t-1)} - a \;\nabla\Sigma\varphi(\mathbf{w}^{(t-1)})$$

όπου η συνάρτηση σφάλματος έχει σαν ελεύθερους παραμέτρους μόνο την μνήμη της

$$\Sigma\phi(w)^{(t)} = \sum_{i=1}^{I} D(R(w^{(t)}, x_i), d_i)$$

συνάρτησης ταξινόμησης:

και η διανυσματική παράγωγος του σφάλματος ως προς τις ελεύθερες μεταβλητές της είναι:

$$\nabla \Sigma \phi(w) = \left(\frac{\partial \Sigma \phi(w)}{\partial w_1}, \frac{\partial \Sigma \phi(w)}{\partial w_2}, \ldots, \frac{\partial \Sigma \phi(w)}{\partial w_K} \right)$$

Έλεγχος σύγκλισης. Αν η σχετική μεταβολή του σφάλματος είναι πολύ μικρή (έχουμε προσεγγίσει ένα τοπικά ελάχιστο ή μία περιοχή στην οποία η κλίση της παραγώγου του σφάλματος είναι πολύ μικρή):

$$\left| \frac{\Sigma \phi(w)^{(t)} - \Sigma \phi(w)^{(t-1)}}{\Sigma \phi(w)^{(t-1)}} \right| < Thr$$

ή ο αριθμός των βημάτων του αλγόριθμου έχει ξεπεράσει το μέγιστο επιτρεπόμενο υπολογιστικό κόστος, τότε ο αλγόριθμος τερματίζει, διαφορετικά επαναλαμβάνονται οι υπολογισμοί από το δεύτερο βήμα.

Η τελευταία τιμή της μνήμης της συνάρτησης ταξινόμησης, η $\mathbf{w}^{(t)}$, ορίζεται να είναι η βέλτιστη λύση που δίνει ο αλγόριθμος.

Η επιλογή της παραμέτρου a, με a $\in \Re^+$, στην εξίσωση επαναπροσδιορισμού της μνήμης του συστήματος ταξινόμησης, που ονομάζεται και ρυθμός εκπαίδευσης, είναι κρίσιμη για την συμπεριφορά του αλγόριθμου.

Αν η τιμή της επιλεγεί μικρή (κοντά στο μηδέν), τότε αν κατά τους διαδοχικούς επαναπροσδιορισμούς περάσουμε από κάποια περιοχή τιμών στην οποία η παράγωγος του σφάλματος έχει μικρή αριθμητική τιμή τότε θα παρατηρήσουμε ότι ο αλγόριθμος "κολλάει", με την έννοια ότι απαιτείται πια πολύ μεγάλος αριθμός επαναλήψεων για να μεταβληθεί η Ευκλείδεια απόσταση των διαδοχικών εκτιμήσεων της μνήμης του συστήματος αισθητά.

Η εξήγηση είναι προφανής διότι η μεταβολή εκφράζεται σαν το άθροισμα του συντελεστή εκπαίδευσης και της πρώτης παραγώγου της συνάρτησης του σφάλματος. Αν και οι δύο αριθμητικές τιμές είναι πολύ μικρές, η αντίστοιχη μεταβολή της εκτίμησης της μνήμης του συστήματος ταξινόμησης ελαχιστοποιείται.

Αντίθετα αν η επιλογή του ρυθμού εκπαίδευσης είναι μία μεγάλη αριθμητικά τιμή τότε υπάρχει περίπτωση, αν η "κοιλάδα" στην οποία βρίσκεται το ελάχιστο έχει απότομες "πλαγιές" (η παράγωγος στην γειτονιά του τοπικά ελάχιστου έχει μεγάλη αριθμητική τιμή), η μεταβολή να προκαλεί διαδοχικές ταλαντώσεις επάνω στην γειτονιά της λύσης, χωρίς όμως να μπορεί και να την προσεγγίσει. Οι συνήθεις τιμές του συντελεστή εκπαίδευσης βρίσκονται συνήθως στο διάστημα των πραγματικών αριθμών (0,1].

Δυστυχώς δεν υπάρχει κάποια μέθοδος η οποία να οδηγεί σε μία βέλτιστη επιλογή του ρυθμού εκπαίδευσης, γιαυτό το λόγο πολλοί ερευνητές έχουν προτείνει διάφορες στρατηγικές προσαρμογής του ρυθμού εκπαίδευσης. Οι σημαντικότερες αυτών είναι οι ακόλουθες:

Από τις πλέον διαδεδομένες τεχνικές επιτάχυνσης του ρυθμού σύγκλισης του επαναληπτικού αλγόριθμου της εκπαίδευσης είναι η προσθήκη όρων στην σχέση επαναπροσδιορισμού της μνήμης του συστήματος, με κυρίαρχη την ακόλουθη:

$$w^{(t)} = w^{(t-1)} - \alpha \nabla \Sigma \phi(w^{(t-1)}) + m \Delta w^{(t-1)}$$

με a,m $\in \Re^{+}$ και

$$\Delta w^{(t-1)} = w^{(t-1)} - w^{(t-2)}$$

Ο επιπρόσθετος όρος είναι ουσιαστικά ένας "αδρανειακός" όρος ο οποίος τείνει να διατηρήσει την μεταβολή της μνήμης σε διεύθυνση ίδια με εκείνη που πραγματοποιήθηκε στο προηγούμενο βήμα του αλγόριθμου.

Με αυτή την τεχνική επιλύουμε ταυτόχρονα δύο μειονεκτήματα που παρουσιάζονται στην απλή εξίσωση επανεκτίμησης της μνήμης του συστήματος ταξινόμησης. Όταν τα διανύσματα επανεκτίμησης είναι παράλληλα τότε η συνολική συνιστώσα μεταβολής της μνήμης αυξάνει σημαντικά τον "βηματισμό" του αλγόριθμου. Επίσης ο αδρανειακός όρος επιτρέπει πολλές φορές να ξεπεράσουμε μικρές "κοιλάδες" τοπικών ελάχιστων μέσα στις οποίες μπορεί να εγκλωβιστεί ο αλγόριθμος.

Η επιλογή της αριθμητικής τιμής των σταθερών συντελεστών ασκεί σημαντική επίδραση στην συμπεριφορά του αλγόριθμου στα φαινόμενα που προαναφέραμε. Για παράδειγμα αν αυξήσουμε την "αδρανειακή" σταθερά m, τότε μπορούμε να ξεπεράσουμε και σχετικά μεγάλες κοιλάδες τοπικών ακρότατων, αν αυτό είναι βέβαια επιθυμητό. Συνήθεις τιμές του αδρανειακού όρου m βρίσκονται μέσα στο διάστημα [0,1].

Μία διαφορετική στρατηγική επιτάχυνσης της σύγκλισης συνίσταται στην αφιέρωση ενός μοναδικού συντελεστή εκπαίδευσης για κάθε μία παράμετρο της μνήμης. Η τροποποίηση αυτή διευκολύνει τον αλγόριθμο να ξεφύγει από "οροπέδια" της συνάρτησης του σφάλματος. Η παράμετρος εκπαίδευσης αυξάνεται όταν κατά δύο διαδοχικούς επαναπροσδιορισμούς το πρόσημο της μεταβολής της παραμέτρου της μνήμης παραμείνει το ίδιο. Αν το πρόσημο μεταβληθεί (η εκπαίδευση της παραμέτρου αλλάξει κατεύθυνση)τότε ο ρυθμός εκπαίδευσης παίρνει αμέσως την ελάχιστη τιμή του.

Ο συνολικός αλγόριθμος εκπαίδευσης έχει την ακόλουθη μορφή:

Αλγόριθμος

Τοποθέτηση αρχικών τιμών. Τοποθετούμε τυχαίες τιμές στις παραμέτρους $\mathbf{w}^{(0)}$

$t \rightarrow 0$, $a^{(0)} = a_0 > 0$.

Νέα εκτίμηση της μνήμης της συνάρτησης ταξινόμησης. $t \rightarrow t+1$. Ακολουθεί ο επαναπροσδιορισμός του συντελεστή εκπαίδευσης και της βέλτιστης λύσης για κάθε παράμετρο w_i:

$$a_i^{(t)} = \begin{cases} a_i^{(t-1)} + a_o, & \dfrac{\partial \Sigma \phi(w^{(t-1)})}{\partial w_i} \dfrac{\partial \Sigma \phi(w^{(t-2)})}{\partial w_i} \geq 0 \\ a_i^{(t)}, & \dfrac{\partial \Sigma \phi(w^{(t-1)})}{\partial w_i} \dfrac{\partial \Sigma \phi(w^{(t-2)})}{\partial w_i} < 0 \end{cases}$$

$$w_i^{(t)} = w_i^{(t-1)} - a_i^{(t)} \frac{\partial \Sigma \phi(w^{(t-1)})}{\partial w_i}$$

Έλεγχος σύγκλισης. Αν η σχετική μεταβολή του σφάλματος είναι πολύ μικρή (έχουμε προσεγγίσει ένα τοπικά ελάχιστο ή μία περιοχή στην οποία η κλίση της παραγώγου του σφάλματος είναι πολύ μικρή):

ή ο αριθμός των βημάτων του αλγόριθμου ξεπεράσει το μέγιστο επιτρεπόμενο υπολογιστικό κόστος, τότε ο αλγόριθμος τερματίζει, διαφορετικά επαναλαμβάνεται ο αλγόριθμος από το δεύτερο βήμα.

Η τελευταία τιμή της μνήμης της συνάρτησης ταξινόμησης, η $\mathbf{w}^{(t)}$ ορίζεται να είναι η βέλτιστη λύση του αλγόριθμου.

$$\left| \frac{\Sigma \phi(w)^{(t)} - \Sigma \phi(w)^{(t-1)}}{\Sigma \phi(w)^{(t-1)}} \right| < Thr$$

Παράδειγμα

Εστω ότι δίνεται η μη-γραμμική συνάρτηση ταξινόμησης

$$R(w,x) = \frac{1}{(x_1 - w_1)^2 + (x_2 - w_2)^4 + 1}$$

Θέλουμε να υπολογίσουμε την αριθμητική τιμή του $w = (w_1, w_2)$ έτσι ώστε η συνάρτηση $R(w,x)$ να διέρχεται όσο το δυνατόν πλησιέστερα από τα σημεία

((1,1),0.3), ((-1,-1),0.4),((0,0),0.8),((1,0),0.6),((0,1),0.67)

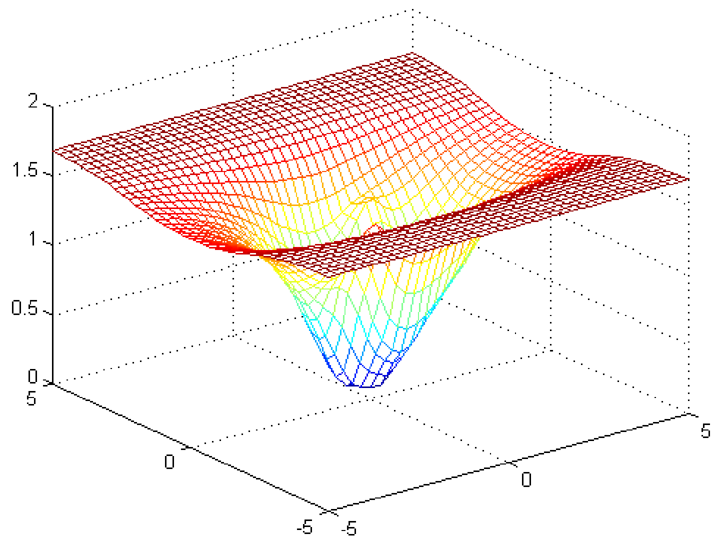
Λύση

Ορίζω σαν συνάρτηση σφάλματος την

$$\Sigma\phi(w) = \sum_{i=1}^{5}(d_i - R(w, x_i))^2$$

όπου (x_i,d_i),i=1,5 είναι οι συντεταγμένες των σημείων στον τριοδιάστατο χώρο από τα οποία θέλω να διέρχεται η συνάρτηση R(w,x). Το πρόβλημα που τελικά πρέπει να λύσω είναι να βρώ το διάνυσμα w για το οποίο η συνάρτηση σφάλματος γίνεται ελάχιστη.

Αν κατασκευάσω την γραφική παράσταση της συνάρτησης Σφ(w), στο διάστημα [-5,5] και για τις δύο ελεύθερες παραμέτρους, θα έχω το ακόλουθο σχήμα.



Για να εφαρμόσω με επιτυχία τον αλγόριθμο εύρεσης του ελάχιστου της συνάρτησης του σφάλματος κατά την διεύθυνση της αρνητικής παραγώγου χρειάζεται να υπολογίσω το μέγεθος $\nabla\Sigma\varphi(w^{(t-1)})$ το οποίο για την συγκεκριμένη συνάρτηση σφάλματος είναι:

$$\nabla\Sigma\phi(w)=\left(\frac{\partial\Sigma\phi(w)}{\partial w_1},\frac{\partial\Sigma\phi(w)}{\partial w_2}\right)=\left(\sum_{i=1}^{5}\frac{\partial}{\partial w_1}(d_i-R(w,x_i))^2,\sum_{i=1}^{5}\frac{\partial}{\partial w_2}(d_i-R(w,x_i))^2\right)=$$

$$\left(\sum_{i=1}^{5}(R(w,x_i)-d_i)\frac{\partial R(w,x_i)}{\partial w_1},\sum_{i=1}^{5}(R(w,x_i)-d_i)\frac{\partial R(w,x_i)}{\partial w_2}\right)=$$

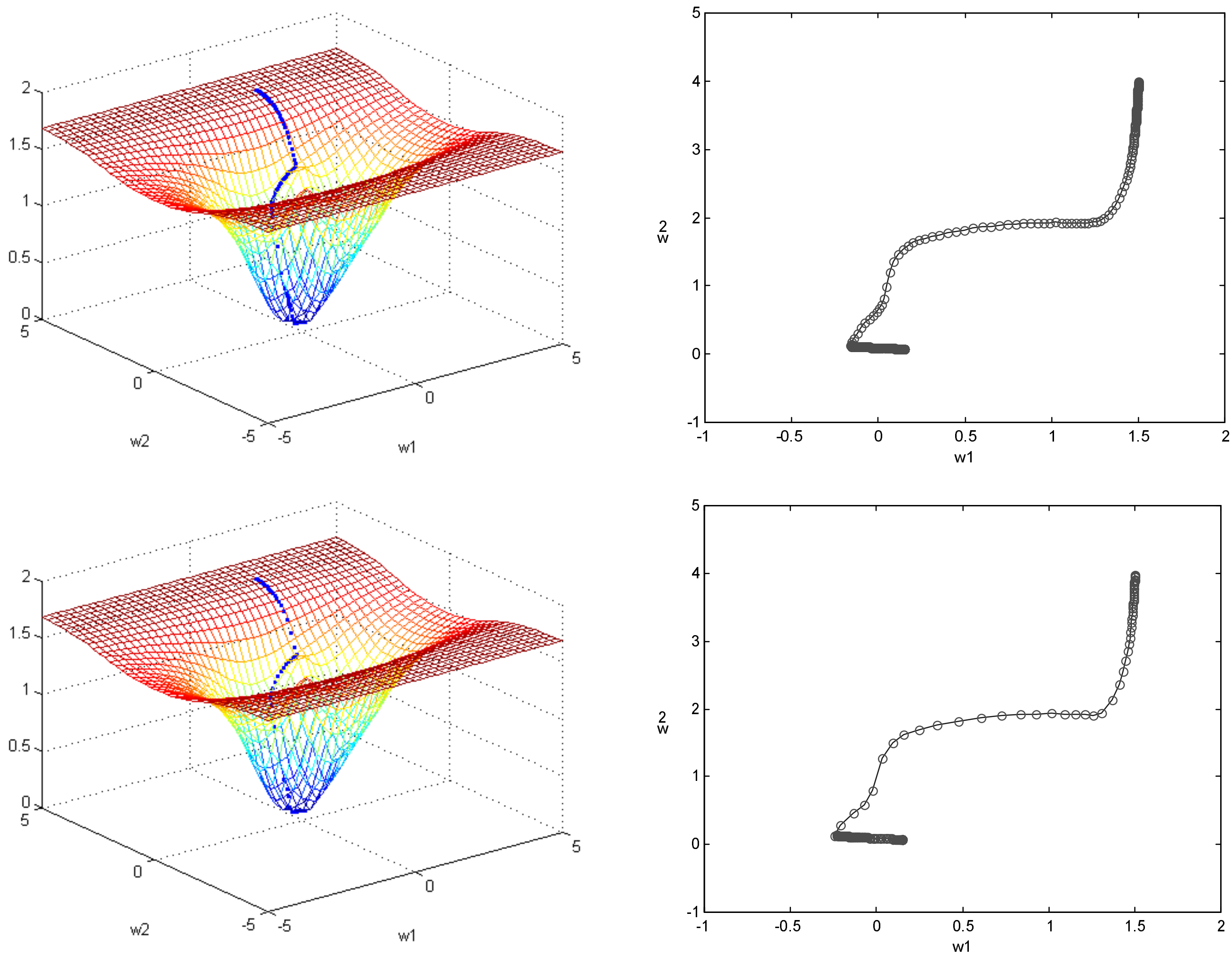
$$\left(\sum_{i=1}^{5}\frac{2(x_{i1}-w_1)(R(w,x_i)-d_i)}{\left((x_{i1}-w_1)^2+(x_{i2}-w_2)^4+1\right)^2},\sum_{i=1}^{5}\frac{4(x_{i2}-w_2)^3(R(w,x_i)-d_i)}{\left((x_{i1}-w_1)^2+(x_{i2}-w_2)^4+1\right)^2}\right)$$

Η τελικά αναδρομική σχέση που θα χρησιμοποιηθεί για την εύρεση του ελάχιστου της συνάρτησης του σφάλματος δίνεται από την σχέση:

$$w^{(t)}=w^{(t-1)}-a\left(\sum_{i=1}^{5}\frac{2(x_{i1}-w_1)(R(w,x_i)-d_i)}{\left((x_{i1}-w_1)^2+(x_{i2}-w_2)^4+1\right)^2},\sum_{i=1}^{5}\frac{4(x_{i2}-w_2)^3(R(w,x_i)-d_i)}{\left((x_{i1}-w_1)^2+(x_{i2}-w_2)^4+1\right)^2}\right)$$

όπου α είναι μικρός θετικός αριθμός.

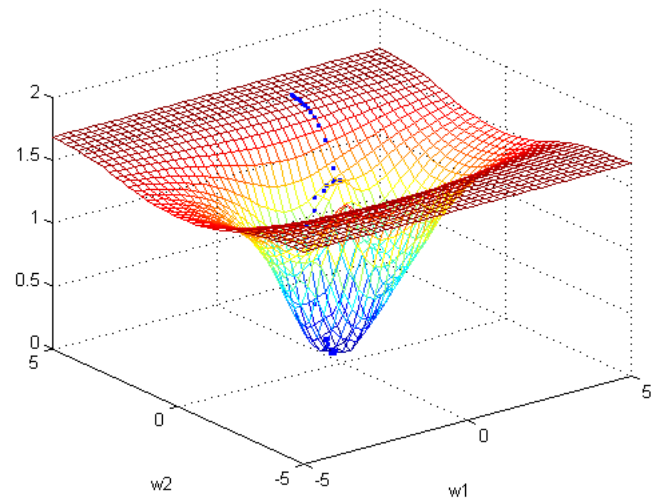
Στο σχήμα που ακολουθεί δίνονται οι διαδοχικές εκτιμήσεις του διανύσματος w με αρχική τιμή το σημείο $w^{(0)}$ = (1.5,4) και διαφορετικές τιμές του συντελεστή εκπαίδευσης. Κάθε σχήμα ακολουθείται και με γραφική απεικόνιση της ακολουθίας των λύσεων στο επίπεδο των ελεύθερων παραμέτρων

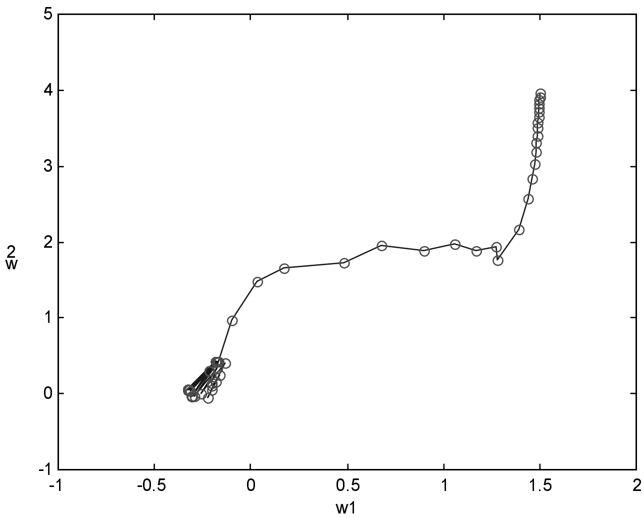


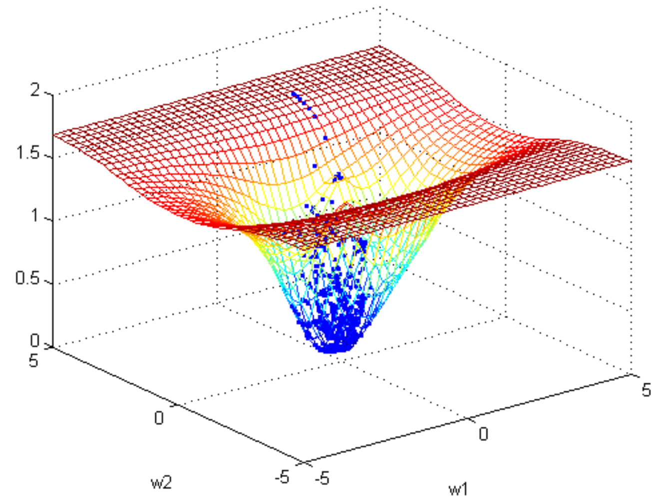
Στην πρώτη σειρά σχημάτων δείχνεται η πορεία σύγκλισης του αλγόριθμου σε τοπικό ελάχιστο (το οποίο στην συγκεκριμένη περίπτωση συμπίπτει με το συνολικό ελάχιστο της συνάρτησης του σφάλματος για τιμή του συντελεστή εκπαίδευσης a=0.2.

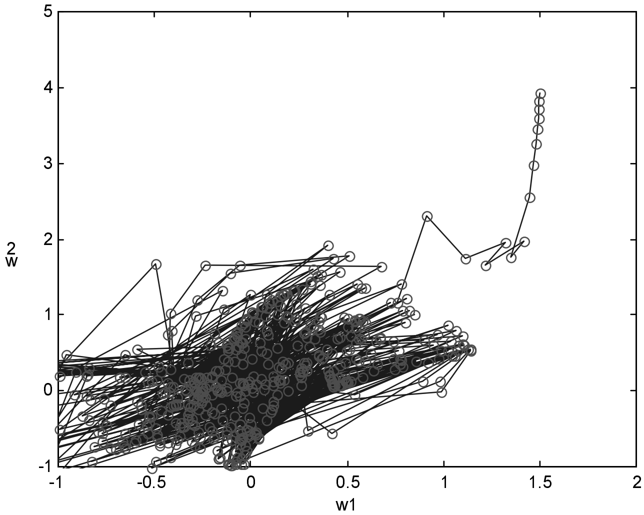
Στην δεύτερη σειρά σχημάτων δείχνεται η πορεία σύγκλισης του αλγόριθμου για τιμή του συντελεστή εκπαίδευσης a=0.5.

Όπως είναι προφανές από την μελέτη των σχημάτων ο αλγόριθμος συγκλίνει ταχύτερα στην περίπτωση που επιλεγεί συντελεστής εκπαίδευσης a=0.5.





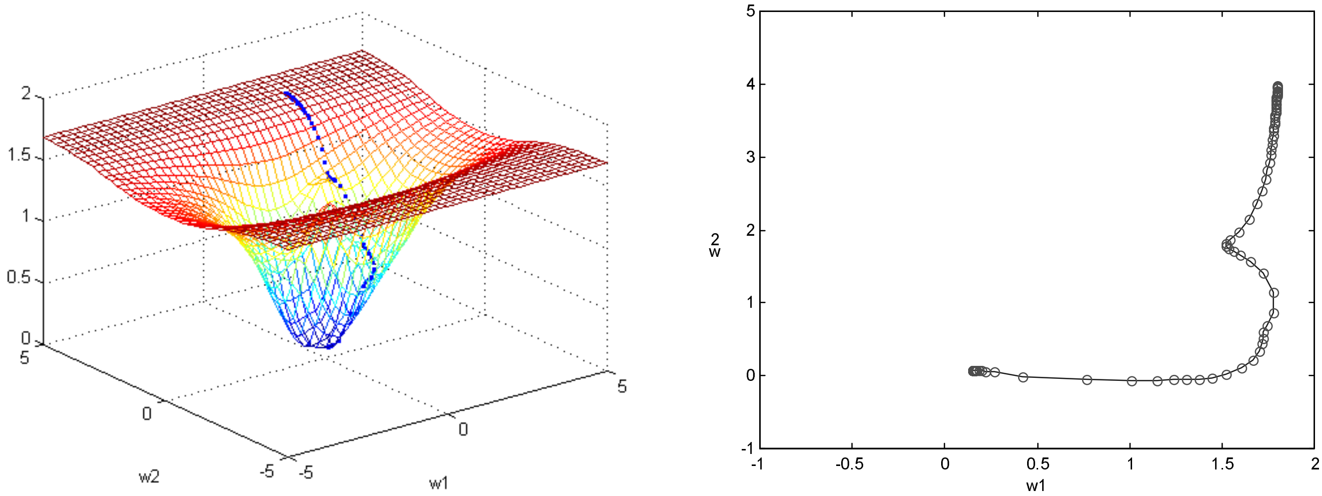




Αυξάνοντας τον συντελεστή εκπαίδευσης σε a=1.0 και a=2.0 και ξεκινώντας από το ίδιο σημείο εκτίμησης της αρχικής τιμής βλέπουμε ότι επιταχύνεται η σύγκλιση του αλγόριθμου στην γειτονιά της βέλτιστης λύσης, αλλά ο αλγόριθμος αδυνατεί να συγκλίνει με ακρίβεια στην τελική τιμή.

Δύο είναι οι λόγοι που εμποδίζουν την σύγκλιση του αλγόριθμου: (α) η παράγωγος του σφάλματος έχει μεγάλη αριθμητική τιμή κοντά στο σημείο σύγκλισης (οι "πλαγιές" της κοιλάδας του ελάχιστου είναι πολύ απότομες), και (β) ο συντελεστής εκπαίδευσης έχει μεγάλη αριθμητική τιμή.

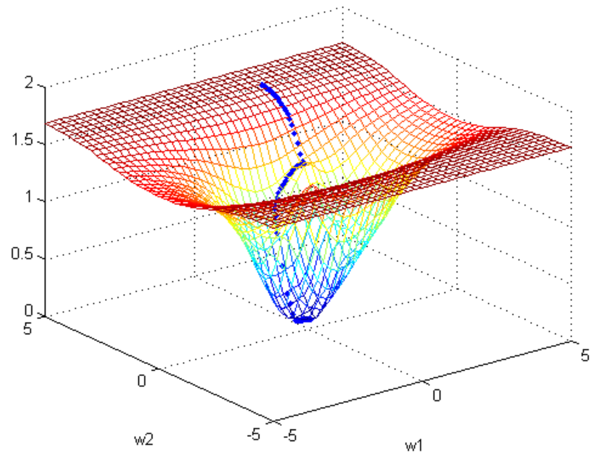
Στο σχήμα που ακολουθεί δίνεται η επίδραση που έχει στην διαδρομή που θα ακολουθήσει ο αλγόριθμος μέχρι η διαδοχική εκτίμηση του ελάχιστου λάβει την τελική της τιμή. Ο αλγόριθμος έχει επαναληφθεί με αρχική τιμή το σημείο $w^{(0)} = (1.8,4)$ αντί του $w^{(0)} = (1.5,4)$ που έχει επιλεγεί σε όλα τα προηγούμενα πειράματα και ο συντελεστής εκπαίδευσης έχει την τιμή a=0.2.

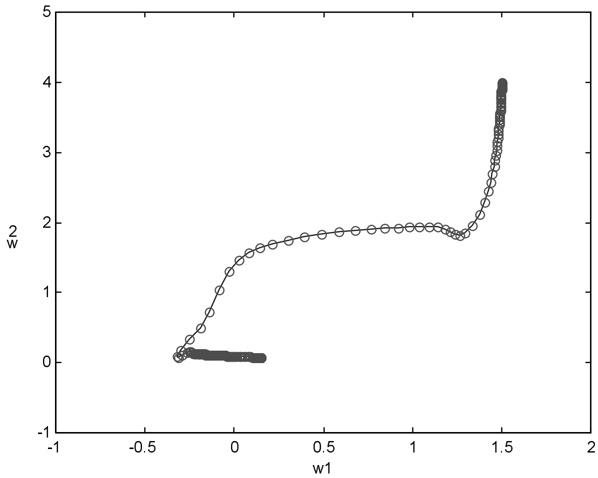


Ας δούμε την συμπεριφορά του αλγόριθμου όταν προσθέσουμε τον "αδρανειακό" όρο ο οποίος τείνει να διατηρήσει την μεταβολή της βέλτιστης λύσης συγγραμμική με την εκτίμηση της μεταβολής κατά το προηγούμενο βήμα του αλγόριθμου, δηλαδή:

$$w^{(t)} = w^{(t-1)} - \alpha \nabla \Sigma \phi(w^{(t-1)}) + m \Delta w^{(t-1)}$$

Διατηρώντας ίδιο το αρχικό σημείο εκτίμησης του ελάχιστου και θέτοντας a=0.2 και m=0.5 έχω την ακόλουθη πορεία εκτίμησης του ελάχιστου:

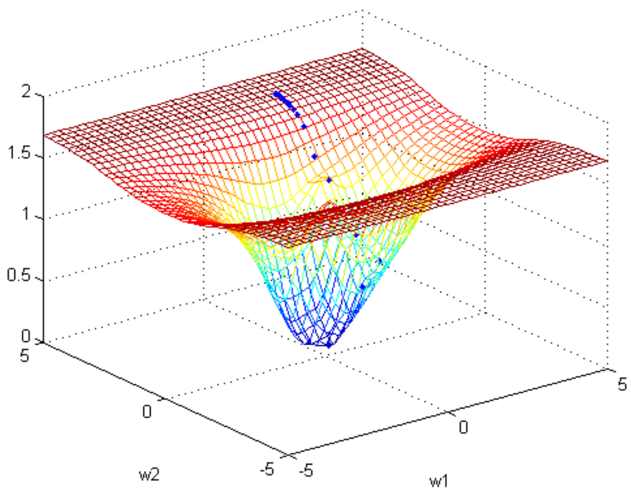


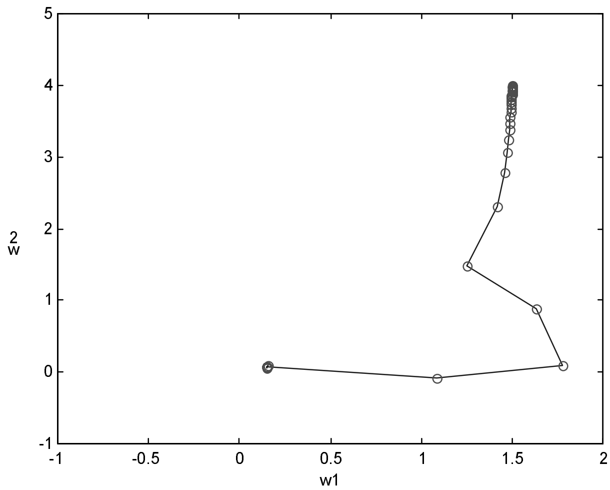


Επιλέγοντας σαν τελευταία παραλαγή του αλγόριθμου την προσαρμοζόμενη μεταβολή του συντελεστή εκπαίδευσης και διατηρώντας αναλοίωτα τα υπόλοιπα χαρακτηριστικά του αλγόριθμου έχω τις ακόλουθες σχέσεις επαναπροσδιορισμού της λύσης του τοπικά ελάχιστου

$$a_i^{(t)} = \begin{cases} a_i^{(t\ 1)} * a_o, & \dfrac{\Sigma\phi(w^{(t\ 1)})}{w_i}\dfrac{\Sigma\phi(w^{(t\ 2)})}{w_i} \geq 0 \\ a_i^{(t)}, & \dfrac{\Sigma\phi(w^{(t\ 1)})}{w_i}\dfrac{\Sigma\phi(w^{(t\ 2)})}{w_i} < 0 \end{cases}$$

$$w_i^{(t)} = w_i^{(t\ 1)} - a_i^{(t)} \frac{\Sigma\phi(w^{(t\ 1)})}{w_i}$$





Παράδειγμα

Έστω ότι η συνάρτηση ταξινόμησης δίνεται από την σχέση:

$$R(w,x) = \left(\sqrt{|w^T x + 1|}, (w-x)^T (w-x)\right)$$

Θέλουμε να κατασκευάσουμε σύστημα ταξινόμησης προτύπων δύο κατηγοριών διαθέτοντας τα εξής παραδείγματα:

$$\Omega_1 = \{(-1.5,0.5),(-1,1.5),(1,1),(2,1),(1,2)\}$$

$$\Omega_2 = \{ (-0.75,0.75),(-0.5,1.2),(0.5,0.5),(2.5,0.5),(2.5,2)\}$$

για τις δύο κατηγορίες αντίστοιχα.

Υπολογίστε τρεις διαδοχικές εκτιμήσεις της μνήμης του συστήματος ταξινόμησης με τον αλγόριθμο εκπαίδευσης κατά την διεύθυνση της αρνητικής παραγώγου.

Η μέθοδος Newton

Με την μέθοδο Newton επιχειρούμε να ελαττώσουμε τον αριθμό των επαναληπτικών εκτιμήσεων του ελάχιστου της συνάρτησης σφάλματος, υποθέτοντας ότι μπορούμε να προσεγγίσουμε την συνάρτηση του σφάλματος με ένα πολυδιάστατο πολυώνυμο δεύτερου βαθμού.

Συγκεκριμένα γνωρίζουμε ότι κάθε παραγωγίσιμη συνάρτηση μπορεί να αναλυθεί σαν άθροισμα πολυωνυμικών όρων Taylor. Αν προσεγγίσουμε την συνάρτηση του

$$\Sigma\phi(w) \cong \Sigma\phi(w_o) + (w - w_o)\nabla\Sigma\phi(w)\,|_{w=w_o} + (w - w_o)^T (w - w_o)\nabla^2\Sigma\phi(w)\,|_{w=w_o}$$

σφάλματος με τους δύο πρώτους όρους της σειράς στο σημείο $\mathbf{w}_o$ έχω:

Η δεύτερη παράγωγος της συνάρτησης ως προς τις ελεύθερες παραμέτρους της ονομάζεται πίνακας Hessian και δίνεται από την σχέση:

$$\nabla^2 \Sigma\phi(w) = \begin{bmatrix} \frac{\partial^2 \Sigma\phi(w)}{\partial w_1^2} & \frac{\partial^2 \Sigma\phi(w)}{\partial w_1 w_2} & \cdots & \frac{\partial^2 \Sigma\phi(w)}{\partial w_1 w_M} \\ \frac{\partial^2 \Sigma\phi(w)}{\partial w_2 w_1} & \frac{\partial^2 \Sigma\phi(w)}{\partial w_2^2} & \cdots & \frac{\partial^2 \Sigma\phi(w)}{\partial w_2 w_M} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial^2 \Sigma\phi(w)}{\partial w_M w_1} & \frac{\partial^2 \Sigma\phi(w)}{\partial w_M w_2} & \cdots & \frac{\partial^2 \Sigma\phi(w)}{\partial w_M^2} \end{bmatrix}$$

Στην θέση του ελάχιστου η πρώτη παράγωγος του σφάλματος πρέπει να δίνει μηδενική τιμή (∇ Σφ(**w**) $|_{w=wo}$=0), συνεπώς πρέπει να ισχύει:

$$\nabla\Sigma\phi(w)\,|_{w=w_o} \cong (w_n - w_o)\nabla^2\Sigma\phi(w)\,|_{w=w_o} \Rightarrow w_n \cong w_o - \frac{\nabla\Sigma\phi(w)\,|_{w=w_o}}{\nabla^2\Sigma\phi(w)\,|_{w=w_o}}$$

Αν θεωρήσουμε ότι $\mathbf{w}_o$ είναι η εκτίμηση του ελάχιστου της συνάρτησης σφάλματος την χρονική στιγμή t-1 η νεώτερη εκτίμηση του ελάχιστου (την χρονική στιγμή t) θα είναι η τιμή της μνήμης $\mathbf{w}_n$:

$$w^{(t)} \cong w^{(t-1)} - \frac{\nabla\Sigma\phi(w)\,|_{w=w^{(t-1)}}}{\nabla^2\Sigma\phi(w)\,|_{w=w^{(t-1)}}} = w^{(t-1)} - \left(\nabla^2\Sigma\phi(w)\,|_{w=w^{(t-1)}}\right)^{-1}\nabla\Sigma\phi(w)\,|_{w=w^{(t-1)}}$$

Η βασική δομή του αλγόριθμου εκτίμησης της μνήμης του συστήματος ταξινόμησης παραμένει η ίδια με τον αλγόριθμο εκπαίδευσης κατά την διεύθυνση της αρνητικής παραγώγου. Τα πλεονεκτήματα της μεθόδου Newton συνίσταται στο μεταβλητό βήμα της επανεκτίμησης της μνήμης της συνάρτησης ταξινόμησης η οποία έχει σαν επακόλουθο την δραστική μείωση των επαναλήψεων του αλγόριθμου. Κύριο μειονέκτημα είναι το υπολογιστικό κόστος του πίνακα Hessian (∇^2 Σφ(**w**)$|_{w = w(t)}$) και η αντιστροφή του. Επειδή οι διαστάσεις του πίνακα είναι ίδιες με τον αριθμό των παραμέτρων ο πίνακας αυξάνει σημαντικά σε προβλήματα κατά τα οποία το μέγεθος της μνήμης του συστήματος ταξινόμησης είναι μεγάλο. Γιαυτό και η μέθοδος αυτή χρησιμοποιείται σπάνια σε προβλήματα εκπαίδευσης νευρωνικών

δικτύων.

Οι κύριες αλγοριθμικές διαφοροποιήσεις που υπάρχουν σε σχέση με την μέθοδο αναζήτησης του ελάχιστου κατά την διεύθυνση της αρνητικής παραγώγου είναι οι ακόλουθες:

1. Η συνάρτηση του σφάλματος πρέπει να είναι συνεχής και παραγωγίσιμη μέχρι τουλάχιστον και την δεύτερη παράγωγο ως προς όλες τις μεταβλητές της μνήμης.

2. Πρέπει να γνωρίζουμε την αναλυτική εξίσωση των παραγώγων πρώτης και δεύτερης τάξης.

3. Η ορίζουσα του πίνακα Hessian πρέπει να μην έχει μηδενική τιμή για να μπορεί να αντιστραφεί. Αυτή η συνθήκη δεν μπορεί να ικανοποιείται πάντα στα διαδοχικά βήματα του αλγόριθμου, γιαυτό καταφεύγουμε συνήθως στην ακόλουθη απλή λύση. Αν η απόλυτη αριθμητική τιμή της ορίζουσας είναι μικρότερη από κάποιο όριο, τότε αυτή τίθεται (κατά απόλυτη αριθμητική τιμή) στην τιμή του κατωφλίου.

Κύριο μειονέκτημα της μεθόδου Newton είναι ο περιορισμός που τίθεται για την αρχική εκτίμηση της μνήμης του συστήματος ταξινόμησης. Συγκεκριμένα η μέθοδος Newton συγκλίνει ικανοποιητικά μόνο όταν η αρχική εκτίμηση της μνήμης του συστήματος ταξινόμησης βρίσκεται στην γειτονιά ενός τοπικά ακρότατου, συνθήκη η οποία δεν μπορεί να ικανοποιηθεί στην πλειονότητα των πρακτικών εφαρμογών.

Το πρόβλημα με την μέθοδο Newton εμφανίζεται διότι υποθέτουμε ότι το σφάλμα είναι κατά προσέγγιση μία πολυωνυμική συνάρτηση δεύτερης τάξης ως προς την μνήμη της συνάρτησης ταξινόμησης. Η ύπαρξη ελάχιστου απαιτεί όλες οι μερικές παράγωγοι δεύτερης τάξης να έχουν θετική τιμή σε όλο το πεδίο ορισμού. Η συνθήκη αυτή δεν μπορεί φυσικά να ικανοποιηθεί για οποιαδήποτε συνάρτηση σφάλματος. Έτσι όταν η δεύτερη παράγωγος παρουσιάσει αρνητική τιμή τότε αυτόματα ο αλγόριθμος αναζητά το μέγιστο της συνάρτησης ταξινόμησης.

Γιαυτό λοιπόν παρουσιάζουμε την πλέον αποτελεσματική παραλλαγή της μεθόδου Newton η οποία μπορεί να θεωρηθεί ότι είναι ταυτόχρονα και μία επέκταση της μεθόδου αναζήτησης του ελάχιστου κατά την αρνητική παράγωγο όταν ο συντελεστής εκπαίδευσης οριστεί σαν συνάρτηση της δεύτερης παραγώγου του σφάλματος.

Αν υποθέσουμε ότι όλοι οι μη διαγώνιοι όροι της συνάρτησης του σφάλματος είναι μηδενικοί και η συνάρτηση του σφάλματος είναι κυρτή σε όλο το πεδίο ορισμού της, (η δεύτερη παράγωγος είναι πάντα θετική), τότε η συνάρτηση επαναπροσδιορισμού της μνήμης της συνάρτησης ταξινόμησης για κάθε συνιστώσα της ξεχωριστά

$$w^{(t)} = w^{(t-1)} - \frac{1}{\left| \frac{\partial^2}{\partial w_i^2} \Sigma\phi(w^{(t-1)}) \right| + S} \frac{\partial}{\partial w_i} \Sigma\phi(w^{(t-1)})$$

απλοποιείται σε:

όπου S $\in \Re^+$ - {0}. Ο θετικός όρος στον παρονομαστή τίθεται έτσι ώστε η μεταβολή της παραμέτρου w_i να μην λαμβάνει πολύ μεγάλη αριθμητική τιμή σε κανένα βήμα του αλγόριθμου.

Η απλοποιημένη και υπολογιστικά ισχυρή μέθοδος εύρεσης του ελάχιστου της συνάρτησης σφάλματος δίνεται από τον ακόλουθο αλγόριθμο:

Αλγόριθμος

- **Τοποθέτηση αρχικών τιμών**. Τοποθετούμε τυχαίες τιμές στις παραμέτρους $\mathbf{w}^0$ και
- t $\rightarrow$ 0.

- **Νέα εκτίμηση της μνήμης της συνάρτησης ταξινόμησης**. t $\rightarrow$ t+1. Ο επαναπροσδιορισμός της μνήμης του συστήματος ταξινόμησης για κάθε παράμετρο w_i δίνεται από την εξίσωση:

$$w^{(t)} = w^{(t-1)} - \frac{1}{\left|\frac{\partial^2}{\partial w_i^2}\Sigma\phi(w^{(t-1)})\right| + S}\frac{\partial}{\partial w_i}\Sigma\phi(w^{(t-1)})$$

- **Έλεγχος σύγκλισης**. Αν η αθροιστική απόλυτη μεταβολή των παραμέτρων w_i είναι μικρότερη από ένα κατώφλι ο αλγόριθμος τερματίζει.

- Η τελευταία τιμή της μνήμης της συνάρτησης ταξινόμησης **w** ορίζεται να είναι η βέλτιστη λύση του αλγόριθμου.

Παράδειγμα. Υπολογίστε τρεις διαδοχικές εκτιμήσεις της μνήμης του συστήματος ταξινόμησης με τον αλγόριθμο εκπαίδευσης Newton για το σύστημα ταξινόμησης και τα παραδείγματα εκπαίδευσης του προηγούμενου παραδείγματος

Παράδειγμα. Τα ακρόταια της συνάρτησης

$$f(x, y) = \log((x - y^2)^2 + (1 - y)^2 + 1)$$

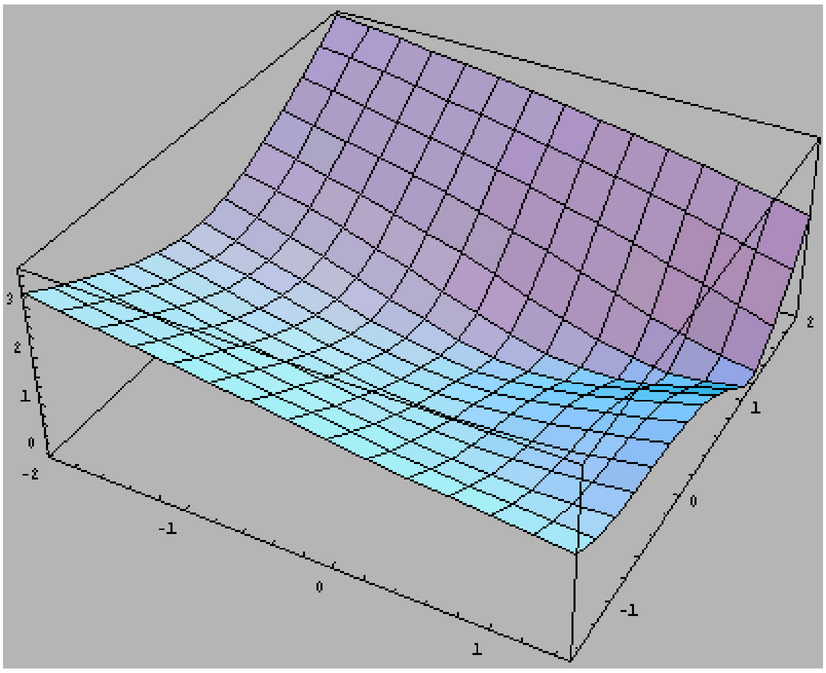
υπολογίζονται εύκολα από την επίλυση των δύο εξισώσεων που προκύπτουν από τις μερικές παραγώγους της συνάρτησης. Τα σημεία των ελαχίστων της συνάρτησης είναι τρία και βρίσκονται στις θέσεις (0.5,0), (1,-1), (1,1).

Εφαρμόστε την μέθοδο μετακίνησης κατά την διεύθυνση της αρνητικής παραγώγου και την μέθοδο Newton ξεκινώντας από το σημείο (1,-1.5).

Βρείτε το πραγματικό ελάχιστο της f(x,y) και μελετήστε την συμπεριφορά της συνάρτησης των δύο αλγορίθμων εύρεσης του ελάχιστου της συνάρτησης.

Λύση:

Η γραφική παράσταση της συνάρτησης f(x,y) στο διάστημα ([-2,1.5],[-1.5,2]) δίνεται από το ακόλουθο σχήμα.

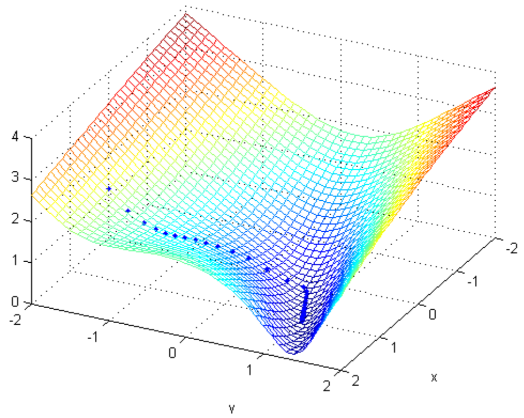


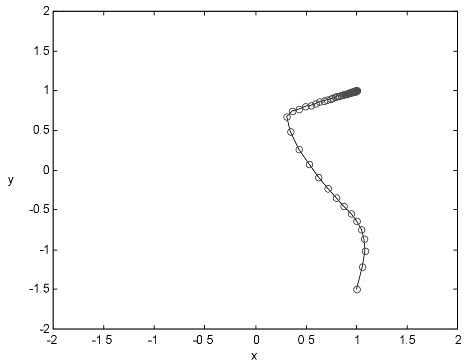
Σχήμα 2. Γραφική απεικόνιση συνάρτησης δύο μεταβλητών που έχει ελάχιστο στο σημείο (1,1).

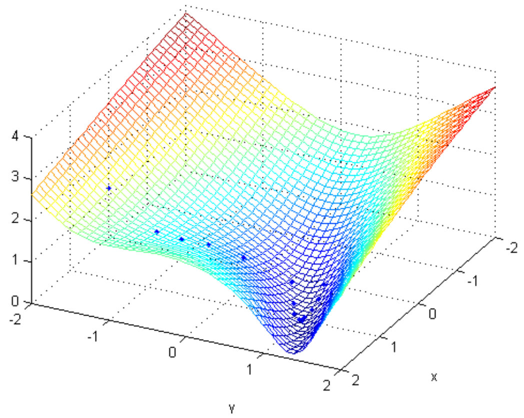
Τα ακρόταια της ***f****(x,y) υπολογίζονται από την επίλυση των ακόλουθων μη-γραμμικών εξισώσεων οι οποίες για την συγκεκριμένη συνάρτηση έχουν αναλυτική λύση:*

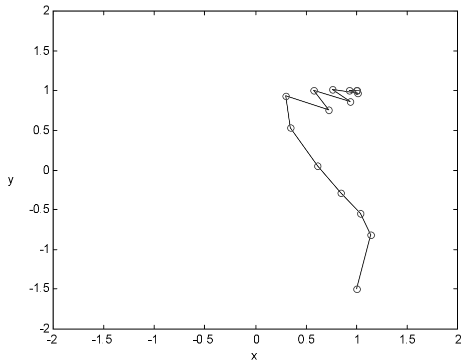
$$\begin{pmatrix} \dfrac{\partial}{\partial x} f(x,y) = 0 \\ \dfrac{\partial}{\partial y} f(x,y) = 0 \end{pmatrix} \Leftrightarrow \begin{pmatrix} \dfrac{2(x-y^2)}{(x-y^2)^2+(1-y)^2+1} = 0 \\ \dfrac{-4y(x-y^2)-2(1-y)}{(x-y^2)^2+(1-y)^2+1} = 0 \end{pmatrix} \Rightarrow \begin{pmatrix} x=1 \\ y=1 \end{pmatrix}$$

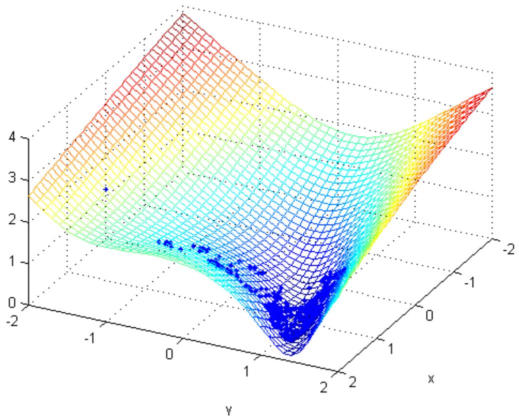
Οι ακόλουθες γραφικές παραστάσεις μας δίνουν τα βήματα του αλγόριθμου μετακίνησης κατά την διεύθυνση της αρνητικής παραγώγου για τιμές του συντελεστή εκπαίδευσης 0.2, 0.5 και 1.0. Στα διαγράμματα που βρίσκονται στα δεξιά των τρισδιάστατων απεικονίσεων δίνονται τα διαδοχικά βήματα προσδιορισμού των σημείων του ελάχιστου.

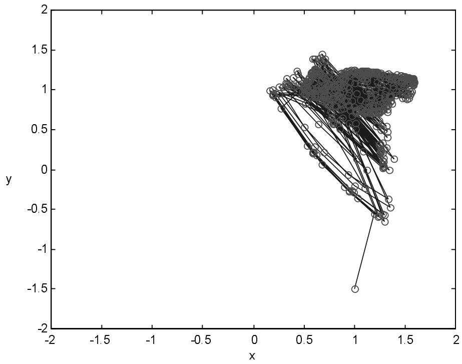












Αν εφαρμόσουμε την μέθοδο Newton χρειαζόμαστε και τον υπολογισμό των δεύτερων παραγώγων της συνάρτησης f(x,y) ως προς τις μεταβλητές της. Οι συναρτήσεις των δεύτερων παραγώγων δίονται από τις ακόλουθες σχέσεις

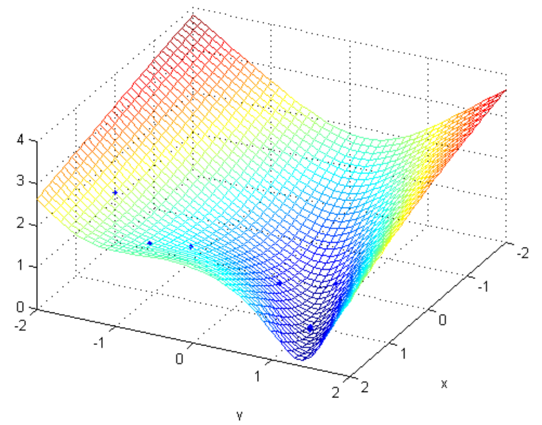
$$\frac{\partial^2}{\partial x^2} f(x,y) = \frac{\partial}{\partial x} \frac{2(x-y^2)}{(x-y^2)^2+(1-y)^2+1} =$$

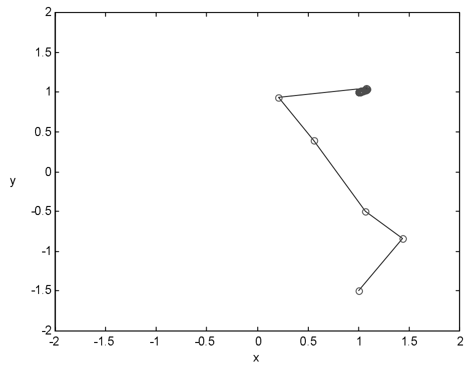
$$\frac{\partial^2}{\partial y^2} f(x,y) = \frac{\partial}{\partial y} \frac{-4y(x-y^2)-2(1-y)}{(x-y^2)^2+(1-y)^2+1} =$$

$$\frac{\partial^2}{\partial x^2} f(x,y) = \frac{2}{(x-y^2)^2+(1-y)^2+1} - \frac{4(x-y^2)^2}{((x-y^2)^2+(1-y)^2+1)^2}$$

$$\frac{\partial^2}{\partial y^2} f(x,y) = \frac{2+8y^2-4(x-y^2)}{(x-y^2)^2+(1-y)^2+1} - \frac{(-4y(x-y^2)-2(1-y))^2}{((x-y^2)^2+(1-y)^2+1)^2}$$

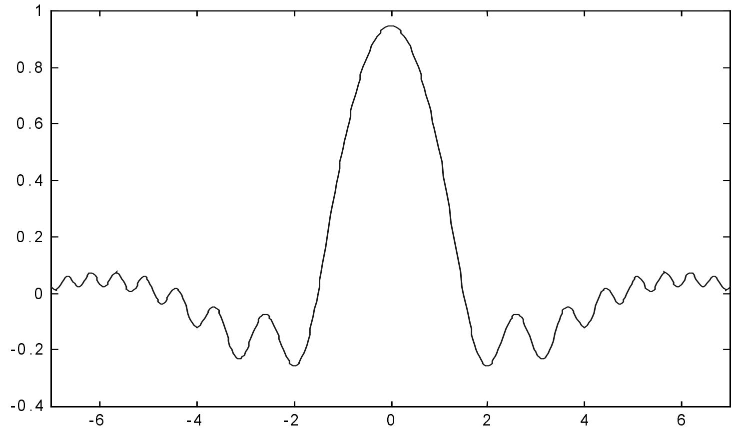
Από τα δύο σχήματα παρατηρούμε ότι, θέτοντας τον σταθερό όρο S=0.5, η σύγκλιση στην γειτονιά του ελάχιστου πραγματοποιείται σε έξι διαδοχικά βήματα του αλγόριθμου Newton.





Αλγόριθμος βραδείας ψύξης (simulated annealing)

Οι αλγόριθμοι εύρεσης του ελάχιστου για την συνάρτηση σφάλματος που σχηματίζεται κατά την διαδικασία εκπαίδευσης συστημάτων ταξινόμησης προτύπων, χρησιμοποιούν πληροφορίες μερικών παραγώγων. Στις περιπτώσεις κατά τις οποίες η συνάρτηση σφάλματος παρουσιάζει πολλά τοπικά ακρότατα ή έχει πολλά σημεία ασυνέχειας, τότε πρέπει να χρησιμοποιήσουμε στοχαστικές μεθόδους, οι σημαντικότερες των οποίων θα περιγραφούν στις επόμενες παραγράφους. Οι αλγόριθμοι αυτοί είναι στοχαστικοί με την έννοια ότι στα διαδοχικά βήματα προσέγγισης της βέλτιστης λύσης δεν επιλέγουν πάντα αυτή που δίνει το μικρότερο σφάλμα. Με αυτό τον τρόπο προσπαθούν να ξεφύγουν από περιπτώσεις σύγκλισης σε τοπικά ελάχιστα.



Εικόνα **Error! Unknown switch argument.** Γραφική παράσταση της συνάρτησης $1/(x^2+3)*(\sin(x^2+1)+2*\cos(x))$ στο διάστημα [-7,7].

Η ύπαρξη ασυνεχειών αλλά και πολλών τοπικών ακρότατων εμποδίζει τις κλασσικές επαναληπτικές μεθόδους εύρεσης ακρότατων που βασίζονται σε πληροφορίες των παραγώγων της συνάρτησης. Παρατηρώντας την γραφική παράσταση της συνάρτησης $f(x)=1/(x^2+3)*(sin(x^2+1)+2*cos(x))$ στο διάστημα [-7,7] βλέπουμε ότι εμφανίζονται πολλά τοπικά ακρότατα. Οι μέθοδοι που βασίζονται σε πληροφορία μερικών παραγώγων θα συγκλίνουν σε κάποιο από τα τοπικά ελάχιστα. Το σημείο σύγκλισης εξαρτάται από τον αλγόριθμο βελτιστοποίησης, την αρχική εκτίμηση του ελάχιστου και την επιλογή των σταθερών παραμέτρων του αλγόριθμου.

Δυστυχώς δεν υπάρχει κάποια τεχνική με την οποία μπορούμε να αποφύγουμε το πρόβλημα σύγκλισης των αλγορίθμων σε τοπικό ελάχιστο. Συνήθης τεχνική που χρησιμοποιείται για να αποφύγουμε το πρόβλημα ύπαρξης τοπικών ελάχιστων είναι να επαναλαμβάνουμε τον αλγόριθμο εύρεσης των ελάχιστων με διαφορετική αρχική εκτίμηση της βέλτιστης λύσης και τοποθέτηση διαφορετικών τιμών στις σταθερές παραμέτρους του. Βέβαια η συγκεκριμένη τροποποίηση της μεθόδου αποτελεί μία απλοποιημένη στοχαστική μέθοδο εύρεσης του ελάχιστου της συνάσρτησης του σφάλματος.

Η πρώτη στοχαστική μέθοδος που θα εξετάσουμε είναι ο αλγόριθμος βραδείας ψύξης (simulated annealing). Ονομάζεται έτσι διότι η βασική ιδέα του αλγόριθμου βασίζεται στην διαδικασία κατασκευής ανθεκτικών μεταλλικών κραμμάτων. Εχει διαπιστωθεί ότι αν ένα μέταλλο ψυχθεί απότομα τότε η μετάβαση από την υγρή, όπου τα μόρια μετακινούνται σχεδόν ελεύθερα, στην στέρεη κατάσταση δεν επιτρέπει την διάταξη των μορίων σε τέτοια τοπολογία η οποία να παρουσιάζει την μέγιστη ανθεκτικότητα. Αντίθετα η μέγιστη ανθεκτικότητα επιτυχάνεται όταν το μέταλλο ψυχθεί αργά με αποτέλεσμα η μετάβαση από την υγρή στην στέρεη κατάσταση να γίνει σε μεγάλο χρονικό διάστημα.

Προσομοιώνοντας την διαδικασία ψύξης μετάλλων με την προσπάθεια εύρεσης του ελάχιστου της συνάρτησης του σφάλματος αποφεύγουμε τα τοπικά ελάχιστα αποδεχόμενοι λύσεις οι οποίες δίνουν μεγαλύτερο σφάλμα με κάποια πιθανότητα η οποία είναι συνάρτηση του αριθμού των επαναλήψεων του αλγόριθμου ή όπως συνιθίζεται να λέξεται της "θερμοκρασίας" της λύσης συσχετίζοντας την λύση του τοπικά ελάχιστου με την θερμοκρασία του μετάλλου.

Η απλοποιημένη μέθοδος εύρεσης του ελάχιστου της συνάρτησης σφάλματος με τον αλγόριθμο βραδείας ψύξης είναι ο ακόλουθος:

Αλγόριθμος

- **Τοποθέτηση αρχικών τιμών**. Τοποθετούμε τυχαίες τιμές στις παραμέτρους $\mathbf{w}^{0}$ και $t \rightarrow 0$.

- **Τυχαία επιλογή νέας μνήμης της συνάρτησης ταξινόμησης στην γειτονιά της τρέχουσας βέλτιστης λύσης w** . $t \rightarrow t+1$. Στην βιβλιογραφία υπάρχουν πολλές τεχνικές τυχαίας επιλογής μίας νέας λύσης στην γειτονιά της τρέχουσας βέλτιστης λύσης. Εδώ θα αναφέρουμε την πλέον διαδεδομένη, την λύση τυχαίου βηματισμού, ενώ σε μερικά παραδείγματα θα βρείτε μερικές από τις εναλλακτικές τεχνικές.

  Η αριθμητική τιμή της νέας λύσης γίνεται με την βοήθεια της σχέσης:

  $$\boldsymbol{w} = \boldsymbol{w}^{(t-1)} + \boldsymbol{e}(t)$$

  e(t) είναι διάνυσμα το οποίο έχει τυχαία κανονική κατανομή με μέση τιμή 0 και διασπορά **αντιστρόφως ανάλογη** του χρόνου t. Οι ιδιότητες αυτές προσδίδουν στο διάνυσμα ***e(t)*** μερικά μοναδικά χαρακτηριστικά. Για μικρές τιμές του χρόνου η νέα λύση επεκτείνεται και σε απομεμακρυσμένα σημεία του πεδίου τιμών των συντελεστών βαρύτητας. Η εκτίμηση αυτή μοιάζει με την τυχαία αναζήτηση. Όταν όμως τα βήματα του αλγόριθμου αυξάνουν το μέτρο του διανύσματος ***e(t)*** στατιστικά μειώνεται οπότε και οι νέες λύσεις αναζητώνται στην γειτονιά της τρέχουσας βέλτιστης λύσης.

- **Αποδοχή της νέας λύσης**. Αν

$$\Sigma\varphi(\boldsymbol{w}) < \Sigma\varphi(\boldsymbol{w}^{(t-1)})$$

η νέα λύση είναι αποδεκτή

$$\boldsymbol{w}^{(t)} = \boldsymbol{w}$$

διαφορετικά, αν

$$U([0,1)) < e^{(\Sigma\phi(w^{(t-1)}-\Sigma\phi(w)))/T(To,t)}$$

πάλι η νέα λύση είναι αποδεκτή

$$\boldsymbol{w}^{(t)} = \boldsymbol{w}.$$

Η συνάρτηση "θερμοκρασίας" T(To,t) είναι μία γνησίως φθίνουσα συνάρτηση του t με πεδίο τιμών το [To,0) και To $\in \Re^+$ και U(x) είναι ένας τυχαίος πραγματικός αριθμός στο διάστημα x. Το φαινομενικά παράδοξο γεγονός να αποδεχόμαστε στοχαστικά χειρότερες λύσεις οδηγεί το αλγόριθμο στην πράξη να ξεπερνά το πρόβλημα των τοπικά ελάχιστων. Η εκθετική σχέση αποδοχής λύσης με μεγαλύτερο σφάλμα εξασφαλίζει ότι λύσεις με μεγάλο σφάλμα θα γίνουν αποδεκτές με πολύ μικρότερη πιθανότητα από εκείνες τις λύσεις πού έχουν μικρή διαφορά από το σφάλμα της τρέχουσας λύσης.

- **Έλεγχος σύγκλισης**. Όταν για μεγάλο αριθμό βημάτων του αλγόριθμου δεν γίνει αποδεκτή νέα λύση ο αλγόριθμος τερματίζει διαφορετικά το επαναληπτικό τμήμα του αλγόριθμου (τα δύο προηγούμενα βήματα) επαναλαμβάνεται.

Παράδειγμα. Έστω ότι η συνάρτηση ταξινόμησης δίνεται από την σχέση:

$$R(w,x) = \left(\log((w_1{}^T x)^2 + 1)^2, (w_2 - x)^T (w_2 - x)\right)^T$$

Θέλουμε να κατασκευάσουμε σύστημα ταξινόμησης προτύπων δύο κατηγοριών διαθέτοντας τα εξής παραδείγματα:

$$\Omega_1 = \{ (1,1),(3,5),(-1,3),(-3,1)\}$$

$$\Omega_2 = \{ (-1,-1),(-2,1),(1,-1),(2,0),(0,0)\}$$

για τις δύο κατηγορίες αντίστοιχα.

Περιγράψτε την διαδικασία ορισμού των παραμέτρων και των βοηθητικών συναρτήσεων που χρειάζεστε για την επιτυχή εφαρμογή του αλγόριθμου βραδείας ψύξης ο οποίος θα υπολογίσει την αριθμητική τιμή των συντελεστών εκπαίδευσης της συνάρτησης ταξινόμησης.

Ποιο είναι το ελάχιστο σφάλμα του συστήματος ταξινόμησης προτύπων;

Παράδειγμα. ***Με την βοήθεια αλγόριθμου βραδείας ψύξης υπολόγισε το ελάχιστο της συνάρτησης***

$$f(x)=1/(x^2+3)*(\sin(x^2+1)+2*\cos(x)).$$

Λύση:

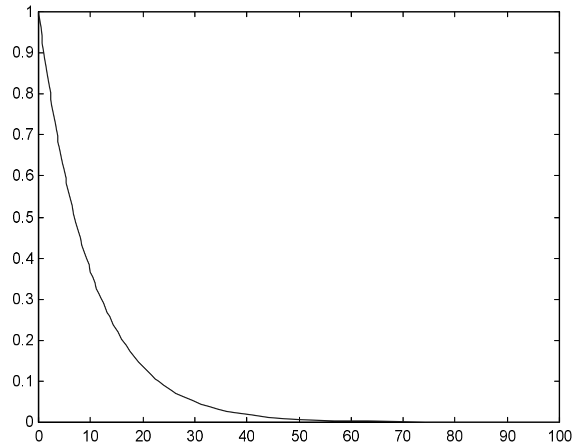
Αρχικά πρέπει να ορίσω τις βοηθητικές συναρτήσεις και τις σταθερές παραμέτρους του αλγόριθμου βραδείας ψύξης.

Θέτω την συνάρτηση πυκνότητας πιθανότητας για το διάνυσμα e(t) την κανονική κατανομή με μέση τιμή 0 και διασπορά 4. Η συνάρτηση πυκνότητας πιθανότητας της μεταβλητής e(t) δίνεται από την ακόλουθη σχέση

$$p(e(t))=N(0,4/t)$$

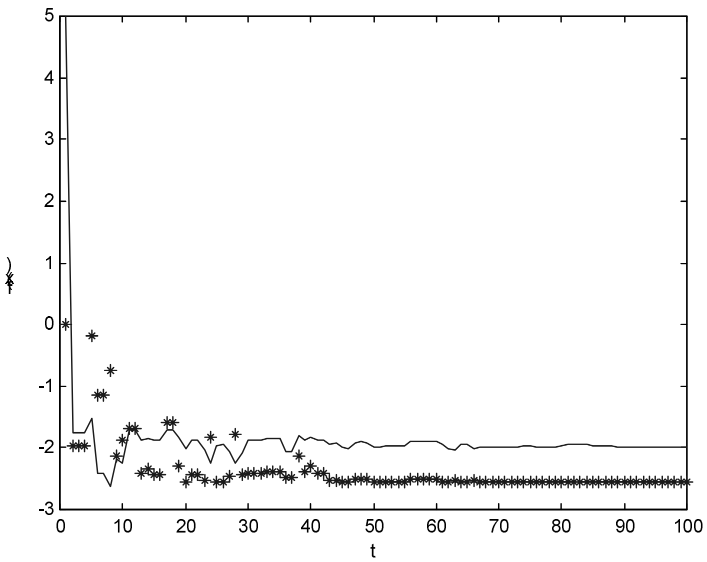
Η συνάρτηση θερμοκρασίας T(To,t) είναι μία γνησίως φθίνουσα συνάρτηση του t με πεδίο τιμών το [1,0) και To = 0.1. Ορίζω σαν συνάρτηση θερμοκρασίας την

$$T(To,t)=e^{-t*To}$$



Το διάγραμμα θερμοκρασίας δείχνει ότι μετά από 40 περίπου επαναλήψεις του αλγόριθμου βραδείας ψύξης δεν θα γίνονται αποδεκτές λύσεις που παρουσιάζουν μεγαλύτερη αριθμητική τιμή από την τρέχουσα βέλτιστη λύση.

Στο σχήμα που ακολουθεί απεικονίζονται τα διαδοχικά βήματα επαναπροσδιορισμού του ελάχιστου της συνάρτησης f(x) για τις πρώτες 100 επαναλήψεις του αλγόριθμου ξεκινώντας με αρχική τιμή την λύση $x^{(o)}$=5.

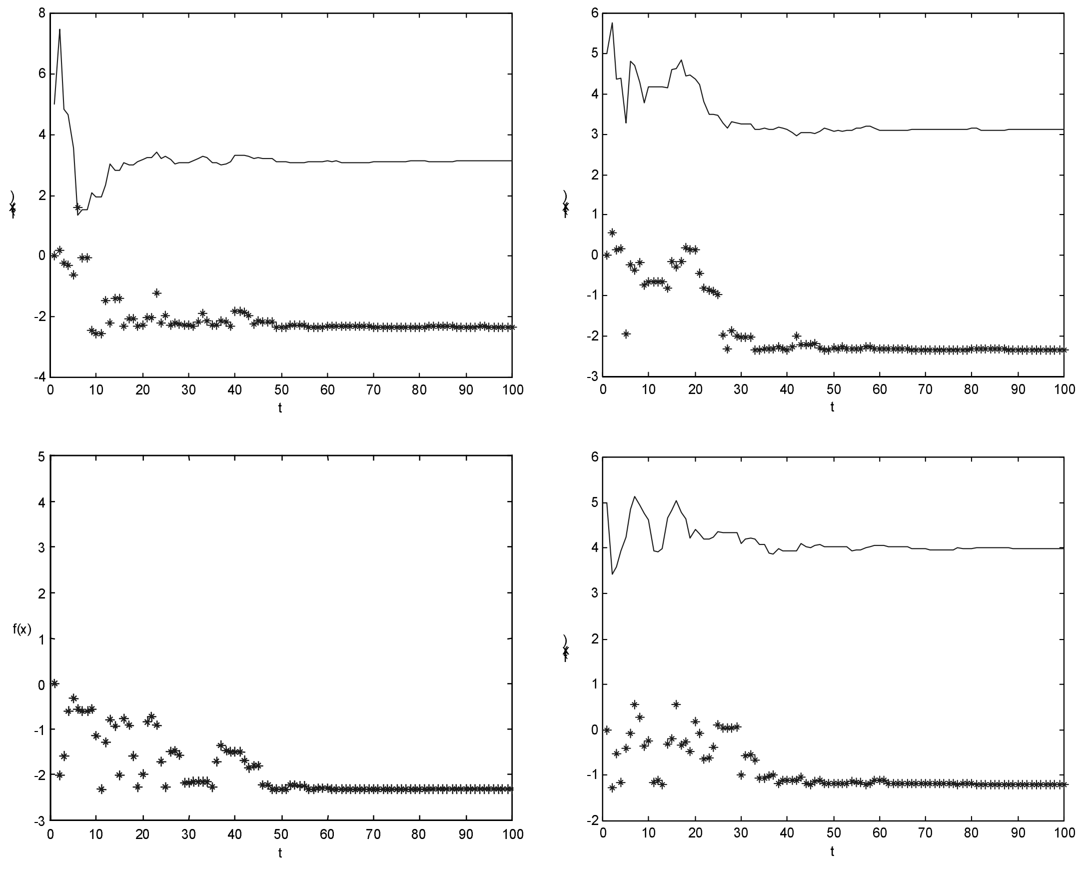


Η συνεχής γραμμή δείχνει την πορεία που ακολουθεί η τρέχουσα εκτίμηση του ελάχιστου που συγκλίνει τελικά στην τιμή x=-1.9797 πού είναι το ένα από τα δύο ελάχιστα της συνάρτησης. Παρατηρούμε στα πρώτα βήματα του αλγόριθμου μία συνεχή μετακίνηση της ρίζας προς την θέση -2. Το γεγονός αυτό είναι τυχαίο. Παρατηρούμε μάλιστα ότι ο αλγόριθμος ξεπερνά με θαυμαστή ευκολία το υπερμεγέθες (τόσο και ύψος όσο και σε πλάτος) μέγιστο στην θέση 0.

Οι αστερίσκοι της γραφικής παράστασης μας δίνουν την αριθμητική τιμή της συνάρτησης 10 * f(x) με x την αριθμητική τιμή της τρέχουσας βέλτιστης λύσης. Ο πολλαπλασιαστικός όρος προστέθηκε για εποπτικούς λόγους. Από την μορφή αυτής της καμπύλης καταλαβαίνουμε και την συμπεριφορά του αλγόριθμου. Οι αύξουσες τιμές αυτής της γραφικής παράστασης υποδηλώνουν την περίπτωση κατά την οποία γίνεται αποδεκτή μία λύση η οποία έχει αριθμητική τιμή μεγαλύτερη από την τρέχουσα βέλτιστη λύση.

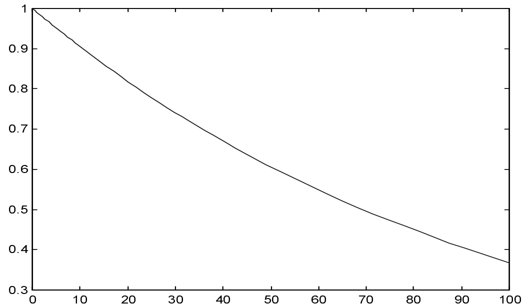
Από τα δεδομένα του σχήματος βλέπουμε ότι μία ακριβής εκτίμηση του ελάχιστου έχει πραγματοποιηθεί σε 50 βήματα του αλγόριθμου βραδείας ψύξης.

Ο αλγόριθμος βραδείας ψύξης είναι στοχαστικός με την έννοια ότι αν επαναλάβουμε τα βήματα του πιθανόν να οδηγηθούμε σε διαφορετική λύση. Στο σχήματα που ακολουθούν δίνονται οι λύσεις του αλγόριθμου, με ίδιες αρχικές συνθήκες και βοηθητικές συναρτήσεις αλλά διαφορετική συμπεριφορά των γεννητριών τυχαίων αριθμών



Παρατηρούμε ότι ο αλγόριθμος δεν συμπεριφέρεται το ίδιο καλά όπως στην πρώτη εφαρμογή του. Στις τρείς από τις τέσσερις φορές συγκλίνει σε ένα τοπικά ακρότατο κοντά στην τιμή x=3 ενώ την τέταρτη φορά συγκλίνει σε τοπικά ελάχιστο κοντά στην τιμή x=4. Σίγουρα αυτή δεν είναι μία αξιόπιστη συμπεριφορά του αλγόριθμου.

Ο αλγόριθμος φαίνεται να δούλεψε ικανοποιητικά την πρώτη φορά, αλλά στην συνέχεια συγκλίνει σε τοπικά ελάχιστα τα οποία βέβαια δίνουν και μικρή αριθμητική τιμή στο τελικό σημείο σύγκλισης , δηλαδή είναι "καλά" τοπικά βέλτιστα για τις τέσσερις από τις πέντε φορές που εκτελέστηκε ο αλγόριθμος. Συγκεκριμένα μπορούμε να πούμε ότι ο αλγόριθμος βραδείας ψύξης σε πέντε επαναλήψεις του έδωσε μία εκτίμηση του ενός από τα δύο ελάχιστα της συνάρτησης, τρείς φορές σύγκλινε στο αμέσως μεγαλύτερο τοπικά ελάχιστο και μία φορά σύγκλινε σε τοπικά ελάχιστο το οποίο έχει τιμή ελαχίστου αρκετά μακρυά από το συνολικό

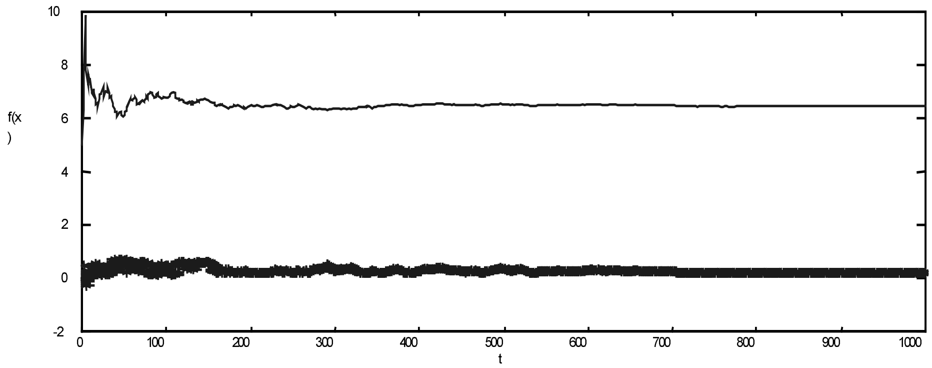


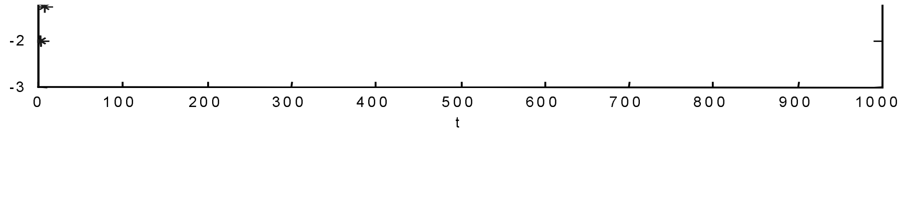
βέλτιστο.

Θα δούμε ότι αν τροποποιήσουμε την συνάρτηση "πτώσης" της θερμοκρασίας, δηλαδή αν η πτώση της θερμοκρασίας γίνει περισσότερο ομαλή τότε όπως και στα μέταλλα παίρνουμε περισσότερο ανθεκτικά κράμματα, έτσι και στην αναζήτηση του ελάχιστου αυξάνονται οι πιθανότητες να βρούμε το πραγματικό ελάχιστο της συνάρτησης f(x). Τροποποιώ λοιπόν τον σταθερό όρο της συνάρτησης θαρμοκρασίας και θέτω To = 0.01. Η γραφική παράσταση της πιθανότητας αποδοχής λύσης η οποία έχει μεγαλύτερη αριθμητική τιμή από τρέχουσα βέλτιστη λύση γίνεται:

Βλέποντας ότι και για την 100η επανάληψη του αλγόριθμου βραδείας ψύξης η πιθανότητα αποδοχής λύσης με μεγαλύτερη αριθμητική τιμή ξεπερνά το 0.4 εκτελούμε 1000 επαναλήψεις του αλγόριθμου και παίρνουμε τα ακόλουθα αποτελέσματα

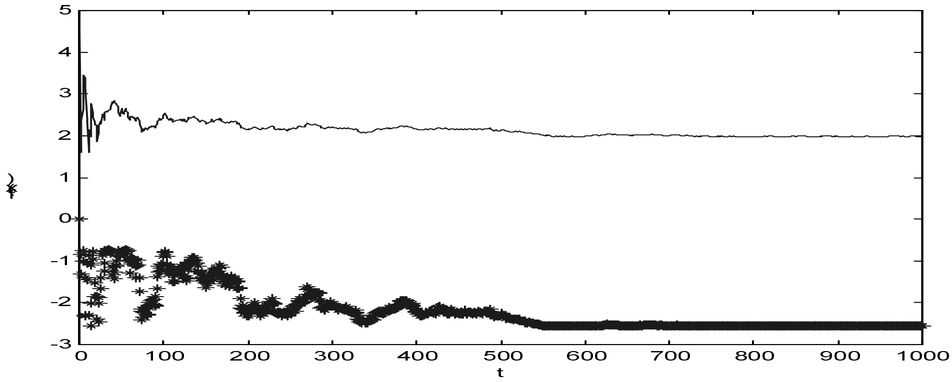
Όπως παρατηρούμε από τέσσερις επαναλήψεις του αλγόριθμου βραδείας ψύξης η σύγκλιση σε τοπικά ελάχιστα επιτυγχάνεται σε διαφορετικό τοπικό ελάχιστο. Μία φορά στις τέσσερις ο αλγόριθμος συγκλίνει στο ολικό ελάχιστο που βρίσκεται στο σημείο x=1.9797.

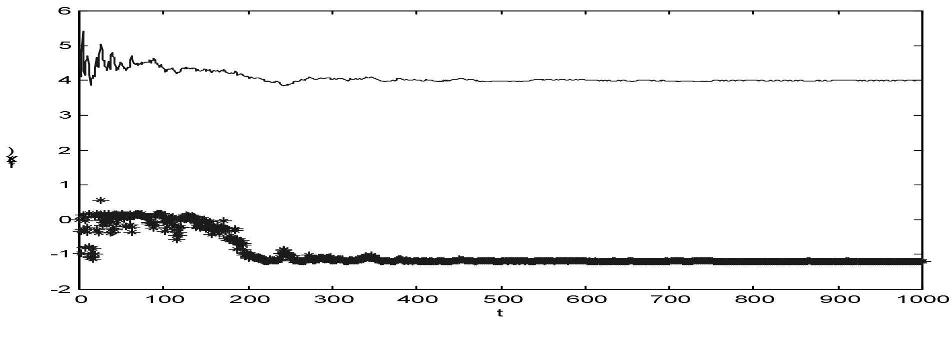




Από τα παραπάνω αποτελέσματα γίνεται φανερό ότι ο αλγόριθμος βραδείας ψύξης δεν μπορεί να βρεί με επιτυχία το ολικό ελάχιστο της συνάρτησης f(x). Σε αυτό το σημείο πρέπει να επισημάνουμε ότι η μορφή της συνάρτησης που ελέγχουμε είναι από τις πλέον δύσκολες διότι παρουσιάζει πολύ μεγάλο αριθμό τοπικών ελάχιστην τα οποία βρίσκονται σε πολύ κοντινές θέσεις. Αυτή η τοπολογία δυσκολεύει και τους πλέον αποτελεσματικούς αλγόριθμους εύρεσης ακρόταιων.

Στο ακόλουθο παράδειγμα θα λύσουμε ένα πρόβλημα εκπαίδευσης συστήματος ταξινόμησης προτύπων με την βοήθεια του αλγόριθμου βραδείας ψύξης.





Παράδειγμα. Έστω ότι η συνάρτηση ταξινόμησης δίνεται από την σχέση:

$$R(w,x) = \left(\frac{1}{(w^T x)^2 + 1)^2}, \log((w-x)^T (w-x)+1) \right)^T$$

Θέλουμε να κατασκευάσουμε σύστημα ταξινόμησης προτύπων δύο κατηγοριών διαθέτοντας τα εξής παραδείγματα:

$Ω_1$ = {(1,1),(3,5),(-1,3),(-3,1)}

$Ω_2$ = {(-1,-1),(-2,1),(1,-1),(2,0),(0,0)}

για τις δύο κατηγορίες αντίστοιχα.

Υπολογίστε την μνήμη της συνάρτησης ταξινόμησης με την βοήθεια του αλγόριθμου βραδείας ψύξης.

Λύση:

Για να εκπαιδεύσω το σύστημα ταξινόμησης πρέπει να υπολογίσω την αριθμητική τιμή του συντελεστή βαρύτητας w. Η τιμή αυτή θα βρεθεί με την βοήθεια της συνάρτησης σφάλματος η οποία ορίζεται να είναι η Ευκλείδεια απόσταση της συνάρτησης ταξινόμησης από τις ιδεατές τιμές που θα θέσουμε.

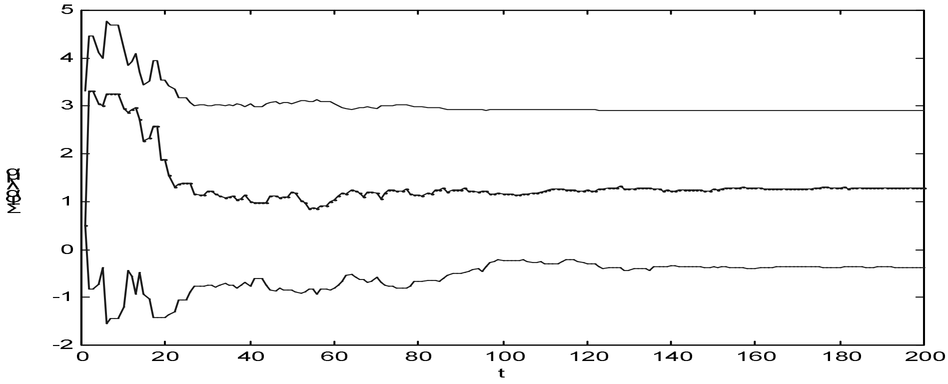
Αυθαίρετα ορίζω σαν επιθυμητές τιμές εξόδου για τα πρότυπα της πρώτης κατηγορίας το διάνυσμα $(1,0)^T$ και $(0,1)^T$ για τα πρότυπα της δεύτερης κατηγορίας. Το σύνολο των παραδειγμάτων εκπαίδευσης διαμορφώνεται ως εξής:

Είσοδος	(1,1)	(3,5)	(-1,3)	(-3,1)	
Εξοδος	(1,0)	(1,0)	(1,0)	(1,0)	
Είσοδος	(-1,-1)	(-2,1)	(1,-1)	(2,0)	(0,0)
Εξοδος	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)

Η συνάρτηση σφάλματος δίνεται από την σχέση:

$$\Sigma\phi(w) = \sum_{i=1}^{9} (d_i - R(w,x_i))^T (d_i - R(w,x_i))$$

Αν χρησιμοποιήσουμε τον αλγόριθμο βραδείας ψύξης για να βρούμε την αριθμητική τιμή του διανύσματος για την οποία το σφάλμα γίνεται ελάχιστο έχουμε την εξής ακολουθία λύσεων σαν συνάρτηση των βημάτων του αλγόριθμου:



Η υψηλότερη γραμμή δίνει την μέση τιμή του σφάλματος ανά παράδειγμα εκπαίδευσης. Παρατηρούμε ότι το σφάλμα μετά την 25 επανάληψη του αλγόριθμου σταθεροποιείται παρόλο που η αριθμητικές τιμές των δύο παραμέτρων που απαρτίζουν τους συντελεστές βαρύτητας της συνάρτησης ταξινόμησης μεταβάλλονται μέχρι περίπου την 120η επανάληψη του αλγόριθμου.

Ας σημειωθεί ότι τέθηκαν οι ακόλουθες βοηθητικές συναρτήσεις και οι εξής σταθερές παράμετροι:

Η συνάρτηση πυκνότητας πιθανότητας για το διάνυσμα e(t) έχει κανονική κατανομή με μέση τιμή 0 και διασπορά 4. Η συνάρτηση πυκνότητας πιθανότητας της μεταβλητής e(t) δίνεται από την ακόλουθη σχέση

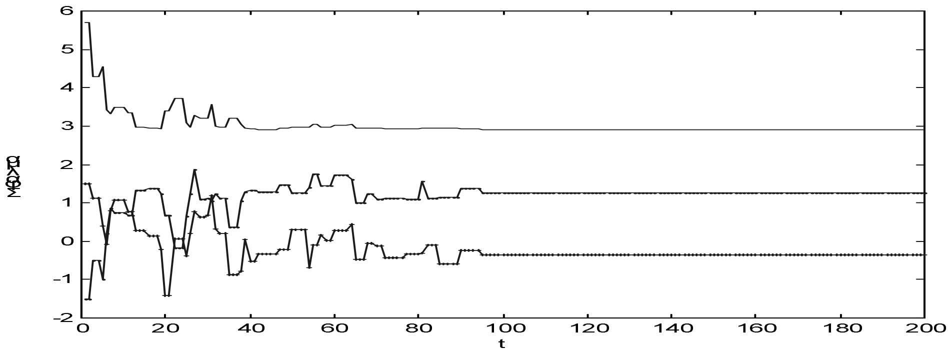
$$p(e(t))=N(0,4/t)$$

Η συνάρτηση θερμοκρασίας T(To,t) είναι μία γνησίως φθίνουσα συνάρτηση του t με πεδίο τιμών το [1,0) και To = 0.05. Ορίζω σαν συνάρτηση θερμοκρασίας την

$$T(To,t) = e^{-t*To}$$

Αν επαναλάβουμε τον αλγόριθμο με ίδιες αρχικές τιμές, παραμέτρους και ίδια μορφή βοηθητικών συναρτήσεων έχω την ακόλουθη πορεία λύσεων. Το τελικό ελάχιστο παρουσιάζει μέσο σφάλμα 2.9135 για συντελεστές βαρύτητας τους w=(-0.3590,1.254).

Το τελικό σημείο σύγκλισης είναι το ίδιο και για τις δύο επαναλήψεις του αλγόριθμου.



Τα ίδια αποτελέσματα θα παρατηρήσουμε αν επαναλάβουμε το πείραμα πολλές φορές ακόμα και αν μεταβάλλουμε τις αρχικές τιμές των συντελεστών βαρύτητας της συνάρτησης ταξινόμησης.

Γενικά μπορούμε να πούμε ότι οι αλγόριθμοι που βασίζονται σε πληροφορίες των πρώτων μερικών παραγώγων της συνάρτησης του σφάλματος είναι πολύ ευαίσθητοι στις αρχική εκτίμηση των συντελεστών βαρύτητας, σε αντίθεση με τον αλγόριθμο βραδείας ψύξης ο οποίος έχει την δυνατότητα να ξεπερνά τοπικά ελάχιστα και δεν επηρεάζεται σημαντικά από την αρχική εκτίμηση των συντελεστών βαρύτητας. Αυτοί είναι και οι σημαντικότεροι λόγοι για τους οποίους τα τελευταία χρόνια διευρύνεται η χρήση του αλγόριθμου βραδείας ψύξης σε δυσεπίλυτα προβλήματα βελτιστοποίησης.

Γενετικοί αλγόριθμοι

Οι γενετικοί αλγόριθμοι αποτελούν μία ιδιαίτερων χαρακτηριστικών μεθοδολογία βελτιστοποίησης. Στηρίζονται σε προσομοίωση των μηχανισμών της φυσικής εξέλιξης όπως παρουσιάστηκαν από τον φυσιοδίφη Darwin (Δαρβίνο). Είναι μία οικογένεια μεθόδων **προσανατολισμένης στοχαστικής αναζήτησης ακρότατων** συνδυάζοντας τα πλεονεκτήματα της τεχνικής τυχαίας αναζήτησης, (μπορούν να εφαρμοστούν σε οποιαδήποτε μορφή συνάρτησης σφάλματος ακόμα και αν αυτή είναι μη παραγωγίσιμη και ασυνεχής στο πεδίο ορισμού της) με τα πλεονεκτήματα των αλγόριθμων που χρησιμοποιούν πληροφορίες για την παράγωγο της συνάρτησης του σφάλματος, (κατά την διαδικασία επανεκτίμησης των παραμέτρων είναι σε θέση να προσεγγίζουν διαδοχικά τοπικά ακρότατα.

Η αναλυτική περιγραφή των μεθόδων βελτιστοποίησης παραμέτρων με γενετικούς αλγόριθμους είναι ένα πολύ εκτεταμένο αντικείμενο, γιαυτό τον λόγο στα πλαίσια αυτού του μαθήματος παρουσιάζεται μονάχα η βασική παραλλαγή του αλγόριθμου που χρησιμοποιείται για προβλήματα εκπαίδευσης συστημάτων αναγνώρισης προτύπων.

Οι βασικές έννοιες και μηχανισμοί που δομούν έναν γενετικό αλγόριθμο είναι οι ακόλουθοι:

Ο πληθυσμός. Σε αντίθεση με τους προηγούμενες προσεγγίσεις, όπου κάθε φορά βελτιστοποιούμε την συμπεριφορά μόνο μίας λύσης, οι γενετικοί αλγόριθμοι βασίζονται στην μαζική μελέτη ενός πεπερασμένου αριθμού Q λύσεων οι οποίες ονομάζονται πληθυσμός του αλγόριθμου:

$$\Pi = \{ \mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_Q \}$$

Σε αντιστοιχία με τους βιολογικούς οργανισμούς, ο πληθυσμός θεωρείται ένα πεπερασμένο σύνολο έμβιων όντων. Κάθε οργανισμός (λύση) έχει μερικά μοναδικά χαρακτηριστικά τα οποία στην περίπτωση της προσομοίωσης ορίζονται αποκλειστικά από την αριθμητική τιμή των παραμέτρων $\mathbf{w}_i$.

Η συνάρτηση επιβίωσης. Η συνάρτηση επιβίωσης αποτελεί ένα μέτρο προσαρμογής της λύσης στις απαιτήσεις της βελτιστοποίησης ή ισοδύναμα περιγράφει την πιθανότητα επιβίωσης της λύσης σε μελλοντικές γενιές. Είναι ένα βαθμωτό μέγεθος με πεπερασμένο άνω όριο. Αν θεωρήσουμε ότι οι απαιτήσεις του περιβάλλοντος περιγράφονται με ακρίβεια από την συνάρτηση της επιβίωσης τότε, αν καταφέρουμε να κατασκευάσουμε έναν μηχανισμό ο οποίος να προσομοιώνει του βασικούς μηχανισμούς που λαμβάνουν χώρα κατά την εξέλιξη των έμβιων οργανισμών, μπορούμε μετά από έναν πεπερασμένο αριθμό γενεών να πετύχουμε λύσεις οι οποίες θα έχουν χαρακτηριστικά προσαρμογής με πολύ μεγάλη αριθμητική τιμή.

Συνεπώς αν έχουμε την συνάρτηση $y = f(\mathbf{x})$, $y \in \Re^+$, γνωρίζουμε ότι είναι άνω φραγμένη και μπορούμε να φτιάξουμε μία διαδικασία που (α) να προσομοιώνει την εξέλιξη των ειδών, (β) να περιγράφει τα χαρακτηριστικά κάθε ανεξάρτητης λύσης σαν συνάρτηση της μεταβλητής $\mathbf{x}$, τότε μετά από ένα πεπερασμένο αριθμό γενεών θα είχαμε λύσεις οι οποίες θα προσέγγιζαν τα τοπικά μέγιστα της συνάρτησης $f(\mathbf{x})$.

Είναι γνωστό ότι πολύ εύκολα μπορούμε μετατρέψουμε ένα πρόβλημα εύρεσης του ελάχιστου μίας συνάρτησης με πεδίο τιμών τους θετικούς αριθμούς σε πρόβλημα

$$g(\mathbf{x}) = \frac{1}{f(\mathbf{x}) + k}$$

εύρεσης μέγιστου. Κάθε τοπικό ελάχιστο της συνάρτησης:

με $k \in \Re^+$, βρίσκεται στην ίδια αριθμητική τιμή του $\mathbf{x}$ για την οποία η συνάρτηση $f(\mathbf{x})$ έχει ένα τοπικό μέγιστο. Με αυτό τον απλό μετασχηματισμό μπορούμε να επιλύσουμε προβλήματα εύρεσης ελάχιστων.

Στα έμβια όντα η συνάρτηση επιβίωσης έχει να κάνει με την ικανότητα του οργανισμού να προσαρμόζεται στο περιβάλλον του. *Είναι γνωστό στην βιολογία ότι από γενιά σε γενιά τα χαρακτηριστικά που κληρονομούνται είναι εκείνα που μεταφέρονται από τα ικανότερα του είδους, δηλαδή από εκείνους τους οργανισμούς που μπορούν να αντιμετωπίσουν με τον καλύτερο δυνατό τρόπο το περιβάλλον τους.*

Το χρωμόσωμα. Το χρωμόσωμα αναφέρεται στα γενετικά χαρακτηριστικά που καθορίζουν με μοναδικό τρόπο την συμπεριφορά της λύσης μας. Από τους περιορισμούς που θέσαμε το χρωμόσωμα εξαρτάται μονάχα από τις ελεύθερες παραμέτρους της συνάρτησης επιβίωσης. Η αντιστοιχία τους πρέπει να είναι αμφιμονοσήμαντη, δηλαδή κάθε χρωμόσωμα πρέπει να έχει μόνο και μόνο ένα ίχνος επάνω στο πεδίο ορισμού της συνάρτησης επιβίωσης. Περιγράφοντας το χρωμόσωμα σαν μία ακολουθία δυαδικών αριθμών είμαστε αναγκασμένοι να κάνουμε μία σειρά παραδοχών οι οποίοι θέτουν όριο στην ακρίβεια υπολογισμού των μέγιστων της συνάρτησης επιβίωσης:

Το πεδίο ορισμού κάθε ανεξάρτητης μεταβλητής της συνάρτησης επιβίωσης είναι άνω και κάτω φραγμένο.

Η δυαδική ακολουθία που περιγράφει κάθε ανεξάρτητη μεταβλητή είναι πεπερασμένη. Η παραδοχή αυτή μας οδηγεί στην ψηφιοποίηση του πεδίου ορισμού κάθε ανεξάρτητης μεταβλητής της συνάρτησης επιβίωσης. Έτσι αν το φραγμένο πεδίο ορισμού είναι γνήσιο απειροσύνολο του συνόλου των πραγματικών αριθμών, κάθε χρωμόσωμα της ελεύθερης μεταβλητής θα αντιστοιχεί σε πεπερασμένο αριθμό στοιχείων του. Υπάρχει συνεπώς η ανάγκη να ψηφιοποιήσουμε το πεδίο ορισμού κάθε ελεύθερης μεταβλητής.

Ας δώσουμε ένα πρακτικό παράδειγμα το οποίο χρησιμοποιείται πολύ συχνά σε πρακτικά προβλήματα. Έστω ότι θέλουμε να φτιάξουμε την αμφιμονοσήμαντη απεικόνιση ανάμεσα στην δυαδική ακολουθία χρωμοσώματος και στην πραγματική τιμή μίας μεταβλητής w_i η οποία έχει πεδίο τιμών το $[w_i^{(Min)}, w_i^{(Max)}]$. Αν θέλουμε τα διακριτά σημεία να έχουν ίση απόσταση μεταξύ τους, τότε αν το μήκος της δυαδικής ακολουθίας του χρωμοσώματος που έχει αφιερωθεί στην μεταβλητή w_i είναι M_i η συνάρτηση που δίνει την αριθμητική τιμή της μεταβλητής για τυχαία ακολουθία δυαδικών αριθμών μήκους M_i είναι:

$$\mathbf{w}_i = \mathbf{w}_i^{(Min)} + \frac{\mathbf{w}_i^{(Max)} - \mathbf{w}_i^{(Min)}}{2^{Mi}} \left(\sum_{j=1}^{Mi} a_j 2^j \right)$$

όπου $a_j \in \{0,1\}$, είναι η δυαδική μεταβλητή στην j θέση του χρωμοσώματος στο οποία φυλάσσεται η δυαδική πληροφορία της μεταβλητής w_i.

Το συνολικό χρωμόσωμα αποτελείται από τις αντίστοιχες δυαδικές ακολουθίες όλων των ελεύθερων μεταβλητών της συνάρτησης επιβίωσης.

Η γενιά. Η γενιά αναφέρεται στον πληθυσμό μίας συγκεκριμένης χρονικής στιγμής. Η μεταβολή των χρωμοσωμάτων κάθε γενιάς πραγματοποιείται με τρεις στοχαστικούς τελεστές οι οποίοι ενεργούν επάνω στον πληθυσμό κάθε γενιάς. Η αναπαραγωγή αντιγράφει τα χρωμοσώματα του πληθυσμού στην επόμενη γενιά, η διασταύρωση ανταλλάσσει τμήματα χρωμοσωμάτων μεταξύ των ατόμων της ίδιας γενιάς και η μετάλλαξη αλλάζει την πληροφορία που μεταφέρεται σε τυχαίες θέσεις του χρωμοσώματος.

Στην συνέχεια ακολουθεί η περιγραφή των τριών βασικών τελεστών οι οποίοι διαφοροποιούν τον πληθυσμό των Q ενεργών λύσεων κάθε γενιάς.

Η αναπαραγωγή. Αναφέρεται στην διαδικασία αντιγραφής χρωμοσώματος στην επόμενη γενιά. Η διαδικασία αντιγραφής είναι στοχαστική. Η πιθανότητα αντιγραφής κάθε διαφορετικού χρωμοσώματος στην επόμενη γενιά έχει μοναδική τιμή που υπολογίζεται από την συνάρτηση επιβίωσης.

Έστω ότι κατά την γενιά t έχουμε Q χρωμοσώματα από τα οποία W από αυτά είναι διαφορετικά μεταξύ τους ($Q \geq W$). Κάθε μοναδικό χρωμόσωμα j, στο οποίο αντιστοιχεί μία μοναδική πραγματική τιμή έστω η $\mathbf{w}_j$ παρουσιάζει την ακόλουθη πιθανότητα επιβίωσης στην επόμενη γενιά:

$$p_j = \frac{1}{\sum_{k=1}^{W} f(\mathbf{w}_k)} f(\mathbf{w}_j)$$

Υπολογίζουμε την αθροιστική πιθανότητα:

$$P_j = \frac{1}{\sum_{k=1}^{W} f(\mathbf{w}_k)} \sum_{k=1}^{j} f(\mathbf{w}_k), \qquad j < W$$

Με την βοήθεια μίας γεννήτριας αριθμών που ακολουθεί ομοιόμορφη πυκνότητας πιθανότητας στο διάστημα [0,1), επιλέγω έναν τυχαίο αριθμό. Έστω ότι ο αριθμός αυτός είναι ο Ta. Το χρωμόσωμα k που θα επιβιώσει στην επόμενη γενιά είναι αυτό που ικανοποιεί την ακόλουθη σχέση:

$$P_{k-1} < \mathbf{T}a < P_k, \quad \mu\varepsilon\ P_0 = 0$$

Η διασταύρωση. Είναι γνωστό ότι κατά την δημιουργία ενός νέου οργανισμού τα χρωμοσώματα δύο διαφορετικών κυττάρων ανταλλάσσουν γενετική πληροφορία με ανταλλαγή μακρομορίων DNA. Η προσομοίωση του μηχανισμού αυτού πραγματοποιείται με τυχαίο τεμαχισμό της δυαδικής συμβολοσειράς του χρωμοσώματος μίας λύσης και την ανταλλαγή του με τυχαίο χρωμόσωμα άλλης λύσης.

Αν έχουμε δύο χρωμοσώματα, τα:

$$X_1 = \{\alpha_1\ \alpha_2\ \alpha_3\ \mathbf{...}\ \alpha_i\ a_{i+1}\ \mathbf{...}\ \alpha_Z\}$$

$$X_2 = \{\beta_1\ \beta_2\ \beta_3\ \mathbf{...}\ \beta_i\ \beta_{i+1}\ \mathbf{...}\ \beta_Z\}$$

Μετά την διασταύρωση τα νέα χρωμοσώματα θα είναι:

$$X_1 = (\alpha_1\ \alpha_2\ \alpha_3\ \mathbf{...}\ \alpha_i\ \beta_{i+1}\ \mathbf{...}\ \beta_Z)$$

$$X_2 = (\beta_1\ \beta_2\ \beta_3\ \mathbf{...}\ \beta_i\ a_{i+1}\ \mathbf{...}\ \alpha_Z)$$

Η μετάλλαξη. Η μετάλλαξη αναφέρεται στους μηχανισμούς εκείνους οι οποίοι ευθύνονται για τον τυχαίο μετασχηματισμό του γενετικού υλικού. Οφείλεται συνήθως σε κάποιους τυχαίους παράγοντες όπως, κοσμική ακτινοβολία και ραδιενέργεια.

Στην προσομοίωση αυτό που κάνουμε είναι να επιλέγουμε ένα τυχαίο δυαδικό στοιχείο του χρωμοσώματος και να του αλλάζουμε την γενετική του πληροφορία, ως εξής:

$$(\alpha_1\ \alpha_2\ \alpha_3\ \ldots\ \alpha_i\ \ldots\ \alpha_Z) \rightarrow (\alpha_1\ \alpha_2\ \alpha_3\ \ldots\ (1\text{-}\alpha_i)\ \ldots\ \alpha_Z)$$

Με δεδομένο ότι η αριθμητική τιμή του α_i είναι 0 ή 1, η ποσότητα $1\text{-}\alpha_i$ αλλάζει την πληροφορία που περιέχει το χρωμόσωμα στην συγκεκριμένη θέση.

Η βασική δομή ενός γενετικού αλγόριθμου ο οποίος μπορεί να χρησιμοποιηθεί για να εκπαιδεύσει ένα σύστημα αναγνώρισης προτύπων έχει ως εξής:

Αλγόριθμος

Τοποθέτηση αρχικών τιμών. Ορίζουμε τα όρια των λύσεων που αναζητάμε για κάθε μία ελεύθερη παράμετρο της συνάρτησης του σφάλματος.

$$L_i = [\ \mathbf{w}_i^{(Min)},\ \mathbf{w}_i^{(Max)}\ ]$$

Ορίζουμε το μήκος του χρωμοσώματος το οποίο θα είναι αφιερωμένο στην παράμετρο w_i, έστω M_i. Ορίζουμε στο πεδίο ορισμού της ελεύθερης μεταβλητής 2^{M_i} διακριτές τιμές στις οποίες αντιστοιχούμε αμφιμονοσήμαντα όλες οι πιθανές ακολουθίες των M_i δυαδικών συμβόλων.

Ορίζουμε την πιθανότητα μετάλλαξης p_μ.

Ορίζουμε την πιθανότητα διασταύρωσης p_δ.

Ορίζουμε το μέγεθος Q του πληθυσμού. t=0.

Τοποθετούμε τυχαίες δυαδικές τιμές στα Q χρωμοσώματα:

$$\Pi^{(0)} = \{\ X_1^{(0)},\ X_2^{(0)},\ X_3^{(0)},\ \ldots,\ X_i^{(0)},\ \ldots,\ X_Q^{(0)})$$

Με την βοήθεια της συνάρτησης σφάλματος ορίζουμε την συνάρτηση επιβίωσης η οποία συνήθως επιλέγεται να είναι:

$$f(\mathbf{w}) = \frac{1}{\Sigma\phi(\mathbf{w}) + k}$$

όπου k είναι μικρός θετικός πραγματικός αριθμός.

Τα επόμενα τέσσερα βήματα του αλγόριθμου εκτελούνται για ένα σταθερό αριθμό επαναλήψεων.

- **Υπολογισμός πιθανοτήτων επιβίωσης**. Για τα Μ διαφορετικά χρωμοσώματα της γενιάς $\Pi^{(t)}$ υπολογίζουμε την συνάρτηση και την πιθανότητα επιβίωσης, ως και την αντίστοιχη αθροιστική πιθανότητα.

- **Αναπαραγωγή Q χρωμοσωμάτων**. Με την στοχαστική διαδικασία που έχουμε ήδη περιγράψει αναπαράγουμε το ίδιο πληθυσμό Q χρωμοσωμάτων.

- **Διασταύρωση**. Με πιθανότητα p_δ διασταυρώνουμε κάθε χρωμόσωμα του πληθυσμού με ένα χρωμόσωμα διαφορετικής δομής από τον ίδιο πληθυσμό.

- **Μετάλλαξη**. Με πιθανότητα p_μ αντιστρέφουμε κάθε δυαδικό σύμβολο για όλα τα χρωμοσώματα του πληθυσμού,

Στον πληθυσμό της τελευταίας γενιάς αναζητούμε το χρωμόσωμα το οποίο έχει την μεγαλύτερη αριθμητική τιμή στην συνάρτηση της επιβίωσης. Η αριθμητική τιμή της μνήμης της συνάρτησης επιβίωσης που αντιστοιχεί στο βέλτιστο χρωμόσωμα είναι εκείνη που ελαχιστοποιεί ταυτόχρονα και την συνάρτηση του σφάλματος όπως μπορεί να αποδειχθεί εύκολα από την συνάρτηση επιβίωσης.

Η αριθμητική τιμή της πιθανότητας μετάλλαξης επηρεάζει σημαντικά την συμπεριφορά του γενετικού αλγόριθμου και πρέπει να τίθεται με προσοχή. Ουσιαστικά αποτελεί ρυθμιστικό παράγοντα απέναντι στους δύο αντιθετικούς παράγοντες που συνυπάρχουν στους γενετικούς αλγόριθμους. Από την μία μεριά υπάρχει η τυχαία αναζήτηση που ελέγχεται από την πιθανότητα μετάλλαξης και από την άλλη ο προσανατολισμός του αλγόριθμου που εκφράζεται με την διαδικασία στοχαστικής επιβίωσης των λύσεων ελέγχεται από την διαδικασία αναπαραγωγής και διασταύρωσης χρωμοσωμάτων.

Όταν η συνάρτηση σφάλματος έχει πολλά τοπικά ελάχιστα ή σημεία ασυνέχειας, τότε πρέπει να δώσουμε έμφαση στον παράγοντα που εκφράζει την τυχαία αναζήτηση πρέπει να αυξήσουμε την πιθανότητα μετάλλαξης ενώ σε αντίθετη περίπτωση, όταν η συνάρτηση σφάλματος είναι συνεχής και έχει λίγα τοπικά ελάχιστα τότε μειώνουμε την πιθανότητα μετάλλαξης. Αποδεκτές πιθανότητες μετάλλαξης είναι συνήθως πραγματικοί αριθμοί στο διάστημα (0.001,0.1).

Επίσης αξίζει να αναφέρουμε ότι η πιθανότητα διασταύρωσης καθορίζει και τον βαθμό της κατευθυνόμενης αναζήτησης.

Στην βιβλιογραφία θα βρει κανείς πολλές παραλλαγές των γενετικών αλγόριθμων, οι οποίοι έχουν κατασκευαστεί για να αντιμετωπίσουν ιδιαιτερότητες που παρουσιάζονται σε πρακτικά προβλήματα.

Η πιο σημαντική παραλλαγή του τρόπου δυαδικής κωδικοποίησης των παραμέτρων της συνάρτησης επιβίωσης, αποτελεί η εκθετική κωδικοποίηση (σε αντιδιαστολή με την ομοιόμορφη κωδικοποίηση που έχουμε ήδη περιγράψει). Η εκθετική κωδικοποίηση χωρίζει την δυαδική ακολουθία σε δύο τμήματα το ακέραιο και το δεκαδικό της τμήμα επιτυγχάνοντας με αυτό τον τρόπο διαφορετική ακρίβεια προσέγγισης για το ακέραιο και το δεκαδικό τμήμα κάθε μεταβλητής που κωδικοποιείται.

Παράδειγμα. Έστω ότι θέλουμε να εκπαιδεύσουμε ένα νευρωνικό δίκτυο το οποίο διαθέτει δύο επίπεδα με πέντε και τρεις νευρώνες αντίστοιχα σε κάθε επίπεδο. Το δίκτυο αναγνωρίζει πρότυπα τα οποία αποτελούνται από μετρήσεις δέκα παραμέτρων. Θεωρώντας ότι όλοι οι συντελεστές του νευρωνικού δικτύου βρίσκονται στο διάστημα [-1,1] υπολογίστε το μήκος του χρωμοσώματος που πρέπει να χρησιμοποιήσετε σε γενετικό αλγόριθμο έτσι ώστε να υπολογίσετε όλους τους συντελεστές του νευρωνικού δικτύου με ακρίβεια μεγαλύτερη των τριών δεκαδικών ψηφίων.

Παράδειγμα. Έστω ότι η συνάρτηση ταξινόμησης δίνεται από την σχέση:

$$\mathbf{R}(\mathbf{w},\,\mathbf{x}) = \left(\frac{1}{\sqrt{|\,\mathbf{w}_1(\mathbf{x}_1 - \mathbf{x}_2) + 1\,|}}\,,\,(\mathbf{w}-\mathbf{x})^T(\mathbf{w}-\mathbf{x}) \right)$$

Θέλουμε να κατασκευάσουμε σύστημα ταξινόμησης προτύπων δύο κατηγοριών διαθέτοντας τα εξής παραδείγματα:

$$\Omega_1 = \{(-1.5,0.5),(-1,1.5),(1,1),(2,1),(1,2)\}$$

$$\Omega_2 = \{(-0.75,0.75),(-0.5,1.2),(0.5,0.5),(2.5,0.5),(2.5,2)\}$$

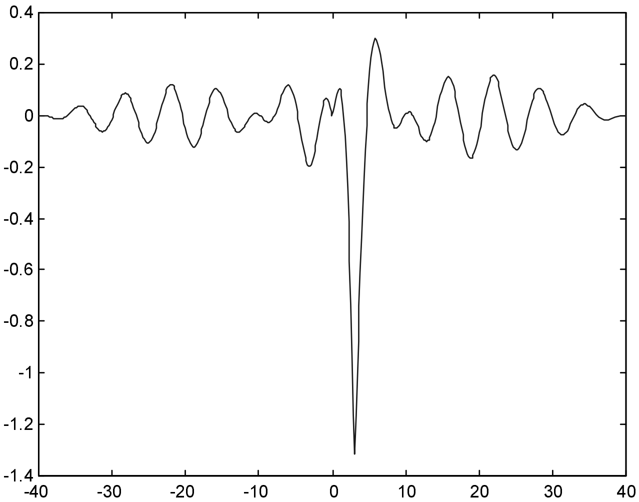
για τις δύο κατηγορίες αντίστοιχα.

Θέσατε τις παραμέτρους γενετικού αλγόριθμου ο οποίος θα υπολογίζει τις παραμέτρους του συστήματος ταξινόμησης λαμβάνοντας υπόψη ότι πρέπει να αναζητήσετε μία βέλτιστη λύση στο διάστημα $w_1 \in [-1000,1000]$ και $w_2 \in [-5,5]$, με ελάχιστη ακρίβεια τεσσάρων δεκαδικών ψηφίων.

Παράδειγμα. Η γραφική παράσταση της συνάρτησης

$$f(\mathbf{x})=\frac{\log(|\mathbf{x}|+1)\sin(\sqrt{\mathbf{x}})\cos(\mathbf{x})}{|\mathbf{x}-3|+1}$$

στο διάστημα [-40,40] δίνεται στο σχήμα που ακολουθεί.



Υπολογίστε την αριθμητική τιμή του ελάχιστου της συνάρτησης f(x) με την βοήθεια γενετικών αλγορίθμων που διδαχθήκατε στο παρόν κεφάλαιο.

Λύση:

Αρχικά πρέπει να ορίσω τις παραμέτρους του γενετικού αλγόριθμου οι οποίοι είναι οι εξής:

Η συνάρτηση επιβίωσης πρέπει να έχει περίο τιμών το σύνολο των θετικών πραγματικών αριθμών και πρέπει να μετατρέψουμε το πρόβλημα ελαχιστοποίησης σε ένα αντίστοιχο πρόβλημα μεγιστοποίησης της συνάρτησης ταξινόμησης. Το πεδίο τιμών της συνάρτησης f(x) στο διάστημα [-40 40] δεν είναι μεγαλύτερο του [-2 1], συνεπώς η αντίστοιχη συνάρτηση:

$$g(\mathbf{x})=\frac{1}{f(\mathbf{x})+3}$$

έχει πεδίο τιμών στο σύνολο των θετικών πραγματικών αριθμών και το μέγιστο της συνάρτησης g(x) αντιστοιχεί στο ελάχιστο της συνάρτησης f(x).

Ορίζω μέγεθος πληθυσμού 20 λύσεων. Το ελάχιστο της συνάρτησης θα αναζητηθεί στο διάστημα [-40 40]. Το διάστημα αυτό θα χωριστεί σε 2^{15} τμήματα. που σημαίνει ότι κάθε ανεξάρτητη λύση θα κωδικοποιηθεί με 15 bits. Θέτω πιθανότητα διασταύρωσης 0.9, πιθανότητα μετάλλαξης 0.2 ανά χρωμόσωμα.

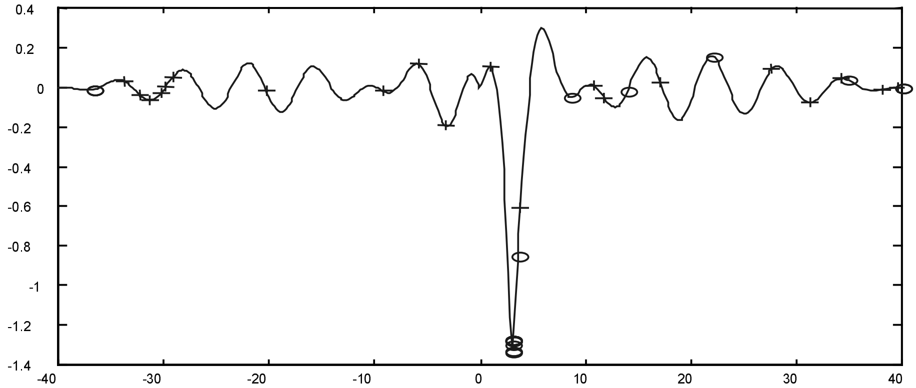
Αρχικά με τυχαίο τρόπο κατασκευάζω ακολουθίες bits σε κάθε ένα από τα είκοσι χρωμοσώματα. Οι αντίστοιχες λύσεις με την αντίστοιχη αριθμητική τιμή της συνάρτησης f(x)

δίνεται στον πίνακα που ακολουθεί.

x	*-25.1514*	*-1.1881*	*22.1632*	*12.3090*	*-38.0147*
f(x)	*0.0075*	*-0.0002*	*-0.0250*	*0.1180*	*-0.0176*
x	*-6.4127*	*17.3175*	*24.2287*	*-36.4573*	*13.8069*
f(x)	*-0.6068*	*0.0526*	*-0.0357*	*-0.0749*	*-0.0092*
x	*8.3424*	*-13.3070*	*-3.1243*	*-36.0303*	*-13.7424*
f(x)	*-0.0621*	*-0.0156*	*0.0034*	*0.0263*	*0.0293*
x	*-15.0067*	*-32.3801*	*3.3875*	*28.3273*	*-20.1818*
f(x)	*-0.1926*	*0.0939*	*0.1054*	*-0.0515*	*0.0438*

Εκτελώ τα επαναληπτικά βήματα του του γενετικού αλγόριθμου για είκοσι γενιές και παρατηρώ την συμπεριφορά των λύσεων όπως αυτή απαικονίζεται στο επόμενο σχήμα και τον επόμενο πίνακα.

Ο πίνακας αποτελεί την γραφική απεικόνιση των λύσεων επάνω στην γραφική παράσταση της συνάρτησης f(x). με '+' είναι οι λύσεις της τυχαίας αρχικής επιλογής λύσεων, ενώ με 'o' επεικονίζονται οι 20 λύσεις μετά από 20 γενιές.



Παρατηρούμε ότι ενώ οι αρχικές λύσεις είναι τυχαία κατανεμημένες στον χώρο στον οποίο αναζητούμε το ελάχιστο της συνάρτησης f(x), οι λύσεις που δίνονται μετά από 20 επαναλήψεις των μετασχηματισμών του γενετικού αλγόριθμου είναι συγκεντρωμένες γύρω από την βέλτιστη λύση, παρά την παρουσία μεγάλου αριθμού τοπικών ελάχιστων. Μερικές μάλιστα από αυτές τις λύσεις (τρείς) βρίσκονται κοντά σε τοπικά ακρότατα.

Συγκεκριμένα η βέλτιστη λύση που μπορούμε να απομονώσουμε από την τελική γενιά των λύσεων είναι η x=2.9878 με f(2.9878)=-1.3337. Παρατηρούμε βάβαια ότι σε απόσταση 0.1 βρίσκονται οι δέκα από ένα σύνολο 20 λύσεων.

x	*-36.5903*	*-36.5838*	*3.0569*	*2.9878*	*3.5522*
f(x)	*-0.0092*	*-0.0091*	*-1.2995*	*-1.3337*	*-0.8515*
x	*3.0832*	*39.9974*	*3.0800*	*3.0808*	*40.0000*
f(x)	*-1.2745*	*-0.0027*	*-1.2775*	*-1.2768*	*-0.0027*
x	*34.8145*	*2.9855*	*14.0122*	*8.6664*	*22.1045*
f(x)	*0.0394*	*-1.3297*	*-0.0159*	*-0.0485*	*0.1552*
x	*3.0831*	*3.0831*	*3.0189*	*3.0872*	*3.0872*
f(x)	*-1.2745*	*-1.2745*	*-1.3362*	*-1.2707*	*-1.2707*

Στην συνέχεια θα δούμε ένα δυσκολότερο παράδειγμα στο οποίο εμφανίζονται πολλαπλά τοπικά ελάχιστα τα οποία διαφέρουν ελάχιστα ως προς την αριθμητική τους διαφορά από το συνολικό ελάχιστο. Σε αυτές τις περιπτώσεις πρέπει να επιλέξουμε προσεκτικά την αριθμητική τιμή των παραμέτρων του γενετικού αλγόριθμου όπως θα φανεί και στο παράδειγμα που ακολουθεί

Παράδειγμα. Υπολογίστε την αριθμητική τιμή του ελάχιστου της συνάρτησης

$$f(\mathbf{x}) = \sqrt{\log\left(\frac{1}{|\mathbf{x}-20|+0.5}\right)\cos(\mathbf{x})}$$

με την βοήθεια γενετικών αλγορίθμων στο διάστημα [0,50]

Λύση:

Χαρακτηριστικό γνώρισμα της συνάρτησης που μελετάμε είναι η ύπαρξη πολλών τοπικών ελάχιστων στην γειτονιά του συνολικά ελάχιστου. Το γεγονός αυτό εμποδίζει τον γενετικό αλγόριθμο να συγκλίνει στο σωστό σημείο διότι όταν κάποια χρωμοσώματα μας δώσουν ένα τοπικά ελάχιστο το οποίο έχει αριθμητική τιμή κοντά στην τιμή του συνολικά ελάχιστου τότε αυτή η λύση θα έχει μεγάλη πιθανότητα να επιζήσει στις επόμενενς γενιές αφού θα παρουσιάζει υψηλή πιθανότητα επιβίωσης. Αν ο αριθμός των χρωμοσωμάτων είναι σχετικά μικρός τότε υπάρχει αρκετά μεγάλη πιθανότητα να εγκλωβιστούν πολλές λύσεις σε αυτό το τοπικά ελάχιστο, με έμμεση συνέπεια να ελαττώνεται αντίστοιχα η πιθανότητα να βρεθεί η βέλτιστη λύση.

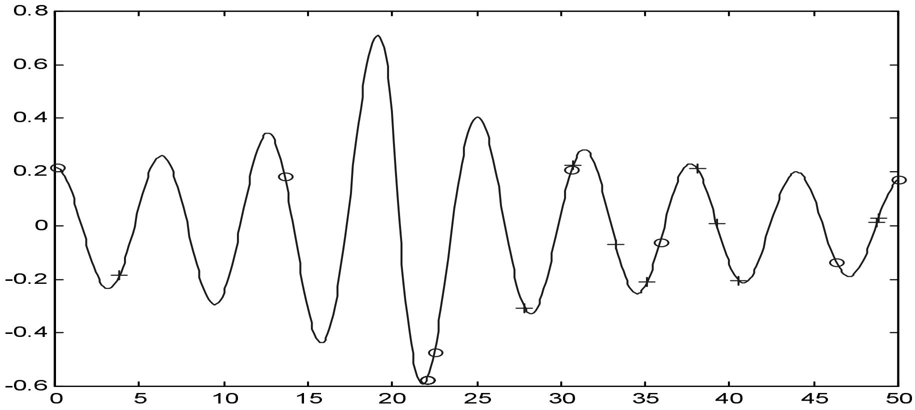
Αν λοιπόν ορίσω τις παραμέτρους του γενετικού αλγόριθμου ως εξής:

Η συνάρτηση επιβίωσης πρέπει να έχει πεδίο τιμών το σύνολο των θετικών πραγματικών αριθμών και πρέπει να αναζητήσουμε την σχέση που υπάρχει ανάμεσα στο ελάχιστο της συνάρτησης f(x) κει της συνάρτησης επιβίωσης μετετρέποντας το πρόβλημα ελαχιστοποίησης σε πρόβλημα μεγιστοποίησης. Το πεδίο τιμών της συνάρτησης f(x) στο διάστημα [0 50] δεν είναι μεγαλύτερο του [-1 1], συνεπώς η αντίστοιχη συνάρτηση επιβίωσης g(x) δίνεται από την σχέση:

$$g(\mathbf{x}) = \frac{1}{f(\mathbf{x})+2}$$

Ορίζω μέγεθος πληθυσμού 20 λύσεων. Το ελάχιστο της συνάρτησης θα αναζητηθεί στο διάστημα [0 50]. Το διάστημα αυτό θα χωριστεί σε 2^{15} τμήματα. που σημαίνει ότι κάθε ανεξάρτητη λύση θα κωδικοποιηθεί με 15 bits. Θέτω πιθανότητα διασταύρωσης 0.9, πιθανότητα μετάλλαξης 0.2 ανά χρωμόσωμα.

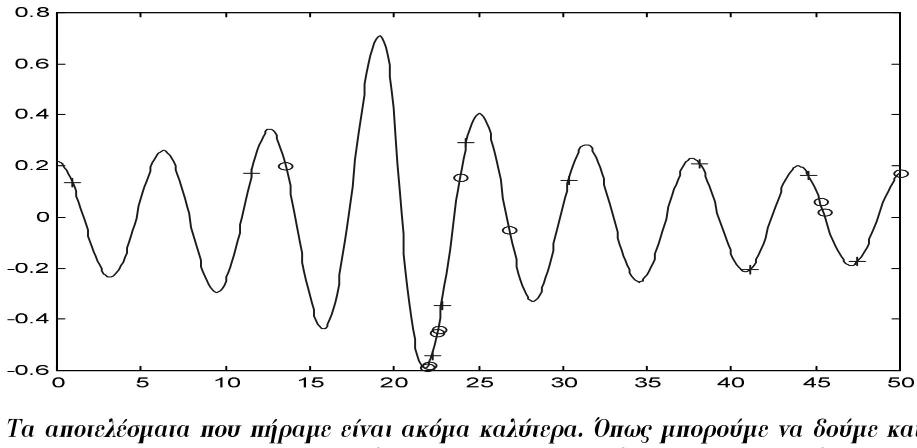
Αρχικά με τυχαίο τρόπο κατασκευάζω ακολουθίες bits σε κάθε ένα από τα είκοσι χρωμοσώματα. Οι αντίστοιχες λύσεις με την αντίστοιχη αριθμητική τιμή της συνάρτησης f(x) να δίνεται στο σχήμα που ακολουθεί. Με "+" απεικονίζονται οι τυχαίες εκτιμήσεις των ελάχιστων πρίν την εκκίνηση του αλγόριθμου και με "o" απεικονίζονται οι λύσεις που προκύπτουν μετά



από πέντε γενιές.

Παρατηρούμε ότι παρόλο που και ο αριθμός των λύσεων κάθε γενιάς είναι μικρός, το γεγονός ότι ο γενετικός αλγόριθμος επαναλαμβάνεται μόνο πέντε φορές και ότι η ακρίβεια της λύσης που αναζητούμε είναι πολύ μεγάλη (ο χώρος αναζήτησης στον δυαδικό χώρο είναι πολύ μεγάλος), βλέπουμε τελικά ότι το συνολικό ελάχιστο της συνάρτησης έχει υπολογιστεί με ακρίβεια 0.75.

Αν εκτελέσουμε για δεύτερη φορά, με ίδιες παραμέτρους, τον γενετικό αλγόριθμο θα πάρουμε τα εξής αποτελέσματα



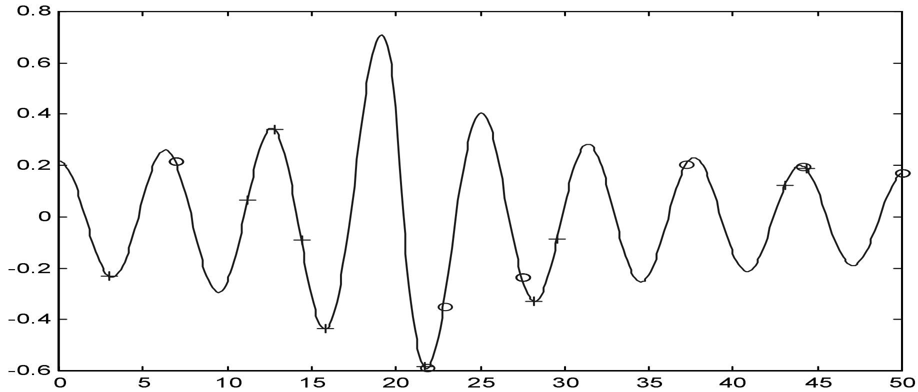
Τα αποτελέσματα που πήραμε είναι ακόμα καλύτερα. Όπως μπορούμε να δούμε και από το σχήμα τέσσερις λύσεις έχουν εγκλωβιστεί στην γειτονιά της βέλτιστης λύσης. Η βέλτιστη λύση που

προκύπτει μετά από πέντε γενιές είναι η x=21.8877 με τιμή της συνάρτησης επιβίωσης g(21.8877)=0.7084.

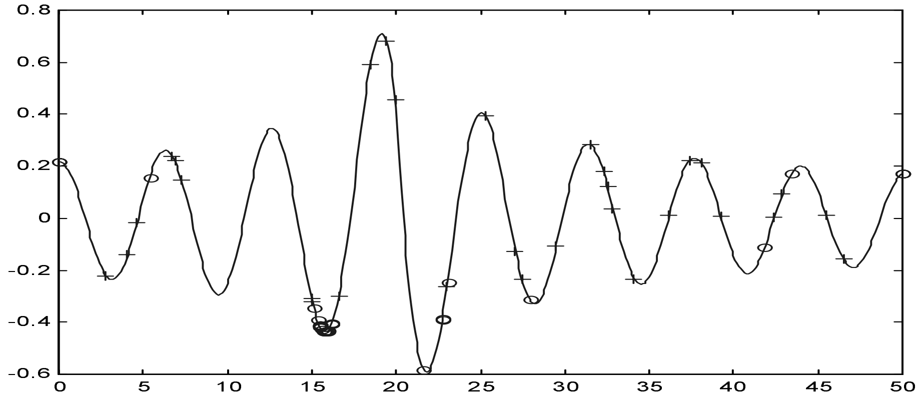
Ας δούμε τι θα συμβεί αν αυξήσουμε τον αριθμό των γενιών.

Στο σχήμα που ακολουθεί δίνονται οι δέκα λύσεις μετά από 30 γενιές. Βλέπουμε ότι στην περιοχή του ολικά ελάχιστου και με ακρίβεια που είναι μεγαλύτερη του 0.1 βρίσκονται τέσσερις από τις δέκα λύσεις.

Συνεπώς και ο αριθμός των λύσεων που βρίσκονται κοντά στο ολικό ελάχιστο έχει αυξηθεί αλλά ταυτόχρονα αυξήθηκε σημαντικά και η ακρίβεια υπολογισμού της θέσης του ελάχιστου.



Τεχνικές βελτίωσης της ακρίβειας με την οποία βρίσκουμε την βέλτιστη λύση υπάρχουν πολλοί. Ο απλούστερος και ο πλέον αποτελεσματικός είναι να αυξήσουμε τον αριθμό των λύσεων ανά γενιά. Σε αυτή την περίπτωση μετά από πέντε γενιές έχουμε τις λύσεις που απεικονίζονται στο σχήμα που ακολουθεί. Βλέπουμε μία θαυμαστή συγκέντρωση 23 λύσεων στο συνολικό και τα δύο καλύτερα τοπικά ελάχιστα της συνάρτησης f(x).



Αν επαναλάβουμε τον γενετικό αλγόριθμο για τον ίδιο αριθμό λύσεων ανά γενιά αλλά τερματίσουμε τους υπολογισμούς μετά από τριάντα γενιές θα αντικρύσουμε μία τελείως διαφορετική εικόνα. Δεκαοκτώ λύσεις βρίσκονται στην γειτονιά του ολικού ελάχιστου όπως

φαίνεται στο σχήμα που ακολουθεί.

Η εκπαίδευση συστημάτων ταξινόμησης προτύπων φαίνεται να είναι μία απλή διαδικασία στην περίπτωση κατά την οποία η ελαχιστοποίηση της συνάρτησης του σφάλματος επιχειρείται να πραγματοποιηθεί με την βοήθεια των γενετικών αλγόριθμων. Ας δούμε ένα απλό παράδειγμα εκπαίδευσης συστήματος ταξινόμησης προτύπων που βασίζεται σε παραδείγματα εκπαίδευσης.

Παράδειγμα. Εστω ότι έχουμε το σύστημα ταξινόμησης προτύπων δύο κατηγοριών το οποίο διαθέτει μία περίεργη συνάρτηση ταξινόμησης.

Το δι-διάστατο πρότυπο x ταξινομείται στην πρώτη κατηγορία όταν

$$\sum_{i=1}^{2} \mathbf{w}_{i1}(\mathbf{x}_i - m_{i1})^2 < 1 \quad \text{είτε} \quad \sum_{i=1}^{2} \mathbf{w}_{i2}(\mathbf{x}_i - m_{i2})^2 < 1 \quad \text{είτε} \quad \sum_{i=1}^{2} \mathbf{w}_{i3}(\mathbf{x}_i - m_{i3})^2 < 1$$

διαφορετικά ταξινομείται στην δεύτερη κατηγορία. Οι συντελεστές w_{ij} είναι θετικοί αριθμοί.

Εστω ότι έχετε τα εξής δεκατρία παραδείγματα εκπαίδευσης για τα πρότυπα που ανήκουν στην πρώτη κατηγορία αντικειμένων

$Ω_1$ = {(1,4),(2,4.5),(3,3.5),(4,4),(5,3.2),(3,3),(3.5,2.5),(2.5,2.5),(1,1),(2,2.5),(3,1),(4,1),(5,1)}

Τα διαθέσιμα παραδείγματα εκπαίδευσης για τα πρότυπα που ανήκουν στην δεύτερη κατηγορία αντικειμένων είναι δεκαπέντε, τα ακόλουθα:

$Ω_2$ = {(1,2.5),(0.5,3),(2,2.5),(3.4,3),(3.5,2.5),(4,2.5),(5,3),(2,6),(4,6),(5,6),(6.5,4),(6,4),(6,1),(6,2),(6,3)}.

Μπορείτε να υπολογίσετε τις σταθερές παραμέτρους της συνάρτησης ταξινόμησης έτσι ώστε να ελαχιστοποιήσετε τα σφάλματα ταξινόμησης;

Τα πρότυπα των κατηγοριών έχουν επιλεγεί με τέτοιο τρόπο έτσι ώστε το σύστημα ταξινόμησης με κατάλληλη επιλογή των παραμέτρων εκπαίδευσης να μπορεί να ταξινομήσει σωστά όλα τα παραδείγματα εκπαίδευσης.

Λύση:

Αρχικά πρέπει να ορίσουμε την συνάρτηση ταξινόμησης και να βρούμε το αντίστοιχο πρόβλημα βελτιστοποίησης που πρέπει να λύσουμε. Το αριθμητικό μέγεθος που πρέπει να ελαχιστοποιήσουμε είναι οι λανθασμένες ταξινομήσεις του συστήματος αναγνώρισης προτύπων. Βλέπουμε λοιπόν ότι δεν είναι τόσο απλό πρόβλημα η κατασκευή της συνάρτησης σφάλματος του συστήματος ταξινόμησης προτύπων. Δεν μπορούμε να βρούμε την αναλυτική έκφραση της συνάρτησης που συνδέει τις ελεύθερες μεταβλητές, τα παραδείγματα εκπαίδευσης με το σφάλμα του συστήματος ταξινόμησης των παραδειγμάτων εκπαίδευσης.

Η σχέση αυτή είναι αλγοριθμική. Δηλαδή μπορεί να υπολογιστεί με πρόγραμμα υπολογιστή. Είναι πολύ δύσκολο να διαπιστώσουμε αν η συνάρτηση αυτή είναι συνεχής και παραγωγίσιμη έτσι ώστε να μπορέσουμε να εφαρμόσουμε κάποιους από τους γνωστούς αλγορίθμους εύρεσης του ελάχιστου της συνάστησης σφάλματος. Είναι εύκολο όμως να αποδείξουμε ότι η σχέση που δίνει το ποσοστό των σωστών ταξινομήσεων στο σύνολο των παραδειγμάτων εκπαίδευσης σε

σχέση με τις μεταβλητές w_{ij} και m_{ij} είναι μία μονοσήμαντη απεικόνιση, δηλαδή για συγκεκριμένο σύνολο παραμέτρων το ποσοστό σωστών ταξινομήσεων υπολογίζεται με μοναδικό τρόπο.

Σχηματικά η σχέση που μας δίνει τον λόγο των σωστών ταξινομήσεων προς το συνολικό αριθμό των παραδειγμάτων εκπαίδευσης είναι συνάρτηση των άγνωστων μεταβλητών τις οποίες θέλουμε να υπολογίσουμε, τα w_{ij} και m_{ij}.

Η συνάρτηση αυτή μπορεί να επιλεγεί σαν συνάρτηση επιβίωσης του γενετικού αλγόριθμου διότι έχει πεδίο τιμών το [0,1] και το μέγιστο της συνάρτησης αντιστοιχεί σε εκείνες τις παραμέτρους του συστήματος αναγνώρισης προτύπων το οποίο δίνει και το μικρότερο σφάλμα ταξινόμησης.

Ας τοποθετήσουμε και τις υπόλοιπες τιμές του γενετικού αλγόριθμου.

Το μήκος του χρωμοσώματος υπολογίζεται σαν συνάρτηση του αριθμού των παραμέτρων και της απαιτούμενης ακρίβειας υπολογισμού.

Ας υποθέσουμε ότι χρειαζόμαστε ακρίβεια υπολογισμού της κάθε μιάς των άγνωστων παραμέτρων η οποία να είναι τουλάχιστων της τάξης τεσσάρων δεκαδικών ψηφίων.

Το πεδίο αναζήτησης λύσεων είναι διαφορετικό για κάθε κατηγορία παραμέτρων. Οι συντελεστές w_{ij} έχουν πεδίο ορισμού το [-5 5], ενώ οι συντελεστές m_{ij} έχουν πεδίο ορισμού το [0,8] (εξετάζοντας τις αριθμητικές τιμές των παραδειγμάτων εκπαίδευσης και γνωρίζοντας ότι οι σχέσεις ταξινόμησης περιγράφουν το εσωτερικό έλλειψης) Το μήκος δυαδικής περιγραφής των συντελεστών w_{ij} είναι

Μήκος περιγραφής w_{ij} = (μήκος πεδίου ορισμού) / (ακρίβεια σε δεκαδικά ψηφία) = $10 * 10^4 = 10^5$

Το μικρότερο μήκος δυαδικής κωδικοποίησης η οποία μπορεί να περιγράψει 10^5 καταστάσεις είναι 17 bits ($2^{16} < 10^5 < 2^{17}$).

Το αντίστοιχο μήκος δυαδικής περιγραφής των συντελεστών m_{ij} είναι

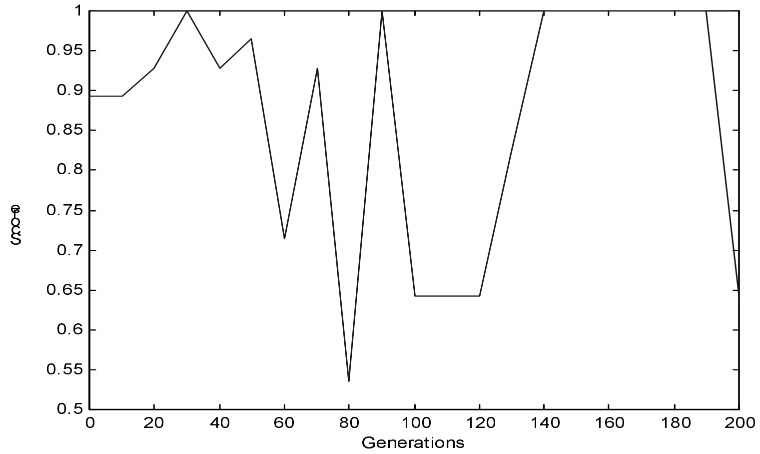
Μήκος περιγραφής m_{ij} = (μήκος πεδίου ορισμού) / (ακρίβεια σε δεκαδικά ψηφία) = $8 * 10^4$

Το μικρότερο μήκος δυαδικής κωδικοποίησης η οποία μπορεί να περιγράψει $8*10^4$ καταστάσεις είναι επίσης 17 bits (($2^{16} < 8*10^4 < 2^{17}$)).

*Το συνολικό μήκος του χρωμοσώματος το οποίο περιγράφει με μοναδικό τρόπο την μνήμη του συστήματος ταξινόμησης στο πεδίο ορισμού των λύσεων του γενετικού αλγόριθμου ισούται με το επιμέρους άθροισμα της δυαδικής αναπαράστασης κάθε ελεύθερης μεταβλητής, δηλαδή 17*6 = 102 bits.*

Θέτουμε πιθανότητα διασταύρωσης 0.6 και πιθανότητα μετάλλαξης ανά χρωμόσωμα 0.6. Η πιθανότητα μετάλλαξης έχει σχετικά μεγάλη αριθμητική τιμή. Συνεπώς ο γενετικός αλγόριθμος ακολουθεί σε μεγαλύτερο βαθμό την συμπεριφορά της τυχαίας αναζήτησης και λιγότερο την συμπεριφορά αναζήτησης λύσης στην γειτονιά των βέλτιστων λύσεων που έχουν επιτευχθεί σε κάθε γενιά χρωμοσωμάτων.

Στο σχήμα που ακολουθεί δείχνεται η πορεία εξέλιξης του γενετικού αλγόριθμου απεικονίζοντας την μεγαλύτερη αριθμητική τιμή της συνάρτησης επιβίωσης των χρωμοσωμάτων που προκύπτουν κάθε 10 γενιές. Ο αριθμός των χρωμοσωμάτων ανά γενιά



είναι 50.

Παρατηρούμε ότι η βέλτιστη τιμή εντοπίζεται μετά από 30 περίπου γενιές δίνοντας αξιοπιστία ταξινόμησης 100%. Η λύση αυτή επιτυγχάνεται διότι τα παραδείγματα εκπαίδευσης έχουν κατασκευαστεί τεχνητά ώστε να υπάρχουν τρείς ελλείψεις οι οποίες να περιέχουν τα πρότυπα της πρώτης κατηγορίας και να μην περιέχουν κανένα πρότυπο της δεύτερης κατηγορίας. Παρατηρούμε επίσης ότι μετά από 60 γενιές οι παράμετροι του συστήματος ταξινόμησης οι οποίοι υπολογίζονται από τον γενετικό αλγόριθμο δεν δίνουν την βέλτιστη λύση. Το φαινόμενο αυτό οφείλεται στην μεγάλη αριθμητική τιμή της πιθανότητας μετάλλαξης. Αν η μετάλλαξη μεταβάλλει σημαντικά την βέλτιστη λύση, συνδυαζόμενη με την ύπαρξη πολλών λύσεων στα χρωμοσώματα της γενιάς τα οποία έχουν αριθμητική τιμή της συνάρτησης επιβίωσης κοντά στην βέλτιστη λύση μπορεί να επιφέρουν την διαγραφή της βέλτιστης λύσης. Το φαινόμενο αυτό είναι εντονώτερο όταν έχουμε επιλέξει περιορισμένο πληθυσμό χρωμοσωμάτων ανά γενιά.

Στον πίνακα που ακολουθεί δίνεται η αριθμητική τιμή των παραμέτρων του συστήματος ταξινόμησης στην βέλτιστη λύση (επιτυχία ταξινόμησης 100% ή ισοδύναμα τιμή της συνάρτησης επιβίωσης 1).

	w_1	m_1	w_2	m_2
Πρώτη έλλειψη	*2.5398*	*5.2589*	*-2.6192*	*6.7778*
Δεύτερη έλλειψη	*1.4107*	*0.1075*	*3.5857*	*3.3335*
Τρίτη έλλειψη	*-4.9959*	*7.2115*	*2.7075*	*7.9036*

Προσοχή πρέπει να δοθεί στις περιπτώσεις κατά τις οποίες ο αριθμός των παραμέτρων που πρέπει να υπολογίσουμε είναι πολύ μεγάλος. Σε αυτές τις περιπτώσεις το μήκος του χρωμοσώματος αυξάνει σημαντικά με συνέπεια η εύρεση της βέλτιστης λύσης να απαιτεί υπολογισμό μεγαλύτερου αριθμού γενιών με μεγαλύτερο αριθμό επίσης χρωμοσωμάτων ανά γενιά.

Σε αυτές τις περιπτώσεις όμως αυξάνει σημαντικά η υπολογιστική πολυπλοκότητα.

Χαρακτηριστική κατηγορία προβλημάτων εμφάνισης πολύ μεγάλου μήκους χρωμοσώματος είναι η εκπαίδευση νευρωνικών δικτύων με γενετικούς αλγόριθμους ιδιαίτερα σε συστήματα ταξινόμησης προτύπων για τα οποία η δομή των νευρωνικών δικτύων είναι γενικά πολύπλοκη.

Ας δούμε ένα χαρακτηριστικό παράδειγμα από την εκπαίδευση νευρωνικών δικτύων τα οποία χρησιμοποιούνται στον εντοπισμό θέσης προσώπου σε φωτογραφίες. Εστω ότι χρησιμοποιούμε νευρωνικό δίκτυο δύο επιπέδων, πλήρως διασυνδεδεμένο, το οποίο αποτελείται από νευρώνες τύπου perceptron για τους οποίους γνωρίζουμε ότι κάθε σύναψη διαθέτει έναν πολλαπλασιαστικό όρο κέρδους για την πληροφορία που περνά μέσω της σύναψης. Ο αριθμός των νευρώνων του πρώτου επιπέδου είναι 30 ενώ του δύτερου επιπέδου είναι δύο νευρώνες. Η ύπαρξη προσώπου ανιχνεύεται όταν στις συνάψεις των νευρώνων του πρώτου επιπέδου θέσουμε την αριθμητική τιμή των ψηφιοποιημένων τιμών της εικόνας ενός πλαισίου το οποίο έχει διαστάσεις 30x30.

Ποιο είναι το μήκος του χρωμοσώματος κάθε λύσης το οποίο πρέπει να κατασκευάσουμε αν θέλουμε να πραγματοποιήσουμε εκπαίδευση (εύρεση των συντελεστών βαρύτητας των συνάψεων) του συστήματος ανίχνευσης προσώπων με γενετικό αλγόριθμο, υποθέτοντας ακρίβεια λύσης 6 δεκαδικών ψηφίων και πεδίο αναζήτησης λύσης το διάστημα [-5,5];

Ακολουθώντας την γνωστή μεθοδολογία έχουμε:

Κάθε σύναψη κωδικοποιείται με 24 bits ($((5-(-5))*10^6 = 10^7 \rightarrow 2^{23} < 10^7 < 2^{24}$).

Συνολικός αριθμός συνάψεων πρώτου επιπέδου νευρώνων: 30*30*30 = 27000.

Συνολικός αριθμός συνάψεων δεύτερου επιπέδου νευρώνων: 2*30 = 60.

Συνολικό μήκος χρωμοσώματος σε bits: (27000+60)*24 = 649440.

Το μήκος του χρωμοσώματος που προκύπτει είναι υπερβολικά μεγάλο με συνέπεια την σημαντική επιβράδυνση της διαδικασίας εκπαίδευσης. Πρακτικά η εκπαίδευση ενός τέτοιου δικτύου από παραδείγματα απαιτεί ισχυρό υπολογιστή και είναι σαφώς βραδύτερος από τους αλγόριθμους εκπαίδευσης που βασίζονται σε πληροφορίες των πρώτων παραγώγων. Αυτός είναι ο σημαντικότερος λόγος για τον οποίο δεν θα συναντήσουμε συχνά στην επιστημονική βιβλιογραφία μεθόδους εκπαίδευσης νευρωνικών δικτύων με γενετικούς αλγόριθμους.

Τεχνικές βελτίωσης της απόδοσης γενετικών αλγορίθμων

Σε αυτό το τμήμα των σημειώσεων θα περιγράψουμε μεθόδους με τις οποίες μπορούμε να λύσουμε προβλήματα υλοποίησης γενετικών αλγόριθμων και βελτίωσης της απόδοσής των. Πρέπει να επισημάνουμε ότι πολύ μεγάλος αριθμός παραλλαγών του βασικού αλγόριθμου έχουν προταθεί προσαρμόζοντας τη τεχνική βελτιστοποίησης σε εξειδικευμένα προβλήματα, όπως τον αυτόματο ορισμό συναρτήσεων, αυτόματη κατασκευή λογισμικού, επίλυση προβλημάτων βελτιστοποίησης κ.ο.κ.

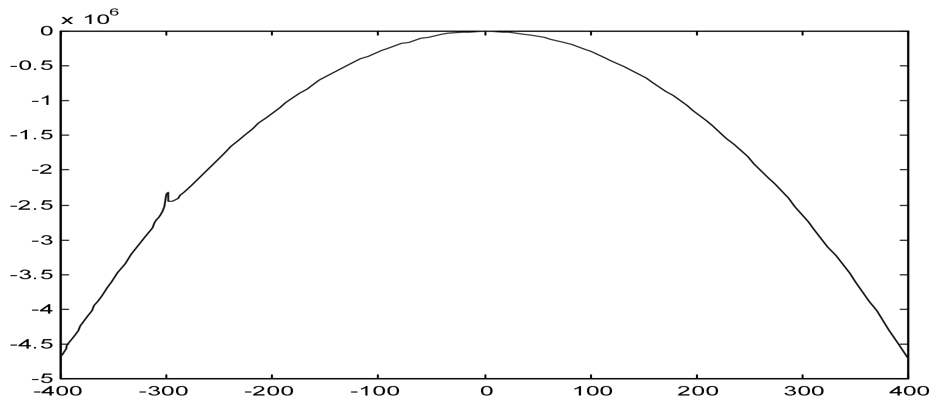
Ο χώρος αναζήτησης λύσεων

Ένα σοβαρό εμπόδιο υλοποίησης γενετικών αλγόριθμων εμφανίζεται όταν δεν έχουμε σαφείς περιορισμούς για τον χώρο αναζήτησης κάθε μιάς παραμέτρου που θέλουμε να υπολογίσουμε με κάποιο βέλτιστο κριτήριο.

Δείτε την περίεργη συμπεριφορά της συνάρτησης

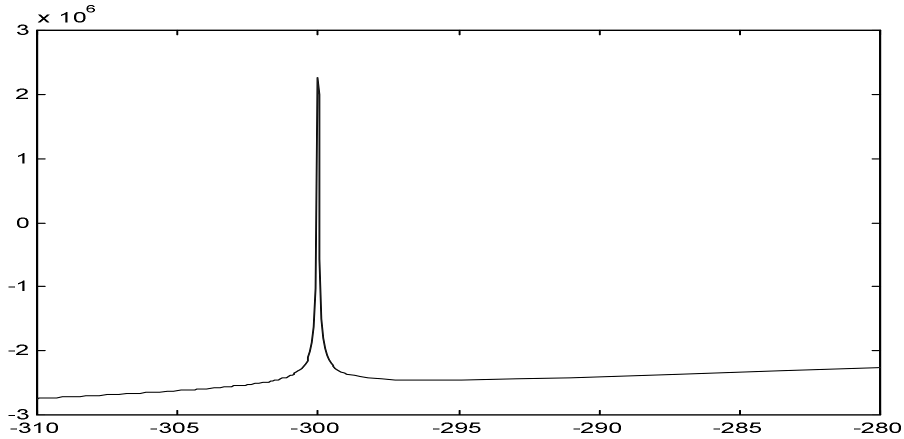
$$f(\mathbf{x}) = (-30\mathrm{x}^2 + 20\,\mathrm{x} - 1200)\ \cos\left(\frac{1}{\log(\sqrt{|\,\mathrm{x}+300\,|}) + 2}\right) + \log(|\,\mathrm{x}\,|)$$

Αν κάνουμε μία γραφική παράσταση στο διάστημα [-400 400] θα πάρουμε την γραφική παράσταση του σχήματος που ακολουθεί.



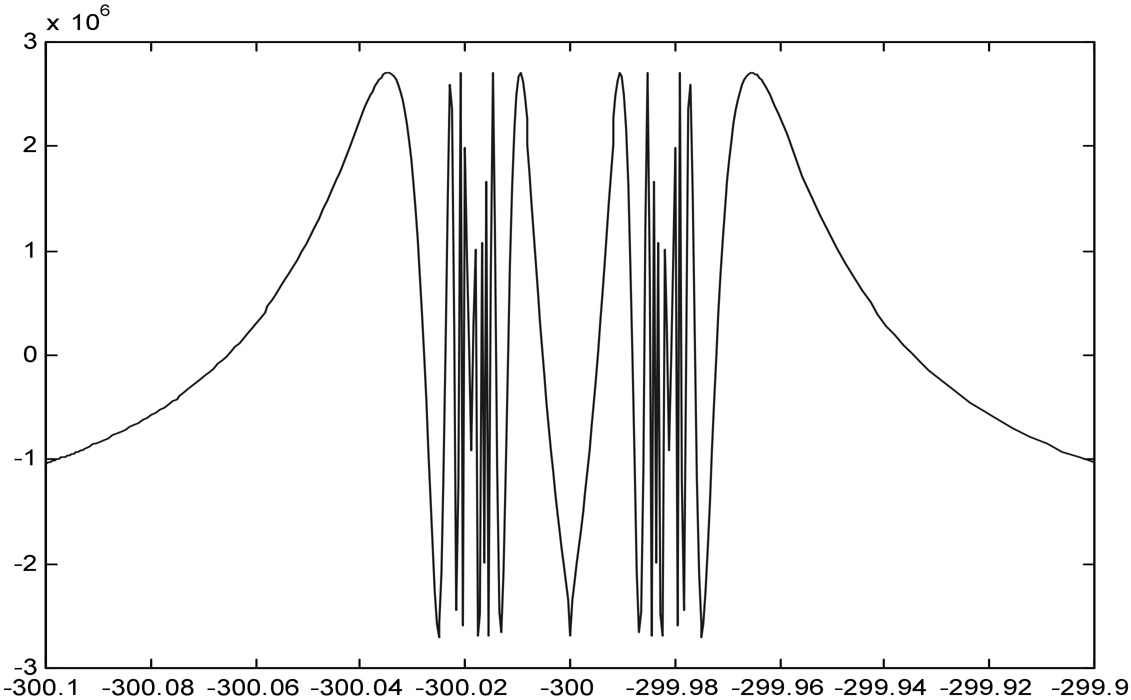
Φαίνεται ότι το μέγιστο της συνάρτησης εμφανίζεται κοντά στο 0 οπότε μπορούμε να αναζητήσουμε λύση στην περιοχή του 0. Μία μικρή περίεργη συμπεριφορά της συνάρτησης έχουμε στην περιοχή του -300 αλλά μακροσκοπικά δεν φαίνεται να υπάρχει ιδιαίτερο πρόβλημα.

Αν κάνουμε μία γραφική παράσταση της συνάρτησης κοντά στην περιοχή του -300 θα δούμε όμως ότι στο σημείο αυτό, μικροσκοπικά, εμφανίζεται ένα μέγιστο το οποίο έχει μεγαλύτερη αριθμητική τιμή από το μέγιστο στην περιοχή του μηδενός. Παρατηρείστε το σχήμα που ακολουθεί που δίνει την γραφική παράσταση της συνάστησης f(x) στο διάστημα [-310, -280].



Από το παράδειγμα αυτό βλέπουμε ότι πολλές φορές, ακόμα και όταν το πρόβλημα είναι σχετικά απλό όπως η εύρεση του μέγιστου συνάρτησης μιάς μόνο μεταβλητής, είναι δύσκολο να ορίσουμε με ακρίβεια την περιοχή στην οποία βρίσκεται το ολικό ακρότατο που θέλουμε να υπολογίσουμε.

Δείτε τι θα συμβεί όταν κάνουμε μία ακόμα λεπτομερέστερη γραφική παράσταση της συνάρτησης f(x) κοντά στην περιοχή του -300. Όπως θα παρατηρήσετε και στο επόμενο σχήμα τα ακρότατα σε αυτή την περιοχή δεν είναι ένα αλλά δεκάδες!.



Για να ξεπεράσουμε αυτό το εμπόδιο κωδικοποιούμε την μεταβλητή x με δύο δυαδικούς αριθμούς έχοντας διαφορετικό μήκος για την εκθετική παράσταση κάθε λύσης. Η κωδικοποίηση αυτή είναι γνωστή και σαν παράσταση δεκαδικού αριθμού με μορφή κινητής υποδιαστολής. Θεωρώντας δηλαδή ότι κάθε αριθμητική τιμή του x εκφράζεται σαν $\alpha*10^{\beta}$ όπου α και β είναι μεταβλητές οι οποίες μπορούν εύκολα να περιοριστούν σε διάστημα τιμών τέτοιο ώστε το αντίστοιχο όριο αναζήτησης της μεταβλητής x να είναι πολύ μεγάλο.

Ας επικεντώσουμε την προσοχή μας στο παράδειγμα με την περίεργη συνάρτηση f(x) η οποία έχει τοπικά μέγιστα στην περιοχή του μηδενός και πολλαπλά τοπικά μέγιατα και ελάχιστα στην περιοχή του x=-300.

Αρχικά ας επιχειρήσουμε να λύσουμε το πρόβλημα χρησιμοποιώντας δυαδική κωδικοποίηση της λύσης. Ορίζω διάστημα αναζήτησης λύσης το [-500, 500] με ακρίχεια 6 δεκαδικών ψηφίων παρατηρώντας ότι στην περιοχή του x=-300 υπάρχει συσσωρευμένος μεγάλος αριθμός ακρότατων σε πολύ μικρό διάστημα τιμών. Το συνολικό μήκος του χρωμοσώματος είναι 30 bits:

$$(500-(-500))*10^6 = 10^9 \rightarrow 2^{29} < 10^9 < 2^{30}$$

Ας δούμε τι μήκος χρωμοσώματος χρειαζόμαστε για την ίδια ακρίβεια κωδικοποίησης σε εκθετική μορφή. Το x εκφράζεται σαν $\alpha*10^{\beta}$ όπου α πραγματικός αριθμός στο διάστημα [-1, 1] και β ακέραιος αριθμός . Επειδή χρειαζόμαστε μεγάλη ακρίβεια υπολογισμών στην περιοχή του -300 τροποποιούμε λίγο την εκθετική μορφή αντιστοίχισης θεωρώντας ότι

$$\mathbf{x} = \alpha*10^{\beta} - 300.$$

Με αυτή την αντιστοίχιση μπορώ να αυξήσω σημαντικά την ακρίβεια των υπολογισμών που εκτελώ μειώνοντας το πεδίο λύσεων της μεταβλητής β.

Ας δούμε λοιπόν το συγκεκριμένο παράδειγμα. Όπως και στην προηγούμενη προσέγγιση θέλω το διάστημα αναζήτησης λύσης για την μεταβλητή x το [-500, 500] με ακρίβεια 6 δεκαδικών ψηφίων. Δυστυχώς στην περίπτωση της δυαδικής κωδικοποίησης των μεταβλητών της εκθετικής αναπαράστασης του x δεν μπορώ να ικανοποιήσω της δεύτερη απαίτηση σε όλο το πεδίο αναζήτησης της λύσης διότι γνωρίζω ότι ο διαμερισμός του χώρου είναι ανομοιόμορφος. Στο συγκεκριμένο πρόβλημα όμως που έχουμε να λύσουμε γνωρίζουμε ότι αυτά είναι συγκεντρωμένα σε περιοχή όπου ο εκθέτης β παίρνει μικρή τιμή. Συνεπώς ας θέσουμε τα εξής όρια

Η περιοχή αναζήτησης της μεταβλητής α να είναι το [-1, 1] με ακρίβεια έξι δεκαδικών ψηφίων και η περιοχή αναζήτησης της μεταβλητής β να είναι το διάστημα [0,3] με ακρίβεια ακέραιου αριθμού. Το μήκος του χρωμοσώματος είναι το άθροισμα του μήκους κωδικοποίησης κάθε μιάς των παραμέτρων:

Το μήκος δυαδικής κωδικοποίησης της μεταβλητής α είναι 21 bits:

$$(1-(-1))*10^6 = 2 * 10^6 \rightarrow 2^{19} < 2 * 10^6 < 2^{21}$$

Το μήκος δυαδικής κωδικοποίησης της μεταβλητής β είναι 2 bits:

Συνολικό μήκος χρωμοσώματος = (Μήκος α) + (Μήκος β) = 23 bits.

Βλέπουμε λοιπόν ότι στην ομοιόμορφη δυαδική κωδικοποίηση του χώρου αναζήτησης της λύσης χρειάζεται χρωμόσωμα μήκους 30 bits, στην περίπτωση κωδικοποίησης με εκθετική αναπαράσταση της λύσης για το ίδιο διάστημα αναζήτησης της λύσης χρειάζομαι 23 bits.

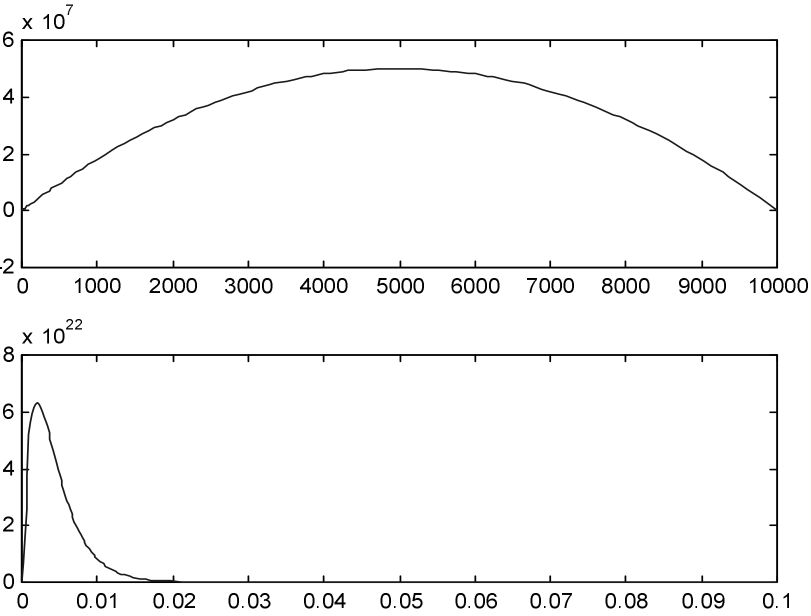
Με αυτήν την τεχνική πετυχαίνω υψηλή ακρίβεια αναζήτησης λύσης σε επιλεγμένες περιοχές. Προσοχή θα πρέπει να δοθεί στο γεγονός ότι οι τεχνικές κωδικοποίησης των μεταβλητών με αριθμούς κινητής υποδιαστολής αφαιρεί διακριτική ικανότητα σε απομεμακρυσμένες του 0 περιοχές.

Ασκηση. Υπολογείστε την τιμή του x για την οποία η συναρτηση f(x)

$$f(\mathbf{x}) = (-\mathbf{x}^2 + 9999\mathbf{x})\mathbf{e}^{\frac{5}{|x-0.001|+0.1}+1}$$

παίρνει την μέγιστη τιμή της στο πεδίο του $\Re^+$ με ακρίβεια επτά δεκαδικών ψηφίων. Χρησιμποποιείστε γενετικό αλγόριθμο που κωδικοποιεί την λύση με αριθμό κινητής υποδιαστολής.

Λάβετε επίσης υπόψιν τις δύο γραφικές παραστάσεις της συνάρτησης που ακολουθούν. Η πρώτη δίνει την μορφή της συνάρτησης στο πεδίο [0,10000], ενώ η δεύτερη δίνει την μορφή της συνάρτησης στο πεδίο [0,0.1].



Το πεδίο τιμών της συνάρτησης επιβίωσης

Ένα από τα πλεόν σοβαρά προβλήματα που επηρεάζουν σημαντικά την συμπεριφορά του γενετικού αλγόριθμου είναι το πρόβλημα του πεδίου τιμών της συνάρτησης επιβίωσης.

Γνωρίζουμε από τους περιορισμούς που θέσαμε ότι η συνάρτηση επιβίωσης πρέπει να έχει πάντα θετική τιμή. Ο περιορισμός αυτός δεν είναι πάντα εφικτός ιδιαίτερα στις περιπτώσεις κατά τις οποίες η μορφή της συνάρτησης ταξινόμησης είναι ιδιαίτερα πολύπλοκη και ο αριθμός των μεταβλητών πολύ μεγάλος.

Ακόμα και αν μπορέσουμε να βρούμε εκείνη την συνάρτηση επιβίωσης η οποία έχει πεδίο τιμών το σύνολο των θετικών πραγματικών αριθμών και επιλύει το πρόβλημα βελτιστοποίησης που έχουμε, μπορούμε εύκολα να διαπιστώσουμε ότι όταν οι αριθμητικέ τιμές της συνάρτησης επιβίωσης είναι πολύ μεγάλες αριθμητικά δεν μπορούμε να έχουμε *επιλεκτική* διάδοση των λύσεων στην επόμενη γενιά. Ένα μικρό παράδειγμα θα μας βοηθήσει να καταλάβουμε αυτό το πρόβλημα.

Εστω ότι η αριθμητική τιμή της συνάρτησης επιβίωσης για τρία χρωμοσώματα είναι

$$v_1 = 2234.1, \quad v_2 = 2235.8, \quad v_3 = 2217.9$$

Η σχέση που μας δίνει την πιθανότητα επιβίωσης της λύσης στην επόμενη γενιά δίνεται από τις ακόλουθες εξισώσεις:

$$p_1 = 2234.1/(2234.1+2235.8+2217.9) = 2234.1/6687.8 = 0.33405$$

$$p_2 = 2235.8/(2234.1+2235.8+2217.9) = 2235.8/6687.8 = 0.33431$$

$$p_3 = 2217.9/(2234.1+2235.8+2217.9) =2217.9/6687.8=0.33163$$

Τα αποτελέσματα αυτής της προσσέγισης δεν φαίνονται να είναι πολύ καλά διότι η πιθανότητα επιβίωσης και για τις τρείς λύσεις είναι μεν θεωρητικά διαφορετική, αλλά στην πράξη πρέπει να υπολογιστούν πολλές γενιές χρωμοσωμάτων μέχρι η λύση η οποία έχει μία μικρή πιθανοτική υπεροχή απέναντι των άλλων να υπερισχύσει.

Ας κάνουμε ένα τέχνασμα αφαιρόντας και από την αριθμητική τιμή της συνάρτησης επιβίωσης των τριών χρωμοσωμάτων το 2217. Οι νέες τιμές της συνάρτησης επιβίωσης για τα τρία χρωμοσώματα θα είναι

$$v_1 = 17.1, \quad v_2 = 18.8, \quad v_3 = 0.9$$

Οι νέες σχέσεις που μας δίνουν την πιθανότητα επιβίωσης κάθε λύσης στην επόμενη γενιά είναι:

$$p_1 = 17.1/(17.1+18.8+0.9) =17.1/36.8=0.46467$$

$$p_2 = 18.8/(17.1+18.8+0.9) =18.8/36.8=0.51086$$

$$p_3 = 0.9/(17.1+18.8+0.9) =0.9/36.8=0.02445$$

Το πρακτικό αποτέλεσμα των νέων πιθανοτήτων που πήραμε μας λέει ότι το τελευταίο χρωμόσωμα έχει πολύ μικρή πιθανότητα να επιβιώσει στην νέα γενιά.

Το τέχνασμα που κάναμε μας βοηθά να ξεπεράσουμε και το δυσεπίλυτο στην πράξη πρόβλημα του πεδίου τιμών της συνάρτησης επιβίωσης. Αν σε κάθε γενιά υπολογίζουμε την τιμή της συνάρτησης επιβίωσης σε κάθε διαφορετικό χρωμόσωμα και στην συνέχεια προσθέτουμε έναν αριθμό ό οποίος θα φέρει όλες τις αριθμητικές τιμές των συναρτήσεων επιβίωσης στο πεδίο των θετικών πραγματικών αριθμών τότε δεν χρειάζεται ο αρχικός περιορισμός που τέθηκε κατά τον ορισμό των συναρτήσεων απόφασης.

Ασκηση. Τροποποιείστε την γενική μορφή του γενετικού αλγόριθμου έτσι ώστε η πιθανότητα επιβίωσης του χρωμοσώματος για το οποίο η συνάρτηση επιβίωσης έχει την μικρότερη αριθμητική τιμή να έχει πάντα τιμή 0.1.

Ασκηση 1.

Μας δίνονται μετρήσεις που αναφέρονται στις θέσεις καρτεσιανών συντεταγμένων που λαμβάνει βλήμα πυροβόλου όπλου το οποίο βρίσκεται στην θέση (0,0). Ο πλανήτης στον οποίο δοκιμάζεται το πυροβόλο έχει ατμόσφαιρα που παρουσιάζει συμπεριφορά πολύ διαφορετική από την γήινη. Το υλικό της ατμόσφαιρας είναι ισότροπο και η συμπεριφορά του δεν αλλάζει με τον χρόνο. Οι μετρήσεις που πήραμε ανά δευτερόλεπτο σε μέτρα είναι οι ακόλουθες:

Οριζόντια απόσταση	15.2	28.3	42.1	53.4	63.9	72.2	80.1	87.4	94.2
κατακόρυφη απόσταση	0	40.5	78.6	97.8	100.5	97.4	88.3	70.3	55.6

Υπάρχει διχογνωμία για την συμπεριφορά της ατμόσφαιρας του πλανήτη. Το αποτέλεσμα είναι να υπάρχουν δύο θεωρίες. Η πρώτη υποθέτει ότι η θέση του βλήματος κατά τον οριζόντιο και τον κατακόρυφο άξονα δίνεται από τις εξισώσεις

$$\mathbf{x}(t) = a * t + b * t^2$$

$$\mathbf{y}(t) = c + \mathbf{d} * t + \mathbf{e}^2$$

ενώ η δεύτερη θεωρία ισχυρίζεται ότι

$$\mathbf{x}(t) = a * t + b * t^{1/2}$$

$$\mathbf{y}(t) = c + \mathbf{d} * t^{1/2} + \mathbf{e}^2$$

Μπορείτε να αποφασίσετε με την βοήθεια των πειραματικών δεδομένων για την ορθότητα της πρώτης ή της δεύτερης θεωρίας υπολογίζοντας ταυτόχρονα και την αριθμητική τιμή των σταθερών παραμέτρων a,b,c,d,e;

Ασκηση 2.

Κατασκευάζετε σύστημα ταξινόμησης προτύπων δύο κατηγοριών και έχετε να αποφασίσετε ανάμεσα σε δύο συστήματα ταξινόμησης. Το πρώτο σύστημα ταξινομεί το πρότυπο το οποίο αποτελείται από μετρήσεις δύο χαρακτηριστικών γνωρισμάτων κάθε αντικειμένου, έστω x_1 και x_2 ως εξής:

Υπολογίζουμε την τιμή εξόδου της συνάρτησης

$$\mathbf{R}(\mathbf{x}) = \sqrt{\frac{(\mathbf{x}_1 - \mathbf{w}_1)^2(\mathbf{x}_2 - \mathbf{w}_2)^2}{(\mathbf{x}_1 - \mathbf{w}_2)^2(\mathbf{x}_2 - \mathbf{w}_1)^2 + 1}}$$

Αν η αριθμητική τιμή της εξόδου είναι μεγαλύτερη από 0.5 το πρότυπο ταξινομείται στην πρώτη κατηγορία, διαφορετικά ταξινομείται στην δεύτερη.

Το δεύτερο σύστημα ταξινόμησης ακολουθεί διαφορετική συλλογιστική. Υπολογίζουμε την τιμή εξόδου δύο συναρτήσεων

$$\mathbf{R}_1(\mathbf{x}) = \frac{2}{(\mathbf{x}_1 - \mathbf{w}_1)^2 + (\mathbf{x}_2 - \mathbf{w}_2)^2 + 1}$$

$$\mathbf{R}_2(\mathbf{x}) = \tanh((\mathbf{x}_1 - \mathbf{w}_1)(\mathbf{x}_2 - \mathbf{w}_2)) + 1$$

Διαθέτοντας τα ακόλουθα παραδείγματα προτύπων για την πρώτη και την δεύτερη κατηγορία προτύπων μπορείτε να εκτιμήσετε την αριθμητική τιμή των παραμέτρων ($\mathbf{w}_1, \mathbf{w}_2$) και να επιλέξετε εκείνο το σύστημα ταξινόμησης το οποίο παρουσιάζει το μικρότερο σφάλμα;

Ω_1	(0.5,1)	(0.5,2)	(2,4)	(2.5,2.5)	(3,2)	(3.5,1)	(4,0)	(3,3)	(2,3)
Ω_2	(1,0)	(2,2)	(2.2,2)	(2.5,2)	(3,1)	(3.5,0)	(5,0)	(2,0)	(1,0)

Ασκηση 3.

Σας δίνεται μία ασπρόμαυρη φωτογραφία η οποία έχει ψηφιοποιηθεί με ακρίβεια 8 bits. Κάθε σημείο της φωτογραφίας χαρακτηρίζεται από έναν ακέραιο αριθμό στο διάστημα [0,255]. Ο αριθμός 0 αντιστοιχεί στο μαύρο και το 255 αντιστοιχεί σε χρώμα λευκό. Οι ενδιάμεσοι αριθμοί αντιστοιχούν σε αποχρώσεις του γκρί. Η ψηφιοποιημένη φωτογραφία έχει διαστάσεις 512x512 σημεία. Γνωρίζετε ότι στην φωτογραφία υπάρχει ένα φωτεινό σημείο το οποίο έχει διαστάσεις 20x20 σημεία και η φωτινότητά του ακολουθεί την μορφή της συνάρτησης κανονικής κατανομής:

$$F(\mathbf{x}, a, b, c) = a\mathbf{e}^{b(x_1 - m_1)^2}\mathbf{e}^{c(x_2 - m_2)^2} + \mathbf{d}$$

Οπου x_1, x_2 είναι οι συντεταγμένες της θέσης m_1, m_2 είναι το κέντρο του φωτεινού σημείου και a,b,c,d είναι άγνωστες σταθερές παράμετροι. Κατασκευάσετε τον αλγόριθμο με τον οποίο θα εντοπίσετε το κέντρο του φωτεινού σημείου το οποίο προσεγγίζει με την μεγαλύτερη ακρίβεια το μοντέλο του φωτεινού σημείου όπως περιγράφεται από την εξίσωση $F(\mathbf{x}, a, b, c, \mathbf{d})$.

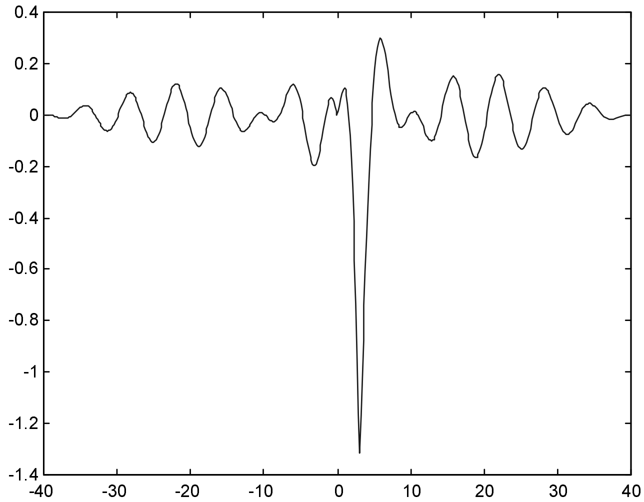
Ασκηση 4.

Η γραφική παράσταση της συνάρτησης

$$f(\mathbf{x}) = \frac{\log(|\mathbf{x}|+1)\sin(\sqrt{\mathbf{x}})\cos(\mathbf{x})}{|\mathbf{x}-3|+1}$$

στο διάστημα [-40,40] δίνεται στο σχήμα που ακολουθεί.

Υπολογίστε την αριθμητική τιμή του ελάχιστου της συνάρτησης $f(\mathbf{x})$ με την βοήθεια των



αλγορίθμων που διδαχθήκατε στο παρόν κεφάλαιο. Ποιος αλγόριθμος νομίζετε ότι θα έδινε τα καλύτερα αποτελέσματα, τόσο από την πλευρά ακρίβειας εύρεσης του ελάχιστου (αποφυγή τοπικών ακρότατων) αλλά και ταχύτητας σύγκλισης;

Υπάρχει αλγόριθμος ο οποίος να ικανοποιεί και οι δύο απαιτήσεις ταυτόχρονα;

Ασκηση 5.

Τροποποιείστε την γενική μορφή του γενετικού αλγόριθμου έτσι ώστε η πιθανότητα επιβίωσης του χρωμοσώματος για το οποίο η συνάρτηση επιβίωσης έχει την μικρότερη αριθμητική τιμή να έχει πάντα τιμή 0.1 και η πιθανότητα επιβίωσης του χρωμοσώματος για το οποίο η συνάρτηση επιβίωσης έχει την μεγαλύτερη αριθμητική τιμή να έχει πάντα τιμή 0.5.

Τέλος Ενότητας

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Σημείωμα Αναφοράς