Αναγνώριση Προτύπων Ι

Ενότητα 3: Στοχαστικά Συστήματα

Αν. Καθηγητής Δερματάς Ευάγγελος

Τμήμα Ηλεκτρολόγων Μηχανικών και Τεχνολογίας Υπολογιστών

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Άδειες Χρήσης

- Το παρόν εκπαιδευτικό υλικό υπόκειται σε άδειες χρήσης Creative Commons.
- Για εκπαιδευτικό υλικό, όπως εικόνες, που υπόκειται σε άλλου τύπου άδειας χρήσης, η άδεια χρήσης αναφέρεται ρητώς.

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. Εισαγωγή
2. Στοχαστικά Κριτήρια Ταξιμόμησης
3. Σχέση Στοχαστικών & Δομικών Κριτηρίων Ταξινόμησης
4. Υπολογισμός της Συνάρτησης Πυκνότητας Πιθανότητας
5. Υπολογισμός των Παραμέτρων της Πυκνότητας Πιθανότητας
6. Επιλογή της Συνάρτησης Πυκνότητας Πιθανότητας
7. Στοχαστκός Υπολογισμός της Πυκνότητας Πιθανότητας

Εισαγωγή (1)

- Τα **στοχαστικά συστήματα** ταξινόμησης προτύπων, μπορούν να αντιμετωπίσουν, με βέλτιστο τρόπο, το πρόβλημα της ταξινόμησης προτύπων. Μπορούν, πιο αναλυτικά, να ελαχιστοποιήσουν την πιθανότητα του σφάλματος στην ταξινόμηση, να αντιμετωπίσουν προβλήματα παρεμβολής θορύβου στις μετρήσεις των προτύπων και επιπλέον, μπορούν να εκμεταλλευτούν δυο επιπρόσθετες πληροφορίες:

1. Την συχνότητα εμφάνισης των προτύπων κάποιας κατηγορίας. Η πληροφορία αυτή είναι σημαντική διότι έτσι μπορεί να κατασκευαστεί στοχαστικό σύστημα ταξινόμησης προτύπων, το οποίο θα αναγνωρίζει με μικρότερη πιθανότητα σφάλματος τα πρότυπα των συχνότερα εμφανιζόμενων κατηγοριών, με αποτέλεσμα η συνολική πιθανότητα σφάλματος να ελαττώνεται.

Εισαγωγή (2)

2. <u>Το γεγονός ότι τα στοχαστικά συστήματα ταξινόμησης μπορούν να ενσωματώσουν την έννοια του κόστους ή της επικινδυνότητας (risk) για την απόφαση που λαμβάνεται</u>. Σε πρακτικές εφαρμογές, το είδος του σφάλματος που μπορεί να συμβεί με την λήψη της απόφασης από το σύστημα ταξινόμησης, έχει διαφορετική σημασία και συνεπώς διαφορετική επικινδυνότητα. Χαρακτηριστικό παράδειγμα διαφορετικής επικινδυνότητας που μπορεί να έχουν τα λάθη της ταξινόμησης είναι οι εφαρμογές επιβεβαίωσης ταυτότητας σε συστήματα ασφαλείας.

➢ Τα παραπάνω χαρακτηριστικά προσδίδουν στα συστήματα ταξινόμησης με στοχαστικές μεθόδους, υψηλή αξιοπιστία, ιδιαίτερα στις εφαρμογές, όπου παρατηρείται σημαντική διαφορά τόσο στην συχνότητα εμφάνισης των προτύπων των κατηγοριών, όσο και στην επικινδυνότητα των σφαλμάτων του συστήματος ταξινόμησης.

Παράδειγμα 1ο

Στοχαστικά Κριτήρια Ταξιμόμησης (1)

1. <u>Το κριτήριο μεγαλύτερης πιθανότητας</u>

Η πιο απλή στοχαστική μέθοδος ταξινόμησης, πραγματοποιείται με την εύρεση της κατηγορίας εκείνης, η οποία έχει και την μεγαλύτερη πιθανότητα να συμβεί, δεδομένης της διαθέσιμης πληροφορίας, δηλαδή της μέτρησης του προτύπου **x**. Το αντίστοιχο κριτήριο ταξινόμησης δίνεται από την

$$\mathrm{x} \in \omega_j \quad \text{ανν} \quad p(\omega_j|\mathrm{x}) \rangle p(\omega_i|\mathrm{x}) \quad , \quad \forall i \neq j$$

Η πιθανότητα $p(\omega_i|\mathrm{x})$ δεν μπορεί να υπολογιστεί, απευθείας, εύκολα. Από τα παραδείγματα μπορεί να υπολογιστεί, κατά προσέγγιση η πιθανότητα Με τον κανόνα του Bayes το κριτήριο της μεγαλύτερης πιθανότητας τροποποιείται ως εξής:

$$\mathrm{x} \in \omega_j \, \text{ανν} \, \frac{p(\mathrm{x}|\omega_j)p(\omega_j)}{p(\mathrm{x})} \rangle \frac{p(\mathrm{x}|\omega_i)p(\omega_i)}{p(\mathrm{x})}, \quad \forall i \neq j$$

$$\mathrm{x} \in \omega_j \quad \text{ανν} \quad p(\mathrm{x}|\omega_j)p(\omega_j) \rangle p(\mathrm{x}|\omega_i)\,p(\omega_i) \quad , \quad \forall i \neq j$$

<u>Παράδειγμα 2ο</u>

Στοχαστικά Κριτήρια Ταξιμόμησης (2)

2. Το κριτήριο του ελαχίστου κόστους

Παρακάτω, εισάγουμε την έννοια του κόστους μιας απόφασης σε στοχαστικά συστήματα, για την περίπτωση ταξινόμησης προτύπου σε δυο κατηγορίες αντικειμένων. Με την επαγωγική μέθοδο, εύκολα, γενικεύονται τα αποτελέσματα, σε πρόβλημα ταξινόμησης, όταν υπάρχουν Ν κατηγορίες αντικειμένων.

- Έστω ότι το σύστημα αναγνωρίζει δυο κατηγορίες προτύπων, τις $\boldsymbol{\omega_1}$ και $\boldsymbol{\omega_2}$.

- Ορίζουμε, παρακάτω, τέσσερα διαφορετικά μεγέθη, $c_{11}, c_{12}, c_{21}, c_{22}$ καθένα από τα οποία ονομάζεται κόστος λήψης απόφασης.

Στοχαστικά Κριτήρια Ταξιμόμησης (3)

- c_{11} είναι το κόστος που προκύπτει, όταν το σύστημα αποφασίσει ότι $X \in \omega_1$, ενώ στην πραγματικότητα $X \in \omega_1$

- c_{12} είναι το κόστος που προκύπτει, όταν το σύστημα αποφασίσει ότι $X \in \omega_2$, ενώ στην πραγματικότητα $X \in \omega_1$

- c_{21} είναι το κόστος που προκύπτει, όταν το σύστημα αποφασίσει ότι $X \in \omega_1$, ενώ στην πραγματικότητα $X \in \omega_2$

- c_{22} είναι το κόστος που προκύπτει, όταν το σύστημα αποφασίσει ότι $X \in \omega_2$, ενώ στην πραγματικότητα $X \in \omega_2$

όπου $c_{11}, c_{12}, c_{21}, c_{22}$ είναι μη αρνητικοί πραγματικοί αριθμοί.

Στοχαστικά Κριτήρια Ταξιμόμησης (4)

- Αν $\mathbf{\Omega_1}$ και $\mathbf{\Omega_2}$ είναι οι δυο περιοχές του πεδίου ορισμού του παραμετρικού διανύσματος των προτύπων, για τις οποίες έχει αποφασιστεί ότι, αν $\mathrm{x} \in \Omega_1 \Rightarrow \mathrm{x} \in \omega_1$ και $\mathrm{x} \in \Omega_2 \Rightarrow \mathrm{x} \in \omega_2$ τότε η βέλτιστη επιλογή των δυο περιοχών μπορεί α υπολογιστεί με κριτήριο την ελαχιστοποίηση του αναμενόμενου κόστους:

$$C = E[\text{κόστος}] =$$

$$\int_{\Omega_1} c_{11} p(\omega_1|\mathrm{x}) p(\mathrm{x}) d\mathrm{x} + \int_{\Omega_1} c_{21} p(\omega_2|\mathrm{x}) p(\mathrm{x}) d\mathrm{x} + \int_{\Omega_2} c_{12} p(\omega_1|\mathrm{x}) p(\mathrm{x}) d\mathrm{x} +$$

$$+ \int_{\Omega_2} c_{22} p(\omega_2|\mathrm{x}) p(\mathrm{x}) d\mathrm{x} = \int_{\Omega_1} \left(c_{11} p(\omega_1) p(\mathrm{x}|\omega_1) + c_{21} p(\omega_2) p(\mathrm{x}|\omega_2) d\mathrm{x} \right)$$

$$+ \int_{\Omega_2} \left(c_{12} p(\omega_1) p(\mathrm{x}|\omega_1) + c_{22} p(\omega_2) p(\mathrm{x}|\omega_2) d\mathrm{x} \right).$$

- Εφόσον οι περιοχές $\mathbf{\Omega_1}$ και $\mathbf{\Omega_2}$ δεν υπερκαλύπτονται και περιέχουν όλο τον χώρο των μετρήσεων, ισχύει το εξής:

$$\int_{\Omega_1} p(\mathrm{x}|\omega_i) d\mathrm{x} + \int_{\Omega_2} p(\mathrm{x}|\omega_i) = 1, i = 1,2$$

Στοχαστικά Κριτήρια Ταξιμόμησης (4)

- Αντικαθιστώντας την παραπάνω εξίσωση στην συνάρτηση του αναμενόμενου κόστους, έχουμε:

$$C = c_{12}\,p(\omega_1) + c_{22}\,p(\omega_2) +$$
$$+ \int_{\Omega_1} \Big(-(c_{12} - c_{11})\,p(\omega_1)\,p(\mathrm{x}|\omega_1) + (c_{21} - c_{22})\,p(\omega_2)\,p(\mathrm{x}|\omega_2) \Big) d\mathrm{x}$$

- Η λύση του προβλήματος δίνεται με την επιλογή της περιοχής Ω_1 για την οποία η ολοκληρωτική παράσταση (που αποτελείται από έναν θετικό και έναν αρνητικό όρο, αφού οι πολλαπλασιαστικοί όροι είναι όλοι θετικοί) είναι ελάχιστη.
- Αν ο αρνητικός όρος είναι μεγαλύτερος του θετικού για όλα τα πρότυπα που ανήκουν στην περιοχή Ω_1, τότε η ολοκληρωτική παράσταση έχει αρνητικές τιμές και το συνολικό ολοκλήρωμα δίνει την ελάχιστη αρνητική της τιμή.
- Η συνθήκη που πρέπει να ισχύει, έτσι ώστε να αποφασίζεται με το ελάχιστο κόστος ότι το άγνωστο πρότυπο x ανήκει στην περιοχή Ω_1 είναι:

$$(c_{12} - c_{11})\,p(\omega_1)\,p(\mathrm{x}|\omega_1) \;\rangle\; (c_{21} - c_{22})\,p(\omega_2)\,p(\mathrm{x}|\omega_2) \Rightarrow \mathrm{x} \in \Omega_1$$

[Παράδειγμα 3ο]

Σχέση Στοχαστικών & Δομικών Κριτηρίων Ταξινόμησης (1)

1. <u>Συναρτήσεις διάκρισης και στοχαστικά κριτήρια ταξινόμησης</u>

Στην περίπτωση του κριτηρίου μεγίστης πιθανότητας και του κριτηρίου του ελαχίστου κόστους, υπολογίζουμε μια αριθμητική ποσότητα για κάθε κατηγορία και στην συνέχεια, ταξινομούμε το άγνωστο πρότυπο στην κατηγορία που έχει την μεγαλύτερη αριθμητική τιμή, ότι, δηλαδή, γίνεται στα δομικά συστήματα ταξινόμησης όταν αυτά χρησιμοποιούν συναρτήσεις διάκρισης.

- Αν ορίσω σαν συνάρτηση διάκρισης την,

$$g_i(\mathrm{x}) = p(\mathrm{x}|\omega_i)p(\omega_i)$$

τότε το στοχαστικό κριτήριο ταξινόμησης της μέγιστης πιθανότητας, γίνεται ισοδύναμο με την διαδικασία ταξινόμησης ενός δομικού συστήματος.

- Με τον ίδιο τρόπο, αν ορίσω σαν συνάρτηση διάκρισης την,

$$g_i = c_i p(\omega_i)p(\mathrm{x}|\omega_i)$$

τότε το στοχαστικό κριτήριο ταξινόμησης του ελαχίστου κόστους, είναι ισοδύναμο με το κριτήριο ταξινόμησης ενός δομικού συστήματος με συναρτήσεις διάκρισης.

Σχέση Στοχαστικών & Δομικών Κριτηρίων Ταξινόμησης (2)

2. <u>Σχέση στοχαστικών και δομικών συστημάτων σύγκρισης προτύπων</u>

Μια από τις σημαντικότερες σχέσεις, ανάμεσα σε στοχαστικές και δομικές μεθόδους ταξινόμησης είναι εκείνη που συνδέει το κριτήριο ταξινόμησης της μικρότερης σταθμισμένης Ευκλείδειας απόστασης και το στοχαστικό κριτήριο ταξινόμησης στην κατηγορία με την μεγαλύτερη πιθανότητα, όταν η πυκνότητα πιθανότητας των προτύπων για κάθε κατηγορία ακολουθεί την κανονική κατανομή.

- Η πυκνότητα πιθανότητας των προτύπων, για πολλές κατηγορίες φυσικών αντικειμένων, μπορεί να προσεγγιστεί, με σημαντική ακρίβεια, με την πυκνότητα πιθανότητας της κανονικής κατανομής:

$$p(\mathrm{x}, y, s) = \frac{1}{(2\pi)^{N/2}\sqrt{|s|}} e^{-\frac{1}{2}(\mathrm{x}-y)^T s^{-1}(\mathrm{x}-y)}$$

Σχέση Στοχαστικών & Δομικών Κριτηρίων Ταξινόμησης (3)

- Στην περίπτωση κατά την οποία υποθέτουμε ότι κάθε μια από τις παραμέτρους του προτύπου είναι στοχαστικά ασυσχέτιστη από τις υπόλοιπες παραμέτρους, τότε η πυκνότητα πιθανότητας για την κανονική κατανομή απλοποιείται και γίνεται:

$$p(\mathrm{x}, y, s) = \frac{1}{(2\pi)^{N/2}\sqrt{|s|}} e^{-\frac{1}{2}\sum_{i=1}^{N}\frac{(\mathrm{x}_i - y_i)^2}{s_i}}$$

όπου y_i και s_i είναι η μέση τιμή και η διασπορά τις i παραμέτρου του διανύσματος των προτύπων.

- Ας υποθέσουμε ότι έχουμε δυο κατηγορίες ($\boldsymbol{\omega_1}$,$\boldsymbol{\omega_2}$), τα πρότυπα των οποίων ακολουθούν την απλοποιημένη κανονική κατανομή. Η ταξινόμηση κάθε προτύπου **x** μπορεί να πραγματοποιηθεί με ένα στοχαστικά βέλτιστο τρόπο, μετρώντας την πυκνότητα πιθανότητας κάθε μιας από τις κατηγορίες στην θέση **x** και ταξινομώντας το άγνωστο πρότυπο στην κατηγορία εκείνη στην οποία η πυκνότητα πιθανότητας μεγιστοποιείται.

Σχέση Στοχαστικών & Δομικών Κριτηρίων Ταξινόμησης (4)

- Για σύστημα ταξινόμησης δυο κατηγοριών, το πρότυπο **x** ταξινομείται στην κατηγορία $\boldsymbol{\omega_1}$, όταν ισχύει:

$$p(\mathrm{x}, y_1, s_1) \rangle p(\mathrm{x}, y_2, s_2) \rightarrow \mathrm{x} \in \omega_1 \Leftrightarrow$$

$$\frac{1}{(2\pi)^{N/2}\sqrt{|s|}} e^{-\frac{1}{2}\sum_{i=1}^{N}\frac{\left(\mathrm{x}_i - y_i^{(1)}\right)^2}{s_i^{(1)}}} \rangle \frac{1}{(2\pi)^{N/2}\sqrt{|s|}} e^{-\frac{1}{2}\sum_{i=1}^{N}\frac{\left(\mathrm{x}_i - y_i^{(2)}\right)^2}{s_i^{(2)}}} \rightarrow \mathrm{x} \in \omega_1$$

- Αν $|s_1| = |s_2|$, τότε η προηγούμενη ανισότητα απλοποιείται σε:

$$e^{-\frac{1}{2}\sum_{i=1}^{N}\frac{\left(\mathrm{x}_i - y_i^{(1)}\right)^2}{s_i^{(1)}}} \rangle e^{-\frac{1}{2}\sum_{i=1}^{N}\frac{\left(\mathrm{x}_i - y_i^{(2)}\right)^2}{s_i^{(2)}}} \rightarrow \mathrm{x} \in \omega_1 \Leftrightarrow$$

$$\sum_{i=1}^{N}\frac{\left(\mathrm{x}_i - y_i^{(1)}\right)^2}{s_i} \langle \sum_{i=1}^{N}\frac{\left(\mathrm{x}_i - y_i^{(2)}\right)^2}{s_i} \rightarrow \mathrm{x} \in \omega_1 \Leftrightarrow$$

$$d_\omega(\mathrm{x}, y_1, m) \langle d_\omega(\mathrm{x}, y_2, m) \rightarrow \mathrm{x} \in \omega_1$$

Σχέση Στοχαστικών & Δομικών Κριτηρίων Ταξινόμησης (5)

➢ Οι παραπάνω συλλογισμοί αποδεικνύουν ότι το στοχαστικό κριτήριο ταξινόμησης με την μεγαλύτερη πυκνότητα πιθανότητας είναι ισοδύναμο με το κριτήριο ταξινόμησης της μικρότερης σταθμισμένης Ευκλείδειας απόστασης, όταν για όλες τις κατηγορίες ισχύει:

$$|s_1| = |s_2|, \ \forall i, j.$$

Σχέση Στοχαστικών & Δομικών Κριτηρίων Ταξινόμησης (6)

3. <u>Σχέση μεταξύ του κριτηρίου ταξινόμησης των k-πλησιέστερων προτύπων και του κριτηρίου ταξινόμησης με την μεγαλύτερη πιθανότητα</u>

Ένας στατιστικός υπολογισμός της συνάρτησης πυκνότητας πιθανότητας των προτύπων των κατηγοριών είναι ο εξής:

$$P_j(|\mathrm{x}-y|\langle\varepsilon)=\int_{|\mathrm{x}-y|\langle\varepsilon} p_j(\mathrm{x})d\mathrm{x} \approx \frac{N_j(|\mathrm{x}-y|\langle\varepsilon)}{N_j}$$

όπου $N_j(|\mathrm{x}-y|\langle\varepsilon)$ είναι ο αριθμός των παραδειγμάτων της κατηγορίας $\boldsymbol{\omega}_j$, τα οποία ικανοποιούν την σχέση $|\mathrm{x}_i-y|\langle\varepsilon$ και **ε** είναι ένας μικρός θετικός πραγματικός αριθμός.

- Αν θεωρήσουμε ότι στην περιοχή που έχει οριστεί από τα σημεία **y** που ικανοποιούν την σχέση $|\mathrm{x}-y|\langle\varepsilon$, βρίσκονται $N_j(|\mathrm{x}-y|\langle\varepsilon)$ παραδείγματα για κάθε μια από τις **M** κατηγορίες, τότε το κριτήριο μεγίστης πιθανότητας γίνεται, όπως φαίνεται παρακάτω:

Σχέση Στοχαστικών & Δομικών Κριτηρίων Ταξινόμησης (7)

- $x \in \omega_j$ ανν $\quad p(x|\omega_j)p(\omega_j) > p(x|\omega_i)p(\omega_i), \ \forall i \neq j \Leftrightarrow$

$$x \in \omega_j \text{ ανν } \quad \frac{N_j(|x-y|<\varepsilon)}{N_j} p(\omega_j) > \frac{N_i(|x-y|<\varepsilon)}{N_i} p(\omega_i), \ \forall i \neq j \Leftrightarrow$$

$$x \in \omega_j \text{ ανν } \quad \frac{N_j(|x-y|<\varepsilon)}{N_j} \frac{N_j}{\sum_{m=1}^{M} N_m} > \frac{N_i(|x-y|<\varepsilon)}{N_i} \frac{N_i}{\sum_{m=1}^{M} N_m} \forall i \neq j \Leftrightarrow$$

$$x \in \omega_j \text{ ανν } \quad N_j(|x-y|<\varepsilon) > N_i(|x-y|<\varepsilon) \quad , \quad \forall i \neq j$$

- Η τελευταία σχέση μας δείχνει ότι το κριτήριο ταξινόμησης της μεγαλύτερης πιθανότητας είναι ισοδύναμο με το κριτήριο ταξινόμησης στην κατηγορία εκείνη, της οποίας τα περισσότερα παραδείγματα βρίσκονται την γειτονία του αγνώστου προτύπου, πρόκειται, δηλαδή, για το κριτήριο ταξινόμησης των k-πλησιέστερων προτύπων.

Υπολογισμός της Συνάρτησης Πυκνότητας Πιθανότητας (1)

- Στις περισσότερες πρακτικές εφαρμογές, δεν είναι γνωστός ο αναλυτικός τύπος της συνάρτησης πυκνότητας πιθανότητας. Συνήθως, αυτό που διατίθεται είναι ένας πεπερασμένος αριθμός παραδειγμάτων. Για το λόγο αυτό, υπάρχουν διάφορες μέθοδοι, οι σημαντικότερες από τις οποίες παρατίθενται στην συνέχεια, που βοηθούν στον υπολογισμό της πυκνότητας πιθανότητας.

1. <u>Μεγιστοποίηση της εντροπίας</u>

Η εντροπία (ποσοτικό μέτρο της αταξίας ενός συστήματος) της συνάρτησης πυκνότητας πιθανότητας δίνει την μέση τιμή της αβεβαιότητας ή ισοδύναμα της πληροφορίας που παίρνεται από την μέτρηση της διανυσματικής παράστασης του προτύπου **x**:

$$\text{Εντροπία} = -\int_{\mathrm{x}} p(\mathrm{x}) \ln\big(p(\mathrm{x})\big) d\mathrm{x}$$

Υπολογισμός της Συνάρτησης Πυκνότητας Πιθανότητας (2)

- Η συνάρτηση εκείνη που μεγιστοποιεί την εντροπία και ικανοποιεί ένα πλήθος περιορισμών που προκύπτουν από τα στατιστικά χαρακτηριστικά των παραδειγμάτων που είναι διαθέσιμα, αποτελεί μια από τις μεθόδους υπολογισμού της πυκνότητας πιθανότητας.

$$\int_{\mathrm{x}} p(\mathrm{x})d\mathrm{x} = 1$$

$$\int_{\mathrm{x}} \mathrm{R}_{\mathrm{i}}(\mathrm{x})p(\mathrm{x})d\mathrm{x} = \rho_{\mathrm{i}}, i = 1, I$$

- Η συνάρτηση *p*(x) που ικανοποιεί τους παραπάνω περιορισμούς και μεγιστοποιεί την εντροπία της, δηλαδή την πληροφορία που λαμβάνεται από τις μετρήσεις, είναι αυτή που αναζητείται.
- Πιο αναλυτικά, αναζητούνται τα ακρότατα μιας συνάρτησης (της εντροπίας), όταν αυτή ικανοποιεί ένα σύνολο περιορισμών. Μια γενική λύση του προβλήματος μπορεί να επιτευχθεί με την βοήθεια ενός θεωρήματος που απέδειξε ο Lagrange.

Υπολογισμός της Συνάρτησης Πυκνότητας Πιθανότητας (3)

➢ **<u>Θεώρημα:</u>** Έστω ότι θέλουμε να υπολογίσουμε τα ακρότατα της συνάρτησης $f(\mathrm{x})$ επάνω στην επιφάνεια που ορίζεται από ένα σύνολο περιορισμών:

$$q_1(\mathrm{x}) = q_2(\mathrm{x}) = \ldots = q_q(\mathrm{x}) = 0$$

Σε κάθε ακρότατο της συνάρτησης ισχύει ότι:

$$\frac{\partial}{\partial \mathrm{x}} f(\mathrm{x}) - \sum_{i=1}^{Q} \lambda_i \frac{\partial}{\partial \mathrm{x}} g_i(\mathrm{x}) = 0$$

➢ Οι συντελεστές λ_i ονομάζονται πολλαπλασιαστές Lagrange. Τα σημεία των ακρότατων της συνάρτησης βρίσκονται με την λύση των εξισώσεων των περιορισμών και της εξίσωσης των πολλαπλασιαστών Lagrange.

[Παράδειγμα 4ο](#)
[Παράδειγμα 5ο](#)

Υπολογισμός της Συνάρτησης Πυκνότητας Πιθανότητας (4)

2. <u>Προσέγγιση της πυκνότητας πιθανότητας με ορθοκανονικές μεθόδους</u>

- Έστω ότι μπορούμε να προσεγγίσουμε την πυκνότητα πιθανότητας των προτύπων κάποιας κατηγορίας με ένα σύνολο Q συναρτήσεων:

$$\hat{p}(\mathrm{x}) = \sum_{i=1}^{Q} c_i \varphi_i(\mathrm{x})$$

- Η ελαχιστοποίηση του σταθμισμένου σφάλματος της προσέγγισης είναι ένα καλό κριτήριο, με το οποίο βρίσκονται οι συντελεστές c_i,i=1,Q, έτσι ώστε η προσέγγιση να δίνει ικανοποιητική αξιοπιστία.

$$Er = \int_{\mathrm{x}} w(\mathrm{x})\left(p(\mathrm{x}) - \hat{p}(\mathrm{x}) \right)^2 d\mathrm{x}$$

Όπου *w*(x) είναι μια συνάρτηση που δίνει διαφορετική βαρύτητα στην προσέγγιση για τις διαφορετικές περιοχές του πεδίου τιμών των προτύπων.

- Ισοδύναμα η σχέση μπορεί να γίνει:

$$Er = \int_{\mathrm{x}} w(\mathrm{x})\left(p(\mathrm{x}) - \sum_{i=1}^{Q} c_i \varphi_i(\mathrm{x}) \right)^2 d\mathrm{x}$$

Υπολογισμός της Συνάρτησης Πυκνότητας Πιθανότητας (5)

- Οι συντελεστές c_i,i=1,Q προκύπτουν από ένα σύστημα Q γραμμικών εξισώσεων:

$$\frac{\partial}{\partial c_i} Er = 0, i = 1, Q$$

- Υπολογίζοντας τις μερικές παραγώγους, έχουμε:

$$\sum_{i=1}^{Q} c_i \int_{\mathrm{x}} w(\mathrm{x})\varphi_i(\mathrm{x})\varphi_j(\mathrm{x}) d\mathrm{x} = \int_{\mathrm{x}} w(\mathrm{x})\varphi_j(\mathrm{x}) p(\mathrm{x}) d\mathrm{x}$$

- Ο όρος των εξισώσεων που βρίσκεται στο δεξί τμήμα, παρατηρούμε ότι περιγράφει την μέση τιμή της παράστασης $w(\mathrm{x})\varphi_j(\mathrm{x})$ συνεπώς αυτός ο όρος μπορεί να προσεγγιστεί με στατιστικές μεθόδους. Από τα Μ παραδείγματα που διαθέτουμε υπολογίζουμε τις αριθμητικές τιμές των παραστάσεων $w(\mathrm{x})\varphi_j(\mathrm{x}_i), i = 1, M$ και στην συνέχεια, υποθέτουμε ότι ο αριθμητικός τους μέσος αποτελεί μια καλή προσέγγιση του αντίστοιχου στοχαστικού μέσου:

$$\int_{\mathrm{x}} w(\mathrm{x})\varphi_j(\mathrm{x}) p(\mathrm{x}) d\mathrm{x} \approx \frac{1}{\mathrm{M}} \sum_{i=1}^{M} w(\mathrm{x}_i)\varphi_j(\mathrm{x}_i)$$

Υπολογισμός της Συνάρτησης Πυκνότητας Πιθανότητας (6)

- Συνδυάζοντας τις δυο τελευταίες εξισώσεις, έχουμε:

$$\sum_{i=1}^{Q} c_i \int_{\mathrm{x}} w(\mathrm{x})\varphi_i(\mathrm{x})\varphi_j(\mathrm{x})d\mathrm{x} \approx \frac{1}{\mathrm{M}} \sum_{i=1}^{M} w(\mathrm{x}_i)\varphi_j(\mathrm{x}_i)$$

- Αν οι συναρτήσεις φ_i(x) έχουν εκλεγεί, έτσι ώστε να είναι ορθοκανονικές ως προς την συνάρτηση βαρύτητας *w*(x) , τότε ισχύει ότι:

$$\int_{\mathrm{x}} w(\mathrm{x})\varphi_i(\mathrm{x})\varphi_j(\mathrm{x})d\mathrm{x} = \begin{cases} 1, i = j \\ 0, i \neq j \end{cases}$$

- Λαμβάνοντας υπόψιν τα παραπάνω, ο τελικός υπολογισμός των συντελεστών γίνεται από τις σχέσεις:

$$c_i \approx \frac{1}{\mathrm{M}} \sum_{j=1}^{M} w(\mathrm{x}_j)\varphi_i(\mathrm{x}_j) \qquad \forall i = 1, M$$

Υπολογισμός της Συνάρτησης Πυκνότητας Πιθανότητας (7)

3. <u>Τα πλαίσια Parzen</u>

Με την μέθοδο των πλαισίων Parzen κάθε παράδειγμα, που αποτελεί από μόνο του μια κατηγορία αντικειμένων, διαθέτει μια πυκνότητα πιθανότητας η οποία ονομάζεται και συνάρτηση πυρήνα (kernel) και η οποία είναι ιδίου τύπου για όλα τα παραδείγματα εκπαίδευσης.

- Η συνολική πυκνότητα πιθανότητας της κατηγορίας, υπολογίζεται από την σχέση:

$$p(\mathrm{x}|\omega_i) \approx \frac{1}{M}\sum_{i=1}^{M} K(\mathrm{x} - \mathrm{x}_i)$$

- Η συνάρτηση πυρήνα πρέπει να ικανοποιεί έναν αριθμό περιορισμών:

$$1. \int_{-\infty}^{+\infty} K(\mathrm{x})d\mathrm{x} = 1 \qquad 2. \int_{-\infty}^{+\infty} |K(\mathrm{x})|d\mathrm{x} = +\infty \qquad 3. \int_{\mathrm{x}\to+\infty} |\mathrm{x}K(\mathrm{x})| = 0$$

Παράδειγμα 6ο

Υπολογισμός της Συνάρτησης Πυκνότητας Πιθανότητας (8)

- Τυπικές συναρτήσεις δίνονται στον πίνακα που ακολουθεί.

Ονομασία συνάρτησης	Πυρήνας
Ομοιόμορφος	$K_o(\mathrm{x}) = \begin{cases} \frac{1}{2h}, \left\lvert \frac{\mathrm{x}}{h} \right\rvert \leq 1 \\ 0, \left\lvert \frac{\mathrm{x}}{h} \right\rvert \rangle 1 \end{cases}$
Τριγωνικός	$K_t(\mathrm{x}) = \begin{cases} \frac{1}{h}\left(1 - \left\lvert \frac{\mathrm{x}}{h} \right\rvert\right), \left\lvert \frac{\mathrm{x}}{h} \right\rvert \leq 1 \\ 0, \left\lvert \frac{\mathrm{x}}{h} \right\rvert \rangle 1 \end{cases}$
Κανονικός	$K_n(\mathrm{x}) = \frac{1}{h\sqrt{2\pi}} e^{-\frac{\mathrm{x}^2}{2h^2}}$
Αντίστροφος	$K_i(\mathrm{x}) = \frac{1}{h\pi} \frac{1}{1+\left(\frac{\mathrm{x}}{h}\right)^2}$
Εκθετικός	$K_e(\mathrm{x}) = \frac{1}{2h} e^{-\left\lvert \frac{\mathrm{x}}{h} \right\rvert}$
Ημιτόνου	$K_s(\mathrm{x}) = \frac{1}{2h\pi} \left(\frac{\sin\left(\frac{\mathrm{x}}{2h}\right)}{\frac{\mathrm{x}}{2h}} \right)^2$

Υπολογισμός των Παραμέτρων της Πυκνότητας Πιθανότητας (1)

• Ορισμένες φορές, παρόλο που είμαστε σε θέση να γνωρίζουμε την γενική μορφή της συνάρτησης που περιγράφει την πυκνότητα πιθανότητας, χρειαζόμαστε και τα παραδείγματα για να υπολογίσουμε τις σταθερές παραμέτρους της.

➢ Στις περιπτώσεις αυτές, εκμεταλλευόμαστε κάποιες από τις ιδιότητες (στατιστικά χαρακτηριστικά) της συνάρτησης πυκνότητας πιθανότητας:

1. Μέση τιμή:

$$Mr = \int_{\mathrm{x}} \mathrm{x}p\left(\mathrm{x}\middle|\omega_i\right)d\mathrm{x} \approx M = \frac{1}{Q}\sum_{j=1}^{Q}\mathrm{x}_j$$

➢ Παρατηρούμε ότι υπάρχει μια διαφορά, ανάμεσα στον στοχαστικό και τον στατιστικό υπολογισμό του διανύσματος της τυχαίας μεταβλητής του προτύπου. Η διαφορά αυτή δημιουργεί κάποια προβλήματα στον υπολογισμό της διασποράς της τυχαίας μεταβλητής, διότι στην πράξη δεν γνωρίζουμε την τιμή της πραγματικής μέσης τιμής (Mr) αλλά μιας προσέγγισης της στατιστικής μέσης τιμής (M).

Υπολογισμός των Παραμέτρων της Πυκνότητας Πιθανότητας (2)

- Η στατιστική διασπορά υπολογίζεται από την ακόλουθη σχέση:

$$\Sigma = \frac{1}{Q}\sum_{j=1}^{Q}\left(\mathrm{x}_j - M\right)^T\left(\mathrm{x}_j - M\right)$$

- Η ακριβέστερη εκτίμηση της στατιστικής διασποράς δίνεται από την σχέση:

$$\Sigma r = \frac{1}{Q}\sum_{j=1}^{Q}\left(\mathrm{x}_j - M\right)^T\left(\mathrm{x}_j - Mr\right)$$

Σε αυτή την περίπτωση, παρατηρούμε, όμως, ότι δεν γνωρίζουμε την πραγματική μέση τιμή Mr.
Το πρόβλημα αντιμετωπίζεται ως εξής:
Η πρώτη εξίσωση μπορεί να γραφτεί ισοδύναμα σαν:

$$\Sigma = \frac{1}{Q}\sum_{j=1}^{Q}\left(\left(\mathrm{x}_j - Mr\right) - \left(M - Mr\right)\right)^T\left(\left(\mathrm{x}_j - Mr\right) - \left(M - Mr\right)\right) \Leftrightarrow$$

$$\Sigma = \frac{1}{Q}\sum_{j=1}^{Q}\left(\mathrm{x}_j - M\right)^T\left(\mathrm{x}_j - M\right) - \left(M - Mr\right)^T\left(M - Mr\right) \Leftrightarrow$$

$$\Sigma r = \Sigma + \left(M - Mr\right)^T\left(M - Mr\right) = \Sigma + \frac{1}{Q}\Sigma r$$

- Παρατηρούμε ότι η πραγματική τιμή της στατιστικής διασποράς είναι μεγαλύτερη από αυτή που, συνήθως, υπολογίζεται με την κλασσική σχέση της διασποράς.

Υπολογισμός των Παραμέτρων της Πυκνότητας Πιθανότητας (3)

➢ Λύνοντας την τελευταία εξίσωση ως προς Σr, προκύπτει:

$$\Sigma r = \frac{Q}{Q-1}\Sigma = \frac{1}{Q-1}\sum_{j=1}^{Q}\left(\mathrm{x}_j - M\right)^T\left(\mathrm{x}_j - M\right)$$

➢ Όταν ο αριθμός των παραδειγμάτων αυξάνει, τότε και η πραγματική στατιστική διασπορά προσεγγίζει την στατιστική διασπορά, η οποία υπολογίζεται με την κλασσική σχέση της εξίσωσης.

• Από τα παραπάνω γίνεται φανερό, ότι η δεύτερη εξίσωση που μας προσφέρει δυνατότητες υπολογισμού των σταθερών παραμέτρων της συνάρτησης πυκνότητας πιθανότητας είναι η ακόλουθη:

2. Διασπορά:

$$\int_{\mathrm{x}}\left(\mathrm{x} - Mr\right)^2 p\left(\mathrm{x}|\omega_i\right)d\mathrm{x} \approx \Sigma r = \frac{1}{Q-1}\sum_{j=1}^{Q}\left(\mathrm{x}_j - M\right)^T\left(\mathrm{x}_j - M\right)$$

Παράδειγμα 7ο

Επιλογή της Συνάρτησης Πυκνότητας Πιθανότητας (1)

- Γνωρίζοντας ένα πλήθος συναρτήσεων, καλούμαστε να λύσουμε το πρόβλημα της επιλογής της πλέον κατάλληλης από αυτές. Πρέπει να βρούμε ένα ποσοτικό μέγεθος με το οποίο θα υπολογίζεται κατά πόσο η συνάρτηση πυκνότητας πιθανότητας που χρησιμοποιούμε 'ταιριάζει' με τα πραγματικά δεδομένα που δεν είναι τίποτα άλλο από τα πρότυπα των παραδειγμάτων.

- Η πλέον διαδεδομένη μέθοδος ελέγχου της ομοιότητας μιας συνεχούς συνάρτησης πυκνότητας πιθανότητας και των αντίστοιχων στατιστικών κατανομών που υπολογίζονται από τα πρότυπα παραδειγμάτων, είναι ο έλεγχος-x^2 (x^2-test). Ο έλεγχος αυτός πραγματοποιείται εκτελώντας τα ακόλουθα βήματα:

1. Ορισμός περιοχών ελέγχου. Χωρίζουμε τον χώρο των προτύπων σε έναν αριθμό περιοχών, έτσι ώστε κάθε περιοχή να περιέχει έναν ελάχιστο αριθμό προτύπων από τα παραδείγματα. Έστω ότι το πεδίο ορισμού των προτύπων χωρίστηκε σε M περιοχές, τις $S_1,S_2,\ldots,S_M$.

Επιλογή της Συνάρτησης Πυκνότητας Πιθανότητας (2)

2. Υπολογισμός των πιθανοτήτων των περιοχών ελέγχου. Έστω ότι ελέγχουμε την συνεχή συνάρτηση πυκνότητας πιθανότητας p(x). Για κάθε μια από αυτές τις περιοχές, υπολογίζουμε την πιθανότητα το πρότυπο της κατηγορίας να ανήκει στην περιοχή. Η αριθμητική τιμή της πιθανότητας υπολογίζεται από την ακόλουθη σχέση:

$$Pt_i = \int_{\mathrm{x} \in S_i} p(\mathrm{x}) d\mathrm{x}$$

3. Στατιστικός υπολογισμός των πιθανοτήτων των περιοχών ελέγχου. Για κάθε μια από τις περιοχές εκτελούμε τον στατιστικό υπολογισμό της πιθανότητας, το πρότυπο να ανήκει στην περιοχή που ελέγχουμε. Αν στην περιοχή S_i βρίσκονται t_i πρότυπα σε ένα σύνολο Τ παραδειγμάτων, τότε η στατιστική πιθανότητα υπολογίζεται από την σχέση:

$$Ps_i = \frac{t_i}{T}$$

Επιλογή της Συνάρτησης Πυκνότητας Πιθανότητας (3)

4. Υπολογισμός της απόκλισης. Υπολογίζουμε την απόκλιση των πιθανοτήτων για όλες τις περιοχές:

$$\text{Απόκλιση} = \sum_{i=1}^{M} \frac{(Ps_i - Pt_i)^2}{Pt_i}$$

5. Σύγκριση. Για κάθε κατανομή που ελέγχουμε βρίσκουμε το αντίστοιχο ποσοτικό μέγεθος της απόκλισης, κρατώντας σταθερές τις περιοχές ελέγχου. Η στοχαστική κατανομή που παρουσιάζει την μικρότερη απόσταση από την στατιστική κατανομή επιλέγεται σαν η πλέον αντιπροσωπευτική.

Παρατηρήσεις-Αδύνατα σημεία του ελέγχου-x^2

i. Ο μικρός αριθμός παραδειγμάτων κάνει ανασφαλή την μέτρηση της στατιστικής πιθανότητας στις περιοχές που ορίζουμε.

ii. Ο ορισμός των περιοχών ελέγχου παίζει πολύ σημαντικό ρόλο στον τελικό υπολογισμό της απόκλισης των κατανομών, ιδιαίτερα όταν ο αριθμός των παραδειγμάτων δεν είναι μεγάλος ή όταν οι διαφορές αποκλίσεων των κατανομών, που ελέγχονται , είναι μικρές.

Παράδειγμα 8ο

Στοχαστκός Υπολογισμός της Πυκνότητας Πιθανότητας (1)

- Σε πολλά προβλήματα οι μετρήσεις που παίρνουμε περιέχουν ‘θόρυβο’, αποτελούν, με άλλα λόγια, προσέγγιση των πραγματικών τιμών εξαιτίας της πεπερασμένης ακρίβειας του μετρητικού οργάνου ή της παρουσίας θορύβου που παρεμβάλλεται.

- Οι συναρτήσεις παλινδρόμησης (regression functions) σε αντίθεση με τις κοινές συναρτήσεις είναι στατιστικές συναρτήσεις. Χαρακτηριστικό γνώρισμα των συναρτήσεων παλινδρόμησης είναι το γεγονός ότι δίνουν κάθε φορά που ζητάμε την τιμή της συνάρτησης σε ένα σημείο διαφορετική αριθμητική τιμή. Οι μέθοδοι, που ακολουθούν, είναι σε θέση να εκτιμήσουν ρίζες και ακρότατα μονοδιάστατων και πολυδιάστατων συναρτήσεων παλινδρόμησης. Οι μέθοδοι αυτές χρησιμοποιούνται ευρύτατα σε προβλήματα αναγνώρισης προτύπων και κυρίως για την εκτίμηση της πυκνότητας πιθανότητας προτύπων.

Στοχαστκός Υπολογισμός της Πυκνότητας Πιθανότητας (2)

1. Ο αλγόριθμος των Robbins-Monro

- Έστω συνάρτηση *f*(x) η οποία έχει μια μοναδική ρίζα *f*(x^)=0. Υποθέτουμε ότι οι τιμές της συνάρτησης είναι αρνητικές στο διάστημα των μικρότερων του x^ τιμών και θετικές στο διάστημα των μεγαλύτερων του x^ τιμών. Η υπόθεση αυτή δεν βλάπτει την γενικότητα διότι αν συμβαίνει το αντίθετο, τότε κάνουμε τους υπολογισμούς μας για την συνάρτηση - *f*(x) η οποία ικανοποιεί τον περιορισμό που θέσαμε.

- Υποθέτουμε ότι το μετρητικό όργανο που έχουμε για να μας μετρά τις τιμές της συνάρτησης για τις διάφορες τιμές του x παρενοχλείται από θόρυβο, ο οποίος αλλοιώνει τις μετρήσεις. Γι'αυτό τον λόγο δεν μπορούμε να υπολογίσουμε απευθείας την ρίζα της συνάρτησης, χρησιμοποιώντας κάποια γνωστή αριθμητική μέθοδο. Έτσι, αντί της πραγματικής τιμής *f*(x) λαμβάνουμε την τιμή της συνάρτησης παλινδρόμησης, την *g*(x). Η διασπορά του σφάλματος εκτίμησης της τιμής της συνάρτησης, δίνεται από την μέση τιμή της ποσότητας |*f*(x)- *g*(x)|. .

Στοχαστκός Υπολογισμός της Πυκνότητας Πιθανότητας (3)

- Κάνουμε δυο υποθέσεις για τα χαρακτηριστικά του θορύβου:

1. Ο θόρυβος είναι τυχαίος. Αν ο θόρυβος που αλλοιώνει τις μετρήσεις είναι στοχαστικά ανεξάρτητος από το σήμα και έχει μηδενική μέση τιμή, τότε ισχύει:

$$E\{g(\mathrm{x},t)\} = f(\mathrm{x})$$

2. Ο θόρυβος είναι πεπερασμένος. Αν η ενέργεια του θορύβου είναι πεπερασμένη, τότε θα ισχύει ότι:

$$\sigma^2 = E\left[E\left\{\left(g(\mathrm{x},t) - f(\mathrm{x})\right)^2\right\}\right] < T_h,\ T_h \in \Re^+$$

- Ο περιορισμός αυτός τίθεται για να επισημάνουμε ότι η ενέργεια του θορύβου που προστίθεται είναι περιορισμένη. Όσο πιο μικρή είναι η ενέργεια του θορύβου, τόσο καλύτερη θα είναι και η εκτίμηση της ρίζας της συνάρτησης.

Στοχαστκός Υπολογισμός της Πυκνότητας Πιθανότητας (4)

- Από τον αλγόριθμο Robbins και Monro αποδεικνύεται ότι αν ο θόρυβος ικανοποιεί τις παραπάνω συνθήκες, τότε ο επαναληπτικός αλγόριθμος που ακολουθεί, συγκλίνει πάντα σε σημείο που βρίσκεται στην γειτονιά της πραγματικής ρίζας.

1. Αρχικές τιμές, k=0. Επιλέγουμε μια τιμή για το x, έστω x_o, τέτοια ώστε να είναι κοντά στην ρίζα της εξίσωσης. Αν αυτό δεν είναι δυνατόν να πραγματοποιηθεί, διαλέγουμε μια οποιαδήποτε τυχαία τιμή x_o.

2. Επαναπροσδιορίζουμε την τιμή της ρίζας της εξίσωσης. Με την σχέση που ακολουθεί επαναπροσδιορίζουμε την εκτίμηση μας για την τιμή της ρίζας της εξίσωσης:

$$\mathrm{x}_{k+1} = \mathrm{x}_k - a(k) g(\mathrm{x}_k)$$

Στοχαστκός Υπολογισμός της Πυκνότητας Πιθανότητας (5)

3. Έλεγχος τερματισμού των επαναλήψεων, k=k+1. Επειδή το φαινόμενο είναι στοχαστικό δεν μπορούμε να θεωρήσουμε ότι τα συνήθη κριτήρια σύγκλισης είναι αξιόπιστα, όπως π.χ το κριτήριο ελέγχου της απόλυτης μεταβολής της εκτίμησης της ρίζας της εξίσωσης. Το κριτήριο που χρησιμοποιούμε σε αυτές τις περιπτώσεις, είναι ο αριθμός των επαναλήψεων του αλγορίθμου ή ισοδύναμα, ελέγχουμε την αριθμητική τιμή της σταθεράς $\alpha(\mathrm{k})$.

- Ο παράγοντας $\alpha(\mathrm{k})$ είναι μια ακολουθία θετικών πραγματικών αριθμών, η οποία πρέπει να ικανοποιεί τις ακόλουθες συνθήκες:

$$\lim_{k\to\infty} a(k)=0 \quad , \quad \sum_{k=1}^{\infty} a(k)=+\infty \quad , \quad \sum_{k=1}^{\infty} a(k)^2 = W, W \in \Re^+$$

- Ο τελευταίος περιορισμός μας λέει ότι το άθροισμα των τετραγώνων της ακολουθίας θα πρέπει να είναι πεπερασμένος θετικός αριθμός.

Στοχαστκός Υπολογισμός της Πυκνότητας Πιθανότητας (6)

- Παράδειγμα ακολουθίας αριθμών που ικανοποιούν τους περιορισμούς και χρησιμοποιούνται πολύ συχνά στις εφαρμογές, λόγω της απλότητας των υπολογισμών, είναι η ακόλουθη:

$$a(k) = \frac{1}{k}, k = 1,2,\ldots$$

- Λίγα χρόνια μετά την παρουσίαση του αλγορίθμου των Robbins και Monro, αποδείχθηκε ότι ο αλγόριθμος συγκλίνει πάντα σε τιμή που βρίσκεται κοντά στην πραγματική ρίζα. Συγκεκριμένα ο Blum απέδειξε ότι:

$$\Pr ob\left(\lim_{k\to\infty} \mathrm{x}_k = \hat{\mathrm{x}}\right) = 1$$

- Ένας αποτελεσματικός τρόπος για να επιτύχουμε την σύγκλιση του αλγορίθμου είναι να μην αλλάζουμε τον συντελεστή α(k), όταν το πρόσημο της συνάρτησης δεν έχει αλλάξει σε δυο διαφορετικά βήματα. Με την τεχνική αυτή, διατηρούμε σταθερό τον συντελεστή επαναπροσδιορισμού της ρίζας, όταν βρισκόμαστε μακριά από την πραγματική τιμή της ρίζας.

Στοχαστκός Υπολογισμός της Πυκνότητας Πιθανότητας (7)

2. Αλγόριθμος των Kiefer-Wolfowitz

- Ο αλγόριθμος των Robbins-Monro μπορεί, εύκολα να προσαρμοστεί, έτσι ώστε να υπολογίζει τα ακρότατα μιας συνάρτησης παλινδρόμησης.

- Όπως είναι γνωστό, τα ακρότατα της συνάρτησης *f*(x) βρίσκονται στις ρίζες της συνάρτησης:

$$\frac{\partial}{\partial \mathrm{x}} f(\mathrm{x}) = 0$$

Έτσι, το πρόβλημα της εύρεσης των ακρότατων μετατρέπεται σε πρόβλημα εύρεσης των ριζών της πρώτης παραγώγου. Επειδή η συνάρτηση παλινδρόμησης είναι στοχαστική, η παράγωγος, σε οποιοδήποτε σημείο, μπορεί να υπολογιστεί μόνο με αριθμητική μέθοδο.

- Η παράγωγος μιας συνάρτησης, μπορεί να προσεγγιστεί από την σχέση που ακολουθεί:

$$\frac{\partial}{\partial \mathrm{x}} f(\mathrm{x}) \approx \frac{f(\mathrm{x}+h) - f(\mathrm{x}-h)}{2h}$$

Στοχαστκός Υπολογισμός της Πυκνότητας Πιθανότητας (8)

- Από τον αλγόριθμο των Robbins-Monro, η αναδρομική σχέση που επαναπροσδιορίζει τις ρίζες της παραγώγου της συνάρτησης που είναι ταυτόχρονα και τα τοπικά ελάχιστα, δίνεται από την σχέση:

$$\mathrm{x}_{k+1} = \mathrm{x}_k - a(k)\frac{g(\mathrm{x}_k + c(k)) - g(\mathrm{x}_k - c(k))}{2c(k)}$$

- Η αναδρομική σχέση που επαναπροσδιορίζει τις τιμές των μεγίστων της συνάρτησης παλινδρόμησης είναι:

$$\mathrm{x}_{k+1} = \mathrm{x}_k + a(k)\frac{g(\mathrm{x}_k + c(k)) - g(\mathrm{x}_k - c(k))}{2c(k)}$$

- Η επαναληπτική εξίσωση ονομάζεται **μέθοδος εύρεσης των ακρότατων παλίνδρομης συνάρτησης των Kiefer-Wolfowitz**.

Στοχαστκός Υπολογισμός της Πυκνότητας Πιθανότητας (9)

- Ισχύουν και εδώ οι περιορισμοί, που τέθηκαν, στον αλγόριθμο των Robbins-Monro και επιπρόσθετα:

$$\lim_{k\to\infty} a(k) = 0$$

$$\lim_{k\to\infty} c(k) = 0$$

$$\sum_{k=1}^{\infty} a(k) = +\infty$$

$$\sum_{k=1}^{\infty} \left(\frac{a(k)}{c(k)}\right)^2 = W, W \in \Re^+$$

- Οι πιο συνηθισμένες ακολουθίες θετικών πραγματικών αριθμών που χρησιμοποιούνται στις εφαρμογές, λόγω της υπολογιστικής τους απλότητας, είναι:

$$a(k) = \frac{1}{k}, k = 1,2,\ldots \quad c(k) = \frac{1}{\log(k)+1}, k = 1,2,\ldots$$

Στοχαστκός Υπολογισμός της Πυκνότητας Πιθανότητας (10)

3. Υπολογισμός κατωφλίων απόφασης με στοχαστικές μεθόδους

➢ Έχει αναφερθεί, ότι η συνάρτηση πιθανότητας μπορεί να προσεγγιστεί με γραμμικό τρόπο, μετασχηματίζοντας τα διανύσματα των παραδειγμάτων. Η συνάρτηση διάκρισης μπορεί να εκφραστεί ως εξής:

$$g(\mathrm{x}) = p(\omega|\mathrm{x}) \approx \sum_{i=1}^{Q+1} c_i \varphi_i(\mathrm{x}) = c^T \mathrm{x}_\varphi$$

➢ Μια απλή μέθοδος για να υπολογιστεί η πιθανότητα *p*(ω|x) είναι η γνώση της τιμής της στα διάφορα σημεία του χώρου των μετρήσεων, π.χ στα σημεία που βρίσκονται τα πρότυπα των παραδειγμάτων. Η πληροφορία που διαθέτουμε για τα σημεία αυτά, κατά την διάρκεια της εκπαίδευσης, είναι η ταυτότητα του προτύπου. Η πληροφορία αυτή, μας δίνει την δυνατότητα να ορίσουμε την παλίνδρομη συνάρτηση *z*(x), η οποία παίρνει τις ακόλουθες τιμές:

$$z(\mathrm{x}_\varphi) = \begin{cases} 1, \mathrm{x}_\varphi \in \omega \\ 0, \alpha\lambda\lambda o\acute{\upsilon} \end{cases}$$

Στοχαστκός Υπολογισμός της Πυκνότητας Πιθανότητας (11)

- Έστω ότι η συνάρτηση $z(\mathrm{x}_\varphi)$ αποτελεί μια προσέγγιση της πραγματικής συνάρτησης πυκνότητας πιθανότητας:

$$z\left(\mathrm{x}_{\varphi}\right)=p\left(\omega \middle| \mathrm{x}_{\varphi}\right)+n$$

όπου n είναι ο θόρυβος, ο οποίος, υποθέτουμε, ότι έχει μηδενική μέση τιμή και πεπερασμένη διαφορά.

- Η συνάρτηση:

$$R\left(z\left(\mathrm{x}_{\varphi}\right),c\right)=\sum_{i=1}^{M}\left|z\left(\mathrm{x}_{\varphi}{}^{(i)}\right)-c^{T}\mathrm{x}_{\varphi}{}^{(i)}\right|$$

είναι και αυτή μια συνάρτηση παλινδρόμησης που εξαρτάται από το διάνυσμα c^T και M είναι ο αριθμός των διαθέσιμων παραδειγμάτων.

Στοχαστκός Υπολογισμός της Πυκνότητας Πιθανότητας (12)

- Αν χρησιμοποιήσουμε την μέθοδο των Kiefer-Wolfowitz για να βρούμε τις τιμές του διανύσματος c για το οποίο η συνάρτηση παλινδρόμησης γίνεται ελάχιστη, τότε έχουμε την ακόλουθη αναδρομική εξίσωση που οδηγεί στην εκτίμηση του ελαχίστου της συνάρτησης:

$$c_{k+1} = c_k - a(k)\left(\frac{\partial R(z(\mathrm{x}_\varphi), c)}{\partial c}\right)_{c=c_k}$$

- Βρίσκοντας την παράγωγο της συνάρτησης ως προς την διανυσματική μεταβλητή x_ϕ έχουμε:

$$\frac{\partial R(z(\mathrm{x}_\varphi), c)}{\partial c} = -\sum_{i=1}^{M} \mathrm{x}_\varphi \operatorname{sgn}\left(z\left(\mathrm{x}_\varphi{}^{(i)}\right) - c^T \mathrm{x}_\varphi{}^{(i)}\right)$$

όπου

$$\operatorname{sgn}(\mathrm{x}) = \begin{cases} 1, \mathrm{x} \rangle 0 \\ -1, \mathrm{x} \le 0 \end{cases}$$

Στοχαστκός Υπολογισμός της Πυκνότητας Πιθανότητας (13)

➢ Με βάσει τα παραπάνω, η αναδρομική εξίσωση που οδηγεί στην εκτίμηση του ελαχίστου της συνάρτησης, παίρνει τελικά την ακόλουθη μορφή:

$$c_{k+1} = \begin{cases} c_k + a(k)\mathrm{x}_\varphi, z(\mathrm{x}_\varphi) \rangle c^T \mathrm{x}_\varphi \\ c_k - a(k)\mathrm{x}_\varphi, z(\mathrm{x}_\varphi) \leq c^T \mathrm{x}_\varphi \end{cases}$$

<u>Παρατήρηση</u>

➢ Η σημαντικότερη διαφορά των δυο αλγορίθμων, βρίσκεται στο γεγονός ότι ο στοχαστικός αλγόριθμος συγκλίνει πάντα, ανεξάρτητα από την κατανομή των προτύπων των παραδειγμάτων στον χώρο των μετρήσεων, σε αντίθεση με τον αλγόριθμο perceptron ο οποίος θέτει σαν απαραίτητη προϋπόθεση για την σύγκλιση, την γραμμική διαχωρισιμότητα των προτύπων.

Στοχαστκός Υπολογισμός της Πυκνότητας Πιθανότητας (14)

4. Στοχαστικός αλγόριθμος ελαχίστου τετραγωνικού σφάλματος

➢ Αν ορίσουμε με διαφορετικό τρόπο την στοχαστική συνάρτηση του σφάλματος:

$$R\left(z\left(\mathrm{x}_{\varphi}\right), c\right) = \sum_{i=1}^{M}\left(z\left(\mathrm{x}_{\varphi}{}^{(i)}\right) - c^{T}\mathrm{x}_{\varphi}{}^{(i)}\right)^{2}$$

ακολουθώντας τον ίδιο τρόπο απόδειξης, μπορούμε να δείξουμε ότι ο προσδιορισμός του ελαχίστου της συνάρτησης $R\left(z\left(\mathrm{x}_{\varphi}\right), c\right)$ μπορεί να επιτευχθεί με την εξής αναδρομική σχέση:

$$c_{k+1} = c_{k} + a(k)\mathrm{x}_{\varphi}\left(z\left(\mathrm{x}_{\varphi}\right) - c^{T}\mathrm{x}_{\varphi}\right)$$

Παράδειγμα 9ο

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