Θεωρία Αποφάσεων
1° Φροντιστήριο

Λύσεις των Ασκήσεων

Πρόβλημα 1

Σε μία κλινική οι ασθενείς οι οποίοι πιθανολογείται ότι πάσχουν από Ηπατίτιδα C περνούν από ένα τεστ ανίχνευσης αντισωμάτων στο αίμα τους.
Έστω:
H «Ο ασθενής πάσχει από τον ιό της Ηπατίτιδας»
nH «Ο ασθενής δεν πάσχει από τον ιό της Ηπατίτιδας»
Pos «Το τεστ του ασθενή είναι θετικό στον ιό»
Neg «Το τεστ του ασθενή είναι αρνητικό στον ιό»

Εμπειρικά, έχουν καταγραφεί οι ακόλουθες πιθανότητες:
P(H) = 0.15
P(Pos | H) = 0.95
P(Pos | nH) = 0.02
Ποια είναι η πιθανότητα ένας ασθενής να πάσχει από τον ιό της Ηπατίτιδας C εάν το αποτέλεσμα του τεστ είναι θετικό;

Λύση
Χρησιμοποιώντας τους κανόνες απόφασης του Bayes προκύπτουν οι παρακάτω εξισώσεις:
πάσχει ενώ έχει θετικό τεστ.

P(H / Pos) = P(Pos / H) . P(H) / P(Pos)   *Σχέση 1*

Επίσης:
P(Pos)= P(Pos / H) . P(H)+ P(Pos /nH) . P(nH) =
= P(Pos / H) . P(H)+ P(Pos /nH) .(1- P(H))

P(Pos) = P(Pos / H) . P(H)+ P(Pos /nH) .(1- P(H))   *Σχέση 2*

P(Pos) = 0.95*0.15 + 0.02*(1-0.15) = 0.1595

Από (1) - (2) προκύπτει:
P(H / Pos) = 0.95*0.15 / 0.1595 = 0.8934

Άρα   P(H / Pos) = 0.8934

Πρόβλημα 2

Θεωρούμε ότι υπάρχουν οι παρακάτω τρεις υποθέσεις σχετικά με ένα κέρμα:
F «Το κέρμα είναι 'fair'»
B «Το κέρμα έχει πειραχτεί ώστε να φέρνει κορώνα στο 60% των ρίψεων»
C «Δεν ισχύει κανένα από τα F, B»
Έστω E το γεγονός ότι το κέρμα έχει έρθει κορώνα 10 φορές σε 10 ρίψεις.

A) Υποθέτουμε ότι αρχικά θεωρείται ότι υπάρχει διπλάσια πιθανότητα το κέρμα να είναι 'fair' απ' ότι να είναι πειραγμένο.

Ποιος είναι ο λόγος των εκ των υστέρων πιθανοτήτων, P(F/E) /P(B/E);
B) Υποθέτουμε ότι P(E) = 0.1, ενώ P(F) = 0.5 και P(B)=0.2. Ποια είναι η εκ των υστέρων πιθανότητα του C δοθέντος του E (δηλ. P(C/E));
C) Υποθέτουμε ότι είτε το F είτε το B πρέπει να είναι αληθές. Ποια είναι η μικρότερη πιθανή τιμή για το P(B) έτσι ώστε για την εκ των υστέρων πιθανότητα του B να ισχύει P(B/E)≥ 0.5;

Λύση

A) Χρησιμοποιώντας το Θεώρημα του Bayes προκύπτει ο παρακάτω λόγος των εκ των υστέρων πιθανοτήτων

$$\frac{P(F / E)}{P(B / E)} = \frac{[P(E / F) P(F)] / P(E)}{[P(E / B) P(B)] / P(E)} = [P(F) / P(B)] * P(E / F) / P(E / B) =$$

$$= 2(0.5)^{10} / (0.6)^{10}$$

το οποίο είναι περίπου ίσο με 0.32, αφού P(E/F)= $(0.5)^{10}$ και P(E/B)= $(0.6)^{10}$.

B) Σύμφωνα με το Θεώρημα του Bayes ισχύει P(C/E) = P(E/C)P(C) / P(E).
Αν και δεν μπορούμε να υπολογίσουμε το P(E/C) απευθείας, γνωρίζουμε από το θεώρημα της συνολικής πιθανότητας ότι

$$P(E/C)\,P(C) + P(E/B)\,P(B) + P(E/F)\,P(F) = P(E)$$

και επίσης P(E) = 0.1.
Επομένως:

$$P(E/C)\,P(C) = 0.1 - (0.6)^{10}(0.2) - (0.5)^{10}(0.5) = 0.098$$

Άρα
P(C/E) = 0.098 / 0.1 = 0.98, το οποίο είναι πολύ κοντά στην τιμή 1.

Γ) Γνωρίζουμε ότι P(B) + P(F) = 1. Από την ισχύουσα συνθήκη με αντικατάσταση προκύπτει

$$P(B/E) >= 0.5,$$

Δηλαδή

$$P(E/B)P(B) / [P(E/B)\,P(B) + P(E/F)\,P(F)] \geq 0.5$$

Αφού

$$P(F) = 1 - P(B),\ P(E / B) = (0.6)^{10} \text{ και } P(E / F) = (0.5)^{10},$$

προκύπτει

$$(0.6)^{10}P(B) / [(0.6)^{10}P(B) + (0.5)^{10}(1 - P(B))] \geq 0.5, \text{ ή}$$
$$[2(0.6)^{10} - (0.6)^{10} + (0.5)^{10}]\,P(B) \geq (0.5)^{10}$$

Επομένως

$$P(B) \geq 0.14.$$

ΠΡΟΒΛΗΜΑ 3

Προσπαθούμε να αποφασίσουμε εάν θα πρέπει να ξεκινήσουμε τις διαδικασίες εύρεσης πετρελαίου σε μία συγκεκριμένη περιοχή των Εμιράτων την οποία κατέχουμε ή όχι. Υπάρχουν δύο πιθανές καταστάσεις της φύσης:
ω1 «η περιοχή έχει πετρέλαιο»,
ω2 «η περιοχή δεν έχει πετρέλαιο».
Θεωρούμε τις τρεις παρακάτω ενέργειες:
α1 «σκάβουμε για πετρέλαιο»,

α2 «πουλάμε την περιοχή»,
α3 «πουλάμε τα μισά δικαιώματα που κατέχουμε στην περιοχή».

Η συνάρτηση κόστους καθορίζεται ως εξής:

	α1	α2	α3
ω1	0	10	5
ω2	12	1	6

Η διεξαγωγή ενός πειράματος για να αποκτήσουμε πληροφορίες σχετικά με την κατάσταση της φύσης μας δίνει τις παρακάτω πληροφορίες σχετικά με την πιθανοφάνεια της τυχαίας μεταβλητής x με πιθανές τιμές 0 και 1:

	Τυχαία Μεταβλητή x (Σχήμα Βράχων)	Τυχαία Μεταβλητή x (Σχήμα Βράχων)
	0	1
ω1	0,3	0,7
ω2	0,6	0,4

Λύση
(x=0)
Η εκ των υστέρων πιθανότητα είναι:

$$p(\omega_j \mid x) = \frac{p(x \mid \omega_j)p(\omega_j)}{p(x)}$$

όπου $p(x) = \sum_{j=1} p(x \mid \omega_j)p(\omega_j)$
Υπολογίζουμε

P(ω1|x) = 0.3*0.2 = 1/9

$$p(\omega_1 \mid x) = \frac{0.3*0.2}{(0.3*0.2)+(0.6*0.8)}$$

και με τον ίδιο τρόπο P(ω2|x) = 8/9.

Το υπό συνθήκη ρίσκο είναι:
$R(a_1|x) = \lambda_{11}\, P(\omega1|x) + \lambda_{12}\, P(\omega2|x) =$
$= 10.67$
$R(a_2|x) = \lambda_{21}\, P(\omega1|x) + \lambda_{22}\, P(\omega2|x) =$
$= 2$
$R(a_3|x) = \lambda_{31}\, P(\omega1|x) + \lambda_{32}\, P(\omega2|x) =$
$= 5.89$
Άρα αφού

$R(a_1|x) \geq R(a_3|x) \geq R(a_2|x)$,

επιλέγουμε το ω_2 (κανόνας ελαχιστοποίησης του ρίσκο Bayes Criterion).

(x=1) Ομοίως υπολογίζουμε:
$R(a_1|x) = 8.35$,
$R(a_2|x) = 3.74$,
$R(a_3|x) = 5.70$,
Οπότε επιλέγουμε το ω_2.

ΠΡΟΒΛΗΜΑ 4

Έστω ένα πρόβλημα ταξινόμησης με τις παρακάτω συναρτήσεις πυκνότητας πιθανότητας :

$$p(x \mid \omega_1) = \frac{1}{\sqrt{2\pi}} e^{-1/2(x-4)^2}$$

$$p(x \mid \omega_2) = \frac{1}{\sqrt{2\pi}} e^{-1/2(x-10)^2}$$

Εάν ισχύει $P(\omega_1)=P(\omega_2)$, να βρείτε τον κανόνα απόφασης βάσει του Maximium Likelihood.

ΛΥΣΗ

Υπολογίζουμε τον λόγο των συναρτήσεων πυκνότητας πιθανότητας :

$$\frac{p(x \mid \omega_1)}{p(x \mid \omega_2)} = \frac{e^{-1/2(x-4)^2}}{e^{-1/2(x-10)^2}} = \frac{(x-10)^2}{(x-4)^2} > 1$$

$$(x-4)^2 - (x-10)^2 < 0$$

$$x^2 + 16 - 8x - (x^2 + 100 - 20x) < 0$$

$$12x - 84 < 0$$

$$x - 7 < 0$$

Επομένως, Εάν $x < 7$ επέλεξε ω1

Εάν $x > 7$ επέλεξε ω2

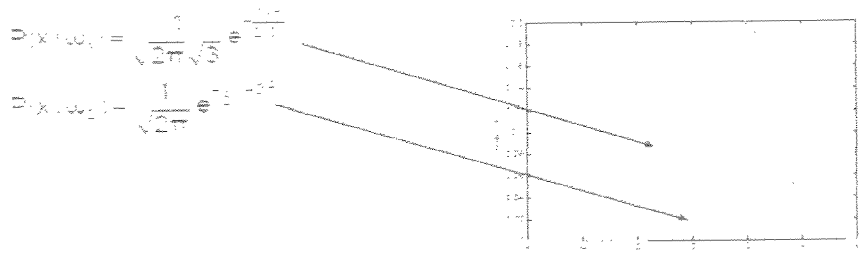
ΠΡΟΒΛΗΜΑ 5

Έστω το παραπάνω πρόβλημα ταξινόμησης με 2 classes ω1 και ω2 που έχουν τις παρακάτω συναρτήσεις πιθανοφάνειας :

$$p(x \mid \omega_1) = \frac{1}{\sqrt{2\pi}\sqrt{3}} e^{-x^2/23}$$

$$p(x \mid \omega_2) = \frac{1}{\sqrt{2\pi}} e^{-1/2(x-2)^2}$$

Ποιος είναι ο κανόνας απόφασης που ελαχιστοποιεί το P[error], δηλαδή την πιθανότητα για ταξινόμηση σε λάθος class ?



ΛΥΣΗ

Για να ελαχιστοποιήσουμε το P[error], σύμφωνα με :

$$P[\text{error}] = \int_{-\infty}^{\infty} P(\text{error},x)dx = \int_{-\infty}^{\infty} P(\text{error}|x)P(x)dx$$

όπου για 2-classes πρόβλημα ισχύει :

$$P(\text{error} \mid x)=\begin{cases} p(\omega_1 \mid x)\cdot \varepsilon\acute{\alpha}\nu \cdot \alpha\pi o\varphi\alpha\sigma\acute{\iota}\sigma o\upsilon\mu\varepsilon \cdot \omega 2 \\ p(\omega_2 \mid x)\cdot \varepsilon\acute{\alpha}\nu \cdot \alpha\pi o\varphi\alpha\sigma\acute{\iota}\sigma o\upsilon\mu\varepsilon \cdot \omega 1 \end{cases}$$

Επομένως αρκεί να επιλέξουμε την κλάση εκείνη για την οποία μεγιστοποιείται η εκ των υστέρων πιθανότητα $P(\omega_i \mid x)$ (δηλ. Maximum A Posteriori Criterion - MAP).

Εφόσον $P(\omega_1)=P(\omega_2)=1/2$,

$$\frac{P(x \mid \omega_1)}{P(x \mid \omega_2)}=\frac{e^{\,x^2/23}}{e^{\,1/2(x\cdot 2)^2}} > \frac{1}{\sqrt{3}}$$

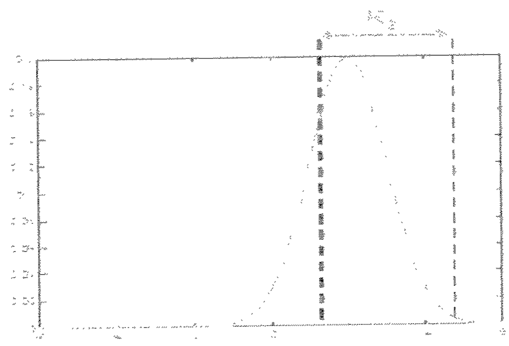
$$\frac{-1}{2}\frac{x^2}{3}+\frac{1}{2}(x-2)^2 >^{\omega 1} 0$$

$$\frac{-1}{2}\frac{x^2}{3}+\frac{1}{2}(x-2)^2 <^{\omega 2} 0$$

Άρα $2x^2-12x+12 >^{\omega 1} 0$

$2x^2-12x+12 <^{\omega 2} 0$

Άρα $x = 4.73,\ 1.27$



Θεωρία Αποφάσεων
2° Φροντιστήριο

Λύσεις των Ασκήσεων

Άσκηση 1

Υποθέστε ότι έχουμε ένα τρισδιάστατο χώρο χαρακτηριστικών **x** = (x,y,z) και δυο κλάσεις ω_1 και ω_2. Για τις δυο κλάσεις ισχύουν τα εξής:

$\mu_1 = [\,0\ 0\ 0\,]^T$
$\mu_2 = [\,1\ 1\ 1\,]^T$

$$\Sigma_1 = \Sigma_2 = \begin{bmatrix} 1/4 & 0 & 0 \\ 0 & 1/4 & 0 \\ 0 & 0 & 1/4 \end{bmatrix}$$

$p(\omega_2) = 2p(\omega_1)$

Α) Αν χρησιμοποιήσουμε τον κανόνα του Bayes ποιές είναι οι συναρτήσεις διαχωρισμού $g_1(x)$ και $g_2(x)$;
Β) Υπολογίστε το όριο απόφασης του Bayes
Γ) Ταξινομήστε το διάνυσμα $x_u = [\,0.1\ 0.7\ 0.8\,]^T$

Λύση

Α) $g_i(x) = -\frac{1}{2\sigma^2}(x-\mu_i)^T(x-\mu_i) + \log P(\omega_i) = -\frac{1}{2}\begin{bmatrix} x-\mu_x \\ y-\mu_y \\ z-\mu_z \end{bmatrix}^T \begin{bmatrix} x-\mu_x \\ y-\mu_y \\ z-\mu_z \end{bmatrix} +$ $\log P(\omega_i)$

$$g_1(x) = -\frac{1}{2}\begin{bmatrix} x-0 \\ y-0 \\ z-0 \end{bmatrix}^T \begin{bmatrix} x-0 \\ y-0 \\ z-0 \end{bmatrix} + \log\frac{1}{3}$$

$$g_2(x) = -\frac{1}{2}\begin{bmatrix} x-1 \\ y-1 \\ z-1 \end{bmatrix}^T \begin{bmatrix} x-1 \\ y-1 \\ z-1 \end{bmatrix} + \log\frac{2}{3}$$

Β) Το όριο απόφασης του Bayes δίνεται από την εξίσωση:

$g_1(x) = g_2(x) \Leftrightarrow$
$-2(x^2 + y^2 + z^2) + \log\frac{1}{3} = -2((x-1)^2 + (y-1)^2 + (z-1)^2) + \log\frac{2}{3} \Leftrightarrow$

Γ) Βρείτε την πιθανότητα λάθους για τον ταξινομητή του ερωτήματος 2 θεωρώντας ότι οι πραγματικές *a priori* πιθανότητες έχουν τις τιμές του ερωτήματος 1.

Λύση

Α) Οι διακρίνουσες συναρτήσεις έχουν την εξής μορφή: $g_i(x) = p(x \mid \omega_i)P(\omega_i)$.
Δηλαδή είναι το γινόμενο των υπό συνθήκη πιθανοτήτων επί τις *a priori* πιθανότητες. Αυτό το γινόμενο έχει ως αποτέλεσμα να πολλαπλασιάζονται οι παραπάνω γραφικές παραστάσεις, που είναι οι γραφικές παραστάσεις των υπό συνθήκη πιθανοτήτων, για τις κλάσεις 1 και 2 επί 0.25 και για την κλάση 3 επί 0.5.

Οι περιοχές ελαχίστου λάθους είναι:

$$a = \begin{cases} 2 \ \alpha\nu \ \ x \le 1/2 \\ 1 \ \alpha\nu \ \ 1/2 \le x \le 97/64 \ \ \acute{\eta} \ \ 127/64 \succ x \le 2 \\ 3 \ \sigma\varepsilon \ \ \kappa\acute{\alpha}\theta\varepsilon \ \ \acute{\alpha}\lambda\lambda\eta \ \ \pi\varepsilon\rho\acute{\iota}\pi\tau\omega\sigma\eta \end{cases}$$

Η πιθανότητα λάθους είναι :

$$P_e = 1/4P_e(\omega_1) + 1/4P_e(\omega_2) + 1/2P_e(\omega_3) = 1/4(1/2 - 1/64) + 1/4(1/8) + 1/2(1/256) = 79/512$$

Β) Αν όλες οι a priori είναι ίσες , τότε τα όρια απόφασης θα είναι :

$$a = \begin{cases} 2 \ \alpha\nu \ \ x \le 1/2 \\ 1 \ \alpha\nu \ \ 1/2 \le x \le 49/32 \ \ \acute{\eta} \ 63/32 \succ x \le 2 \\ 3 \ \sigma\varepsilon \ \ \kappa\acute{\alpha}\theta\varepsilon \ \ \acute{\alpha}\lambda\lambda\eta \ \ \pi\varepsilon\rho\acute{\iota}\pi\tau\omega\sigma\eta \end{cases}$$

Γ) *Η πιθανότητα λάθους για αυτά τα όρια απόφασης*, θεωρώντας ότι οι πραγματικές *a priori* πιθανότητες έχουν τις τιμές του ερωτήματος 1, είναι :

$$P_e = 1/4P_e(\omega_1) + 1/4P_e(\omega_2) + 1/2P_e(\omega_3) = 1/4(1/2 - 1/32) + 1/4(1/8) + 1/2(1/64) = 20/128$$

Το οποίο επιβεβαιώνει το ότι ο βέλτιστος ταξινομητής πρέπει να χρησιμοποιεί τις «σωστές» (πραγματικές) τιμές των *a priori* πιθανοτήτων, τις τιμές δηλαδή με βάση τις οποίες κατασκευάστηκε και ο ίδιος.

$x + y + z = \dfrac{6 - \log 2}{4} = 1.32$

Γ) $0.1 + 0.7 + 0.8 = 1.6 > 1.32$

Άρα το x_u ταξινομείται στην κλάση ω_2.

Άσκηση 2

Έστω το πρόβλημα ταξινόμησης προτύπων σε δύο κατηγορίες όπου τα πρότυπα είναι δισδιάστατα $\mathbf{x} = (x_1\ x_2)$. Οι δύο κατηγορίες είναι οι ω_1 και ω_2 ενώ ισχύουν τα εξής:

$$p(\mathbf{x}/\omega_1) = N(0, \boldsymbol{\Sigma}_1),$$

$$p(\mathbf{x}/\omega_2) = N\left(\begin{bmatrix}1\\1\end{bmatrix}, \boldsymbol{\Sigma}_2\right),$$

$$P(\omega_1) = P(\omega_2) = 1/2 \text{ και}$$

$$\boldsymbol{\Sigma}_1 = \begin{bmatrix}3 & 0\\0 & 3\end{bmatrix},\ \boldsymbol{\Sigma}_2 = \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}.$$

Α)Υπολογίστε το όριο απόφασης του Bayes.
Β)Υπολογίστε το όριο λάθους Bhattacharya.
Γ) Αν για 1000 πρότυπα της ω_1 ταξινομούνται 308 στην ω_2 και για 1000 πρότυπα της ω_2 ταξινομούνται 250 στην ω_1. Βρείτε το ρυθμό λάθους ταξινόμησης.

Λύση

Α) Το όριο απόφασης του Bayes δίνεται από την σχέση

$P(\omega_1|x) = P(\omega_2|x) \Leftrightarrow g_1(x) = g_2(x) \Leftrightarrow$

$$-\frac{1}{2}\log|\Sigma_1| - \frac{1}{2}(x - M_1)^T \Sigma_1^{-1} (x - M_1) + \log(P(\omega_1)) =$$

$$-\frac{1}{2}\log|\Sigma_2| - \frac{1}{2}(x - M_2)^T \Sigma_2^{-1} (x - M_2) + \log(P(\omega_2)) \Leftrightarrow$$

(είναι $\log(P(\omega_1) = \log(P(\omega_2)$ αφού $P(\omega_1) = P(\omega_2)$)

$$-\frac{1}{2}\log 9 - -\frac{1}{2}\cdot\frac{1}{3}(x_1^2 + x_2^2) = -\frac{1}{2}\log 1 - \frac{1}{2}((x_1 - 1)^2 + (x_2 - 1)^2) \Leftrightarrow$$

$$\frac{1}{3}x_1^2 + \frac{1}{3}x_2^2 - x_1 - x_2 + 1 - \log 3 = 0$$

Β) Για Gaussian περίπτωση έχουμε:

$$k(1/2) = \frac{1}{8}\,(\mu_2 - \mu_1)^T \cdot \left[\frac{\Sigma_1 + \Sigma_2}{2}\right]^{-1} \cdot (\mu_2 - \mu_1) + \frac{1}{2}\cdot \log \frac{\frac{\Sigma_1 + \Sigma_2}{2}}{\sqrt{|\Sigma_1||\Sigma_2|}} \Leftrightarrow$$

$$= \frac{1}{8}\begin{bmatrix}1 & 1\end{bmatrix}\cdot\begin{bmatrix}1/2 & 0\\ 0 & 1/2\end{bmatrix}\cdot\begin{bmatrix}1\\ 1\end{bmatrix} + \frac{1}{2}\cdot\log\frac{4}{3} = \frac{1}{8} + \frac{1}{2}\log\frac{4}{3}$$

και το όριο λάθους Bhattacharya δίνεται από την σχέση : (*εξίσωση 2.75 των σημειώσεων*)

$$P(\lambda\acute{\alpha}\theta o\upsilon\varsigma) \le \sqrt{P(\omega_1)P(\omega_2)}\cdot e^{\,k(1/2)}$$ οπότε έχουμε :

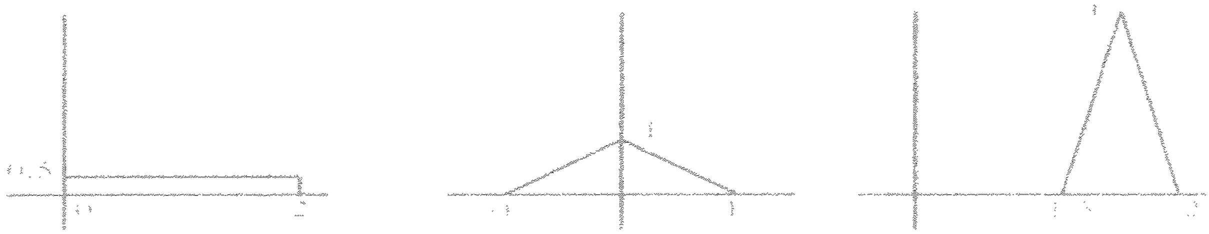
$$P(\lambda\acute{\alpha}\theta o\upsilon\varsigma) \le \sqrt{\frac{1}{2}\cdot\frac{1}{2}}\cdot e^{\,k(1/2)} = 0.382$$

Γ) Από τα 1000 πρότυπα που παράγονται για την κλάση ω_1, τα 308 ταξινομούνται λάθος.
Ομοιώς, από τα 1000 πρότυπα που παράγονται για την κλάση ω_2, τα 250 ταξινομούνται λάθος.
Ο ρυθμός λάθους ταξινόμησης είναι :

$$Error\ rate = \frac{1}{2}\cdot\frac{308}{1000} + \frac{1}{2}\cdot\frac{250}{1000} = 27.9\%$$

Άσκηση 3

Οι παρακάτω γραφικές παραστάσεις δείχνουν τις υπό συνθήκη πυκνότητες πιθανότητας $p(\mathbf{x}/\omega_i)$ για τις κλάσεις ω_i και το μονοδιάστατο διάνυσμα παρατήρησης x.



Α) Βρείτε τις περιοχές ελάχιστου λάθους και τις αντίστοιχες πιθανότητες λάθους, θεωρώντας ότι $p(\omega_1) = 0.25$,, $p(\omega_2) = 0.25$, $p(\omega_3) = 0.5$.

Β) Βρείτε τις περιοχές απόφασης ελάχιστου λάθους θεωρώντας ότι $p(\omega_1) = p(\omega_2) = p(\omega_3)$.

ΘΕΩΡΙΑ ΑΠΟΦΑΣΕΩΝ
3° ΦΡΟΝΤΙΣΤΗΡΙΟ

ΛΥΣΕΙΣ ΤΩΝ ΑΣΚΗΣΕΩΝ

ΑΣΚΗΣΗ 1

Σύμφωνα με στοιχεία από το Πανεπιστήμιο της Οξφόρδης η πιθανότητα ένας φοιτητής να αποφοιτήσει μέσα σε 5 χρόνια από την ημέρα εγγραφής του στο Πανεπιστήμιο είναι 0.8.
Το Oxford University Testing Service score (OUT score) όσων φοιτητών αποφοιτούν μέσα σε 5 χρόνια ακολουθεί κανονική κατανομή με μέση τιμή 26 και τυπική απόκλιση 2. Τα scores όσων δεν αποφοιτούν μέσα σε 5 χρόνια ακολουθούν επίσης κανονική κατανομή, αλλά με μέση τιμή 22 και τυπική απόκλιση 3. Στο Πανεπιστήμιο για την ταξινόμηση των φοιτητών χρησιμοποιείται επιπροσθέτως ο βαθμός τάξης (παράμετρος y) (η βαθμολογία που παίρνει ένας φοιτητής σε κάποιο τεστ σε σύγκριση με τη βαθμολογία των συμφοιτητών του).
Έστω G το γεγονός κατά το οποίο ένας φοιτητής ολοκληρώνει τη φοίτησή του στα 5 έτη και nG το γεγονός κατά το οποίο ένας φοιτητής δεν ολοκληρώνει τη φοίτησή του στα 5 έτη. Θεωρούμε μία τυχαία τιμή x του OUT score.
Υποθέτουμε ότι οι υπό συνθήκη πιθανοφάνειες τόσο του x όσο και του y είναι κανονικές, δύο μεταβλητών.

$P(x,y|G)=N(\mu_G|\Sigma_G)$ με $\mu_G=[26,85]^t$ και $\Sigma_G=\begin{pmatrix}4 & 6\\ 6 & 40\end{pmatrix}$

$P(x,y|nG)=N(\mu_{nG}|\Sigma_{nG})$ με $\mu_{nG}=[26,85]^t$ και $\Sigma_{nG}=\begin{pmatrix}9 & 12\\ 12 & 58\end{pmatrix}$

A) Ποια είναι η πιθανότητα ένας φοιτητής με OUT score 21.5 και βαθμό τάξης 72 να αποφοιτήσει μέσα σε 5 έτη;
B) Να βρείτε την εξίσωση του βέλτιστου ορίου απόφασης για τις δύο κατηγορίες G και nG.

ΛΥΣΗ

A) Αναζητούμε την πιθανότητα P(G | x=21.5, y=72).
Σύμφωνα με τον κανόνα του Bayes ισχύει :

$$P(G|x,y)=P(x,y|G)\frac{P(G)}{P(x,y)}$$

Σύμφωνα με τον τύπο της συνάρτησης πυκνότητας πιθανότητας πολλών μεταβλητών ισχύει :

$$P(x,y|G)=\frac{1}{(2\pi)^{d/2}}\frac{1}{|\Sigma_G|^{1/2}}\exp[-1/2(x-\mu)^t\sum\nolimits_G^{-1}(x-\mu)]$$, όπου d = 2.

Υπολογίζουμε τον αντίστροφο του πίνακα συνδιασποράς Σ_G :

$$\sum\nolimits_G^{-1}=\begin{pmatrix}0.3226 & -0.0484\\ -0.0484 & 0.0323\end{pmatrix} \quad \text{και} \quad |\Sigma_G|=124$$

Αντικαθιστώντας στον τύπο της συνάρτησης πυκνότητας πιθανότητας προκύπτει :

$$P(x,y\mid G) = 0.0143\exp[-1/2(x-\mu_G)^t \textstyle\sum_G^{1}(x-\mu_G)]$$

P(21.5, 72 | G) = $6.0591*10^{-4}$

Σύμφωνα με το Θεώρημα Ολικής Πιθανότητας :

P(x,y) = P(x,y | G) * P(G) + P(x,y | nG) * P(nG)

Υπολογίζουμε τον αντίστροφο του πίνακα συνδιασποράς Σ_{nG} :

$$\sum_{nG}^{1} = \begin{pmatrix} 0.1534 & -0.0317 \\ -0.0317 & 0.0238 \end{pmatrix} \quad \text{και} \quad |\Sigma_{nG}| = 378$$

Αντικαθιστώντας στον τύπο της συνάρτησης πυκνότητας πιθανότητας προκύπτει :

$$P(x,y\mid nG) = \frac{1}{(2\pi)^{d/2}}\frac{1}{|\Sigma_{nG}|^{1/2}}\exp[-1/2(x-\mu_{nG})^t \textstyle\sum_{nG}^{1}(x-\mu_{nG})] \quad \text{όπου } d = 2$$

$$P(x,y\mid nG) = 0.0082\exp[-1/2(x-\mu_{nG})^t \textstyle\sum_{nG}^{1}(x-\mu_{nG})]$$

Άρα

P(21.5 , 72 | nG) = 0.0074

Αντικαθιστώντας στο Θεώρημα Ολικής Πιθανότητας προκύπτει :

P(21.5, 72 | nG) = 0.0020

Άρα η πιθανότητα να αποφοιτήσει ο φοιτητής στα 5 έτη είναι :

P(G | x=21.5, y=72) = 0.2460

Β) Η εξίσωση απόφασης πρέπει να ικανοποιεί την παρακάτω εξίσωση:

$$P(G\mid x,y) - P(nG\mid x,y) \Rightarrow$$

$$\frac{P(x,y\mid G)P(G)}{P(x,y)} = \frac{P(x,y\mid nG)P(nG)}{P(x,y)} \Rightarrow$$

$$0.0143\exp[-1/2(x-\mu_G)^t \textstyle\sum_G^{1}(x-\mu_G)]0.8 = 0.0082\exp[-1/2(x-\mu_{nG})^t \textstyle\sum_{nG}^{1}(x-\mu_{nG})]0.2 \Rightarrow$$

$$\exp[-1/2(x-\mu_G)^t \textstyle\sum_G^{1}(x-\mu_G) + 1/2(x-\mu_{nG})^t \textstyle\sum_{nG}^{1}(x-\mu_{nG})] = \frac{0.0082}{0.0143*4} = 0.1434$$

$$\Rightarrow$$

$$(-1/2)(x-\mu_G)^t \textstyle\sum_G^{1}(x-\mu_G) + 1/2(x-\mu_{nG})^t \textstyle\sum_{nG}^{1}(x-\mu_{nG}) = \log(0.1434) \Rightarrow$$

$$-(x-\mu_G)^t \textstyle\sum_G^{1}(x-\mu_G) + (x-\mu_{nG})^t \textstyle\sum_{nG}^{1}(x-\mu_{nG}) + 3.8848 = 0$$

Έστω το διάνυσμα **x** = [x y]. Τότε :

-([x-26 y-85]) $\sum_G^{1}$ ([x-26 y-85] $)^t$ + ([x-22 y-70]) $\sum_{nG}^{1}$ ([x-22 y-70] $)^t$ + 3.8848 = 0

από το οποίο προκύπτει η ζητούμενη εξίσωση :

$0.476x^2 + 0.0561y^2 - 0.1602xy - 10.1588x - 4.9114y + 334.6539 = 0$

ΑΣΚΗΣΗ 2 *Ταξινομητής ελάχιστης απόστασης*

Θεωρούμε ένα πρόβλημα classification σε δύο διαστάσεις, με 3 classes, όπου: $p(x|\omega_i)=N(\mu_i|\Sigma_i)$, με i=1,2,3

με $\mu_1=[\ 0\ \ 2\ ]^t$, $\mu_2=[\ 3\ \ 1\ ]^t$ και $\mu_3=[\ 1\ \ 0\ ]^t$ και $\Sigma_i=\Sigma=\begin{pmatrix} 1 & 0 \\ 0 & 1/3 \end{pmatrix}$

και στο οποίο υποθέτουμε ίσες εκ των προτέρων πιθανότητες: $P(\omega_1)= P(\omega_2)= P(\omega_3)$.

Α) Να υπολογίσετε τη διακρίνουσα συνάρτηση $g_i(x)$ για κάθε μία class.

Β) Να καθορίσετε τα όρια απόφασης. Πόσα είναι; Να αναφέρετε πώς πιστεύετε ότι θα είναι η γεωμετρική τους μορφή.

ΛΥΣΗ

Α) Υπολογίζουμε $\sum=\begin{pmatrix} 1 & 0 \\ 0 & 1/3 \end{pmatrix}$, $\sum^{-1}=\begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}$ και $|\Sigma| = 1/3$, όπου d = 2.

Εφόσον Sum(P) = 1 ισχύει $P(\omega_1) = P(\omega_2) = P(\omega_3)$.

Επιπλέον, αφού $p(x|\omega_i) \sim N(\mu_i|\Sigma_i)$, οι συναρτήσεις πυκνότητας πιθανότητας είναι κανονικής κατανομής συναρτήσεις πολλών μεταβλητών και η διακρίνουσα συνάρτηση θα είναι της μορφής :

$$g_i(x) = \frac{1}{2}(x-\mu_i)^t\sum_i^{-1}(x-\mu_i)-\frac{d}{2}\log 2\pi-\frac{1}{2}\log|\sum_i|+\log P(\omega)$$

$$g_i(x) = \frac{1}{2}(x-\mu_i)^t\sum_i^{-1}(x-\mu_i)-2.3872$$

Έστω το διάνυσμα $\mathbf{x} = [x\ \ y]^t$. Τότε :

$g_i(x) = 0.5\ (\ [x\ \ y]^t\ -\ \mu_i)^t\ \Sigma^{-1}\ (\ [x\ \ y]^t\ -\ \mu_i) - 2.3872$

Για i=1 έχουμε:

$\mu_1=[\ 0\ \ 2\ ]^t$

$$g_1(x)\ =\ 0.5\ (\ [x\ (y-2)])\ \Sigma^{-1}(\ [x\ (y-2)])^t\ -\ 2.3872$$

$$g_1(x)\ =\ 0.5\ (\ [x\ (y-2)])\begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}(\ [x\ (y-2)])^t\ -\ 2.3872$$

Άρα

$$g_1(x)=\frac{1}{2}x^2+\frac{3}{2}y^2-6y+3.6128$$

Για i=2 έχουμε:

$\mu_2=[\ 3\ \ 1\ ]^t$

$$g_2(x)\ =\ 0.5\ (\ [x-3\ y-1])\ \Sigma^{-1}(\ [x-3\ y-1])^t\ -\ 2.3872$$

$$g_2(x)\ =\ 0.5\ (\ [x-3\ y-1])\begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}(\ [x-3\ y-1])^t\ -\ 2.3872$$

Άρα

$$g_2(x)=\frac{1}{2}x^2-3x+\frac{3}{2}y^2-3y+3.6128$$

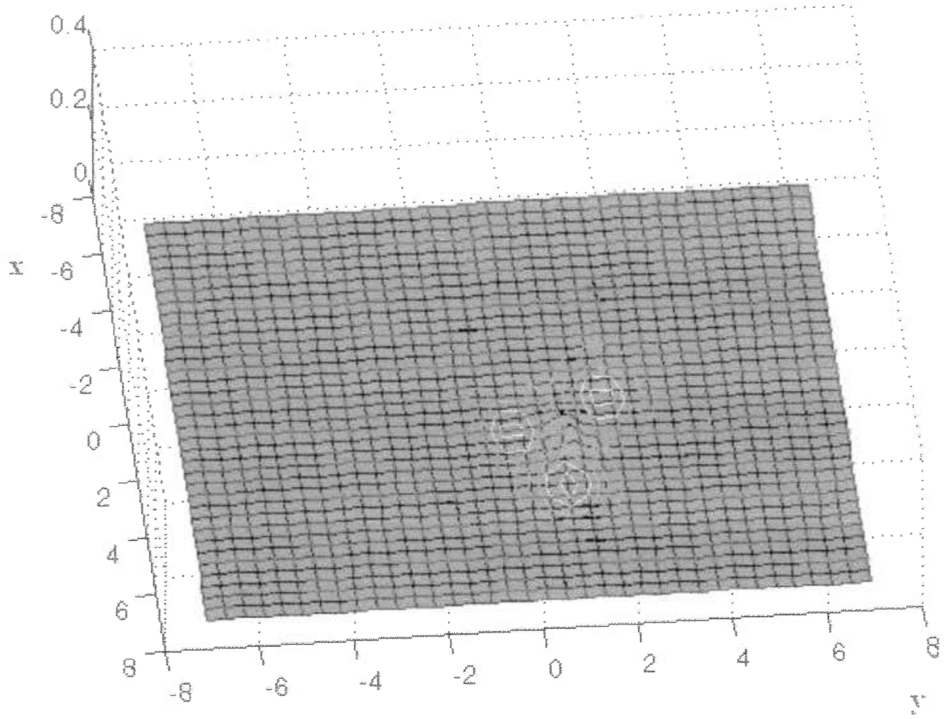
Για i=3 έχουμε:

$\mu_3=[\ 1\ \ 0\ ]^t$

$$g_3(x)\ =\ 0.5\ (\ [x-1\ y])\ \Sigma^{-1}(\ [x-1\ y])^t\ -\ 2.3872$$

Άρα

$$g_3(x) = \frac{1}{2}x^2 - x + \frac{3}{2}y^2 - 1.8872$$



B) Οι επιφάνειες απόφασης θα είναι υπερεπίπεδα επειδή οι τρεις συναρτήσεις πυκνότητας πιθανότητας έχουν τους ίδιους πίνακες συνδιασποράς (οι οποίοι είναι ίσοι και ανάλογοι του ταυτοτικού πίνακα). Είμαστε στις d=2 διαστάσεις άρα τα όρια είναι γενικευμένα υπερεπίπεδα των (d-1)=1 διαστάσεων (κάθετα στις γραμμές που χωρίζουν τα μέσα των κατανομών). Άρα έχουμε γραμμές (ευθείες) ως επιφάνειες των ορίων.

Οι επιφάνειες απόφασης ορίζονται από τις γραμμικές εξισώσεις

$g_i(x) = g_j(x)$, με $i \neq j$, $i,j \in \{1,2,3\}$

(Προσοχή! Εάν είχαμε $P(\omega_1) \neq P(\omega_2) \neq P(\omega_3)$ θα διαλέγαμε για την παραπάνω εξίσωση τις δύο κατηγορίες με τις μεγαλύτερες εκ των υστέρων πιθανότητες.)

Θέλουμε να βρούμε 3 ημιευθείες και 3 σημεία ελέγχου.

Έχουμε:

$$g_i(x) = g_j(x)$$

$$w_i^t x + w_{i0} = g_j(x)$$

όπου $w_i = \frac{1}{\sigma^2}\mu_i$ και $w_{i0} = (\frac{-1}{2\sigma^2})\mu_i^t \mu_i + \ln P(w_i)$ (Το κατώφλι της i-οστής κατηγορίας).

Αφού οι περιοχές απόφασης είναι γειτονικές τα όρια απόφασης μεταξύ τους έχουν την εξίσωση:

$$W_{ij}^t (x - X_{ij}) = 0$$

$$W_{ij} = \Sigma^{-1}(\mu_i - \mu_j) \quad \text{και}$$

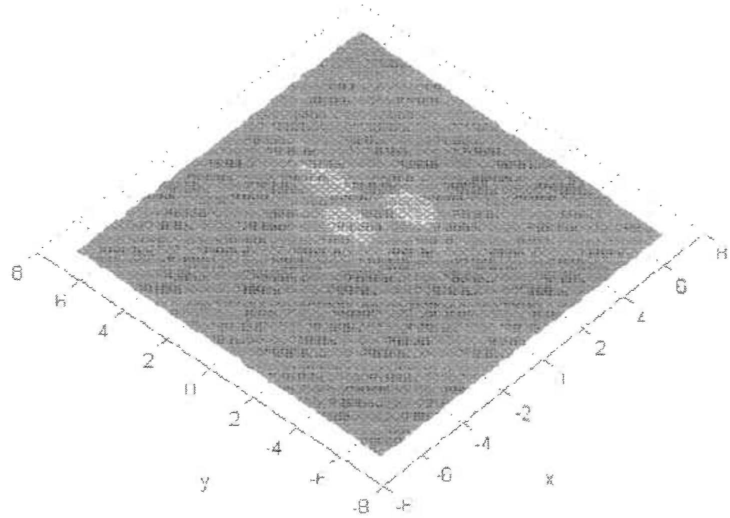
$$x_{ij} = 0.5(\mu_i + \mu_j) - \frac{(\mu_i - \mu_j)\log[P(w_i)/P(w_j)]}{(\mu_j - \mu_i)^t \Sigma^{-1}(\mu_j - \mu_i)}$$

Επειδή $P(W_i)= P(W_j)$ ο δεύτερος όρος ισούται με μηδέν και έχουμε $x_{ij} = 0.5(\mu_i + \mu_j)$.
Ορίσαμε λοιπόν ένα υπερεπίπεδο που περνάει από το σημείο x_0 και είναι ορθογώνιο στο διάνυσμα W.
Έχουμε:
$W_{ij}^t(X_c - X_{ij}) = 0$ και $W_{ij+1}^t(X_c - X_{ij+1}) = 0$

Υπολογίζουμε με τη βοήθεια της Matlab:

```
x12 =
    1.5000
    1.5000

w12 =
   -3.0000
    3.0000
x13 =

    0.5000
    1.0000
w13 =

   -1.0000
    6.0000
x23 =

    2.0000
    0.5000

w23 =
    2.0000
    3.0000
xc12_13 =
    1.1000
    1.1000
xc12_23 =
    1.1000
    1.1000
xc13_23 =
    1.1000
    1.1000
```

Άρα έχουμε 3 ημιευθείες που περνούν από το ίδιο σημείο [1.1 1.1], αφού τα σημεία ελέγχου συμπίπτουν.



ΑΣΚΗΣΗ 3

Έστω ότι η τυχαία μεταβλητή x έχει εκθετική πυκνότητα πιθανότητας

$$p(x|\theta) = \begin{cases} \theta e^{-\theta x} & x \geq 0 \\ 0 & \alpha\lambda\lambda o\acute{\upsilon} \end{cases} \qquad \text{με } \theta>0$$

Υποθέστε ότι τα δείγματα $x_1, x_2, \ldots, x_n$ σχηματίζονται ανεξάρτητα σύμφωνα με το $p(x|\theta)$. Να δείξετε ότι η εκτίμηση μέγιστης πιθανοφάνειας του θ δίνεται από το

$$\hat{\theta} = \frac{1}{\frac{1}{n}\sum_{k=1}^{n} x_k}$$

ΛΥΣΗ

Έστω τα n δείγματα που σχηματίζονται ανεξάρτητα, $D=\{x_1, x_2, \ldots, x_n\}$.
Η εκτίμηση της μέγιστης πιθανοφάνειας του θ είναι εξ ορισμού η τιμή $\hat{\theta}$ που μεγιστοποιεί το $p(D|\theta)$, με $p(D|\theta) = \prod_{k=1}^{n} p(x_k|\theta)$.

Η log-συνάρτηση πιθανοφάνειας είναι:

$$l = \ln(p(D|\theta)) = \ln\left(\prod_{k=1}^{n} p(x_k|\theta)\right) = \sum_{k=1}^{n}\ln(p(x_k|\theta)) = \sum_{k=1}^{n}\ln(\theta e^{-\theta x_k}) = \sum_{k=1}^{n}(\ln\theta - \theta x_k)$$

$$= n^* \ln\theta - \theta\sum_{k=1}^{n} x_k$$

Το $\hat{\theta}$ μπορεί να βρεθεί με μεθόδους διαφορικής λογικής. Παίρνοντας το ανάδελτα ως προς θ προκύπτει : $\nabla l = \frac{\partial}{\partial\theta}(n\ln\theta - \theta\sum_{k=1}^{n} x_k) = \frac{n}{\theta} - \sum_{k=1}^{n} x_k$.

Για να βρούμε τη εκτίμηση μέγιστης πιθανοφάνειας εξισώνουμε το παραπάνω με μηδέν και προκύπτει :

$$\forall l = 0 \Rightarrow \frac{n}{\hat{\theta}_{MLE}} - \sum_{k=1}^{n} x_k = 0 \Rightarrow \hat{\theta}_{MLE} = \frac{n}{\sum_{k=1}^{n} x_k} = \frac{1}{\frac{1}{n}\sum_{k=1}^{n} x_k}$$

Μία λύση $\hat{\theta}_{MLE}$ της παραπάνω εξίσωσης μπορεί να αντιπροσωπεύει ένα αληθινό, ολικό μέγιστο, ένα τοπικό μέγιστο ή ελάχιστο ή σπανιότερα ένα σημείο του l(θ).

ΑΣΚΗΣΗ 4

Έστω ότι το θ αναπαριστά την πιθανότητα να έρθει κορώνα σε ένα πρόβλημα ρίψης ενός κέρματος. Έχουμε διαθέσιμο ένα σύνολο από n ανεξάρτητες και ομοιόμορφα κατανεμημένες τυχαίες παρατηρήσεις για να υπολογίσουμε το θ.

A) Να βρείτε την Bayes εκτίμηση του θ για την εκ των προτέρων πυκνότητα εάν $p(\theta) = 1,\ 0 \le \theta \le 1$ (Δηλαδή όλες οι τιμές του θ μεταξύ 0 και 1 είναι εξίσου πιθανές).

B) Ποια είναι η MLE του θ? Πώς διαφέρει από την Bayes εκτίμηση του πρώτου ερωτήματος?

Δίνεται $\int_0^1 \theta^p (1-\theta)^q d\theta = \frac{p!q!}{(p+q+1)!}$ και να υποθέσετε συνάρτηση τετραγωνικού σφάλματος για το θ.

ΛΥΣΗ

A) Έστω θ η πιθανότητα να έρθει κορώνα σε ένα πρόβλημα ρίψης κέρματος.

Έστω n_H το πλήθος των φορών που εμφανίστηκε κορώνα σε n ρίψεις του νομίσματος.

Η πιθανότητα να πάρουμε ακριβώς n_H επιτυχίες σε n προσπάθειες δίνεται από την probability mass function :

$$p(n_H|\omega) = p(n_H|\theta) = \binom{n}{n_H} \theta^{n_H} (1-\theta)^{n-n_H}$$

Η εκ των υστέρων πιθανότητα του θ δεδομένου ότι n_H φορές έχει έρθει κορώνα είναι σύμφωνα με τον κανόνα του Bayes:

$$p(\theta|n_H) = \frac{p(n_H|\theta)p(\theta)}{\int_0^1 p(n_H|\theta)p(\theta)}$$

Οπότε προκύπτει

$$\int_0^1 p(n_H|\theta)p(\theta) = \int_0^1 \binom{n}{n_H} \theta^{n_H} (1-\theta)^{n-n_H} =$$

$$= \binom{n}{n_H} \frac{n_H!(n-n_H)!}{(n+1)!} = \frac{n!}{(n+1)!} = \frac{n!}{n!(n+1)} = \frac{1}{n+1}$$

Άρα

$$p(\theta|n_H) = (n+1)\binom{n}{n_H} \theta^{n_H} (1-\theta)^{n-n_H}$$

Για συνάρτηση κόστους τετραγωνικού σφάλματος αναζητούμε την τιμή του $\hat{\theta}$ η οποία ελαχιστοποιεί την ποσότητα $(\theta-\hat{\theta})^2$. Επομένως:

$$\hat{\theta}_{Bayes} = E_{\theta|n_H}[\theta] = \int \theta p(\theta \mid n_H) d\theta$$

Άρα

$$\hat{\theta}_{Bayes} = (n+1)\binom{n}{n_H}\int_0^1 \theta^{n_H+1}(1-\theta)^{n-n_H} d\theta = \frac{n_H+1}{n+2}$$

Β) Η log-συνάρτηση πιθανοφάνειας είναι:

$$l(\theta) = \log p(\theta \mid n_H)$$

$$l(\theta) = \log\binom{n}{n_H} + (n-n_H)\log(1-\theta) + (n_H+1)\log\theta$$

όπου $C = \binom{n}{n_H}$ σταθερά ανεξάρτητη του θ.

Η MLE του θ δίνεται αν μεγιστοποιήσουμε το l(θ).

Διαφορίζοντας το l(θ) ως προς θ και εξισώνοντας με μηδέν την παράγωγο έχουμε:

$$\frac{\partial l(\theta)}{\partial \theta} = \frac{n_H+1}{\theta} - \frac{(n-n_H)}{(1-\theta)} = 0$$

$$\theta = \frac{n_H}{n}$$

Άρα $\hat{\theta}_{MLE} - \frac{n_H}{n}$.

Η εκτίμηση Bayes στο πρώτο ερώτημα λαμβάνει υπόψη την εκ των προτέρων πληροφορία σχετικά με το θ. Κατά συνέπεια, η εκτίμηση Bayes υπάρχει ακόμα και όταν δεν έχουμε διαθέσιμα δείγματα εκπαίδευσης δηλαδή όταν n = 0.

Η Bayes εκτίμηση είναι σχεδόν ίση με την MLE όταν η εκ των προτέρων πιθανότητα είναι flat και η $p(n_H \mid \theta)$ έχει peaks (αυτό συνήθως συμβαίνει όταν $n \to \infty$)

Θεωρία Αποφάσεων
4° Φροντιστήριο

Λύσεις των Ασκήσεων

Άσκηση 1

Έστω ένα πρόβλημα ταξινόμησης μιας διάστασης με δύο κατηγορίες, όπου για κάθε κατηγορία έχουν συλλεχθεί τα παρακάτω δεδομένα:

$$\mathbf{D}_1 = \{-1,-2,3,3,6,7\} \text{ και } \mathbf{D}_2 = \{-3,-2,3,5,8\}.$$

Έστω ότι αποφασίζουμε να χρησιμοποιήσουμε παράθυρα Parzen με μέγεθος παραθύρου ίσο με $h = 2$ και συνάρτηση παραθύρου $\varphi(x)$ που ορίζεται ως:

$$\varphi(u) = \begin{cases} -\frac{3}{4}u^2 + \frac{3}{4} & \text{εάν } -1 \le u \le 1 \\ 0 & \text{αλλού} \end{cases}$$

A) Ταξινομήστε το δείγμα $x = 4$ χρησιμοποιώντας τον *ML* ταξινομητή.
B) Αποδείξτε ότι η εκτιμήσεις των συναρτήσεων πυκνότητας πιθανότητας που γίνονται με χρήση των παράθυρων Parzen με την φ(x) να έχει τη μορφή που ορίστηκε παραπάνω είναι στην πραγματικότητα συναρτήσεις πυκνότητας πιθανότητας.

Λύση

A) $p_1(4) = \frac{1}{6}\sum_{i=1}^{6}\frac{1}{2}\varphi\left(\frac{4-x_i}{2}\right) = \frac{1}{12}\left(\varphi\left(\frac{1}{2}\right)+\varphi\left(\frac{1}{2}\right)\right) = \frac{\varphi\left(\frac{1}{2}\right)}{6}$

$$p_2(4) = \frac{1}{5}\sum_{i=1}^{5}\frac{1}{2}\varphi\left(\frac{4-x_i}{2}\right) = \frac{1}{10}\left(\varphi\left(\frac{1}{2}\right)+\varphi\left(-\frac{1}{2}\right)\right) = \frac{\varphi\left(\frac{1}{2}\right)}{5}$$

όπου το τελευταίο προκύπτει διότι $\varphi\left(\frac{1}{2}\right) = \varphi\left(-\frac{1}{2}\right)$.

Επομένως, το σημείο $x = 4$ πρέπει να ταξινομηθεί στην κατηγορία 2.

B) *Για να αποδείξουμε ότι οι συναρτήσεις που υπολογίζονται είναι στην* πραγματικότητα συναρτήσεις πυκνότητας πιθανότητας πρέπει να δείξουμε ότι η συνάρτηση φ(u) είναι μη αρνητική και ότι το ολοκλήρωμά της ισούται με 1.
Προφανώς, $\varphi(u) \ge 0$, αφού το $-\frac{3}{4}u^2 + \frac{3}{4}$ παίρνει μη αρνητικές τιμές στο διάστημα [-1, 1].

$$\int_{-\infty}^{\infty}\varphi(u)du = \int_{-1}^{1}\left(-\frac{3}{4}u^2 + \frac{3}{4}\right)du = -\frac{3}{12}u^3 + \frac{3}{4}u\Big|_{-1}^{1} = \frac{6}{12} + \frac{6}{12} = 1.$$

Άσκηση 2

Θεωρήστε έναν k-NN classifier και τα δείγματα δύο διαστάσεων :

ω_1	(6,5)	(12,3)	(6,2)
ω_2	(2,7)	(3,14)	(0,16)
ω_3	(2,10)	(1,9)	(1,17)

Κατατάξετε το δείγμα (5,15) για k=1.

Λύση

Υπολογίζουμε τις αποστάσεις του δείγματος απ΄όλα τα διανύσματα των κλάσεων ω_1, ω_2 και ω_3. Το πλησιέστερο στο δείγμα διάνυσμα είναι το (3,14) (ω_2), με Ευκλείδια απόσταση d = 2.23. Οι αποστάσεις όλων των άλλων σημείων είναι μεγαλύτερες. Σύμφωνα με τον ταξινομητή k=1 πλησιέστερου γείτονα ταξινομούμε το δείγμα στην κλάση ω_2, εφόσον ο πλησιέστερος γείτονας ανήκει στην ω_2.

Άσκηση 3 (Παραλλαγή κανόνα πλησιέστερου γείτονα)

Δίνονται τα ακόλουθα παραδείγματα δύο κατηγοριών :

Κατηγορία	π_1	π_2	π_3	π_4
ω_1	(-3,-1)	(1,1)	(2,0)	(3,1)
ω_2	(-2,0)	(0,0)	(1,-1)	

Κατασκευάστε σύστημα αναγνώρισης προτύπων χρησιμοποιώντας
a) Την συνάρτηση της Ευκλείδιας απόστασης
b) Κριτήριο ταξινόμησης την μικρότερη απόσταση προτύπων
c) Διαδικασία εκπαίδευσης κατά την οποία να επιλέξετε σαν πρωτότυπο εκείνο το οποίο έχει την μικρότερη αθροιστική απόσταση από τα υπόλοιπα της κατηγορίας του

Ταξινομείστε το δείγμα (1,2).

Λύση

Αρχικά πρέπει να βρούμε το πρότυπο εκείνο το οποίο έχει την μικρότερη αθροιστική απόσταση από τα υπόλοιπα της κατηγορίας του. Υπολογίζουμε την Ευκλείδια απόσταση των προτύπων των κατηγοριών:

Ευκλείδιες αποστάσεις των προτύπων της ω_1

Πρότυπο	π_1	π_2	π_3	π_4	Άθροισμα
π_1	0.0	4.472	5.099	6.324	15.895
π_2	4.472	0.0	1.414	2.0	7.886
π_3	5.099	1.414	0.0	1.414	7.927
π_4	6.324	2.0	1.414	0.0	9.738

Το άθροισμα των γραμμών του πίνακα μας δίνει την αθροιστική απόσταση κάθε προτύπου από τα υπόλοιπα πρότυπα της κατηγορίας του. Το δεύτερο πρότυπο (π_2), το οποίο έχει την μικρότερη αθροιστική απόσταση από τα υπόλοιπα πρότυπα της κατηγορίας του, θα επιλέξουμε σαν πρωτότυπο της κατηγορίας ω_1.

Εκτελούμε την ίδια εργασία και για τα παραδείγματα της δεύτερης κατηγορίας:

Ευκλείδιες αποστάσεις των προτύπων της ω_2

Πρότυπο	π_1	π_2	π_3	Άθροισμα
π_1	0.0	2.0	3.162	5.162
π_2	2.0	0.0	1.414	3.414
π_3	3.162	1.414	0.0	4.576

Με το ίδιο κριτήριο επιλέγουμε το παράδειγμα (0,0) σαν πρωτότυπο της κατηγορίας ω_2.

Για να κατατάξουμε το δείγμα (1,2) υπολογίζουμε τις Ευκλείδιες αποστάσεις του δείγματος από τα πρωτότυπα της κάθε κατηγορίας. Χρησιμοποιώντας σαν κριτήριο ταξινόμησης την μικρότερη απόσταση προτύπων το δείγμα ταξινομείται στην κλάση ω_1 από το πρωτότυπο της οποίας απέχει την μικρότερη απόσταση.

Άσκηση 4

Έστω ένα πρόβλημα ταξινόμησης μιας διάστασης με τρεις κατηγορίες, όπου για κάθε κατηγορία έχουν συλλεχθεί τα παρακάτω δεδομένα:

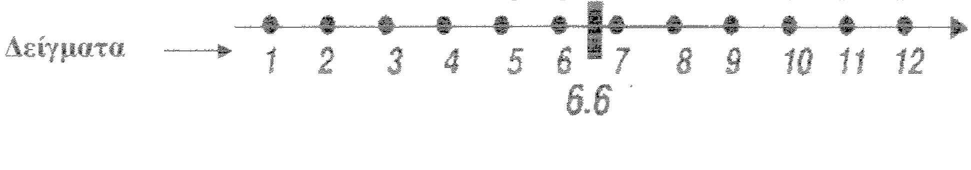
$$\mathbf{D}_1 = \{1,7,8,9,10\},\ \mathbf{D}_2 = \{2,3,4,11,12\} \text{ και } \mathbf{D}_3 = \{5,6\}.$$

Α) Ταξινομήστε το δείγμα 6.6 χρησιμοποιώντας των αλγόριθμο των *k* πλησιέστερων γειτόνων με *k = 6*.

Β) *Εφαρμόστε τον αλγόριθμο περικοπής και υπολογίστε τα καινούρια σύνολα* δειγμάτων που προκύπτουν.

Λύση

Α)



Οι 6 πλησιέστεροι γείτονες του δείγματος 6.6 είναι οι 7, 8 ,9 από την κατηγορία 1, οι 5, 6 από την κατηγορία 3 και ο 4 από την κατηγορία 2.

Επομένως ο μεγαλύτερος αριθμός γειτόνων ανήκει στην κατηγορία 1 και επομένως το δείγμα 6.6 πρέπει να ταξινομηθεί στην κατηγορία 1.

Β) Τα δείγματα 3, 8, 9, 12 έχουν όλους τους γείτονές τους μέσα στο διάγραμμα Voronoi της ίδιας κατηγορίας και επομένως μπορούν να περικοπούν.
Τα ελαττωμένα σύνολα δειγμάτων που προκύπτουν είναι τα εξής:

$\mathbf{D}_1 = \{1,7,10\}$, $\mathbf{D}_2 = \{2,4,11\}$ και $\mathbf{D}_3 = \{5,6\}$.

Άσκηση 5

Δίνονται τα παρακάτω δεδομένα εκπαίδευσης. Ταξινομήστε το δείγμα χρησιμοποιώντας τον αλγόριθμο των k πλησιέστερων γειτόνων με k = 5 .

Δείγμα: ηλικία ≤ 30, εισόδημα = μεσαίο, μαθητής = ναι, αξιολόγηση φερεγγυότητας = μέτρια

Χρησιμοποιήστε σαν μετρική ομοιότητας την εξής συνάρτηση:

$$\text{Ομοιότητα(A,B)} = \frac{\sum_{i=1}^{4} w_i \partial(a_i, b_i)}{4}$$

Όπου $\partial(a_i, b_i)$ είναι 1 αν a_i είναι ίσο με b_i και 0 αλλιώς.
Τα a_i και b_i είναι είτε η ηλικία, το εισόδημα, ο μαθητής ή η αξιολόγηση φερεγγυότητας.
Τα βάρη w_i είναι 1 για όλα τα χαρακτηριστικά εκτός από το εισόδημα που είναι 2.

<u>Δεδομένα εκπαίδευσης</u>

Δεδομένο	ηλικία	εισόδημα	μαθητής	Αξιολόγηση φερεγγυότητας	Κλάση:αγοράζει υπολογιστή
1	≤30	υψηλό	οχι	μέτρια	οχι
2	≤30	υψηλό	οχι	υψηλή	οχι
3	31...40	υψηλό	οχι	μέτρια	ναι
4	>40	μέτριο	οχι	μέτρια	ναι
5	>40	χαμηλό	ναι	μέτρια	ναι
6	>40	χαμηλό	ναι	υψηλή	οχι
7	31...40	χαμηλό	ναι	υψηλή	ναι
8	≤30	μέτριο	οχι	μέτρια	οχι
9	≤30	χαμηλό	ναι	μέτρια	ναι
10	>40	μέτριο	ναι	μέτρια	ναι
11	≤30	μέτριο	ναι	υψηλή	ναι
12	31...40	μέτριο	οχι	υψηλή	ναι
13	31...40	υψηλό	ναι	μέτρια	ναι
14	>40	μέτριο	οχι	υψηλή	οχι

Λύση

Για κάθε δεδομένο εκπαίδευσης μετράμε την απόσταση που απέχει από το δείγμα.

Δεδομένο	Κλάση	Απόσταση απο δείγμα
1	οχι	(1+0+0+1)/4 = 0.5
2	οχι	(1+0+0+0)/4 = 0.25
3	ναι	(0+0+0+1)/4 = 0.25
4	**ναι**	**(0+2+0+1)/4 = 0.75**
5	ναι	(0+0+1+1)/4 = 0.5
6	οχι	(0+0+1+0)/4 = 0.25
7	ναι	(0+0+1+0)/4 = 0.25
8	**οχι**	**(1+2+0+1)/4 = 1**
9	**ναι**	**(1+0+1+1)/4 = 0.75**
10	**ναι**	**(0+2+1+1)/4 = 1**
11	**ναι**	**(1+2+1+0)/4 = 1**
12	ναι	(0+2+0+0)/4 = 0.5
13	ναι	(0+0+1+1)/4 = 0.5
14	οχι	(0+2+0+0)/4 = 0.5

Από τους 5 κοντινότερους γείτονες 4 ανήκουν στην κλάση *ναι* και 1 στην κλάση *οχι*. Άρα ο ταξινομητής θα κατατάξει το νέο δείγμα στην κλάση αγοράζει υπολογιστή = ναι.

5^o Φροντιστήριο

Λύσεις των Ασκήσεων

Άσκηση 1

Έστω μια τυχαία μεταβλητή x που έχει την παρακάτω συνάρτηση πυκνότητας πιθανότητας :

$$p(x \mid \theta) = \begin{cases} \theta \cdot e^{-\theta x} & x \geq 0 \\ 0 & \alpha\lambda\lambda o\acute{\upsilon} \end{cases}$$

Α) Σχεδιάστε τη γραφική παράσταση της $p(x/\theta)$ ως προς το x για $\theta=1$.
Σχεδιάστε την $p(x/\theta)$ ως προς το θ για $0 \leq \theta \leq 5$ και $x=2$.

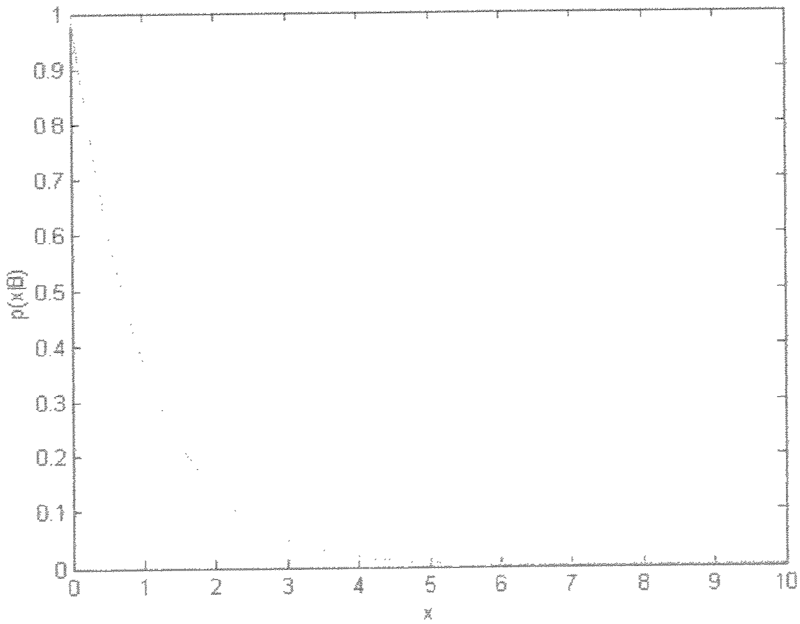
Β) Έστω ένα σύνολο από n δείγματα x_1, x_2,... x_n που επιλέγονται ανεξάρτητα με βάση την $p(x/\theta)$.Καθορίστε την εκτίμηση μέγιστης πιθανοφάνειας για την άγνωστη παράμετρο θ.

Λύση

Α) Για θ=1 είναι :

$$p(x \mid \theta) = \begin{cases} e^{-x} & x \geq 0 \\ 0 & \alpha\lambda\lambda o\acute{\upsilon} \end{cases}$$

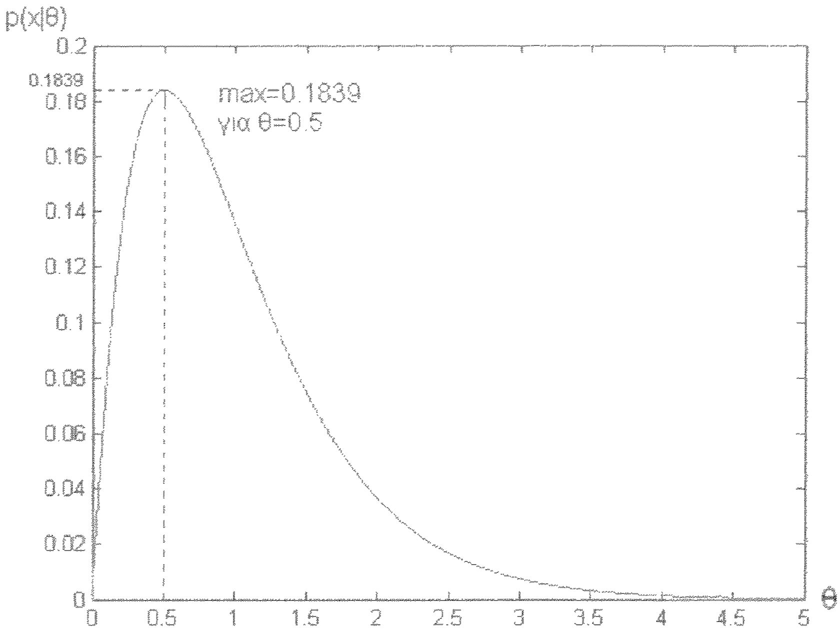
Η γραφική της παράσταση είναι :



Β)) Για $0 \leq \theta \leq 5$ και $x=2$ είναι :

$$p(x \mid \theta) = \begin{cases} \theta \cdot e^{2\theta} & 0 \le \theta \le 5 \\ 0 & \alpha\lambda\lambda o\acute{\upsilon} \end{cases}$$

Η γραφική της παράσταση είναι :



$$l(\theta) = \ln p(D \mid \theta) = \ln[p(x_1 \mid \theta) \cdot p(x_2 \mid \theta) \cdot \ldots \cdot p(x_n \mid \theta)] \Rightarrow$$

$$l(\theta) = \sum_{k=1}^{n} \ln(x_\kappa \mid \theta) = \sum_{k=1}^{n} \ln[\theta \cdot e^{\theta x_\kappa}] \Rightarrow \sum_{k=1}^{n} [\ln\theta - \theta \cdot x_\kappa]$$

gradient:

$$\nabla l(\theta) = \sum_{k=1}^{n} [\frac{1}{\theta} - x_\kappa]$$

Η MLE (εκτίμηση μέγιστης πιθανοφάνειας) βρίσκεται για $\nabla l(\theta) = 0$ και είναι:

$$\nabla l(\theta^*) = 0 \Rightarrow \sum_{k=1}^{n} [\frac{1}{\theta^*} - x_\kappa] = 0 \Rightarrow \frac{n}{\theta^*} - \sum_{k=1}^{n} x_\kappa = 0 \Rightarrow \frac{n}{\theta^*} = \sum_{k=1}^{n} x_\kappa \Rightarrow \theta^* = \frac{n}{\sum_{k=1}^{n} x_\kappa} \Rightarrow \theta^* = \frac{1}{\mu_n}$$

όπου μ_n η μέση τιμή των n δειγμάτων.

Για να βεβαιωθούμε ότι αυτή η τιμή αντιστοιχεί στο ολικό μέγιστο παίρνουμε τη δεύτερη παράγωγο:

$\nabla^2 l(\theta) = \sum_{k=1}^{n} \frac{-1}{\theta^2} < 0$ οπότε η τιμή που βρήκαμε για το θ είναι ολικό μέγιστο.

Άσκηση 2

Έστω μια τυχαία μεταβλητή που ακολουθεί την ακόλουθη συνάρτηση πυκνότητας πιθανότητας (κατανομή Erlang):

$$p(x \mid \theta) = \begin{cases} \theta^2 \cdot x \cdot e^{-\theta x}, & x \in [0,1] \\ 0, & \alpha\lambda\lambda o\acute{\upsilon} \end{cases}$$

Έστω n ανεξάρτητα δείγματα $x_1, x_2, \ldots, x_n$ που επιλέγονται ανεξάρτητα με βάση την *p(x/θ)*. Καθορίστε την εκτίμηση μέγιστης πιθανοφάνειας για την άγνωστη παράμετρο θ.

Λύση

$$l(\theta) = \ln p(D \mid \theta) = \ln(p(x_1 \mid \theta) \cdot p(x_2 \mid \theta) \ldots p(x_n \mid \theta)) \Rightarrow$$

$$l(\theta) = \sum_{k=1}^{n} \ln p(x_k \mid \theta) = \sum_{k=1}^{n} \ln[\theta^2 x_k e^{-\theta x_k}] = \sum_{k=1}^{n} (2 \ln \theta + \ln x_k - \theta \cdot x_k)$$

Το gradient ισούται με:

$$\nabla l(\theta) = \sum_{k=1}^{n} \left(\frac{2}{\theta} - x_k\right)$$

Η εκτίμηση μέγιστης πιθανοφάνειας βρίσκεται για $\nabla l(\theta^*) = 0$

$$\nabla l(\theta^*) = 0 \Rightarrow \sum_{k=1}^{n} \left(\frac{2}{\theta^*} - x_k\right) \Rightarrow \frac{2n}{\theta^*} = \sum_{k=1}^{n} x_k \Rightarrow \theta^* = \frac{2n}{\sum_{k=1}^{n} x_k} \Rightarrow \theta^* = \frac{2}{\mu_n}$$

όπου μ_n η μέση τιμή των n δειγμάτων.

Για να βεβαιωθούμε ότι αυτή η τιμή αντιστοιχεί στο ολικό μέγιστο παίρνουμε τη δεύτερη παράγωγο:

$\nabla^2 l(\theta) = \sum_{k=1}^{n} \frac{-2}{\theta^2} < 0$ οπότε η τιμή που βρήκαμε για το θ είναι ολικό μέγιστο.

Άσκηση 3

Η εκτίμηση μέγιστης πιθανοφάνειας δουλεύει για ομοιόμορφα κατανεμημένα σύνολα εκπαίδευσης;

Λύση

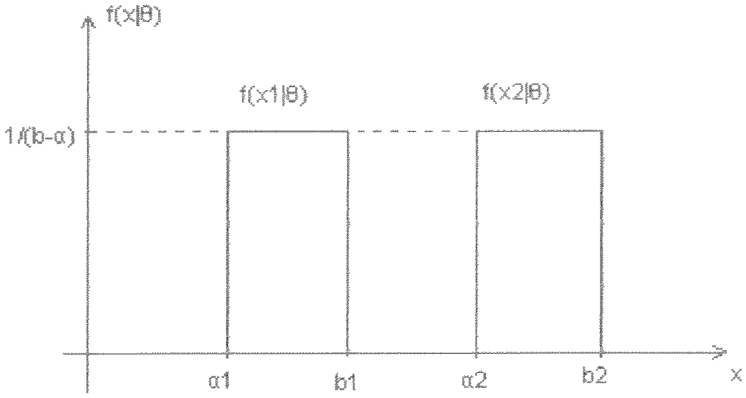
Η μέθοδος εκτίμησης μέγιστης πιθανοφάνειας είναι η εξής:
Δοκιμάζουμε διάφορες τιμές για την παράμετρο θ : $\theta_1, \theta_2, \ldots, \theta_r$ για να πάρουμε την μέγιστη τιμή για την πιθανότητα κατανομής στον χώρο τον δειγμάτων $(x_1, x_2, \ldots, x_n)$. Θεωρούμε ότι οι τυχαίες μεταβλητές είναι $X_1, X_2, \ldots, X_n$ ανεξάρτητες, επομένως κατανέμονται όπως η *X*. Εάν η κατανομή των *n* δειγμάτων έχει συνάρτηση πυκνότητας πιθανότητας *f(x/θ)* για όλες τις τυχαίες μεταβλητές $X_1, X_2, \ldots, X_n$. Η κοινή πυκνότητα πιθανότητας ισούται με $f(x_1/\theta)*\ldots*f(x_n/\theta)$.
Οι τιμές $(x_1, x_2, \ldots, x_n)$ εξαρτώνται από το θ. Για τον υπολογισμό της τιμής του θ παίρνουμε την συνάρτηση πιθανοφάνειας :
$L(\theta) = f(x_1/\theta)*\ldots*f(x_n/\theta)$
Ψάχνουμε το θ στο διάστημα $(x_1, x_2, \ldots, x_n)$ για το οποίο η $L(\theta)$ είναι μέγιστη.

Για ομοιόμορφες κατανομές:
Εάν τα *α* και *b* είναι πραγματικές τιμές και *α<b* επιλέγουμε ένα τυχαίο σημείο ανάμεσα στο *α* και *b*. Το τυχαίο σημείο αποτελεί μια τιμή της τυχαίας μεταβλητής *X* η οποία είναι ομοιόμορφα κατανεμημένη και

$$f(x)=\begin{cases}\dfrac{1}{b-a}, & \alpha\nu \quad \alpha<x<b\\ 0 & \alpha\lambda\lambda o\acute{\upsilon}\end{cases}$$

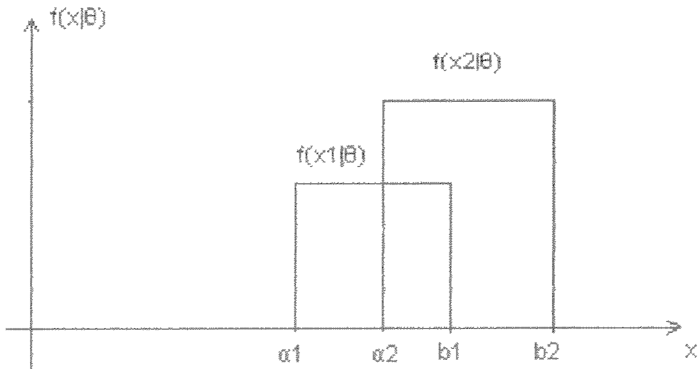
Οπότε έχουμε:

α)



Είναι $L(\theta)= f(x_1/\theta)* f(x_2/\theta)=0$ για κάθε x

β)



Είναι $L(\theta)= f(x_1/\theta)* f(x_2/\theta)=\begin{cases} f(x_1\mid\theta)\cdot f(x_2\mid\theta) & \gamma\iota\alpha \quad \alpha_2<x<b_1\\ 0 \quad \alpha\lambda\lambda o\acute{\upsilon} \end{cases}$

Άσκηση 4

Να δείξετε ότι δείγματα που ακολουθούν Gaussian κατανομές η εκτίμηση μέγιστης πιθανοφάνειας για την μέση τιμή ισούται με τη μέση τιμή των δειγμάτων. (Θεωρούμε ότι $\Sigma_i=\Sigma$ για κάθε i)

Λύση

Θεωρούμε ότι $\mu_i=\theta_i=\mu$ όπου $p(x \mid \propto) = \dfrac{1}{\sqrt{(2\pi)^d |\Sigma|^{1/2}}} e^{-\frac{1}{2}(x-\propto)^T \Sigma^{-1}(x-\propto)}$

Οι κλάσεις εκπαίδευσης είναι οι H_i για τα x_κ και επίσης έχουμε:

$$\log\{p(x_\kappa \mid \propto)\} = -\frac{1}{2}(x-\propto)^T \Sigma^{-1}(x-\propto) - \frac{1}{2}\log\{(2\pi)^d |\Sigma|\}$$

Χρησιμοποιούμε επίσης την παρακάτω ιδιότητα για το x και τον πίνακα A και το γεγονός ότι ο A είναι συμμετρικός ($A^T=A$) :

$$\frac{d(x^T Ax)}{dx} = (A + A^T)x = 2Ax$$

Άρα το gradient ισούται με:

$$\nabla_{\propto} \log\{p(x_\kappa \mid \propto)\} = -\frac{1}{2}[2\Sigma^{-1}(x-\propto)] = -(\Sigma^{-1}x_\kappa - \Sigma^{-1}\propto)$$

Για να βρούμε την τιμή του $\hat{\propto}$ εξισώνουμε το gradient με το 0:

$$\nabla_{\propto} l(\hat{\propto}) = \Sigma^{-1}\left(\sum_{k=1}^{n}(\hat{\propto} - x_\kappa)\right) = 0 \Rightarrow$$

Θεωρούμε ότι ο Σ είναι αντιστρέψιμος (θετικά ορισμένος)

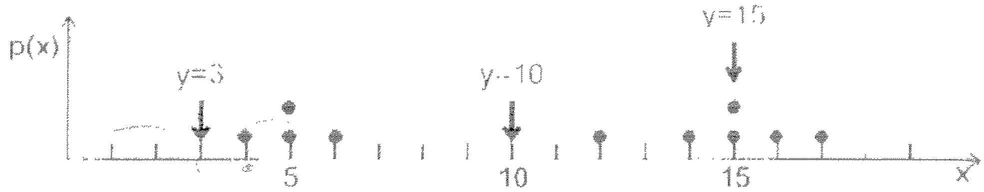
$$\Rightarrow \sum_{k=1}^{n} x_\kappa = n \cdot \hat{\propto} \Rightarrow \hat{\propto} = \frac{1}{n} \cdot \sum_{k=1}^{n} x$$

6^o Φροντιστήριο – Μη παραμετρικές τεχνικές

N=10 (πλήθος)

1η Άσκηση

Δεδομένου του συνόλου $X = \{x^1, x^2, ..., x^N\} = \{4,5,5,6,12,14,15,15,16,17\}$, χρησιμοποιήστε παράθυρα Parzen για να υπολογίσετε τη συνάρτηση πυκνότητας πιθανότητας $p(x)$ για $y=3$, 10 και 15. Χρησιμοποιήστε μέγεθος παραθύρου ίσο με $h = 4$.



Λύση

Αρχικά θα σχεδιάσουμε το σύνολο των σημείων για να έχουμε μια ιδέα των αποτελεσμάτων που θα πάρουμε:

Για $y=3$ υπολογίζουμε την $p(y=3)$:

$$p_{KDE}(y=3) = \frac{1}{Nh^D}\sum_{n=1}^{N} K\left(\frac{y-x^n}{h}\right) = \frac{1}{10\cdot 4^1}\left[\underbrace{K\left(\frac{3-4}{4}\right)}_{1/4} + \underbrace{K\left(\frac{3-5}{4}\right)}_{1/2} + \underbrace{K\left(\frac{3-5}{4}\right)}_{1/2} + \underbrace{K\left(\frac{3-6}{4}\right)}_{1} + ... + \underbrace{K\left(\frac{3-17}{4}\right)}_{13.4}\right] =$$

$$\frac{1}{10\cdot 4^1}[1+0+0+0+0+0+0+0+0+0] = \frac{1}{10\cdot 4^1} = 0.025$$

Ομοίως προκύπτουν τα εξής:

$$p_{KDE}(y=10) = \frac{1}{10\cdot 4^1}[0+0+0+0+0+0+0+0+0+0] = \frac{0}{10\cdot 4^1} = 0.$$

$$p_{KDE}(y=15) = \frac{1}{10\cdot 4^1}[0+0+0+0+0+1+1+1+1+0] = \frac{4}{10\cdot 4^1} = 0.1.$$

Άσκηση 2η

Έστω ένα πρόβλημα ταξινόμησης μιας διάστασης με δύο κατηγορίες, όπου για κάθε κατηγορία έχουν συλλεχθεί τα παρακάτω δεδομένα:

$$\mathbf{D}_1 = \{-1,-2,3,3,6,7\} \text{ και } \mathbf{D}_2 = \{-3,-2,3,5,8\}.$$

Έστω ότι αποφασίζουμε να χρησιμοποιήσουμε παράθυρα Parzen με μέγεθος παραθύρου ίσο με $h = 2$ και συνάρτηση παραθύρου $\varphi(x)$ που ορίζεται ως:

$$\varphi(u) = \begin{cases} -\dfrac{3}{4}u^2 + \dfrac{3}{4} & \text{εάν } -1 \le u \le 1 \\ 0 & \text{αλλού} \end{cases}$$

A) Ταξινομήστε το δείγμα $x = 4$ χρησιμοποιώντας τον *ML* ταξινομητή.
B) Αποδείξτε ότι η εκτιμήσεις των συναρτήσεων πυκνότητας πιθανότητας που γίνονται με χρήση των παράθυρων Parzen με την φ(x) να έχει τη μορφή που ορίστηκε παραπάνω είναι στην πραγματικότητα συναρτήσεις πυκνότητας πιθανότητας.

Λύση

A) $$p_1(4) = \frac{1}{6}\sum_{i=1}^{6}\frac{1}{2}\varphi\left(\frac{4-x_i}{2}\right) = \frac{1}{12}\left(\varphi\left(\frac{1}{2}\right) + \varphi\left(\frac{1}{2}\right)\right) = \frac{\varphi\left(\frac{1}{2}\right)}{6}$$

$$p_2(4) = \frac{1}{5}\sum_{i=1}^{5}\frac{1}{2}\varphi\left(\frac{4-x_i}{2}\right) = \frac{1}{10}\left(\varphi\left(\frac{1}{2}\right) + \varphi\left(-\frac{1}{2}\right)\right) = \frac{\varphi\left(\frac{1}{2}\right)}{5}$$

όπου το τελευταίο προκύπτει διότι $\varphi\left(\frac{1}{2}\right) = \varphi\left(-\frac{1}{2}\right)$.

Επομένως, το σημείο *x = 4 πρέπει να ταξινομηθεί στην κατηγορία 2.*

B) Για να αποδείξουμε ότι οι συναρτήσεις που υπολογίζονται είναι στην πραγματικότητα συναρτήσεις πυκνότητας πιθανότητας πρέπει να δείξουμε ότι η συνάρτηση φ(u) είναι μη αρνητική και ότι το ολοκλήρωμά της ισούται με 1.

Προφανώς, $\varphi(u) \ge 0$, αφού το $-\frac{3}{4}u^2 + \frac{3}{4}$ παίρνει μη αρνητικές τιμές στο διάστημα [-1, 1].

$$\int_{-\infty}^{\infty}\varphi(u)du = \int_{-1}^{1}\left(-\frac{3}{4}u^2 + \frac{3}{4}\right)du = -\frac{3}{12}u^3 + \frac{3}{4}u\Big|_{-1}^{1} = \frac{6}{12} + \frac{6}{12} = 1.$$

$3^η$ Άσκηση

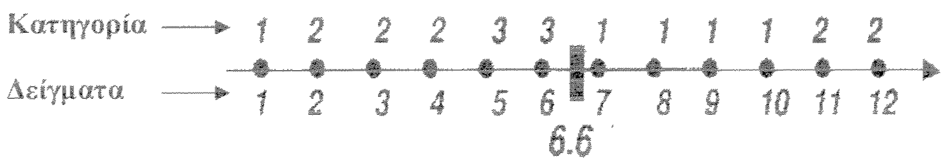
Έστω ένα πρόβλημα ταξινόμησης μιας διάστασης με τρεις κατηγορίες, όπου για κάθε κατηγορία έχουν συλλεχθεί τα παρακάτω δεδομένα:

$$\mathbf{D}_1 = \{1,7,8,9,10\},\ \mathbf{D}_2 = \{2,3,4,11,12\} \text{ και } \mathbf{D}_3 = \{5,6\}.$$

A) Ταξινομήστε το δείγμα 6.6 χρησιμοποιώντας των αλγόριθμο των *k* πλησιέστερων γειτόνων με *k = 6*.
B) Εφαρμόστε τον αλγόριθμο περικοπής και υπολογίστε τα καινούρια σύνολα δειγμάτων που προκύπτουν.

Λύση

A)



Οι 6 πλησιέστεροι γείτονες του δείγματος 6.6 είναι οι 7, 8 ,9 από την κατηγορία 1, οι 5, 6 από την κατηγορία 3 και ο 4 από την κατηγορία 2.
Επομένως ο μεγαλύτερος αριθμός γειτόνων ανήκει στην κατηγορία 1 και επομένως το δείγμα 6.6 πρέπει να ταξινομηθεί στην κατηγορία 1.

B) Τα δείγματα 3, 8, 9, 12 έχουν όλους τους γείτονές τους μέσα στο διάγραμμα Voronoi της ίδιας κατηγορίας και επομένως μπορούν να περικοπούν.
Τα ελαττωμένα σύνολα δειγμάτων που προκύπτουν είναι τα εξής:

$$\mathbf{D}_1 = \{1,7,10\},\ \mathbf{D}_2 = \{2,4,11\} \text{ και } \mathbf{D}_3 = \{5,6\}.$$