

TUTORIAL 3 – December 14, 2023**PROBLEM 1**

Θεωρήστε ένα μονοδιάστατο πρόβλημα ταξινόμησης 2 κλάσεων, όπου έχουμε τα παρακάτω δεδομένα για κάθε κλάση : $D_1 = \{-1, -2, 3, 3, 6, 7\}$ και $D_2 = \{-3, -2, 3, 5, 8\}$. Θεωρήστε ότι αποφασίσαμε να χρησιμοποιήσουμε παράθυρα Parzen, με πλάτος παραθύρου $h = 2$ και συνάρτηση παραθύρου $\phi(x)$ που ορίζεται ως :

$$\phi(x) = \begin{cases} -\frac{3}{4}u^2 + \frac{3}{4} & , \text{αν } -1 \leq x \leq 1 \\ 0 & , \text{αλλού} \end{cases}$$

(α) Ταξινομήστε το δείγμα $x = 4$ χρησιμοποιώντας τον ταξινομητή μέγιστης πιθανοφάνειας ML.

(β) Να δείξετε ότι οι εκτιμήσεις των συναρτήσεων πυκνότητας πιθανότητας που γίνονται με χρήση των παραθύρων Parzen με την $\phi(x)$ να έχει την μορφή που ορίστηκε παραπάνω είναι στην πραγματικότητα συναρτήσεις πυκνότητας πιθανότητας.

PROBLEM 2

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

$$D_1 = \{1, 7, 8, 9, 10\} \quad , \quad D_2 = \{2, 3, 4, 11, 12\} \quad \text{και} \quad D_3 = \{5, 6\}$$

(α) Ταξινομήστε το δείγμα 6.6 χρησιμοποιώντας τον αλγόριθμο των k πλησιέστερων γειτόνων με $k=6$

(β) Εφαρμόστε τον αλγόριθμο περικοπής και υπολογίστε τα καινούρια σύνολα δειγμάτων που προκύπτουν.

PROBLEM 3

Διαθέτουμε δύο διαφορετικές κλάσεις ω_1 και ω_2 και θεωρούμε ότι οι συναρτήσεις πυκνότητας

$$\text{πιθανότητας } P(x|\omega_1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x}{4}} \quad \text{και} \quad P(x|\omega_2) = \frac{1}{8} e^{-\frac{x}{2}}.$$

Να υπολογίσετε για ποιες τιμές της παρατήρησης x κατηγοριοποιούμε το στοιχείο στην κλάση ω_1 , όταν τα κόστη για λανθασμένη απόφαση είναι 4 όταν γίνεται λάθος ταξινόμηση στοιχείου κλάσης ω_1 και 2 όταν γίνεται λάθος ταξινόμηση στοιχείου κλάσης ω_2 . Θεωρήστε πως $P(\omega_1) = 0.2$

PROBLEM 4

Θεωρήστε ένα πρόβλημα ταξινόμησης 2 κλάσεων με πιθανοφάνειες αντίστοιχα :

$$P(x|C_1) = x + \frac{1}{2} \quad \text{και} \quad P(x|C_2) = \frac{3x^2}{4} + \frac{3}{4} \quad \text{με } x \in [0, 1]. \quad \text{Υποθέστε ότι } P(C_2) = 0.25 \quad \text{και ότι ο πίνακας κόστους}$$

είναι $\begin{bmatrix} 1 & 2 \\ 3 & 1 \end{bmatrix}$. Να βρείτε τον ταξινομητή που ελαχιστοποιεί το υποθετικό ρίσκο .

PROBLEM 5

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

$$p(x;\theta) = \begin{cases} \theta^2 x e^{-\theta x} & , \text{αν } x \in [0,1] \\ 0 & , \text{αλλιού} \end{cases}$$

Έστω n ανεξάρτητα δείγματα $D=x_1, x_2, \dots, x_n$ που επιλέγονται ανεξάρτητα με βάση την $p(x;\theta)$. Καθορίστε την εκτίμηση μέγιστης πιθανοφάνειας του θ .