ΔΗΜΟΚΡΙΤΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΡΑΚΗΣ

ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ & ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ

Φαναρίδου Κυριακούλα

9ο εξάμηνο

ΑΕΜ: 57830

ΑΣΚΗΣΗ 1

Οι κώδικες για την συνάρτηση διάκρισης , την ευκλείδεια απόσταση και την mahalanobis βρίσκονται στο αρχείο code3_1_2

Γ) Η Mahalanobis απόσταση είναι ένας ταξινομητής που χρησιμοποιείται για μη διαγώνιους πίνακες συνδιασποράς όπου η μεγιστοποίηση της $g_i(x)$ είναι ισοδύναμη με την ελαχιστοποίηση της Σ^{-1}, δηλαδη:

$d_m = ((x-\mu_i))^T\Sigma^{-1}(x-\mu_i))^{1/2}$

ΑΣΚΗΣΗ 2

Οι κώδικες βρίσκονται στο αρχείο code3_1_2

1. Το σημείο διαχωρισμού για τις δύο κατηγορίες είναι το $-5.0855<x<4.33746$
   Για την εύρεση του αφαιρούμε τις συναρτήσεις διάκρισης για $P(\omega_1) = P(\omega_2) = ½$
2. Υπολογίζουμε το εμπειρικό σφάλμα κατάρτισης με τα δείγματα για x_1, δηλαδή τα δείγματα που ταξινομήθηκαν λάθος είναι 0.3 για την κλάση ω_1 και 0.3 για την κλάση ω_2.
3. Ομοίως, υπολογίζουμε τα σφάλματα για x_1 και x_2 είναι 0.5 για την κλάση ω_1 και 0.4 για την κλάση ω_2.
4. Υπολογίζουμε τα σφάλματα για x_1 και x_2 και x_3 είναι 0.2 για την κλάση ω_1 και 0.1 για την κλάση ω_2.
5. Παρατηρούμε πως η προσθήκη περισσότερων χαρακτηριστικών δεν μειώνει απαραίτητα το σφάλμα της ταξινόμησης. Είναι λογικό πως για να πετύχουμε καλύτερη ταξινόμηση των δειγμάτων πρέπει να αυξηθούν τόσο τα χαρακτηριστικά όσο και το πλήθος των δειγμάτων.
6. Οι συναρτήσεις διαχωρισμού των κατηγοριών είναι 3 για [x1,x2] , [x2,x3], [x1,x3] και παρατίθενται στο matlab αρχείο.

ΑΣΚΗΣΗ 3

Αρχείο code3_3

1. Το A υπολογίζεται π/2 ή 1.5708
   Από την αρχική κατανομή του θ που μας δίνεται ολοκληρώνοντας με όρια το 0,1 βρίσκουμε την τιμή του A.
   $\int_0^1 A * \sin(\theta * \pi)\, d\theta = 1 \quad => A * \int_0^1 \sin(\theta * \pi)\, d\theta => A\frac{1}{\pi}[-cos\pi + cos0] => A = \pi/2$
   Το A βρίσκεται ομοίως και με κώδικα matlab.
2. Για να σχεδιασθεί το $p(\theta|D^1)$ για ένα νόμισμα που πετάμε N=10 φορές και φέρνουμε {κ,κ,γ,κ,γ,κ,κ,κ,γ,κ} χρησιμοποιούμε τον τύπο
   $P(\theta|D^n) = \frac{\theta^k(1-\theta)^{N-k}p(\theta|D^0)}{integral(\theta^k(1-\theta)^{N-k}p(\theta|D^0)d\theta}$

Από εκφώνηση το $p(\theta|D^0) = (\pi/2)\sin(\pi\theta)$

Και από θεωρία $p(x_n|\theta) = \begin{cases} \theta, αν\ έρθει\ κεφάλι \\ 1-\theta, αν\ έρθουν\ γράμματα \end{cases}$

Άρα ,

το $p(\theta|D^1) = \frac{\theta^1(1-\theta)^0\left(\frac{\pi}{2}\right)\sin(\theta\pi)}{integral(\theta^1(1-\theta)^0(\frac{\pi}{2})\sin(\pi\theta)d\theta}$

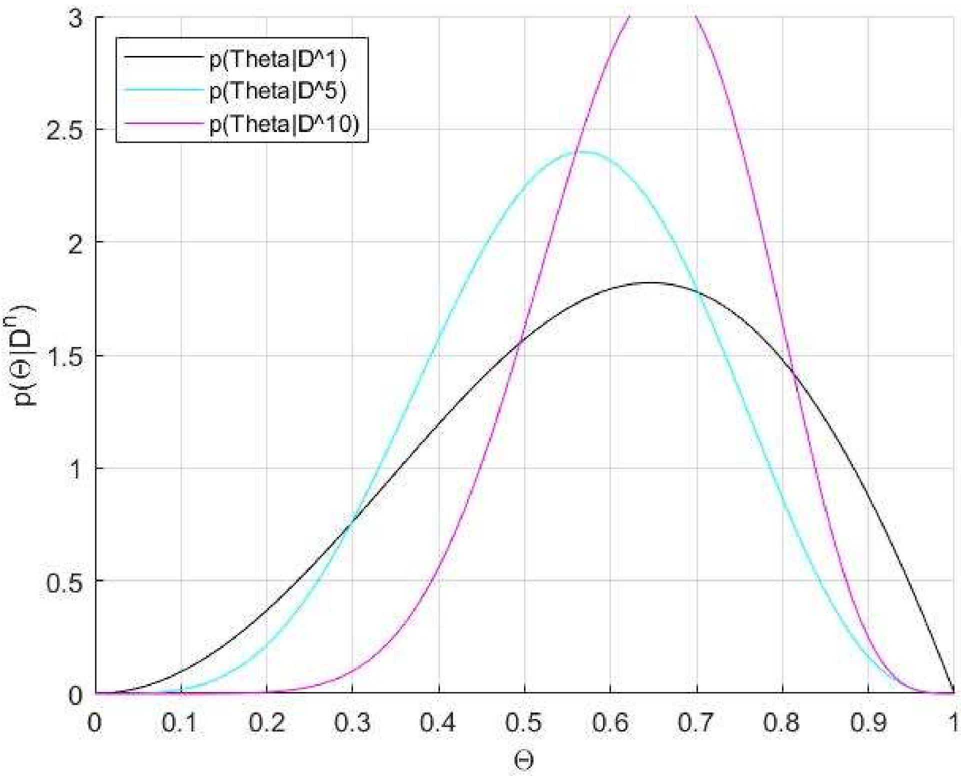
ομοίως για

το $p(\theta|D^5) = \frac{\theta^5(1-\theta)^3\left(\frac{\pi}{2}\right)\sin(\theta\pi)}{integral(\theta^5(1-\theta)^3(\frac{\pi}{2})\sin(\pi\theta)d\theta}$

και

το $p(\theta|D^{10}) = \frac{\theta^{10}(1-\theta)^7\left(\frac{\pi}{2}\right)\sin(\theta\pi)}{integral(\theta^{10}(1-\theta)^7(\frac{\pi}{2})\sin(\pi\theta)d\theta}$

Τα διαγράμματα παρατίθενται παρακάτω



3. Το $p(x=k|D^{10})$ βρέθηκε αριθμητικά μέσω matlab ίσο με 3.1244 για x=0.66

ΑΣΚΗΣΗ 4

1. Λύσεις σε matlab αρχεία code3_4_2, code 3_4_3,code_3_4_4_initial_values και code3_4_4_diff_values

2. Παρατηρώ πως οι πίνακες συνδιασποράς είναι ίσοι αλλά όχι διαγώνιοι οπότε δεν μπορώ να χρησιμοποιήσω τον ευκλείδειο ταξινομητή. Παρόλα αυτά στον κώδικα υπολογίζεται.
   Τα σφάλματα υπολογίζονται:
   Την καλύτερη απόδοση έχουν ο bayes και ο mahalanobis ταξινομητές.

Ευκλείδιος	**0.11011**
Mahalanobis	0.09710
Bayes	0.09710

3. Με την μέθοδο της μέγιστης πιθανοφάνειας υπολογίζουμε τις πιθανότητες λάθους:

Ευκλείδιος	**0.11011**
Mahalanobis	0.10210
Bayes	0.10410

   Παρατηρούμε πως ο καλύτερος ταξινομητής είναι ο Mahalanobis αλλά με πολύ μικρή διαφορά από τον Bayes. Διαλέγουμε τον Mahalanobis.
   Παρατηρούμε πως η ευκλείδιος που δεν θα έπρεπε να χρησιμοποιηθεί εδώ δίνει το ίδιο σφάλμα και μεγαλύτερο από τους άλλους δυο ταξινομητές.

4. Επαναλαμβάνουμε τα βήματα 2,3 αλλά με διαφορετικούς πίνακες συνδιασποράς.
   Για τα αρχικά δεδομένα τα σφάλματα είναι:

Ευκλείδιος	**0.07415**
Mahalanobis	0.06914
Bayes	0.06313

   Παρατηρούμε πως ο καλύτερος ταξινομητής είναι ο Bayes αλλά με πολύ μικρή διαφορά από τον Mahalanobis. Διαλέγουμε τον Bayes.

   Εφαρμόζουμε ομοίως την μέθοδο μέγιστης πιθανοφάνειας και τα αποτελέσματα είναι τα εξής:

Ευκλείδιος	**0.07114**
Mahalanobis	0.06713
Bayes	0.06012

   Αυτή τη φορά βλέπουμε πως ο bayes έδωσε πολύ καλύτερα αποτελέσματα από τον Mahalanobis.

   Συμπαιρένουμε, πώς όταν ο πίνακας συνδιασποράς είναι ίδιος για όλες τις κλάσεις τότε οι ταξινομητές έχουν σχεδόν ίδια αποτελέσματα ενώ όταν έχουμε διαφορετικούς πίνακες συνδιασποράς οι διαφορά στους ταξινομητές είναι αισθητή.