LECTURE I

Pattern Recognition [#1]

Ταξινομητές Βασισμένοι στην Μπεϋζιανή Θεωρία Αποφάσεων

Στόχος: Σχεδιασμός ενός ταξινομητή σε ένα σύστημα αναγνώρισης προτύπων

- Η προσέγγιση που θα ακολουθήσουμε βασίζεται στην κατανόηση της πιθανοτικής φύσης των παραγόμενων χαρακτηριστικών.

- Η συγκεκριμένη επιλογή υπαγορεύεται τόσο από την στατιστική διακύμανση των προτύπων όσο και από την παρουσία θορύβου κατά την διενέργεια των μετρήσεων.

- Με βάση τα παραπάνω, είναι εύδηλη η ανάγκη για τον σχεδιασμό ενός ταξινομητή που θα αναθέτει ένα άγνωστο πρότυπο στην πιθανότερη κλάση.

- Επομένως, στόχος μας είναι ο καθορισμός της έννοιας της πιθανότερης κλάσης.

* * * *

Έστω ένα Μ-ταξινό πρόβλημα ταξινόμησης μεταξύ των κλάσεων $\omega_1, \omega_2, \ldots, \omega_M$ όπου ένα άγνωστο πρότυπο αναπαρίσταται από το διάνυσμα των χαρακτηριστικών $\underline{x} \in \mathbb{R}^{\ell}$. Για το συγκεκριμένο πρόβλημα σχηματίζουμε τις παρακάτω δεσμευμένες πιθανότητες:

$$\boxed{P(\omega_i \mid \underline{x}), \quad \forall i \in [M] \quad (1)}$$

(i): Conditional Probabilities

(ii): A-posteriori Probabilities

Οι συγκεκριμένες πιθανότητες αναφέρονται και ως εκ των υστέρων πιθανότητες.

Στην πραγματικότητα καθεμιά από τις προηγούμενες ποσότητες συγκεκριμένα την πιθανότητα ενός άγνωστου προτύπου να ανήκει στην σχετική κλάση ω_i, δοθέντος ότι το αντίστοιχο διάνυσμα χαρακτηριστικών λαμβάνει την τιμή $\underline{x}$.

Είναι προφανές πως οι εκ των υστέρων πιθανότητες των κλάσεων $\omega_1, \omega_2, \ldots, \omega_M$ αποτελούν την πιο εύλογη επιλογή για την ποσοτικοποίηση του όρου πιθανότερη κλάση.

Πράγματι, οι ταξινομητές που θα μελετηθούν στην παρούσα ενότητα λειτουργούν υπολογίζοντας είτε την μέγιστη τιμή αυτών των (M) εκ των υστέρων πιθανοτήτων είτε μία συνάρτηση αυτών.

* * * *

Μπεϋζιανή Θεωρία Αποφάσεων

Έστω ένα δυαδικό πρόβλημα ταξινόμησης μεταξύ των κλάσεων ω_1 και ω_2. Υποθέτουμε πως οι εκ των προτέρων πιθανότητες των δύο κλάσεων $P(\omega_1)$ και $P(\omega_2)$ είναι γνωστές. Η συγκεκριμένη υπόθεση είναι εύλογη καθώς οι εκ των προτέρων πιθανότητες των δύο κλάσεων θα μπορούσαν να υπολογισθούν από το σύνολο των διανυσμάτων χαρακτηριστικών που είναι διαθέσιμο κατά την φάση της εκπαίδευσης του ταξινομητή.

Αν (N) είναι το πλήθος των διαθέσιμων διανυσμάτων χαρακτηριστικών κατά την εκπαίδευση του ταξινομητή και (N_1), (N_2) είναι το πλήθος των διανυσμάτων χαρακτηριστικών που προέρχονται από τις κλάσεις ω_1 και ω_2, τότε μπορούμε να υπολογίσουμε τις εκ των προτέρων πιθανότητες των δύο κλάσεων ως:

$$P(\omega_1) \approx \frac{N_1}{N} \quad \text{και} \quad P(\omega_2) = \frac{N_2}{N} \,.$$ [A-priori Probabilities]

Οι άλλες ποσότητες του προβλήματος που υποθέτουμε πως είναι γνωστές είναι οι δεσμευμένες συναρτήσεις πυκνότητας πιθανότητας των κλάσεων

$$\boxed{p(\underline{x} \mid \omega_i), \quad \forall i \in [2] \quad (2)}$$

Class-Conditional Probability Density Functions.

(*) Αν το ($\underline{x}$) ήταν ένα μονοδιάστατο διάνυσμα χαρακτηριστικών (x) τότε οι δεσμευμένες συναρτήσεις πυκνότητας πιθανότητας, $p(x \mid \omega_i)$ $\forall i \in [M]$, μπορούν να χρησιμοποιηθούν για τον υπολογισμό της πιθανότητας του x να βρίσκεται σε μια περιοχή του χώρου $R \subset \mathbb{R}$ δοθέντος ότι ανήκει στην κλάση ω_i, $i \in [M]$.

$$\boxed{P_r(x \in R \mid \omega_i) = \int_R p(x \mid \omega_i)\,dx, \quad \forall i \in [2] \quad (3)}$$

Class-Conditional pdf.

(*) Οι δεσμευμένες συναρτήσεις πυκνότητας πιθανότητας περιγράφουν την κατανομή των διανυσμάτων χαρακτηριστικών σε καθεμιά από τις κλάσεις του προβλήματος.

(*) Οι δεσμευμένες συναρτήσεις πυκνότητας πιθανότητας εάν δεν είναι γνωστές μπορούν να εκτιμηθούν από τα δεδομένα εκπαίδευσης που έχουμε στην διάθεσή μας.

(*) Η συνάρτηση πυκνότητας πιθανότητας $p(\underline{x} \mid \omega_i)$ αναφέρεται συχνά ως η συνάρτηση πιθανοφάνειας της κλάσης ω_i (likelihood function) σχετικά με το διάνυσμα $\underline{x}$.

(*) Η ανάλυσή μας βασίζεται στην υπόθεση ότι τα διανύσματα χαρακτηριστικών μπορούν να λάβουν μια οποιαδήποτε τιμή μέσα στον ℓ-διάστατο χώρο των χαρακτηριστικών.

(*) Στην περίπτωση κατά την οποία τα διανύσματα χαρακτηριστικών μπορούν να λάβουν μόνο διακριτές τιμές, τότε οι συναρτήσεις πυκνότητας πιθανότητας $p(\underline{x}|\omega_i)$ μετατρέπονται σε πιθανότητες και επισημαίνονται ως $P(\underline{x}|\omega_i)$.

Σύμφωνα με το Θεώρημα του Μπέυζ, οι δεσμευμένες πιθανότητες των κλάσεων $P(\omega_i|\underline{x})$, $\forall i \in [M]$, μπορούν να εκφραστούν ως εξής:

$$\boxed{P(\omega_i|\underline{x}) = \frac{p(\underline{x}|\omega_i)P(\omega_i)}{p(\underline{x})}, \forall i \in [2] \quad (4)}$$

όπου $p(\underline{x})$ είναι η συνάρτηση πυκνότητας πιθανότητας για το διάνυσμα χαρακτηριστικών $\underline{x}$.

Για την συνάρτηση πυκνότητας πιθανότητας $p(\underline{x})$ ισχύει ότι:

$$\boxed{p(\underline{x}) = \sum_{i=1}^{2} p(\underline{x}|\omega_i)P(\omega_i) \quad (5)}$$

* * * *

<u>Κανόνας Ταξινόμησης Μπέυζ</u>: [Bays Classification Rule]

Εάν $P(\omega_1|\underline{x}) > P(\omega_2|\underline{x})$, ανάθεση του $\underline{x}$ στην κλάση ω_1.

Εάν $P(\omega_2|\underline{x}) > P(\omega_1|\underline{x})$, ανάθεση του $\underline{x}$ στην κλάση ω_2.

Η περίπτωση της ισότητας είναι άνευ σημασίας καθώς το πρότυπο θα μπορούσε να ταξινομηθεί εξίσου στην κάθε κλάση.

Λαμβάνοντας υπόψιν το θεώρημα του Μπέυζ, ο παραπάνω κανόνας απόφασης μπορεί να εκφραστεί ως εξής:

$$\boxed{p(\underline{x}|\omega_1)P(\omega_1) \gtrless p(\underline{x}|\omega_2)P(\omega_2) \quad (6)}$$

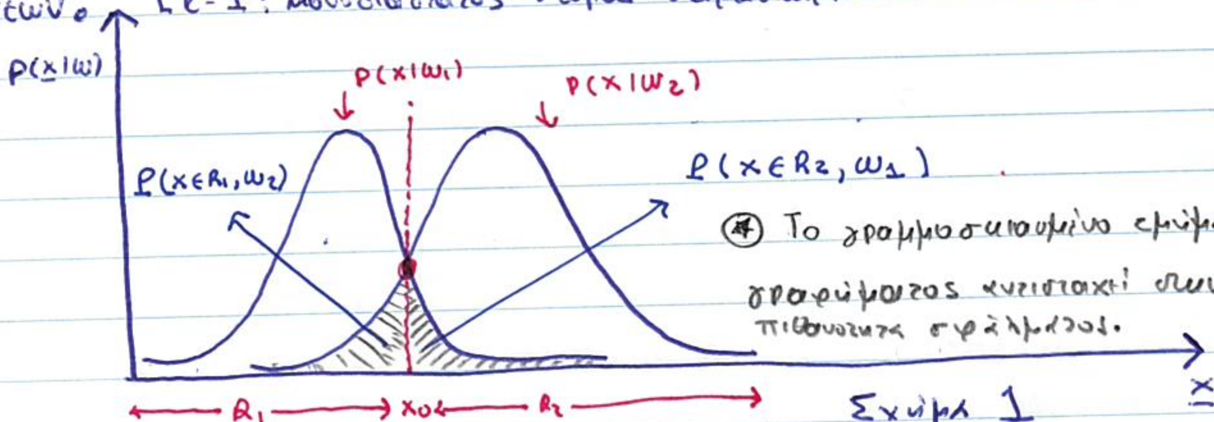
όπου η ποσότητα $p(\underline{x})$ δεν παίζει ρόλο καθώς εμφανίζεται και στα δύο μέλη της ανίσωσης.

Επιπλέον, αν υποθέσουμε πως οι εκ των προτέρων πιθανότητες των δύο κλάσεων είναι ίδιες, τότε ο κανόνας απόφασης του Μπέυζ λαμβάνει την μορφή:

$$\boxed{p(x|\omega_1) \gtrless p(x|\omega_2) \quad (7)}$$

Σύμφωνα με τα παραπάνω, ο κανόνας του Μπέυζ ανάγεται στην αναζήτηση της κλάσης με την μέγιστη πιθανοφάνεια. Η πιθανοφάνεια της κάθε κλάσης αποτελείται υπολογίζοντας την τιμή της αντίστοιχης δεσμευμένης συνάρτησης πυκνότητας πιθανότητας σε ένα δεδομένο σημείο $\underline{x}$.

Έστω η παρακάτω σχηματική αναπαράσταση για τις ποσότητες $p(\underline{x}|\omega_1)$ και $p(\underline{x}|\omega_2)$ για την περίπτωση δύο ισοπίθανων κλάσεων. [$l=1$: Μονοδιάστατος χώρος χαρακτηριστικών]



(*) Το γραμμοσκιασμένο εμβαδόν του γραφήματος αντιστοιχεί στην συνολική πιθανότητα σφάλματος.

Σχήμα 1

Η διακεκομμένη γραμμή στο $x = x_0$ αναπαριστά το σύνορο στον χώρο των χαρακτηριστικών μεταξύ των περιοχών απόφασης R_1 και R_2. Οι περιοχές απόφασης R_1 και R_2 ορίζονται ως εξής:

$$R_1 = \{ x \in \mathbb{R} : p(x|\omega_1) > p(x|\omega_2) \} \quad (8)$$
$$R_2 = \{ x \in \mathbb{R} : p(x|\omega_2) > p(x|\omega_1) \} \quad (9)$$

Επομένως, ο κανόνας του Μπέυζ λειτουργεί ως εξής:

$$\boxed{\begin{aligned} &\bullet\ \forall x \in R_1, \ x \in \omega_1 \\ &\bullet\ \forall x \in R_2, \ x \in \omega_2 \end{aligned} \quad (10)}$$

είναι προφανές από την προηγούμενη σχηματική αναπαράσταση πως η πιθανότητα εσφαλμένων αποφάσεων είναι αναπόφευκτη. Είναι εύκολο να παρατηρήσει κανείς πως:

(i): υπάρχει μια πεπερασμένη πιθανότητα το x να βρίσκεται στην περιοχή R_2 και ταυτόχρονα να ανήκει στην κλάση ω_1,

(ii): υπάρχει μια πεπερασμένη πιθανότητα το x να βρίσκεται στην περιοχή R_1 και ταυτόχρονα να ανήκει στην κλάση ω_2.

Η συνολική πιθανότητα σφάλματος για τον Μπεϋζιανό ταξινομητή μπορεί να υπολογισθεί ως εξής:

MEMO: $P(X,Y) = P(X|Y)P(Y)$

$$P_e = \underbrace{P(x \in R_2, \omega_1)}_{\text{Από κοινού πιθανότητα (JOINT PROBABILITY)}} + \underbrace{P(x \in R_1, \omega_2)}_{\text{Από κοινού πιθανότητα (JOINT PROBABILITY)}} \Leftrightarrow$$

$$P_e = \underbrace{P(\omega_1)}_{1/2} \cdot \underbrace{P(x \in R_2 | \omega_1)} + \underbrace{P(\omega_2)}_{1/2} \cdot \underbrace{P(x \in R_1 | \omega_2)} \Leftrightarrow$$

Μπορούν να υπολογισθούν από τις δεσμευμένες (ως προς την κλάση) συναρτήσεις πυκνότητας πιθανότητας [class-conditional probability density functions]

$$\boxed{P_e = \frac{1}{2}\int_{R_1} p(x|\omega_2)\,dx + \frac{1}{2}\int_{R_2} p(x|\omega_1)\,dx = \frac{1}{2}\int_{-\infty}^{x_0} p(x|\omega_2)\,dx + \frac{1}{2}\int_{x_0}^{\infty} p(x|\omega_1)\,dx \quad (11)}$$

Λαμβάνοντας υπόψιν πως η τιμή του ορίου x_0 μεταξύ των περιοχών απόφασης R_1 και R_2 μπορεί να υπολογισθεί από την λύση της εξίσωσης $f(x) = 0$ με $f(x) = p(x|\omega_1) - p(x|\omega_2)$, μπορούμε να γράψουμε τις περιοχές απόφασης ως $R_1 = (-\infty, x_0]$ και $R_2 = [x_0, +\infty)$