# LECTURE II

## Ελαχιστοποίηση της Πιθανότητας Ταξινομητικού Σφάλματος

Σύμφωνα με τα συμπεράσματα της προηγούμενης ενότητας η λειτουργία του Μπεϋζιανού ταξινομητή μπορεί να περιγραφεί ως μια συνάρτηση διάκρισης (discrimination function) της μορφής:

$$\boxed{g : \mathbb{R}^{\ell} \rightarrow \{\omega_1, \omega_2\} \quad (1)}$$

Στην ειδική περίπτωση όπου ο θεωρούμενος χώρος των χαρακτηριστικών είναι μονοδιάστατος ($\ell = 1$) (*) δείξαμε πως ο Μπεϋζιανός κανόνας απόφασης παίρνει την μορφή:

$$\boxed{g(x) = \begin{cases} \omega_1, & x \leq x_0; \\ \omega_2, & x > x_0. \end{cases} \quad (2)}$$

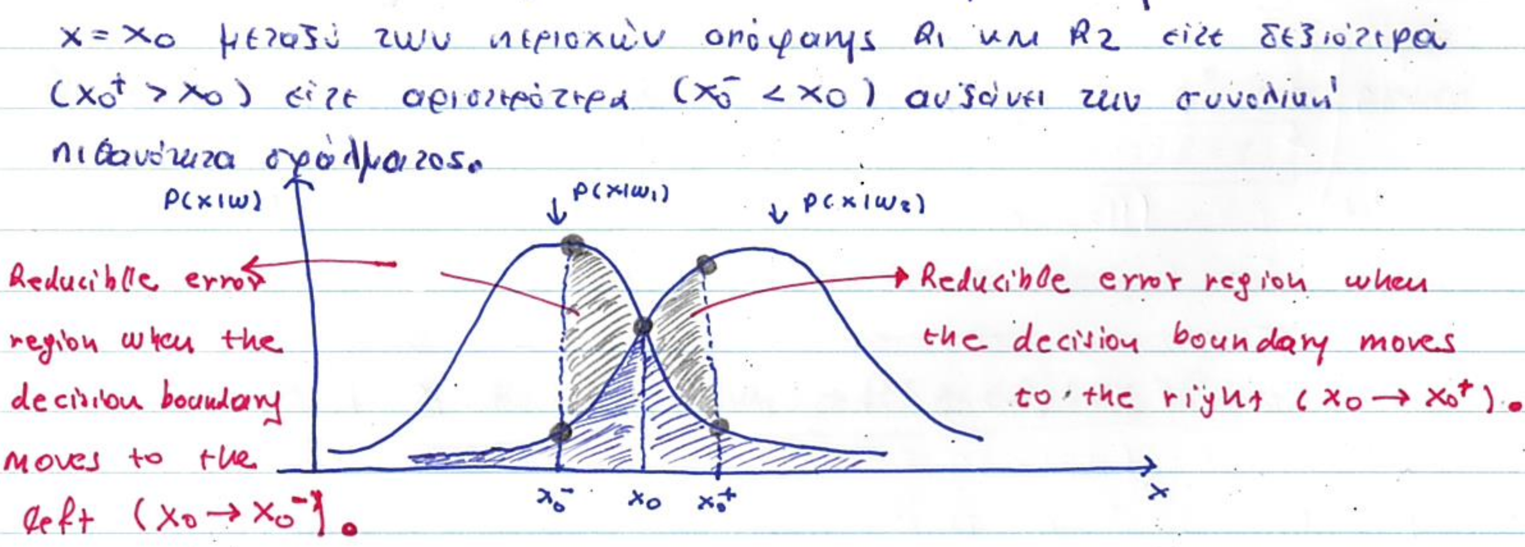
(*) Η εξίσωση (2) ισχύει στην περίπτωση κατά την οποία οι εκ των προτέρων πιθανότητα των κλάσεων $\omega_1$ και $\omega_2$ είναι ίσες: $P(\omega_1) = P(\omega_2) = 1/2$.

Η παραπάνω μορφή προκύπτει από την γενικότερη διατύπωση του Μπεϋζιανού κανόνα απόφασης:

$$\boxed{g(x) = \begin{cases} \omega_1, & p(x|\omega_2) \leq p(x|\omega_1); \\ \omega_2, & p(x|\omega_1) < p(x|\omega_2). \end{cases} \quad (3)}$$

(*) Η εξίσωση (3) είναι γενικότερη της (2) που ισχύει και πάλι υπό την προϋπόθεση $P(\omega_1) = P(\omega_2)$.

Είναι εύκολο να αντιληφθεί κανείς πως η μετακίνηση του συνόρου $x = x_0$ μεταξύ των περιοχών απόφασης $R_1$ και $R_2$ είτε δεξιότερα ($x_0^+ > x_0$) είτε αριστερότερα ($x_0^- < x_0$) αυξάνει την συνολική πιθανότητα σφάλματος.



Θα αποδείξουμε πως ο Μπεϋζιανός ταξινομητής είναι βέλτιστος υπό το πρίσμα της ελαχιστοποίησης της πιθανότητας του ταξινομητικού σφάλματος.

Απόδειξη: Έστω $R_1$ η περιοχή του χώρου των χαρακτηριστικών εντός της οποίας ο ταξινομητής αποφασίζει υπέρ της κλάσης $\omega_1$ και $R_2$ η περιοχή του χώρου των χαρακτηριστικών εντός της οποίας ο ταξινομητής αποφασίζει υπέρ της κλάσης $\omega_2$.

$$\boxed{R_1 = \{ \underline{x} \in \mathbb{R}^{\ell} : g(\underline{x}) = \omega_1 \}} \text{ και } \boxed{R_2 = \{ \underline{x} \in \mathbb{R}^{\ell} : g(\underline{x}) = \omega_2 \}} \quad (4)$$

Σύμφωνα με τα παραπάνω, ένα σφάλμα πραγματοποιείται στις εξής περιπτώσεις:

(i): το διάνυσμα χαρακτηριστικών $\underline{x} \in R_1$ (ανήκει στην περιοχή απόφασης $R_1$) αλλά προέρχεται από την κατηγορία $\omega_2$;

(ii): το διάνυσμα χαρακτηριστικών $\underline{x} \in R_2$ (ανήκει στην περιοχή απόφασης $R_2$) αλλά προέρχεται από την κατηγορία $\omega_1$.

Σύμφωνα με τα παραπάνω μπορούμε να γράψουμε ότι:

$$\boxed{P_e = P(\underline{x} \in R_2, \omega_1) + P(\underline{x} \in R_1, \omega_2) \quad (4)}$$

Note once again that for the joint probability of two events $X, Y$ we may write that $\boxed{P(X,Y) = P(X|Y)P(Y)}$

Αυτό συμπαίνει πως η σχέση (4) μπορεί να γραφεί στην μορφή:

$$\boxed{P_e = P(\underline{x} \in R_2 | \omega_1) P(\omega_1) + P(\underline{x} \in R_1 | \omega_2) P(\omega_2) \quad (5)}$$

Η σχέση (9) αναλύεται περισσότερο ως εξής:

$$\boxed{P_e = P(\omega_1)\int_{R_2} p(\underline{x}|\omega_1)dx + P(\omega_2)\int_{R_1} p(\underline{x}|\omega_2)dx \quad (10)}$$

Κάνοντας χρήση του κανόνα του <u>Bayes</u> η σχέση (10) μπορεί να γραφεί ως:

$$\boxed{P_e = \int_{R_2} P(\omega_1|\underline{x})p(\underline{x})dx + \int_{R_1} P(\omega_2|\underline{x})p(\underline{x})dx \quad (11)}$$

(*) Ιδανικά, θα θέλαμε να απαλλαγούμε από αυτόν τον όρο.

Σύμφωνα με την παραπάνω σχέση είναι εύκολο να αντιληφθεί κανείς πως η συνολική πιθανότητα σφάλματος ελαχιστοποιείται όταν οι περιοχές διαμέρισης του χώρου των χαρακτηριστικών $R$, $R_1$ και $R_2$ επιλέγονται έτσι ώστε να ισχύει η παρακάτω σχέση: [όπου $R = \mathbb{R}^{\ell}$]

$$R_1 = \{\underline{x} \in R : P(\omega_1|\underline{x}) > P(\omega_2|\underline{x})\}$$
$$R_2 = \{\underline{x} \in R : P(\omega_2|\underline{x}) > P(\omega_1|\underline{x})\}.$$

<u>Οριακή Συνάρτηση Πυκνότητας Πιθανότητας</u> (Marginal Probability Density Function)

$$p_x(x) = \int_y p_{x,y}(x,y)dy = \int_y p_{x|y}(x|y)p_y(y)dy = E_y\left[p_{x|y}(x|y)\right] \quad (12)$$

Με βάση τον ορισμό της οριακής πιθανότητας για την εκάστοτε κλάση $\{\omega_i, i \in [M]\}$ του προβλήματος Μ-ταξικής ταξινόμησης μπορούμε να γράψουμε ότι:

$$P(\omega_i) = \int_R P(\omega_i|\underline{x})p(\underline{x})dx = E\left[p(\omega_i|\underline{x})\right] \quad (13)$$

Λαμβάνοντας υπόψιν το γεγονός ότι οι περιοχές διακρίσεως $R_1$ και $R_2$ καλύπτουν το σύνολο του χώρου των χαρακτηριστικών $R$ έτσι ώστε:

$$R = R_1 \cup R_2 \quad \text{και} \quad R_1 \cap R_2 = \emptyset \qquad (14)$$

μπορούμε να εκφράσουμε την εκ των προτέρων πιθανότητα της κάθε κλάσης ως:

$$\boxed{P(\omega_i) = \int_{R_1} P(\omega_i|\underline{x})p(\underline{x})d\underline{x} + \int_{R_2} P(\omega_i|\underline{x})p(\underline{x})d\underline{x} \qquad (15)}$$

Κάνοντας χρήση της εξίσωσης (15) στο πλαίσιο του δυαδικού προβλήματος ταξινόμησης που αναλύουμε, έχουμε ότι:

$$\boxed{P(\omega_1) = \int_{R_1} P(\omega_1|\underline{x})p(\underline{x})d\underline{x} + \int_{R_2} P(\omega_1|\underline{x})p(\underline{x})d\underline{x} \qquad (16)}$$

Από την εξίσωση (16) προκύπτει ότι:

$$\boxed{\int_{R_2} P(\omega_1|\underline{x})p(\underline{x})d\underline{x} = P(\omega_1) - \int_{R_1} P(\omega_1|\underline{x})p(\underline{x})d\underline{x} \qquad (17)}$$

Συνδυάζοντας τις εξισώσεις (11) και (17), έχουμε ότι:

$$P_e = P(\omega_1) + \int_{R_1} p(\omega_2|\underline{x})p(\underline{x})d\underline{x} - \int_{R_1} P(\omega_1|\underline{x})p(\underline{x})d\underline{x} \Rightarrow$$

$$\boxed{P_e = P(\omega_1) - \int_{R_1} [P(\omega_1|\underline{x}) - p(\omega_2|\underline{x})]p(\underline{x})d\underline{x} \qquad (18)}$$

(*) Προφανώς, η πιθανότητα σφάλματος ελαχιστοποιείται όταν η ποσότητα που εμφανίζεται στο ολοκλήρωμα μεγιστοποιείται γεγονός που πραγματοποιείται όταν η περιοχή απόφασης $R_1$ επιλέγεται έτσι ώστε: $P(\omega_1|\underline{x}) > P(\omega_2|\underline{x})$.

Ο κανόνας απόφασης του Μπέυζ (Bayes) γενικεύεται σε προβλήματα Μ-ταξικής ταξινόμησης. Μάλιστα, σε ένα τέτοιο πρόβλημα διάκρισης μεταξύ των κατηγοριών $\omega_1, \omega_2, \ldots, \omega_M$ ένα άγνωστο πρότυπο ανατίθεται στην κλάση $\omega_i$ όταν:

$$\boxed{P(\omega_i|\underline{x}) > P(\omega_j|\underline{x}), \ \forall j \neq i. \quad (19)}$$

Μια εναλλακτική προσέγγιση για την εκτίμηση της συνολικής πιθανότητας σφάλματος του Μπεϋζιανού ταξινομητή είναι η εξής:

$$\boxed{P(\text{error}) = \int_R P(\text{error}, \underline{x})\,d\underline{x} = \int_R P(\text{error}|\underline{x})\,p(\underline{x})\,d\underline{x} \quad (20)}$$

Η ελαχιστοποίηση της συνολικής πιθανότητας σφάλματος πραγματοποιείται όταν η ποσότητα $P(\text{error}|\underline{x})$ γίνεται οσοδήποτε μικρή ώστε να ελαχιστοποιείται αντίστοιχα και η τιμή του ολοκληρώματος.
Ο μπεϋζιανός ταξινομητής εξασφαλίζει την εν λόγω συνθήκη καθώς προσδιορίζει την εκ των υστέρων πιθανότητα σφάλματος (posterior error probability / conditional error probability) ως εξής:

$$\boxed{P(\text{error}|\underline{x}) = \min\left\{ P(\omega_1|\underline{x}), P(\omega_2|\underline{x}) \right\} \quad (21)}$$

Η γενική μορφή του υπό συνθήκη σφάλματος (δεσμευμένου σφάλματος) για τον Μπεϋζιανό ταξινομητή δίνεται από την σχέση:

$$P(\text{error}|\underline{x}) = \begin{cases} P(\omega_1|\underline{x}), & \text{όταν η απόφαση είναι } \omega_2; \\ P(\omega_2|\underline{x}), & \text{όταν η απόφαση είναι } \omega_1. \end{cases}$$