ΚΕΣ 03 – Αναγνώριση Προτύπων και Ανάλυση Εικόνας

Μπαεσιανοί Ταξινομητές (Bayesian Classifiers)

Τμήμα Επιστήμης και Τεχνολογίας Τηλεπικοινωνιών

Πανεπιστήμιο Πελοποννήσου

Περιεχόμενα – Βιβλιογραφία

- Περιεχόμενα Ενότητας
  - Εισαγωγή
  - Θεωρία Bayes και ταξινόμηση
  - Συναρτήσεις διαχωρισμού
  - Ταξινομητές Bayes για κανονικές κατανομές

- Βιβλιογραφία:
  - Παπαμάρκος [2005]: Κεφάλαιο 7
  - Duda [2004]: Chapter 2
  - Theodoridis [2002]: Chapter 2
  - Bow [2002]: Chapter 2

Εισαγωγή

- Τυπικές εφαρμογές αναγνώρισης προτύπων
  - Υπολογιστική όραση
  - Οπτική αναγνώριση χαρακτήρων (OCR)
  - Διάγνωση με τη βοήθεια υπολογιστή
  - Αναγνώριση ομιλίας
  - Αναγνώριση προσώπων
  - Ταυτοποίηση προσώπων από βιομετρικά χαρακτηριστικά (δακτυλικά αποτυπώματα, γεωμετρία παλάμης, κ.ο.κ)

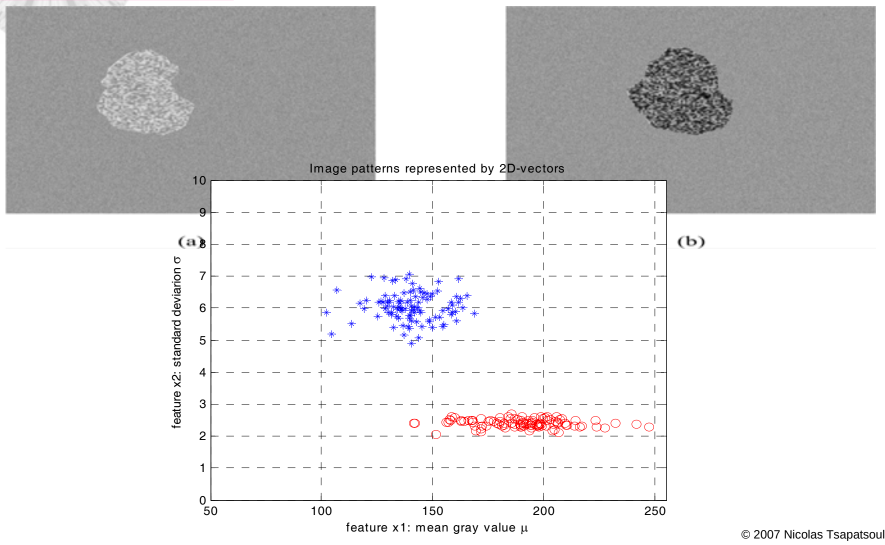
- Στόχος: Ταξινόμησε άγνωστα αντικείμενα – πρότυπα – στη σωστή κατηγορία. Η διαδικασία αυτή είναι γνωστή και ως ταξινόμηση ή κατηγοριοποίηση (classification)

Αναπαράσταση προτύπων

- Χαρακτηριστικά:
  - Μετρήσιμες ποσότητες οι οποίες εξάγονται από τα πρότυπα. Η ταξινόμηση των προτύπων σε κάποια κατηγορία βασίζεται στις τιμές των χαρακτηριστικών

- Διανύσματα χαρακτηριστικών:
  - Ένα σύνολο από χαρακτηριστικά $x_1, \ldots, x_l$, συνιστούν ένα διάνυσμα χαρακτηριστικών: $\underline{x} = [x_1, \ldots, x_l]^T \in R^l$
  - Τα διανύσματα χαρακτηριστικών θεωρούνται ως τυχαία διανύσματα δεδομένου ότι οι τιμές τους δεν είναι σταθερές για το πλήθος των προτύπων μιας κατηγορίες. Επιπλέον τα σφάλματα κατά τη διάρκεια των μετρήσεων ενισχύουν τη θεώρηση αυτή

Παράδειγμα

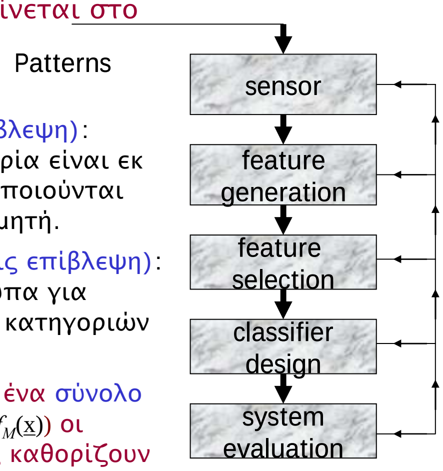


ΚΕΣ 03: Αναγνώριση Προτύπων και Ανάλυση Εικόνας

★ Εισαγωγή
☐ Θεωρία Bayes και ταξινόμηση
☐ Συναρτήσεις διαχωρισμού
☐ Ταξινομητές Bayes για κανονικές κατανομές

Ταξινομητές

- Η δομή ενός ταξινομητή εμφαίνεται στο διπλανό σχήμα:
- Δύο κατηγορίες ταξινόμησης:
  - Supervised (ταξινόμηση με επίβλεψη): Πρότυπα των οποίων η κατηγορία είναι εκ των προτέρων γνωστή χρησιμοποιούνται για την εκπαίδευση του ταξινομητή.
  - Unsupervised (ταξινόμηση χωρίς επίβλεψη): Δεν υπάρχουν διαθέσιμα πρότυπα για εκπαίδευση και το πλήθος των κατηγοριών είναι άγνωστο
- Ο ταξινομητής απαρτίζεται από ένα σύνολο συναρτήσεων του $\underline{x}$ ($f_1(\underline{x}), f_2(\underline{x}),\ldots, f_M(\underline{x})$) οι οποίες ανάλογα με την τιμή τους καθορίζουν σε ποια κατηγορία ανήκει το πρότυπο με διάνυσμα αναπαράστασης $\underline{x}$.



Θεωρία Bayes και ταξινόμηση

- Η στοχαστική υφή των διανυσμάτων αναπαράστασης των προτύπων οδηγεί στη θεωρία Bayes, η οποία προτείνει την ταξινόμηση του διανύσματος:

$$\underline{x} = [x_1, x_2, \dots, x_l]^T$$

στην πιθανότερη από τις κατηγορίες: $\omega_1, \omega_2, \dots, \omega_M$

δεδομένης της τιμής του διανύσματος x: $\underline{x} \rightarrow \omega_i : P(\omega_i | \underline{x})$ maximum

Υπολογισμός της εκ-των υστέρων (a-posteriori) πιθανότητας

- Η εκ των υστέρων πιθανότητες $P(\omega_i | \underline{x})$ $i = 1,2, \dots M$ υπολογίζονται με βάση τον κανόνα του Bayes
- Κανόνας Bayes (M=2)

$$p(\underline{x})P(\omega_i | \underline{x}) = p(\underline{x} | \omega_i)P(\omega_i) \Rightarrow$$

$$P(\omega_i | \underline{x}) = \frac{p(\underline{x} | \omega_i)P(\omega_i)}{p(\underline{x})}$$

όπου

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

  - Εκ των προτέρων (a-priori) πιθανότητες: $P(\omega_1), P(\omega_2) \dots, P(\omega_M)$
  - Πιθανοφάνεια (likelihood) του $\underline{x}$ ως προς τη κλάση ω_i:

$$p(\underline{x} | \omega_i), i = 1,2 \dots M$$

Κανόνας ταξινόμησης Bayes

- Για M = 2 (δύο κατηγορίες-κλάσεις):
  - Δεδομένου του x ταξινόμησε το σύμφωνα με τον κανόνα:

$$\boxed{\begin{array}{l} If\ P(\omega_1|\underline{x}) > P(\omega_2|\underline{x})\ \ \underline{x} \rightarrow \omega_1 \\ If\ P(\omega_2|\underline{x}) > P(\omega_1|\underline{x})\ \ \underline{x} \rightarrow \omega_2 \end{array}}$$

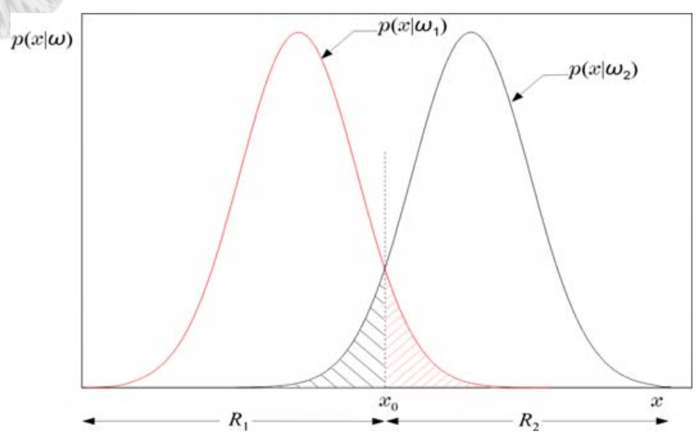
ή ισοδύναμα:

$$\boxed{p(\underline{x}|\omega_1)P(\omega_1)(><)p(\underline{x}|\omega_2)P(\omega_2)}$$

- Για ισοπίθανες κατηγορίες ο κανόνας απλοποιείται σε:

$$\boxed{p(\underline{x}|\omega_1)(><)p(\underline{x}|\omega_2)}$$

Κανόνας ταξινόμησης Bayes (II)



$$R_1(\rightarrow \omega_1)\ and\ \ R_2(\rightarrow \omega_2)$$

Κανόνας ταξινόμησης Bayes (III)

- Με βάση το προηγούμενο σχήμα:
  - Διαιρούμε το πεδίο τιμών του διανύσματος x σε δύο περιοχές R_1, R_2.

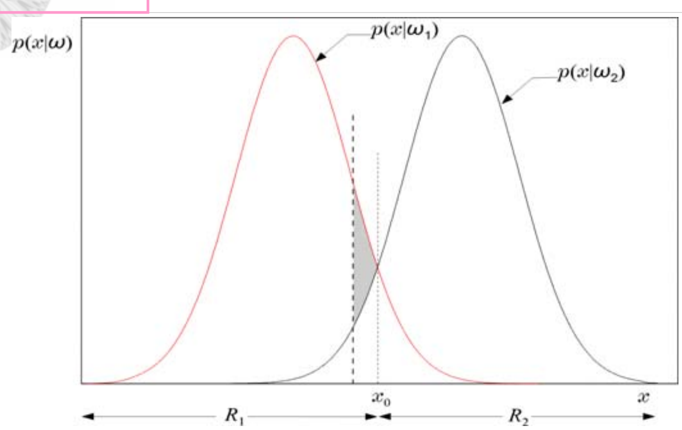
$$\boxed{\begin{array}{l} \text{If } \underline{x} \in R_1 \Rightarrow \underline{x} \text{ in } \omega_1 \\ \text{If } \underline{x} \in R_2 \Rightarrow \underline{x} \text{ in } \omega_2 \end{array}}$$

- Πιθανότητα εσφαλμένης ταξινόμησης:
  - Το εμβαδόν της σκιασμένης περιοχής:

$$P_e = \int_{-\infty}^{x_0} p(x|\omega_2)dx + \int_{x_0}^{+\infty} p(x|\omega_1)dx$$

- Ο ταξινομητής Bayes είναι βέλτιστος όσον αφορά την ελαχιστοποίηση του σφάλματος ταξινόμησης

Κανόνας ταξινόμησης Bayes (IV)



Πραγματικά: μετακινώντας το σημείο διαίρεσης των περιοχών (x_0) η σκιασμένη περιοχή αυξάνει κατά το εμβαδό της γκρι περιοχής

Κανόνας ταξινόμησης Bayes (VI)

- Για Μ > 2 (περισσότερες από δύο κατηγορίες-κλάσεις):
  - Δεδομένου του x ταξινόμησε το σύμφωνα με τον κανόνα:

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

  - Η ανωτέρω επιλογή ελαχιστοποιεί επίσης το σφάλμα ταξινόμησης

Ελαχιστοποίηση του μέσου ρίσκου (average risk) ταξινόμησης

- Σε ορισμένες περιπτώσεις η εσφαλμένη ταξινόμηση σε κάποιες κατηγορίες είναι περισσότερο σημαντική.
  - Για παράδειγμα η ταξινόμηση σολομού ως λαβράκι δεν είναι το ίδιο επιβαρυντική για τον καταναλωτή (ο οποίος είναι ωφελημένος σε αυτή την περίπτωση) από το αντίστροφο.
- Έστω η ταξινόμηση σε δύο κατηγορίες (Μ=2):
  - Ορίζουμε τον πίνακα απωλειών: $L = \begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{pmatrix}$

    όπου λ_{ij} το κόστος ταξινόμησης του x στη κλάση ω_j ενώ στη πραγματικότητα αυτό ανήκει στη κλάση ω_i.
- Το ρίσκο ταξινόμησης r_1 ως προς την κατηγορία ω_1, δίνεται από τη σχέση:

$$r_1 = \lambda_{11}\int_{R_1} p(\underline{x}|\omega_1)d\underline{x} + \boxed{\lambda_{12}\int_{R_2} p(\underline{x}|\omega_1)d\underline{x}}$$

ΚΕΣ 03: Αναγνώριση Προτύπων και Ανάλυση Εικόνας

Ελαχιστοποίηση του μέσου ρίσκου ταξινόμησης (II)

- Ομοίως, το ρίσκο ταξινόμησης r_2 ως προς την κατηγορία ω_2, δίνεται από τη σχέση:

$$r_2 = \boxed{\lambda_{21} \int_{R_1} p(\underline{x}|\omega_2)\, d\underline{x}} + \lambda_{22} \int_{R_2} p(\underline{x}|\omega_2)\, d\underline{x}$$

=> πιθανότητες εσφαλμένης ταξινόμησης πολλαπλασιασμένες με το αντίστοιχο κόστος ταξινόμησης

- Το μέσο ρίσκο ταξινόμησης δίνεται από τη σχέση:

$$r = r_1 P(\omega_1) + r_2 P(\omega_2)$$

ΚΕΣ 03: Αναγνώριση Προτύπων και Ανάλυση Εικόνας

☑ Εισαγωγή
★ Θεωρία Bayes και ταξινόμηση
☐ Συναρτήσεις διαχωρισμού
☐ Ταξινομητές Bayes για κανονικές κατανομές

Ελαχιστοποίηση του μέσου ρίσκου ταξινόμησης (III)

- Για την ελαχιστοποίηση του μέσου ρίσκου ταξινόμησης r επιλέγουμε κατάλληλα τις περιοχές R_1 και R_2.

- Η ταξινόμηση του $\underline{x}$ στην κατηγορία ω_1 πραγματοποιείται εφόσον πληρείται η σχέση:

$$\ell_1 \equiv \lambda_{11} p(\underline{x}|\omega_1) P(\omega_1) + \lambda_{21} p(\underline{x}|\omega_2) P(\omega_2) \quad <$$
$$\ell_2 \equiv \lambda_{12} p(\underline{x}|\omega_1) P(\omega_1) + \lambda_{22} p(\underline{x}|\omega_2) P(\omega_2)$$

- Ισοδύναμα: Ταξινόμησε το $\underline{x}$ στην κατηγορία ω_1 (ω_2) αν:

$$\boxed{\ell_{12} \equiv \frac{p(\underline{x}|\omega_1)}{p(\underline{x}|\omega_2)} > (<) \frac{P(\omega_2)}{P(\omega_1)} \frac{\lambda_{21} - \lambda_{22}}{\lambda_{12} - \lambda_{11}}}$$

όπου l_{12} ο λόγος πιθανοφανειών

Ελαχιστοποίηση του μέσου ρίσκου ταξινόμησης (IV)

- Αν $P(\omega_1) = P(\omega_2) = \frac{1}{2}$ και $\lambda_{11} = \lambda_{22} = 0$

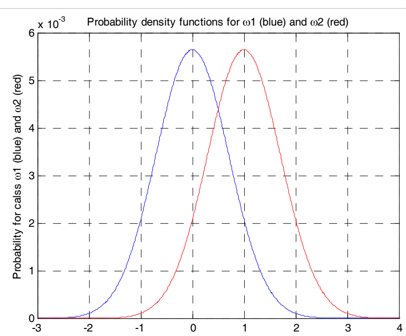
$$\underline{x} \to \omega_1 \text{ if } P(\underline{x}|\omega_1) > P(\underline{x}|\omega_2)\frac{\lambda_{21}}{\lambda_{12}}$$

$$\underline{x} \to \omega_2 \text{ if } P(\underline{x}|\omega_2) > P(\underline{x}|\omega_1)\frac{\lambda_{12}}{\lambda_{21}}$$

$$\text{if } \lambda_{21} = \lambda_{12} \Rightarrow \text{Ελαχιστοποίηση σφάλματος ταξινόμησης}$$

Παράδειγμα

- $p(x|\omega_1) = \frac{1}{\sqrt{\pi}}\exp(-x^2)$
- $p(x|\omega_2) = \frac{1}{\sqrt{\pi}}\exp(-(x-1)^2)$
- $P(\omega_1) = P(\omega_2) = \frac{1}{2}$
- $L = \begin{pmatrix} 0 & 0.5 \\ 1.0 & 0 \end{pmatrix}$



- Για ελαχιστοποίηση του σφάλματος ταξινόμησης:

$$x_0 \text{ for minimum } P_e:$$

$$x_0: \exp(-x^2) = \exp(-(x-1)^2) \Rightarrow$$

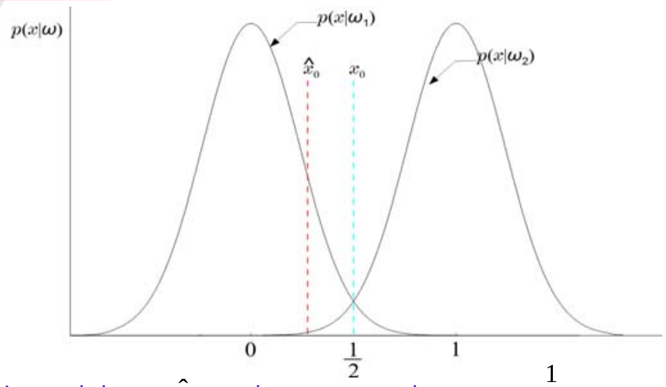
$$x_0 = \frac{1}{2}$$

- Για ελαχιστοποίηση του μέσου ρίσκου:

$$\hat{x}_0: \exp(-x^2) = 2\exp(-(x-1)^2) \Rightarrow$$

$$\hat{x}_0 = \frac{(1-\ell n2)}{2} < \frac{1}{2}$$

Παράδειγμα (συν.)



- Είναι φανερό ότι το $\hat{x}_0$ κινείται αριστερότερα του $\frac{1}{2} = x_0$
- Γιατί; Τι θα συνέβαινε αν είχαμε $\lambda_{12} > \lambda_{21}$;

Συναρτήσεις διαχωρισμού

- Αν οι περιοχές R_1 και R_2 γειτνιάζουν τότε η εξίσωση:

$$g(\underline{x}) \equiv P(\omega_i|\underline{x}) - P(\omega_j|\underline{x}) = 0$$

$$R_i : P(\omega_i|\underline{x}) > P(\omega_j|\underline{x})$$

$$+$$

$$g(\underline{x}) = 0$$

$$-$$

$$R_j : P(\omega_j|\underline{x}) > P(\omega_i|\underline{x})$$

ορίζει την επιφάνεια διαχωρισμού των περιοχών. Στη μια πλευρά της επιφάνειας η συνάρτηση g(x) παίρνει θετικές τιμές (+) και στην άλλη αρνητικές (-). Η επιφάνεια διαχωρισμού ονομάζεται και επιφάνεια απόφασης

Συναρτήσεις διαχωρισμού

- Αν $f(\cdot)$ είναι μια μονοτονική συνάρτηση τότε ταξινόμηση μπορεί να επιτευχθεί και με τη βοήθεια του κανόνα:

$$\underline{x} \rightarrow \omega_i \;\; \text{if: } f(P(\omega_i|\underline{x})) > f(P(\omega_j|\underline{x})) \;\forall i \neq j$$

- Η συνάρτηση $g_i(\underline{x}) \equiv f(P(\omega_i|\underline{x}))$ είναι μια συνάρτηση διαχωρισμού

- Συναρτήσεις διαχωρισμού μπορούν να ορισθούν ανεξάρτητα του κανόνα του Bayes οδηγώντας όμως σε υποβέλτιστη ταξινόμηση (όχι ελάχιστο σφάλμα ταξινόμησης)

ΚΕΣ 03: Αναγνώριση Προτύπων και Ανάλυση Εικόνας

Ταξινομητές Bayes για κανονικές κατανομές

- Μια πολυδιάστατη κανονική κατανομή (κατανομή Gauss) ορίζεται από τις σχέσεις:

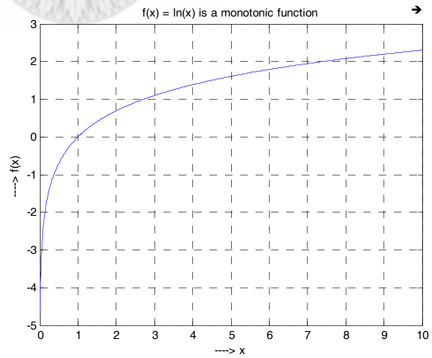
$$p(\underline{x}|\omega_i) = \frac{1}{(2\pi)^{\frac{\ell}{2}}|\Sigma_i|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\underline{x}-\underline{\mu}_i)^{\mathrm{T}}\Sigma_i^{-1}(\underline{x}-\underline{\mu}_i)\right)$$

$$\underline{\mu}_i = E[\underline{x}] \;\ell \times \ell \text{ matrix in } \omega_i$$

$$\Sigma_i = E\left[(\underline{x}-\underline{\mu}_i)(\underline{x}-\underline{\mu}_i)^{\mathrm{T}}\right]$$

Ο πίνακας Σ_i ονομάζεται πίνακας συμμεταβλητότητας.

Ταξινομητές Bayes για κανονικές κατανομές



- Η συνάρτηση f(x) = ln(x) για x>0 είναι μια μονοτονική συνάρτηση.

$$g_i(\underline{x}) = \ln(p(\underline{x}|\omega_i)P(\omega_i)) =$$
$$\ln p(\underline{x}|\omega_i) + \ln P(\omega_i)$$

$$g_i(\underline{x}) = -\frac{1}{2}(\underline{x}-\underline{\mu}_i)^T \Sigma_i^{-1}(\underline{x}-\underline{\mu}_i) + \ln P(\omega_i) + C_i$$

$$C_i = -(\frac{\ell}{2})\ln 2\pi - (\frac{1}{2})\ln|\Sigma_i|$$

ΚΕΣ 03: Αναγνώριση Προτύπων και Ανάλυση Εικόνας

☑ Εισαγωγή
☑ Θεωρία Bayes και ταξινόμηση
☑ Συναρτήσεις διαχωρισμού
★ Ταξινομητές Bayes για κανονικές κατανομές

Παράδειγμα

- Έστω $\Sigma_i = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}$

  τότε $g_i(\underline{x}) = -\frac{1}{2\sigma^2}(x_1^2 + x_2^2) + \frac{1}{\sigma^2}(\mu_{i1}x_1 + \mu_{i2}x_2)$

  $$-\frac{1}{2\sigma^2}(\mu_{i1}^2 + \mu_{i2}^2) + \ln(P\omega_i) + C_i$$

- Επομένως οι συναρτήσεις διαχωρισμού $g_i(\underline{x})$ είναι τετραγωνικής μορφής.
- Οι επιφάνειες διαχωρισμού $g_i(\underline{x}) - g_j(\underline{x}) = 0$ είναι ελλειψοειδή, παραβολοειδή, υπερβολοειδή ή ζεύγη γραμμών

Υπερεπίπεδα απόφασης

- Τετραγωνικοί όροι: $\underline{x}^T \Sigma_i^{-1} \underline{x}$

αν ισχύει $\Sigma_i = \Sigma$

δηλαδή όλοι οι πίνακες συμμεταβλητότητας είναι ίσοι τότε οι τετραγωνικοί όροι δεν συμμετέχουν στις συγκρίσεις. Οι συναρτήσεις διαχωρισμού απλοποιούνται, τότε, στη μορφή:

$$\boxed{\begin{aligned} g_i(\underline{x}) &= \underline{w}_i^T \underline{x} + w_{io} \\ \underline{w}_i &= \Sigma^{-1} \underline{\mu}_i \\ w_{i0} &= \ln P(\omega_i) - \frac{1}{2} \underline{\mu}^T{}_i \Sigma^{-1} \underline{\mu}_i \end{aligned}}$$

Επομένως οι συναρτήσεις διαχωρισμού $g_i(\underline{x})$ είναι γραμμικές.

Υπερεπίπεδα απόφασης (II)

- Αν επιπλέον οι πίνακες συμμεταβλητότητας είναι διαγώνιοι:

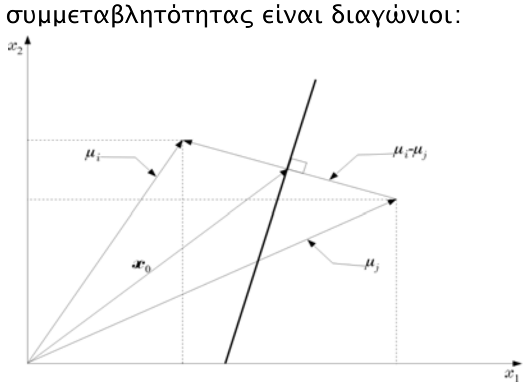
$$\Sigma = \sigma^2 I, \text{τότε}$$

$$g_i(\underline{x}) = \frac{1}{\sigma^2} \underline{\mu}_i^T \underline{x} + w_{i0}$$

$$g_{ij}(\underline{x}) = g_i(\underline{x}) - g_j(\underline{x}) = 0$$

$$= \underline{w}^T (\underline{x} - \underline{x}_o)$$

$$\underline{w} = \underline{\mu}_i - \underline{\mu}_j,$$



$$\underline{x}_o = \frac{1}{2}(\underline{\mu}_i + \underline{\mu}_j) - \sigma^2 \ln \frac{P(\omega_i)}{P(\omega_j)} \frac{\underline{\mu}_i - \underline{\mu}_j}{\left\| \underline{\mu}_i - \underline{\mu}_j \right\|^2}$$

Υπερεπίπεδα απόφασης (III)

- Αν οι πίνακες συμμεταβλητότητας είναι μη διαγώνιοι:

$$\Sigma \neq \sigma^2 I$$

$$g_{ij}(\underline{x}) = \underline{w}^T(\underline{x} - \underline{x}_0) = 0$$

$$\underline{x}_0 = \frac{1}{2}(\underline{\mu}_i + \underline{\mu}_j) - \ell n\left(\frac{P(\omega_i)}{P(\omega_j)}\right) \frac{\underline{\mu}_i - \underline{\mu}_j}{\left\|\underline{\mu}_i - \underline{\mu}_j\right\|^2_{\Sigma^{-1}}}$$

$$\left\|\underline{x}\right\|_{\Sigma^{-1}} \equiv (\underline{x}^T \Sigma^{-1} \underline{x})^{\frac{1}{2}}$$

- Δηλαδή το επίπεδο απόφασης δεν είναι πια κάθετο στην ευθεία $\underline{\mu}_i$-$\underline{\mu}_j$ αλλά στην ευθεία $\Sigma^{-1}(\underline{\mu}_i - \underline{\mu}_j)$

Ταξινομητές ελάχιστης απόστασης

- Έστω $P(\omega_i) = \frac{1}{M}$ (όλες οι κλάσεις ισοπίθανες)

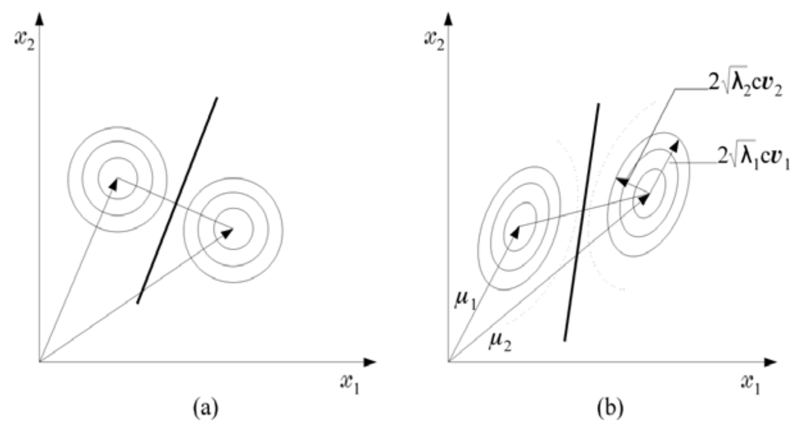
$$g_i(\underline{x}) = -\frac{1}{2}(\underline{x} - \underline{\mu}_i)^T \Sigma^{-1} (\underline{x} - \underline{\mu}_i)$$

- Αν $\Sigma = \sigma^2 I : \text{Ταξινόμησε } \underline{x} \rightarrow \omega_i :$
  σύμφωνα με την ελάχιστη Ευκλείδια απόσταση $d_E \equiv \left\|\underline{x} - \underline{\mu}_i\right\|$

- Αν $\Sigma \neq \sigma^2 I : \text{Ταξινόμησε } \underline{x} \rightarrow \omega_i :$
  σύμφωνα με την ελάχιστη απόσταση Mahalanobis:

$$d_m = ((\underline{x} - \underline{\mu}_i)^T \Sigma^{-1} (\underline{x} - \underline{\mu}_i))^{\frac{1}{2}}$$

Ταξινομητές ελάχιστης απόστασης (II)



Παράδειγμα

Κλάσεις ω_1, ω_2 : $P(\omega_1) = P(\omega_2)$ και $p(\underline{x}|\omega_1) = N(\underline{\mu}_1, \Sigma)$,

$p(\underline{x}|\omega_2) = N(\underline{\mu}_2, \Sigma)$, $\underline{\mu}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, $\underline{\mu}_2 = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$, $\Sigma = \begin{bmatrix} 1.1 & 0.3 \\ 0.3 & 1.9 \end{bmatrix}$

να ταξινομηθεί το διάνυσμα $\underline{x} = \begin{bmatrix} 1.0 \\ 2.2 \end{bmatrix}$ με τον κανόνα του Bayes :

- $\Sigma^{-1} = \begin{bmatrix} 0.95 & -0.15 \\ -0.15 & 0.55 \end{bmatrix}$

- Η απόσταση Mahalanobis d_m από μ_1, μ_2 : $d^2_{m,1} = [1.0, \quad 2.2]\, \Sigma^{-1} \begin{bmatrix} 1.0 \\ 2.2 \end{bmatrix} = 2.952$, $d^2_{m,2} = [-2.0, \quad -0.8]\, \Sigma^{-1} \begin{bmatrix} -2.0 \\ -0.8 \end{bmatrix} = 3.672$

- Ταξινόμηση $\underline{x} \rightarrow \omega_1$. Παρατηρήστε ότι $d_{E,2} < d_{E,1}$