

# ❖ ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ ❖ (PATTERN RECOGNITION)

**Σέργιος Θεοδωρίδης**  
**Κωνσταντίνος Κουτρούμπας**

# ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ (PATTERN RECOGNITION)

- ❖ Τυπικές περιοχές εφαρμογής
  - Μηχανική όραση (Machine vision)
  - Αναγνώριση χαρακτήρων (Character recognition (OCR))
  - Ιατρική διάγνωση υποβοηθούμενη από Η/Υ( Computer aided medical diagnosis)
  - Αναγνώριση ομιλίας (Speech recognition)
  - Αναγνώριση προσώπου (Face recognition)
  - Βιομετρία (Biometrics)
  - Ανάσυρση εικόνων από Βάσεις Δεδομένων (Image Data Base retrieval)
  - Εξόρυξη δεδομένων (Data mining)
  - Βιοπληροφορική (Bionformatics)
- ❖ **Το πρόβλημα:** Καταχώρηση άγνωστων αντικειμένων – **προτύπων** – στη σωστή κατηγορία (κλάση). Το πρόβλημα είναι γνωστό ως **ταξινόμηση (classification)**.

❖ Αναγνώριση προτύπων με επίβλεψη (supervised) – χωρίς επίβλεψη (unsupervised):

➤ **Με επίβλεψη:**

- Γνωστός αριθμός κλάσεων (κατηγοριών)
- Διαθέσιμα αντικείμενα για τα οποία είναι γνωστή η κλάση στην οποία ανήκουν.

➤ **Χωρίς επίβλεψη:**

- Άγνωστος αριθμός κλάσεων (γενικά).
- Διαθέσιμα αντικείμενα για τα οποία δεν είναι γνωστή οποιαδήποτε πληροφορία σχετική με κλάση.

## Παραδείγματα:

- ❖ Εφαρμογή στην κυτταρολογία.
  - Αντικείμενα: πυρήνες κυττάρων
  - Κατηγορίες: καλοήθη, κακοήθη
- ❖ Εφαρμογή σε αλυσίδα παραγωγής στη βιομηχανία.
  - Αντικείμενα: Π.χ. βίδες
  - Κατηγορίες: Ελαττωματική, μη ελαττωματική
- ❖ Εφαρμογή στην αναγνώριση χαρακτήρων κειμένου
  - Αντικείμενα: Αλφαριθμητικοί χαρακτήρες, σημεία στίξης
  - Κατηγορίες: 26 κατηγορίες πεζών+26 κατ. κεφαλαίων+10 κατ. ψηφίων +κατηγορίες σημείων στίξης.

❖ **Χαρακτηριστικά (Features)**: Πρόκειται για μετρήσιμες ποσότητες που λαμβάνονται από τα προς ταξινόμηση αντικείμενα. Η διαδικασία της ταξινόμησης βασίζεται στις αντίστοιχες τιμές τους.

❖ **Διανύσματα χαρακτηριστικών (Feature vectors)**: Ένας αριθμός χαρακτηριστικών

$$x_1, \dots, x_l,$$

συνιστούν το **διάνυσμα χαρακτηριστικών**

$$\underline{x} = [x_1, \dots, x_l]^T \in R^l$$

Τα διανύσματα χαρακτηριστικών θεωρούνται ως **τυχαίες μεταβλητές (random vectors)**.

Αναπαράσταση αντικειμένων σε σύστημα αναγνώρισης προτύπων:

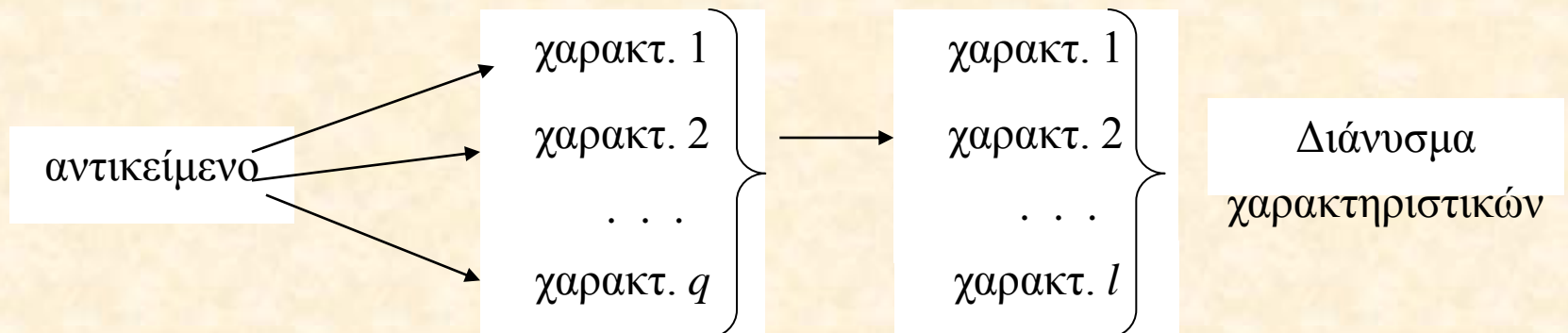
Αντικείμενο  $\leftarrow \equiv \equiv \rightarrow$  Διάνυσμα χαρακτηριστικών μετρήσεων (feature vector)

Αντικείμενο  $\leftarrow = \rightarrow$  Σημείο στο χώρο

## Τυπική δομή ενός συστήματος ταξινόμησης προτύπων



## Πώς δουλεύει



- ❖ Ο ταξινομητής (classifier) αποτελείται από ένα σύνολο συναρτήσεων, των οποίων οι τιμές, υπολογισμένες στο  $\underline{x}$ , καθορίζουν την κλάση στην οποία το αντίστοιχο πρότυπο θα καταχωρηθεί.

## Σημαντικές παρατηρήσεις:

- ❖ Στην πράξη, είναι απαραίτητη η ύπαρξη ενός αντιπροσωπευτικού δείγματος προτύπων από κάθε κατηγορία (συνήθως, όσο περισσότερα είναι διαθέσιμα, τόσο καλύτερη εικόνα έχουμε για τις κλάσεις).
- ❖ Στη συνέχεια ακολουθεί μελέτη των προτύπων ώστε να προσδιορίσουμε τα χαρακτηριστικά που διαφοροποιούν πιο πολύ τα πρότυπα των δύο κλάσεων



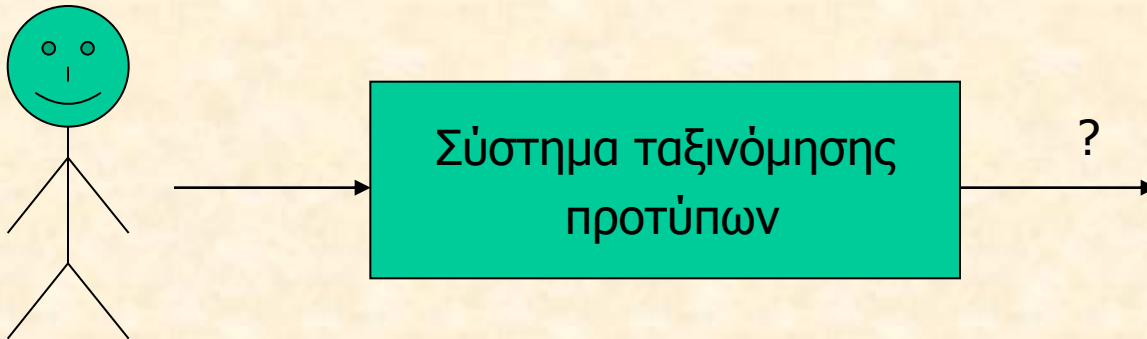
## Προβλήματα:

- (A) Ποια χαρακτηριστικά θα επιλέξω;
- (B) Πώς ο ταξινομητής διαχωρίζει τις κλάσεις;

## Παράδειγμα:

Να δημιουργηθεί ένα σύστημα ταξινόμησης προτύπων που θα διαχωρίζει καλαθοσφαιριστές (B) από χορευτές (D).

- Αντικείμενα: Αθλητές (καλαθοσφαιριστές, χορευτές)
- Κατηγορίες: Καλαθοσφαιριστές (B), χορευτές (D).



## (A) Χαρακτηριστικά

*ύψος, βάρος.*

Περιμένουμε ότι

	D	B
Βάρος	μικρό	μεγάλο
Ύψος	μικρό	μεγάλο

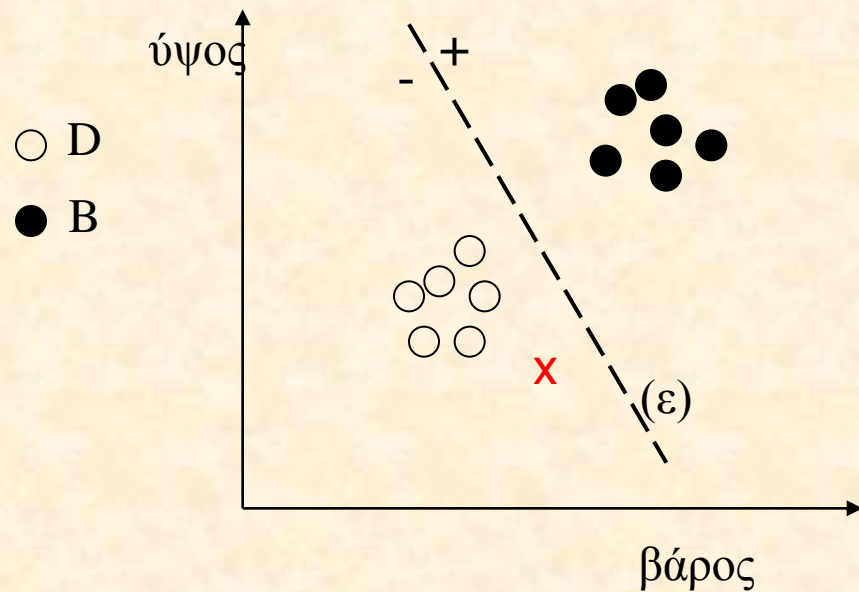
Πώς καθορίζεται ποσοτικά όμως το «**μεγάλο**» και το «**μικρό**»;

Βάσει ενός αντιπροσωπευτικού δείγματος που περιλαμβάνει αντικείμενα από τις δύο κλάσεις

Όσο πιο πλούσιο είναι το δείγμα, τόσο πιο ακριβές είναι το νόημα που αποκτούν οι παραπάνω χαρακτηρισμοί.

## (B) Ταξινομητής

Σύνολο συναρτήσεων/κανόνων  
βάσει των οποίων γίνεται η  
ταξινόμηση.

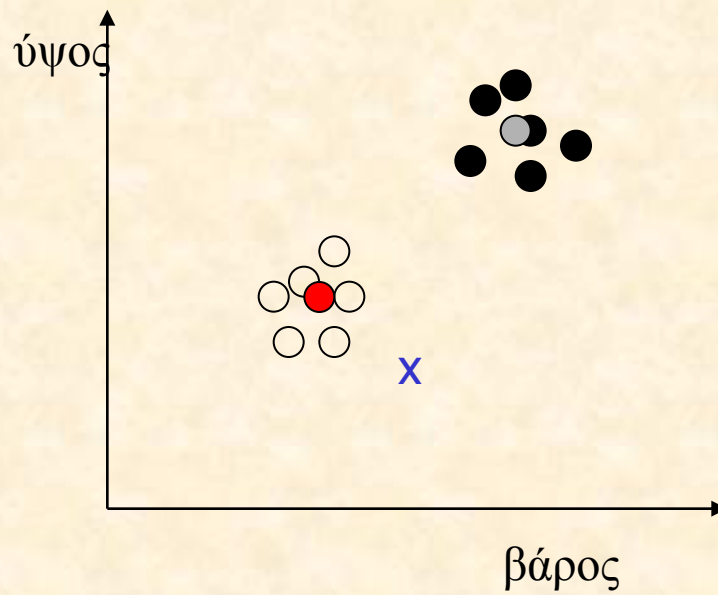


### Παράδειγμα 1:

Δοθέντος ενός χαρακτηριστικού διανύσματος, ο ταξινομητής εξετάζει αν αυτό βρίσκεται

(α) στην αρνητική πλευρά της  $(\epsilon)$ , οπότε το κατατάσσει στην κατηγορία “D”

(β) στη θετική πλευρά της  $(\epsilon)$ , οπότε το κατατάσσει στην κατηγορία “B”

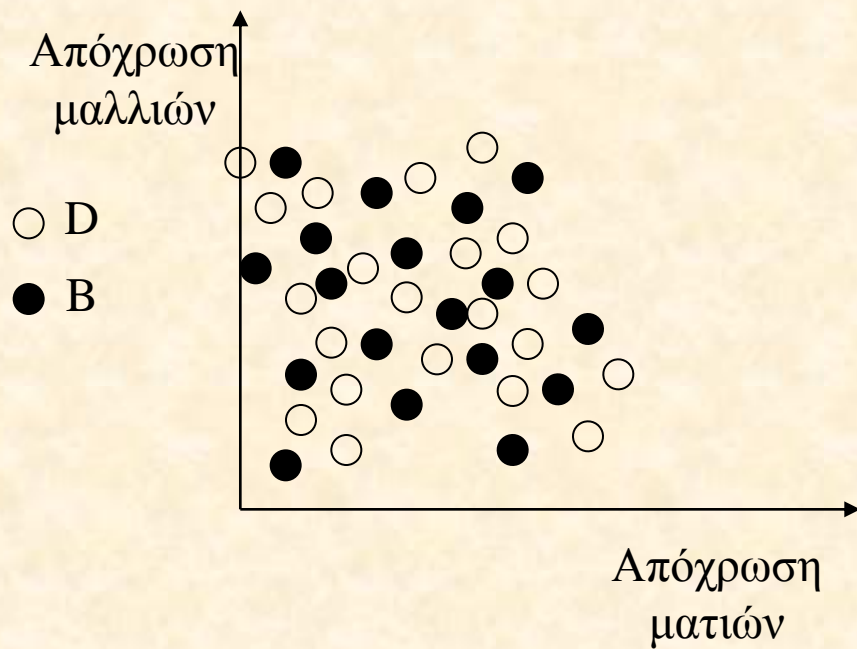


## Παράδειγμα 2:

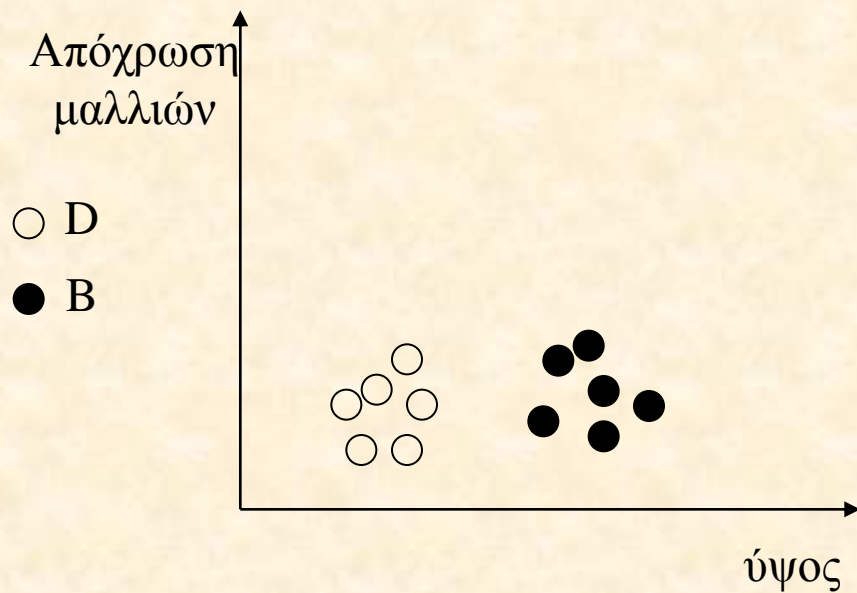
Δοθέντος ενός χαρακτηριστικού διανύσματος, ο ταξινομητής προσδιορίζει το πλησιέστερο μέσο διάνυσμα και ταξινομεί το διάνυσμα στην αντίστοιχη κατηγορία.

# Χαρακτηριστικά (ξανά)

*απόχρωση μαλλιών, απόχρωση ματιών*



# Χαρακτηριστικά (ξανά) απόχρωση μαλλιών, ύψος



# Χαρακτηριστικά (ξανά) ύψος



Η επιλογή χαρακτηριστικών είναι ζωτικής σημασίας για την απόδοση του συστήματος.

Γενικά είναι επιθυμητή η χρήση χαρακτηριστικών που διαχωρίζουν καλά τις κατηγορίες. Αυτό (συνήθως) σημαίνει

- Οι μέσες τιμές του χαρακτηριστικού να διαφέρουν σημαντικά μεταξύ τους από κατηγορία σε κατηγορία.
- Οι διασπορές γύρω από τις παραπάνω μέσες τιμές να είναι μικρές.

**Ερώτηση:** Αφού με τα παραπάνω μπορούμε να δημιουργήσουμε ένα σύστημα ταξινόμησης χρειάζεται περαιτέρω ανάπτυξη θεωριών; Αν ναι γιατί;

**Απάντηση:** Πρόβλημα διάστασης – Πολυπλοκότητα προβλήματος

## Μερικά προκαταρκτικά

Έστω  $\mathcal{A}_i, i = 1, 2, \dots, M$ ,  $M$  ενδεχόμενα έτσι ώστε  $\sum_{i=1}^M P(\mathcal{A}_i) = 1$

Η πιθανότητα για ένα αυθαίρετο ενδεχόμενο  $\mathcal{B}$  είναι

$$P(\mathcal{B}) = \sum_{i=1}^M P(\mathcal{B}|\mathcal{A}_i)P(\mathcal{A}_i)$$

Όπου  $P(\mathcal{B}|\mathcal{A})$  η υπό συνθήκη πιθανότητα του  $\mathcal{B}$  δοθέντος του  $\mathcal{A}$  :

$$P(\mathcal{B}|\mathcal{A}) = \frac{P(\mathcal{B},\mathcal{A})}{P(\mathcal{A})}$$

και  $P(\mathcal{B}, \mathcal{A})$  η από κοινού πιθανότητα των  $\mathcal{A}$  και  $\mathcal{B}$

## Θεώρημα συνολικής πιθανότητας



## Μερικά προκαταρκτικά (συν.)

Είναι

$$P(\mathcal{B}|\mathcal{A})P(\mathcal{A}) = P(\mathcal{A}|\mathcal{B})P(\mathcal{B})$$

Για τυχαίες μεταβλητές ή διανύσματα τυχαίων μεταβλητών που περιγράφονται από συναρτήσεις πυκνότητας πιθανότητας είναι

$$p(\mathbf{x}|\mathcal{A})P(\mathcal{A}) = P(\mathcal{A}|\mathbf{x})p(\mathbf{x})$$

$$p(\mathbf{x}|\mathbf{y})p(\mathbf{y}) = p(\mathbf{y}|\mathbf{x})p(\mathbf{x})$$

$$p(\mathbf{x}) = \sum_{i=1}^M p(\mathbf{x}|\mathcal{A}_i)P(\mathcal{A}_i)$$

$$P(\mathcal{A}_j|\mathbf{x}) = \frac{p(\mathbf{x}|\mathcal{A}_j)P(\mathcal{A}_j)}{\sum_{i=1}^M p(\mathbf{x}|\mathcal{A}_i)P(\mathcal{A}_i)}$$

**A posteriori (εκ των υστέρων) πιθανότητα**

## Μερικά προκαταρκτικά (συν.)

Αν

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1l} \\ a_{21} & a_{22} & \cdots & a_{2l} \\ \vdots & \vdots & \vdots & \vdots \\ a_{l1} & a_{l2} & \cdots & a_{ll} \end{bmatrix} \quad \mathbf{x} = [x_1, x_2, \dots, x_l]^T$$

ΤΟΤΕ

$$\mathbf{x}^T A \mathbf{x} = \sum_{i=1}^l \sum_{j=1}^l a_{ij} x_i x_j$$

Αν ο  $A$  είναι συμμετρικός τότε

$$\mathbf{x}^T A \mathbf{x} = \sum_{i=1}^l a_{ii} x_i^2 + 2 \sum_{i=1}^l \sum_{j>i}^l a_{ij} x_i x_j$$

Αν ο  $A$  είναι διαγώνιος τότε

$$\mathbf{x}^T A \mathbf{x} = \sum_{i=1}^l a_{ii} x_i^2$$

## Μερικά προκαταρκτικά (συν.)

$$Av \quad A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1l} \\ a_{21} & a_{22} & \cdots & a_{2l} \\ \vdots & \vdots & \vdots & \vdots \\ a_{l1} & a_{l2} & \cdots & a_{ll} \end{bmatrix} \quad \mathbf{x} = [x_1, x_2, \dots, x_l]^T$$

Ο  $A$  είναι **οριστικά θετικός (μη αρνητικός)** αν

$$\mathbf{x}^T A \mathbf{x} > (\geq) 0$$

# Μερικά προκαταρτικά (συν.)

Μονοδιάστατη Gaussian/κανονική κατανομή

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Πολυδιάστατη Gaussian/κανονική κατανομή  $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

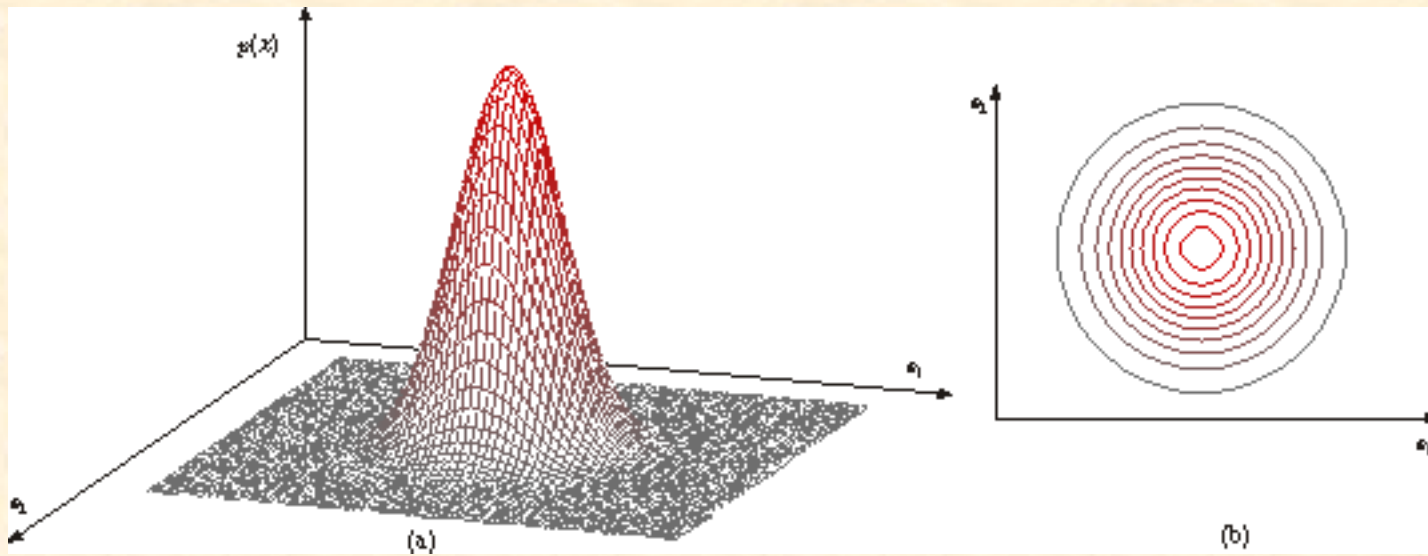
$$p(\mathbf{x}) = \frac{1}{(2\pi)^{l/2} |\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right)$$

όπου

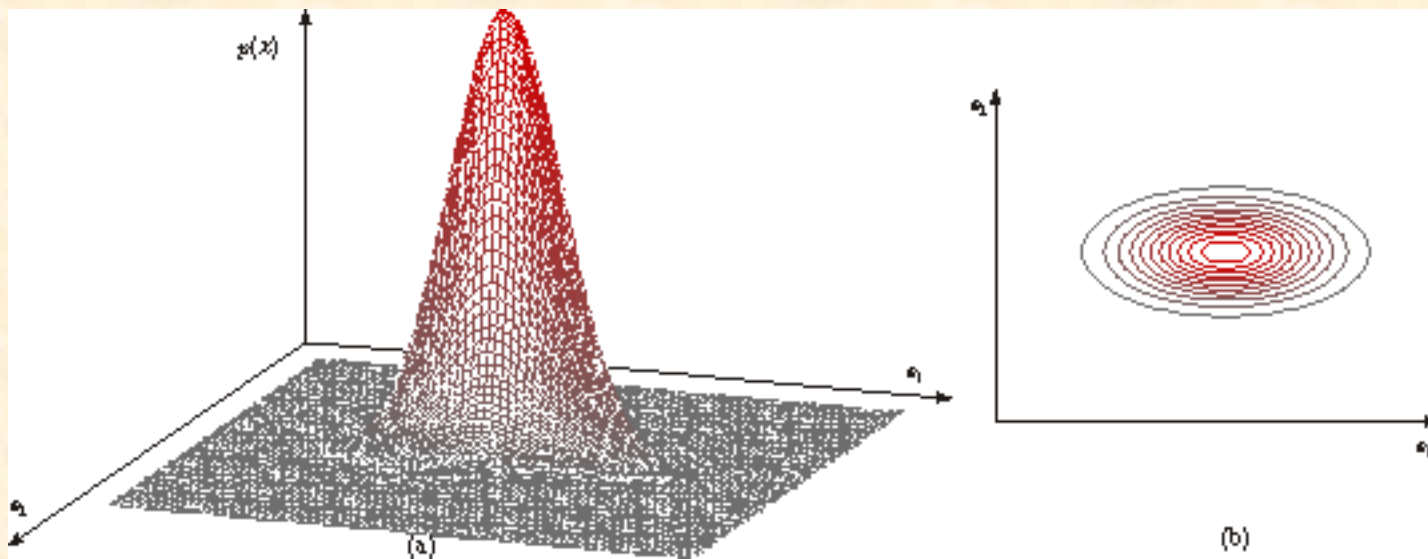
$$E(\mathbf{x}) \equiv E\left[[x_1, x_2, \dots, x_l]^T\right] = [\mu_1, \mu_2, \dots, \mu_l]^T \equiv \boldsymbol{\mu}$$

$$\boldsymbol{\Sigma} = E[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T] = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1l} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2l} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{l1} & \sigma_{l2} & \cdots & \sigma_l^2 \end{bmatrix}$$

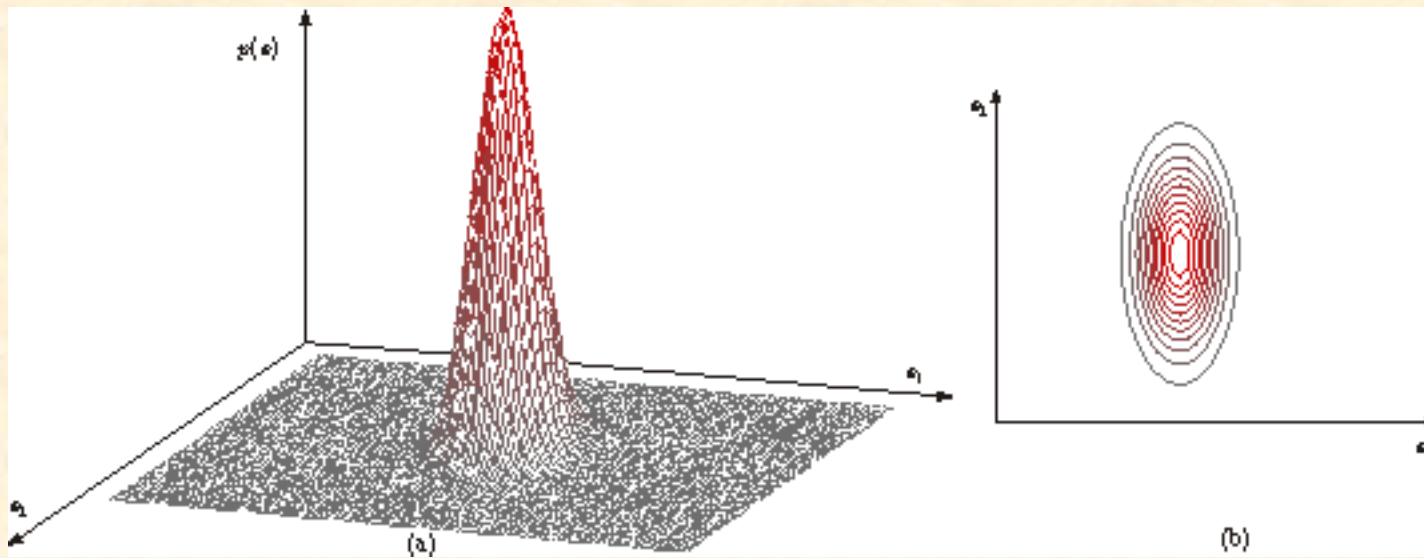
$$\sigma_i^2 = E[(x_i - \mu_i)^2], \sigma_{ij} = \sigma_{ji} = E[(x_i - \mu_i)(x_j - \mu_j)]$$



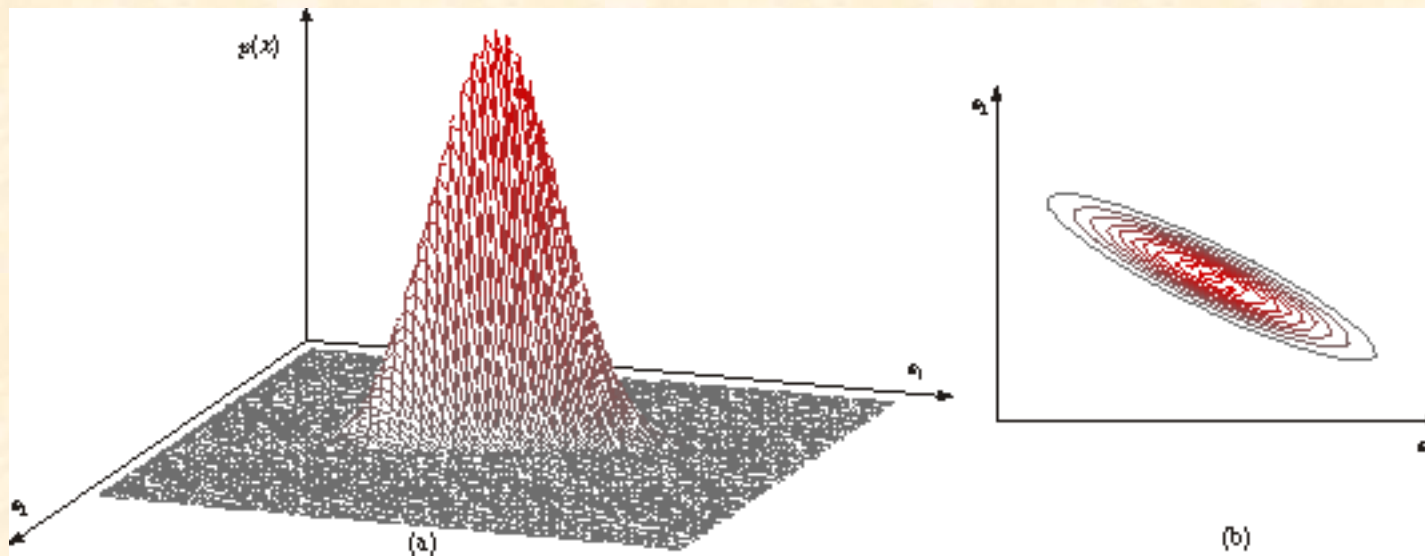
$\Sigma$ : διαγώνιος  
με ίσα  
διαγώνια  
στοιχεία



$\Sigma$ : διαγώνιος  
με  $\sigma_1^2 \gg \sigma_2^2$

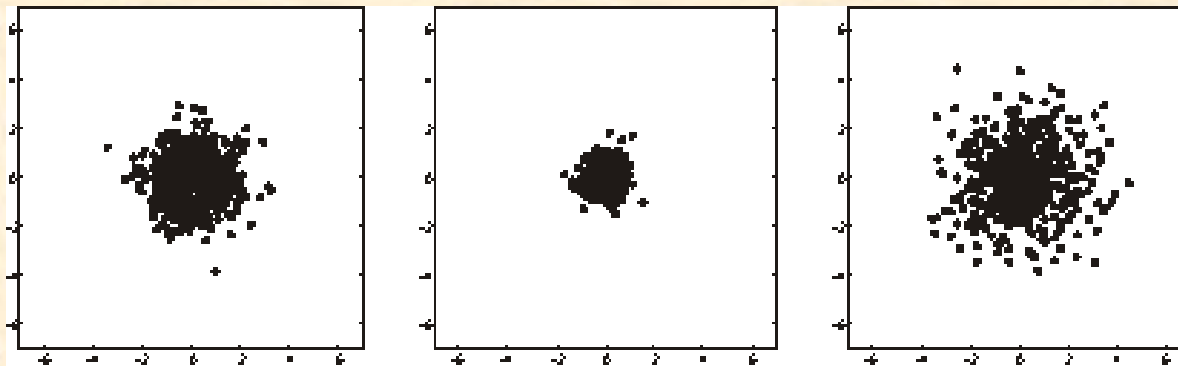


$\Sigma$ : διαγώνιος  
 με  $\sigma_1^2 \ll \sigma_2^2$



$\Sigma$ : μη διαγώνιος

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$$



(a)

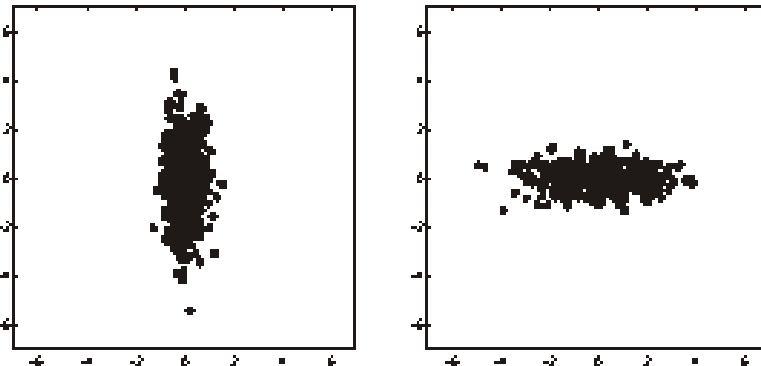
(b)

(c)

(a)  $\sigma_1^2 = \sigma_2^2 = 1, \sigma_{12} = 0$

(b)  $\sigma_1^2 = \sigma_2^2 = 0.2, \sigma_{12} = 0$

(c)  $\sigma_1^2 = \sigma_2^2 = 2, \sigma_{12} = 0$

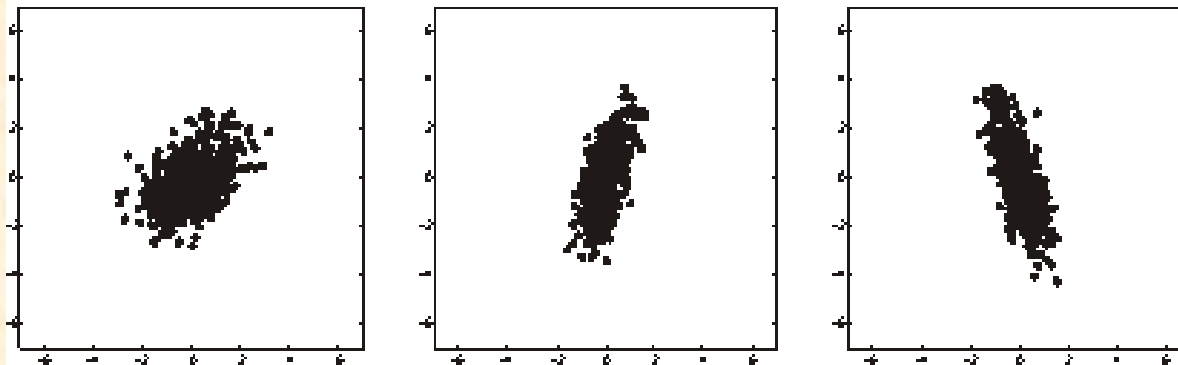


(d)

(e)

(d)  $\sigma_1^2 = 0.2, \sigma_2^2 = 2, \sigma_{12} = 0$

(e)  $\sigma_1^2 = 2, \sigma_2^2 = 0.2, \sigma_{12} = 0$



(f)

(g)

(h)

(f)  $\sigma_1^2 = \sigma_2^2 = 1, \sigma_{12} = 0.5$

(g)  $\sigma_1^2 = 0.3, \sigma_2^2 = 2, \sigma_{12} = 0.5$

(h)  $\sigma_1^2 = 0.3, \sigma_2^2 = 2, \sigma_{12} = -0.5$

## Μερικά προκαταρκτικά (συν.)

Αν τα  $x_i, x_j$  είναι ανά δύο στατιστικά ανεξάρτητες τυχ. μεταβλητές,

$$\sigma_{ij} = \sigma_{ji} = E[(x_i - \mu_i)(x_j - \mu_j)] = 0$$

τότε ο πίνακας  $\Sigma$  είναι διαγώνιος

**Στην περίπτωση της πολυδιάστατης κανονικής κατανομής ισχύει:**

$\Sigma$  διαγώνιος  $\Leftrightarrow$  οι συνιστώσες του  $x$  είναι στατιστικά ανεξάρτητες

$$p(\mathbf{x}) = \prod_{i=1}^l p_i(x_i)$$

$$p_i(x_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{(x_i - \mu_i)^2}{2\sigma_i^2}\right)$$



Στη συνέχεια θα μιλήσουμε μόνο για στατιστικές μεθόδους ταξινόμησης προτύπων (statistical pattern classification).

**Υπόθεση:** Σε μια μεγάλη αίθουσα γυμναστηρίου βρίσκονται 100 χορευτές και 200 καλαθοσφαιριστές

**Ερώτηση 1:** «Αν επιλέξουμε κάποιον από την αίθουσα, ποια είναι η πιθανότητα να είναι χορευτής;»

**Απάντηση:** Για χορευτή:  $P(D)=100/(100+200)=1/3$ .

Για καλαθοσφαιριστή:  $P(B)=200/(100+200)=2/3$

**Επιπλέον δεδομένα:** Οι κατανομές πυκνότητας πιθανότητας των βαρών των δύο κατηγοριών:

$$p(x / D) = \frac{1}{\sqrt{300\pi}} e^{-\frac{(x-50)^2}{300}} \quad p(x / B) = \frac{1}{\sqrt{300\pi}} e^{-\frac{(x-90)^2}{300}} .$$

**Ερώτηση 2:** «Αν επιλέξουμε κάποιον από την αίθουσα, ποια είναι η πιθανότητα να είναι χορευτής;»

**Απάντηση:** Ίδια με την ερώτηση 1.

**Ερώτηση 3:** «Αν επιλέξουμε κάποιον από την αίθουσα, ποια είναι η πιθανότητα να είναι χορευτής, δοθέντος ότι το ύψος του είναι 155 εκ;»

**Απάντηση:** Ίδια με την ερώτηση 1.

**Ερώτηση 4:** «Αν επιλέξουμε κάποιον από την αίθουσα, ποια είναι η πιθανότητα να είναι χορευτής, δοθέντος ότι το βάρος του είναι 60 κιλά;»

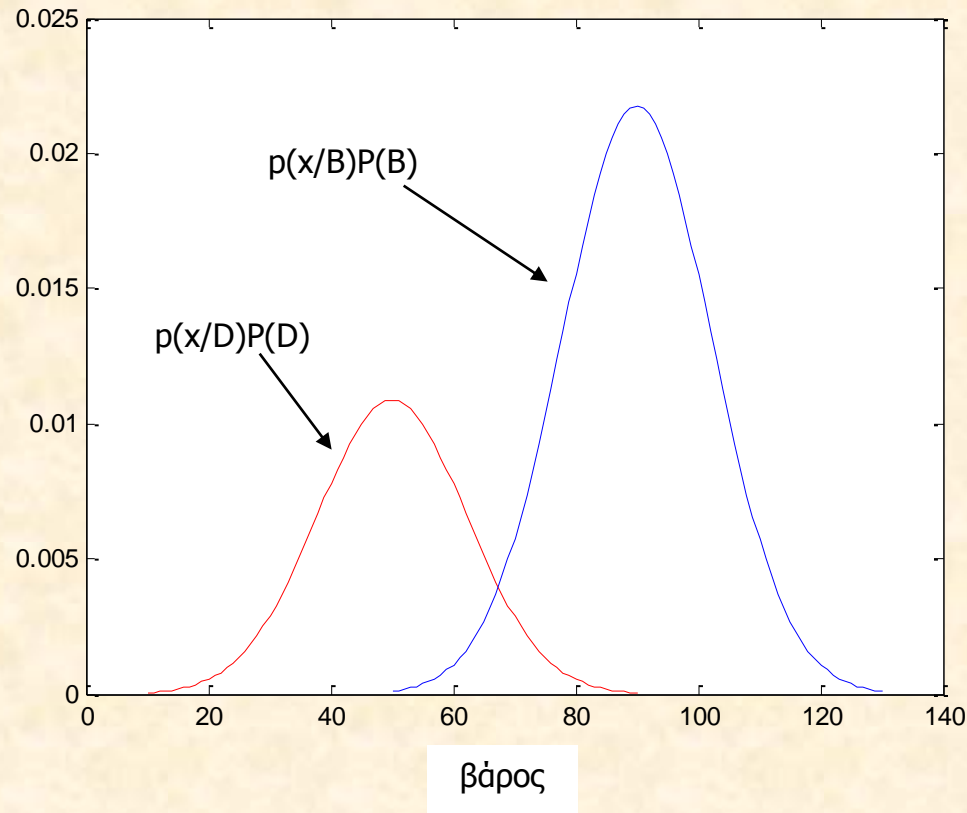
**Απάντηση:**

$$P(D / \beta = 60) = \frac{p(60 / D)P(D)}{p(60 / D)P(D) + p(60 / B)P(B)} = 0.8780$$

$$P(\kappa / \beta = 60) = \frac{p(60 / \kappa)P(\kappa)}{p(60 / \alpha)P(\alpha) + p(60 / \kappa)P(\kappa)} = 0.1220$$

**Προβληματισμός:** Γιατί η κατάσταση άλλαξε τόσο δραματικά με την επιπλέον πληροφορία;

**Εξήγηση:** Οι τιμές του βάρους διαφοροποιούνται πολύ στις δύο κατηγορίες.



**Υπόθεση:** Ας θεωρήσουμε ότι το επιπλέον χαρακτηριστικό είναι η απόχρωση των μαλλιών (η ίδια ομοιόμορφη κατανομή και για τις δύο κατηγορίες)

$$p(x/D) = \begin{cases} \frac{1}{\chi_2 - \chi_1}, & x \in [\chi_1, \chi_2] \\ 0 & x \notin [\chi_1, \chi_2] \end{cases} \quad p(x/B) = \begin{cases} \frac{1}{\chi_2 - \chi_1}, & x \in [\chi_1, \chi_2] \\ 0 & x \notin [\chi_1, \chi_2] \end{cases}$$

**Ερώτηση 5:** «Αν επιλέξουμε κάποιον από την αίθουσα, ποια είναι η πιθανότητα να είναι χορευτής, δοθέντος ότι η απόχρωση των μαλλιών του είναι  $y \in [\chi_1, \chi_2]$ ».

**Απάντηση:**

$$P(D / \chi = y) = \frac{p(y/D)P(D)}{p(y/D)P(D) + p(y/B)P(B)} = \frac{\frac{1}{\chi_2 - \chi_1} \frac{1}{3}}{\frac{1}{\chi_2 - \chi_1} \frac{1}{3} + \frac{1}{\chi_2 - \chi_1} \frac{2}{3}} = \frac{1}{3} = P(D)$$

$$P(B / \chi = y) = \frac{p(y/B)P(B)}{p(y/D)P(D) + p(y/B)P(B)} = \frac{\frac{1}{\chi_2 - \chi_1} \frac{2}{3}}{\frac{1}{\chi_2 - \chi_1} \frac{1}{3} + \frac{1}{\chi_2 - \chi_1} \frac{2}{3}} = \frac{2}{3} = P(B).$$

**Συμπέρασμα:** Αν έχουμε τη δυνατότητα να επιλέξουμε χαρακτηριστικά ας διαλέγουμε εκείνα που διαφοροποιούνται όσο το δυνατόν περισσότερο στις υπό εξέταση κατηγορίες (τα καταλληλότερα χαρακτηριστικά επιλέγονται με τη βοήθεια ενός ειδικού στην υπό μελέτη εφαρμογή).

Έστω και πάλι ότι το επιπλέον χαρακτηριστικό είναι το βάρος

**Ερώτηση 6:** «Αν επιλέξω κάποιον αθλητή από την αίθουσα, τι είδους αθλητής είναι αυτός, δεδομένου ότι το βάρος του είναι  $x$  κιλά;»

**Καλύτερη δυνατή απάντηση:** Είναι αθλητής από την κατηγορία με τη μεγαλύτερη εκ των υστέρων (a posteriori) πιθανότητα.

**Κανόνας του Bayes (1):**

Αν  $P(D/x) > P(B/x)$  τότε καταχώρησε τον αθλητή στην κατηγορία «χορευτής»

Διαφορετικά καταχώρησε τον αθλητή στην κατηγορία «καλαθοσφαιριστής».

Λαμβανομένου υπόψη ότι

$$P(D/x) = \frac{p(x/D)P(D)}{p(x/D)P(D) + p(x/B)P(B)}$$

$$P(B/x) = \frac{p(x/B)P(B)}{p(x/D)P(D) + p(x/B)P(B)}$$

**Κανόνας του Bayes (2):**

Αν  $P(D)p(x/D) > P(B)p(x/B)$  τότε καταχώρησε τον αθλητή στην κατηγορία «χορευτής»

Διαφορετικά καταχώρησε τον αθλητή στην κατηγορία «καλαθοσφαιριστής».

**Υπόθεση:** Οι a priori πιθανότητες των κατηγοριών είναι ίσες.

**Κανόνας του Bayes (ειδ):**

Αν  $p(x/D) > p(x/B)$  τότε καταχώρησε τον αθλητή στην κατηγορία «χορευτής»

Διαφορετικά καταχώρησε τον αθλητή στην κατηγορία «καλαθοσφαιριστής».

**Σημαντική παρατήρηση:** ΠΑΝΤΑ υπάρχει και η πιθανότητα λάθους (ένας «αφύσικα» αδύνατος καλαθοσφαιριστής η ένας «αφύσικα» παχουλός χορευτής μπορούν να ταξινομηθούν λάθος).

## Ταξινόμηση κατά Bayes και γεωμετρική ταξινόμηση

- Τα σημεία όπου

$$p(x/D)P(D) = p(x/B)P(B)$$

οριοθετούν τις περιοχές του χώρου των χαρακτηριστικών (βάρος) που αντιστοιχούν στις δύο κατηγορίες

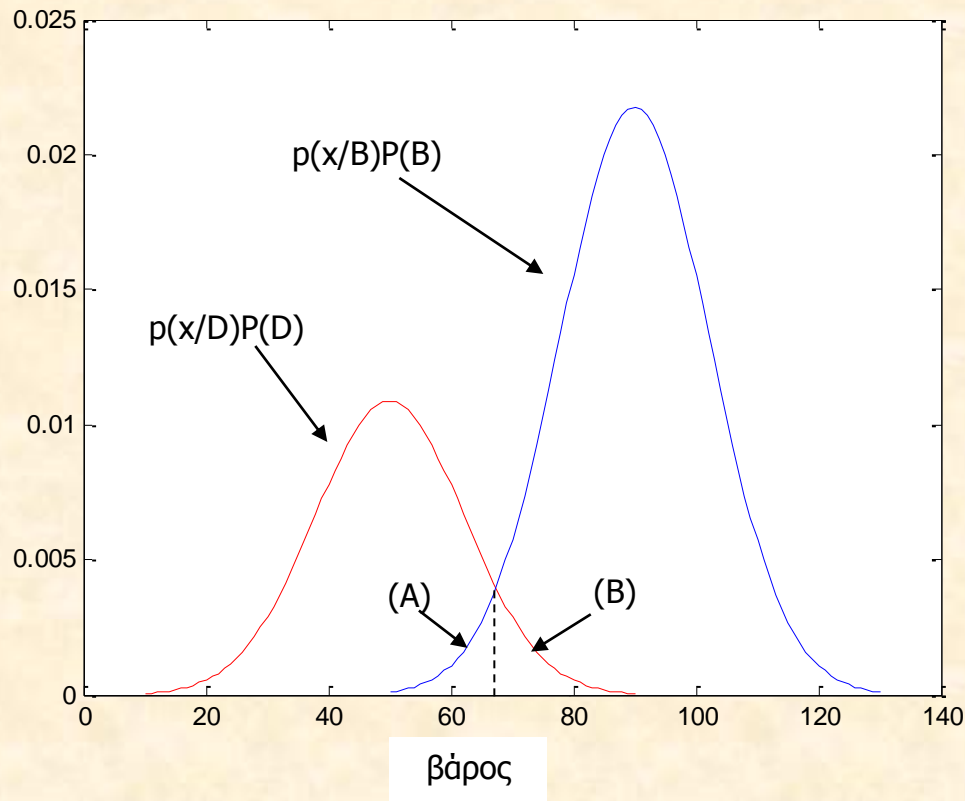
- Ταξινόμηση οντότητας με συγκεκριμένη τιμή βάρους ανάλογα με την περιοχή στην οποία αυτή ανήκει.

**Ορισμός περιοχών:** Λύνοντας την εξίσωση

$$p(x/D)P(D) = p(x/B)P(B)$$

$$\frac{1}{3} \frac{1}{\sqrt{300\pi}} e^{-\frac{(x-50)^2}{300}} = \frac{2}{3} \frac{1}{\sqrt{300\pi}} e^{-\frac{(x-90)^2}{300}} \Leftrightarrow -(x-50)^2 = -(x-90)^2 + 300 \ln 2 \Leftrightarrow x \approx 67.4 \equiv x_0$$





**Χορευτής:**  $R_D = \{x: x < 64.7\}$

**Καλαθοσφ.**:  $R_B = \{x: x > 64.7\}$

**Ερώτηση:** Ποιά είναι η πιθανότητα λάθους;

**Απάντηση:** 
$$P_\lambda = E\mu\beta(A) + E\mu\beta(B) = \int_{x \in R_B} P(D) p(x/D) dx + \int_{x \in R_D} P(B) p(x/B) dx.$$

**Για το παράδειγμά μας:** 
$$P_\lambda = \frac{1}{3} 0.0778 + \frac{2}{3} 0.0322 = 0.0474.$$

## Σχεδιασμός ταξινομητή

### Συμβολισμός:

- $\omega_1, \omega_2, \dots, \omega_c$  οι κατηγορίες στις οποίες θα ταξινομηθεί μια οντότητα
- $l$  το πλήθος των χαρακτηριστικών που αναπαριστούν μια οντότητα (διάσταση του χώρου των χαρακτηριστικών)
- $\mathbf{x} = [x_1, x_2, \dots, x_l]^T$  Διάνυσμα αναπαράστασης οντότητας ( $x_i$  η τιμή του  $i$  χαρακτηριστικού γι' αυτήν την οντότητα)
- $P(\omega_i)$  η *a priori* πιθανότητα η υπό εξέταση οντότητα να ανήκει στην  $i$  κατηγορία
- $P(\omega_i/\mathbf{x})$  η *a posteriori* πιθανότητα η υπό εξέταση οντότητα να ανήκει στην  $i$  κατηγορία δοθέντος του διανύσματος μετρήσεων  $\mathbf{x}$
- $p(\mathbf{x}/\omega_i)$  η συνάρτηση πυκνότητας πιθανότητας που περιγράφει την κατανομή των χαρακτηριστικών διανυσμάτων  $\mathbf{x}$  για την κατηγορία  $\omega_i$ .

**Ερώτηση 7:** «Δοθέντος ενός διανύσματος μετρήσεων  $\mathbf{x}$  μιας υπό εξέταση οντότητας να προσδιοριστεί η κατηγορία  $\omega_i$  στην οποία ανήκει η οντότητα»

**Bayes 1:** «καταχώρησε την υπό εξέταση οντότητα στην κατηγορία  $\omega_i$  για την οποία

$$P(\omega_i/\mathbf{x}) = \max_{j=1, \dots, c} P(\omega_j/\mathbf{x}).»$$

Δεδομένου ότι:

$$P(\omega_i / \mathbf{x}) = \frac{p(\mathbf{x} / \omega_i)P(\omega_i)}{\sum_{j=1}^c p(\mathbf{x} / \omega_j)P(\omega_j)}$$

**Bayes 2:** «καταχώρησε την υπό εξέταση οντότητα στην κατηγορία  $\omega_i$  για την οποία

$$p(\mathbf{x}/\omega_i)P(\omega_i) = \max_{j=1, \dots, c} p(\mathbf{x}/\omega_j)P(\omega_j).»$$

**Bayes (ειδ):** «Αν οι κατηγορίες είναι ισοπίθανες, καταχώρησε την υπό εξέταση οντότητα στην κατηγορία  $\omega_i$  για την οποία

$$p(\mathbf{x}/\omega_i) = \max_{j=1, \dots, c} p(\mathbf{x}/\omega_j).»$$

**Παράδειγμα:** Σ' ένα πρόβλημα τριών κλάσεων, χρησιμοποιείται μόνο ένα χαρακτηριστικό για την ταξινόμηση των δειγμάτων. Οι αντίστοιχες πυκνότητες πιθανότητας και οι a priori πιθανότητες για τις κλάσεις είναι:

$$\omega_1: P(\omega_1)=1/2, p(x|\omega_1)=(1/\sqrt{2\pi})*\exp(-x^2/2)$$

$$\omega_2: P(\omega_2)=1/3, p(x|\omega_2)=(1/\sqrt{2\pi})*\exp(-(x-1)^2/2)$$

$$\omega_3: P(\omega_3)=1/6, p(x|\omega_3)=(1/\sqrt{2\pi})*\exp(-(x-2)^2/2)$$

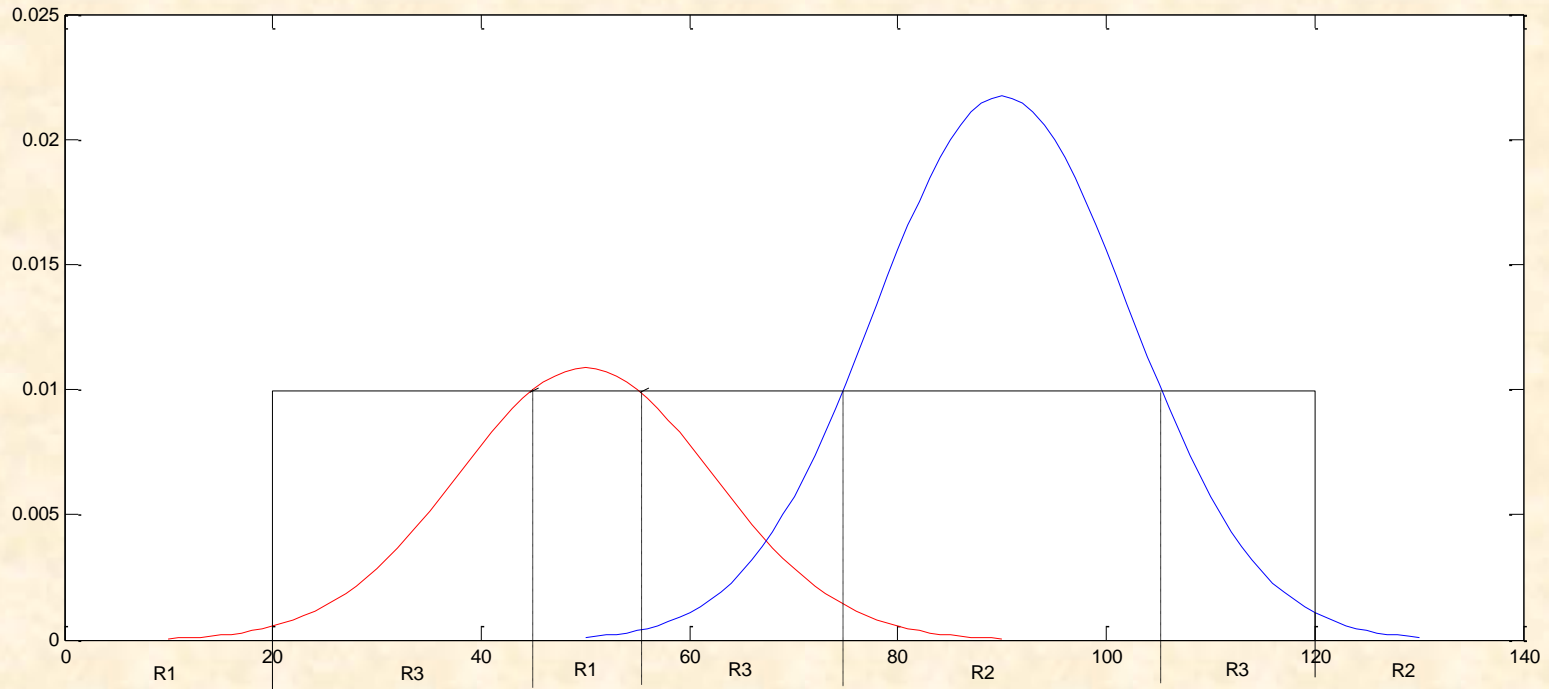
Ταξινομήστε σε μία από τις παραπάνω κατηγορίες άγνωστο δείγμα με τιμή χαρακτηριστικού  **$x=1.6$** .

Είναι:

$$P(\omega_1)p(x|\omega_1)= \mathbf{0.0555}, P(\omega_2)p(x|\omega_2)= \mathbf{0.1111}, P(\omega_3)p(x|\omega_3)= \mathbf{0.0614}$$

Συνεπώς το δείγμα καταχωρείται στην κατηγορία  **$\omega_2$** .

### Παράδειγμα 3 τριών κατανομών



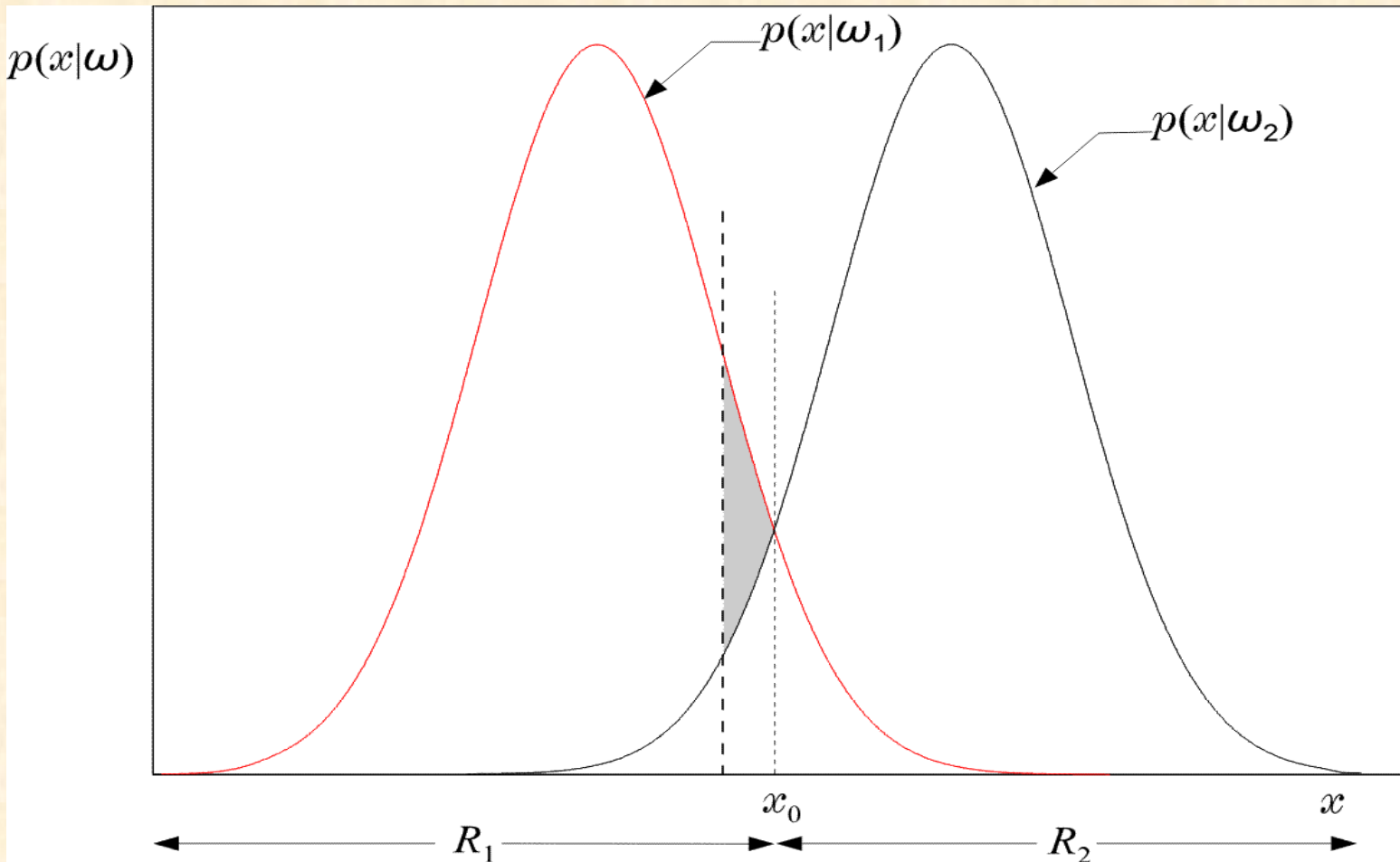
Πιθανότητα λάθους:

$$P_\lambda = \sum_{i=1}^c \int_{R_i} \left( \sum_{k=1, k \neq i}^c p(x / \omega_k) P(\omega_k) \right) dx$$

**Ερώτηση:** Γιατί τόση επιμονή στον ταξινομητή Bayes;

**Απάντηση:**

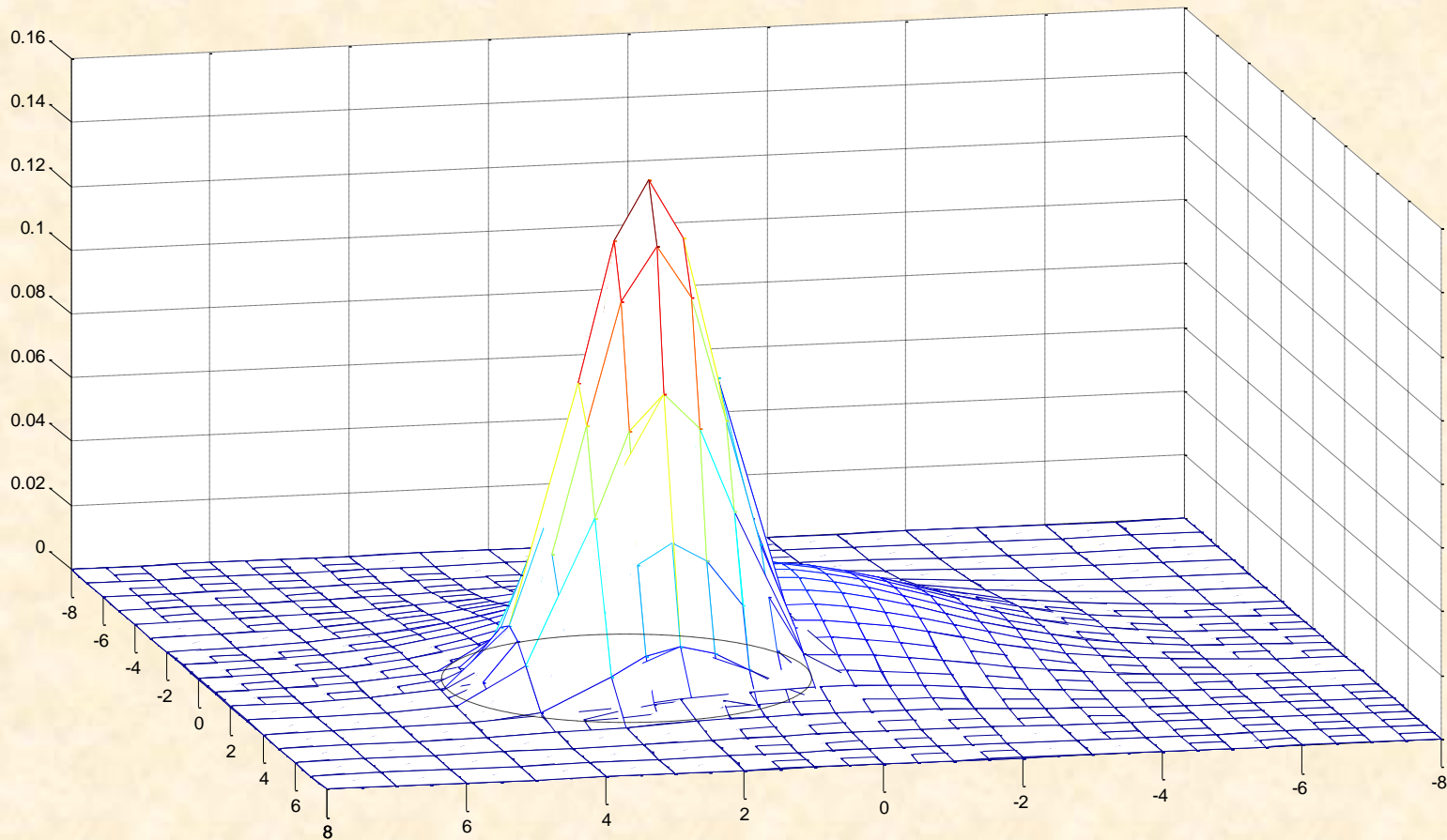
- γιατί η λογική του δεν έρχεται σε αντίθεση με την διαίσθησή μας.
- Ο κανόνας ταξινόμησης κατά Bayes είναι βέλτιστος με την έννοια ότι ελαχιστοποιεί την πιθανότητα λάθους.



**Παρατήρηση:** οι πυκνότητες πιθανότητας των κατηγοριών μαζί με τις a priori πιθανότητές τους ορίζουν **μονοσήμαντα** μια διαμέριση του χώρου του  $l$ -διάστατου χώρου των χαρακτηριστικών, σε (όχι απαραίτητα ενιαίες) περιοχές ( $R_i$ ) όπου κάθε μια αντιστοιχεί και σε μια κατηγορία ( $\omega_i$ ).

Έτσι, η ταξινόμηση ενός διάνυσματος μπορεί να γίνει

- (α) είτε με χρήση του κανόνα του Bayes όπως αυτός διατυπώνεται στον *Bayes 1* ή *Bayes 2*
- (β) είτε με προσδιορισμό των περιοχών των χώρου των χαρακτηριστικών που αντιστοιχούν στις υπό εξέταση κατηγορίες και για κάθε νέο διάνυσμα χαρακτηριστικών να εξετάζει απλώς σε ποια περιοχή ανήκει.



*Παράδειγμα δύο κατηγοριών στον 2-διάστατο χώρο.*



# ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ (DISCRIMINANT FUNCTIONS ) ΕΠΙΦΑΝΕΙΕΣ ΑΠΟΦΑΣΗΣ (DECISION SURFACES)

❖ Αν  $R_i, R_j$  είναι παρακείμενες έχουν κοινό σύνορο που ορίζεται από την  $g(\underline{x}) \equiv P(\omega_i|\underline{x}) - P(\omega_j|\underline{x}) = 0$

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

$$\frac{+}{-} \quad \underline{g(\underline{x}) = 0}$$

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

Η επιφάνεια αυτή ( $g(x)$ ) διαχωρίζει τις περιοχές. Από τη μία πλευρά είναι θετική (+) και από την άλλη αρνητική (-). Είναι γνωστή ως **επιφάνεια απόφασης (Decision Surface)**

❖ Αν  $f(\cdot)$  αύξουσα, ο κανόνας μένει αναλλοίωτος αν γράψουμε:

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

❖  $g_i(\underline{x}) \equiv f(P(\omega_i|\underline{x}))$  is a **συνάρτηση διάκρισης (discriminant function)**

❖ Γενικά, οι συναρτήσεις διάκρισης μπορούν να οριστούν **ανεξάρτητα** από τον κανόνα του Bayes. Σ' αυτή την περίπτωση οδηγούν σε **υποβέλτιστες (suboptimal)** λύσεις. Παρόλα αυτά, αν επιλεγούν κατάλληλα μπορούν να οδηγήσουν σε υπολογιστικές απλουστεύσεις.

# ΤΑΞΙΝΟΜΗΤΗΣ BAYES ΓΙΑ ΚΑΝΟΝΙΚΕΣ ΚΑΤΑΝΟΜΕΣ

❖ Πολυδιάστατη Gaussian pdf

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

$\underline{\mu}_i = E[\underline{x}]$   $\ell \times 1$  μέσο διάνυσμα για την  $\omega_i$

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

Πίνακας συνδιασποράς (covariance matrix)

❖ Η  $\ln(\cdot)$  είναι μονότονη. Ορίζουμε:

$$\begin{aligned} \text{➤ } g_i(\underline{x}) &= \ln(p(\underline{x}|\omega_i)P(\omega_i)) = \\ & \ln p(\underline{x}|\omega_i) + \ln P(\omega_i) \end{aligned}$$

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

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

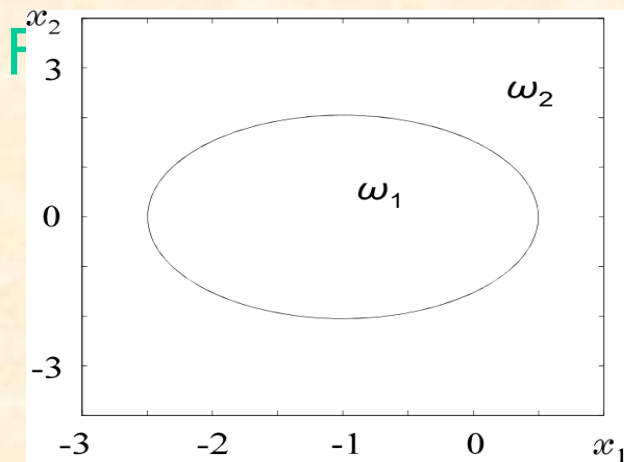
$$\text{➤ Παράδειγμα: } \Sigma_i = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}$$

$$\begin{aligned} \triangleright \quad g_i(\underline{x}) &= -\frac{1}{2\sigma^2} (x_1^2 + x_2^2) + \frac{1}{\sigma^2} (\mu_{i1}x_1 + \mu_{i2}x_2) \\ &\quad - \frac{1}{2\sigma^2} (\mu_{i1}^2 + \mu_{i2}^2) + \ln P(\omega_i) + C_i \end{aligned}$$

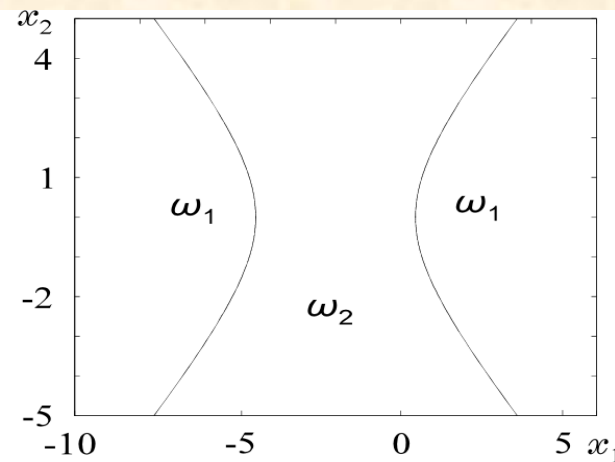
Δηλαδή, η  $g_i(\underline{x})$  είναι **τετραγωνική**. Σημειώστε ότι οι επιφάνειες

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

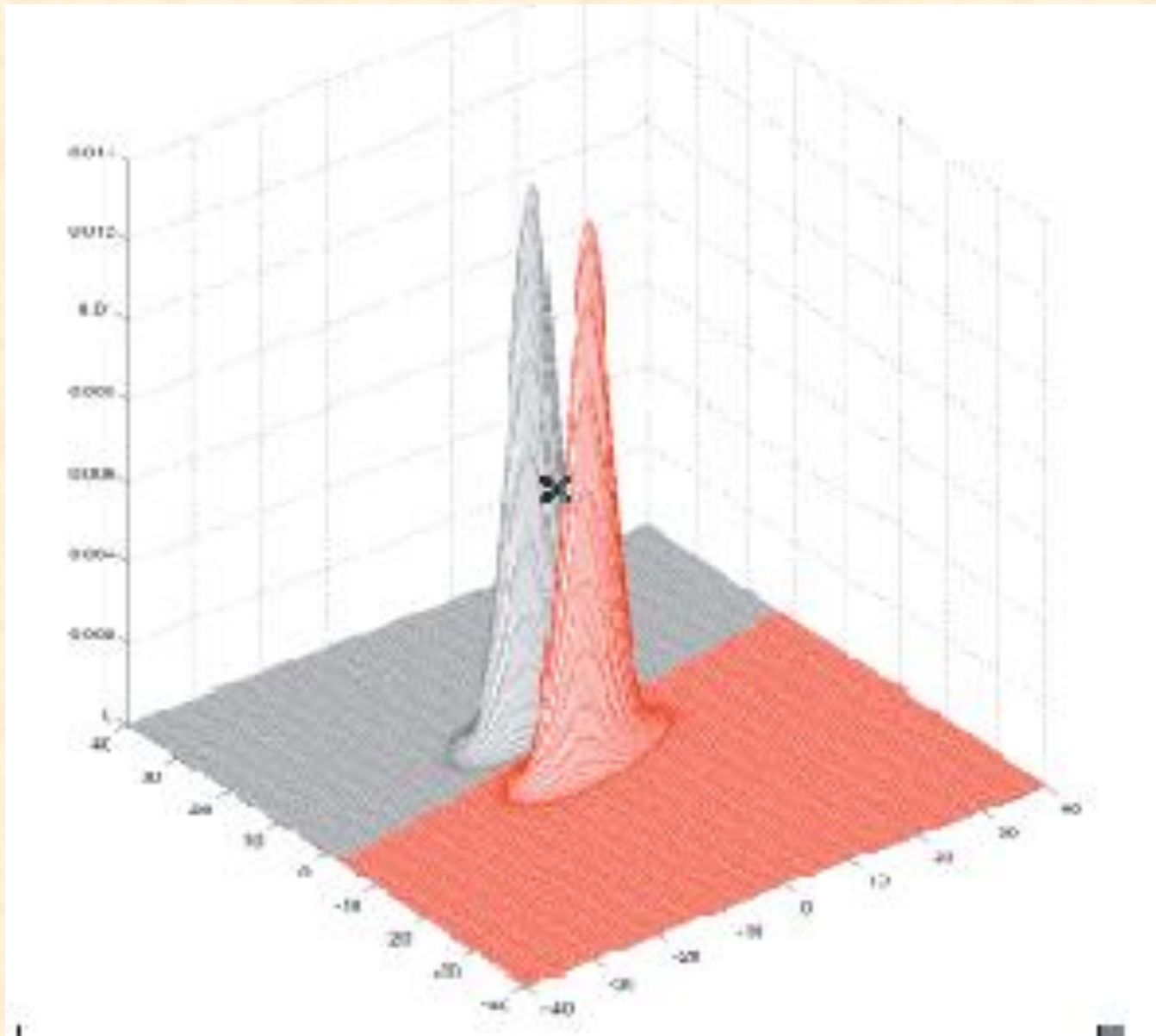
**ΔΕΝ** είναι, **απαραίτητα**, τετραγωνικής μορφής (ελλείψεις, παραβολές, υπερβολές, ζεύγη ευθειών).

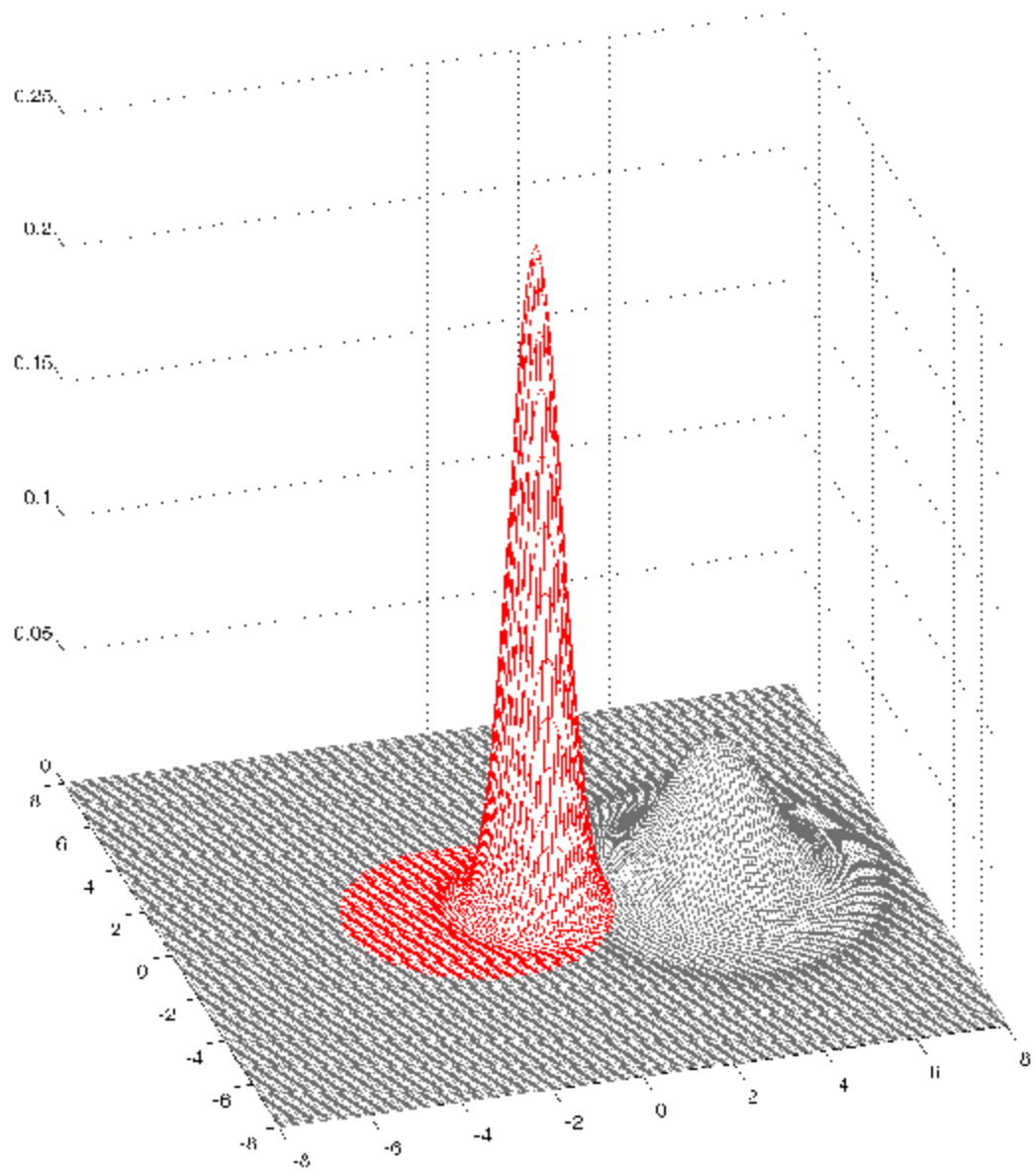


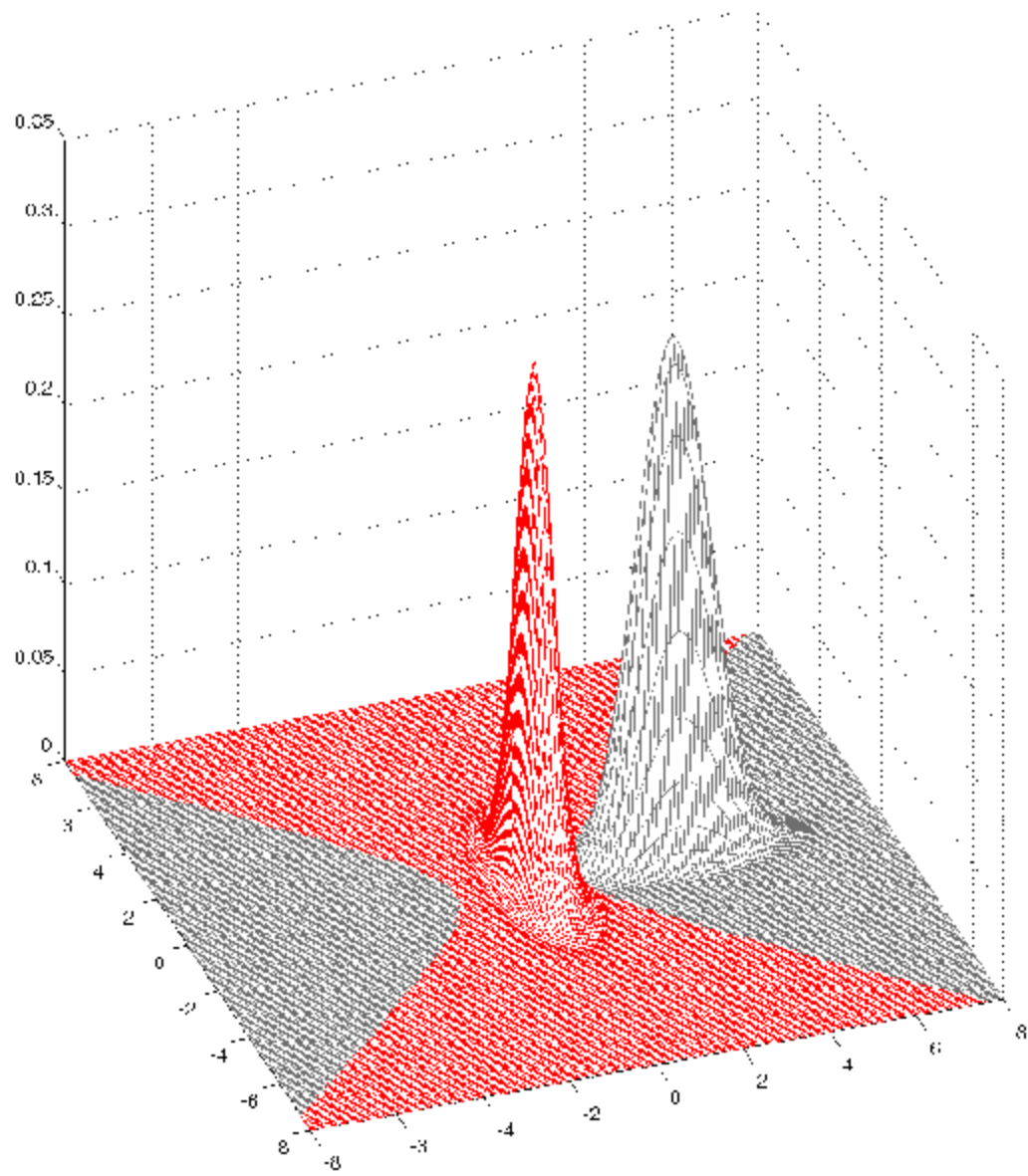
(a)



(b)









## ❖ Υπερεπίπεδα απόφασης

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

Αν **ΟΛΟΙ**  $\Sigma_i = \Sigma$  (είναι ίδιοι) οι τετραγωνικοί όροι δεν επηρεάζουν, αφού δεν εμπλέκονται στις συγκρίσεις. Τότε μπορούμε να γράψουμε ισοδύναμα:

$$g_i(\underline{x}) = \underline{w}_i^T \underline{x} + w_{i0}$$

$$\underline{w}_i = \Sigma^{-1} \underline{\mu}_i$$

$$w_{i0} = \ln P(\omega_i) - \frac{1}{2} \underline{\mu}_i^T \Sigma^{-1} \underline{\mu}_i$$

Οι συναρτήσεις διάκρισης είναι **ΓΡΑΜΜΙΚΕΣ**

➤ Έστω επιπλέον ότι:

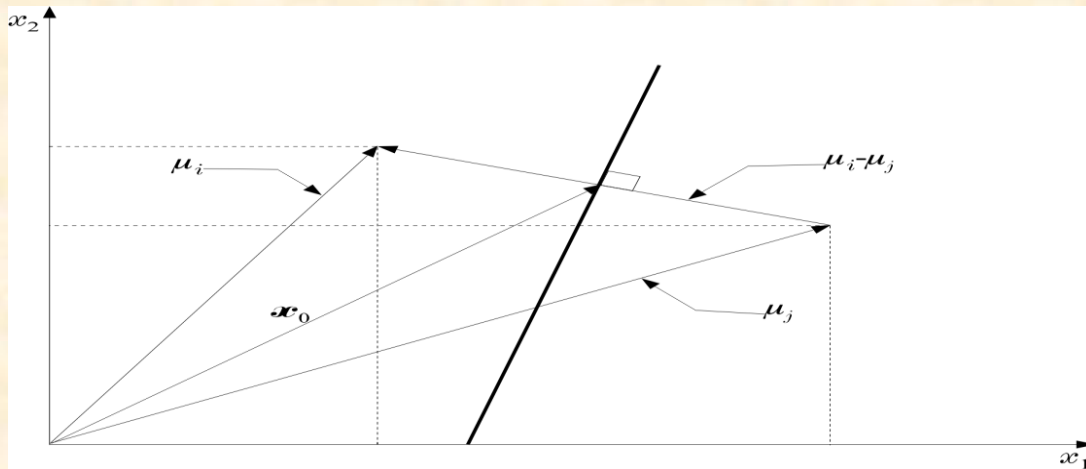
- $\Sigma = \sigma^2 I$ . Τότε

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

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

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

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




➤ Μη διαγώνιος:  $\Sigma \neq \sigma^2 I$

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

- $\underline{w} = \Sigma^{-1}(\underline{\mu}_i - \underline{\mu}_j)$

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

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

➤ Υπερεπίπεδο απόφασης 

όχι κάθετο στο  $\underline{\mu}_i - \underline{\mu}_j$   
κάθετο στο  $\Sigma^{-1}(\underline{\mu}_i - \underline{\mu}_j)$

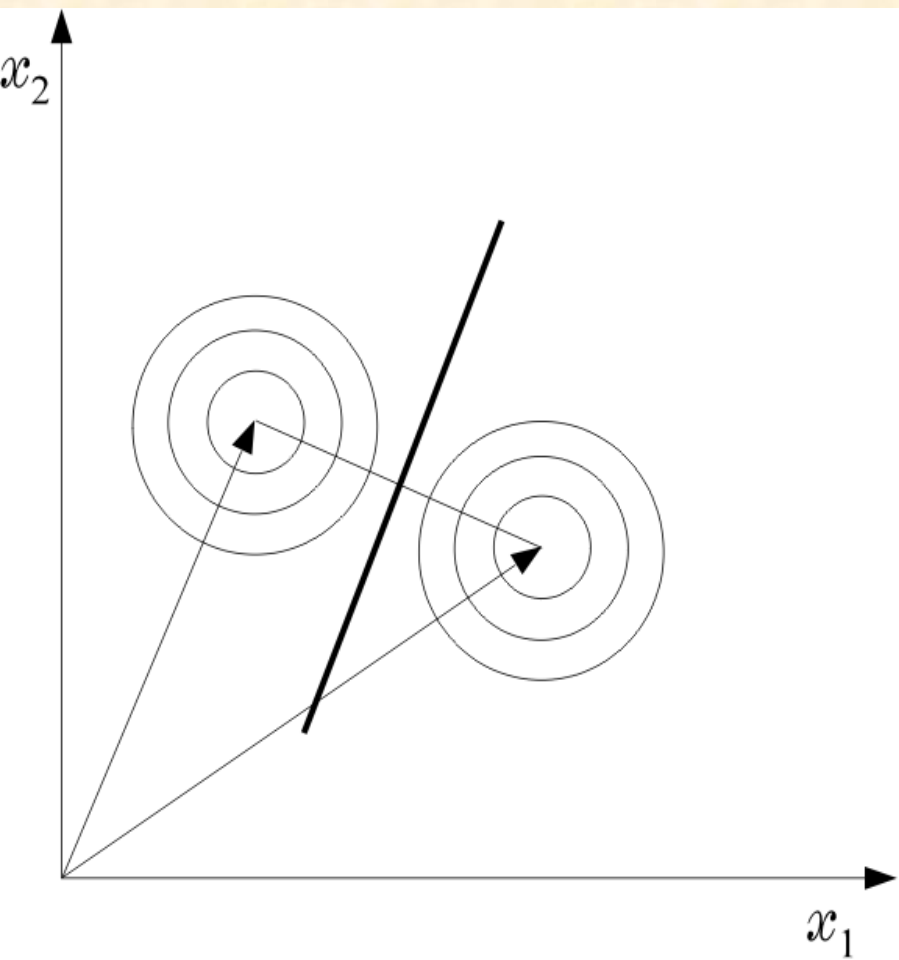
## ❖ Ταξινομητές ελάχιστης απόστασης

- $P(\omega_i) = \frac{1}{M}$  ισοπίθανες
- $g_i(\underline{x}) = -\frac{1}{2}(\underline{x} - \underline{\mu}_i)^T \Sigma^{-1}(\underline{x} - \underline{\mu}_i)$
- $\Sigma = \sigma^2 I$ : Καταχώρησε  $\underline{x} \rightarrow \omega_i$ :

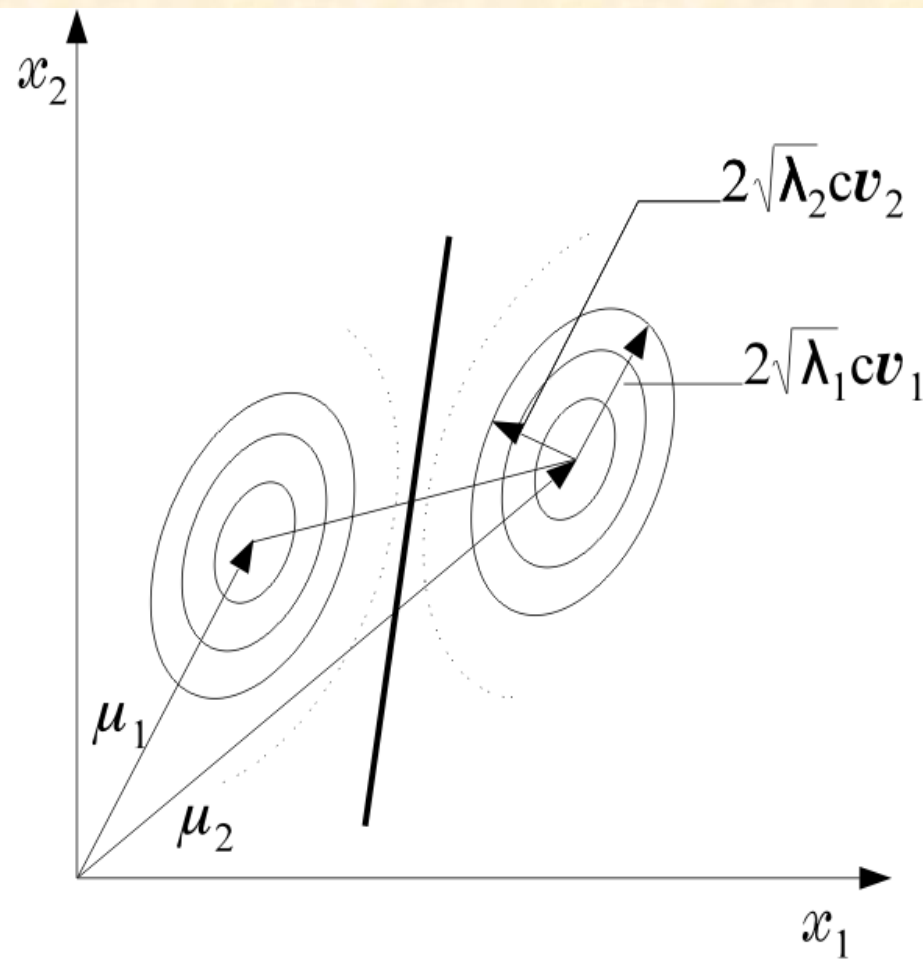
**Ευκλείδεια απόσταση:**  $d_E \equiv \left\| \underline{x} - \underline{\mu}_i \right\|$   
μικρότερη

- $\Sigma \neq \sigma^2 I$ : Καταχώρησε  $\underline{x} \rightarrow \omega_i$ :

**Mahalanobis απόσταση:**  $d_m = \left( (\underline{x} - \underline{\mu}_i)^T \Sigma^{-1} (\underline{x} - \underline{\mu}_i) \right)^{\frac{1}{2}}$   
μικρότερη



(a)



(b)

## ❖ Παράδειγμα:

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

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

κατηγοριοποίησε το διάνυσμα  $\underline{x} = \begin{bmatrix} 1.0 \\ 2.2 \end{bmatrix}$  με χρήση Bayesian ταξινομητή:

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

- Υπόλ. Mahalanobis απόστ.  $d_m$  από  $\mu_1, \mu_2$  :  $d_{m,1}^2 = [1.0, 2.2]$

$$\Sigma^{-1} \begin{bmatrix} 1.0 \\ 2.2 \end{bmatrix} = 2.952, \quad d_{m,2}^2 = [-2.0, -0.8] \Sigma^{-1} \begin{bmatrix} -2.0 \\ -0.8 \end{bmatrix} = 3.672$$

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

# ΕΚΤΙΜΗΣΗ ΑΓΝΩΣΤΩΝ ΣΥΝΑΡΤΗΣΕΩΝ ΠΥΚΝΟΤΗΤΑΣ ΠΙΘΑΝΟΤΗΤΑΣ

## ❖ Μέγιστη πιθανοφάνεια (Maximum Likelihood)

(\*) Έστω  $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_N$  γνωστά και ανεξάρτητα

(\*) Έστω  $p(\underline{x})$  γνωστής μορφής με άγνωστο διάνυσμα παραμέτρων  $\underline{\theta}$ :  $p(\underline{x}) \equiv p(\underline{x}; \underline{\theta})$

(\*)  $X = \{ \underline{x}_1, \underline{x}_2, \dots, \underline{x}_N \}$

(\*)  $p(X; \underline{\theta}) \equiv p(\underline{x}_1, \underline{x}_2, \dots, \underline{x}_N; \underline{\theta})$

$$= \prod_{k=1}^N p(\underline{x}_k; \underline{\theta})$$

που είναι γνωστή ως η πιθανοφάνεια (Likelihood)  
του  $\underline{\theta}$  ως προς το  $X$

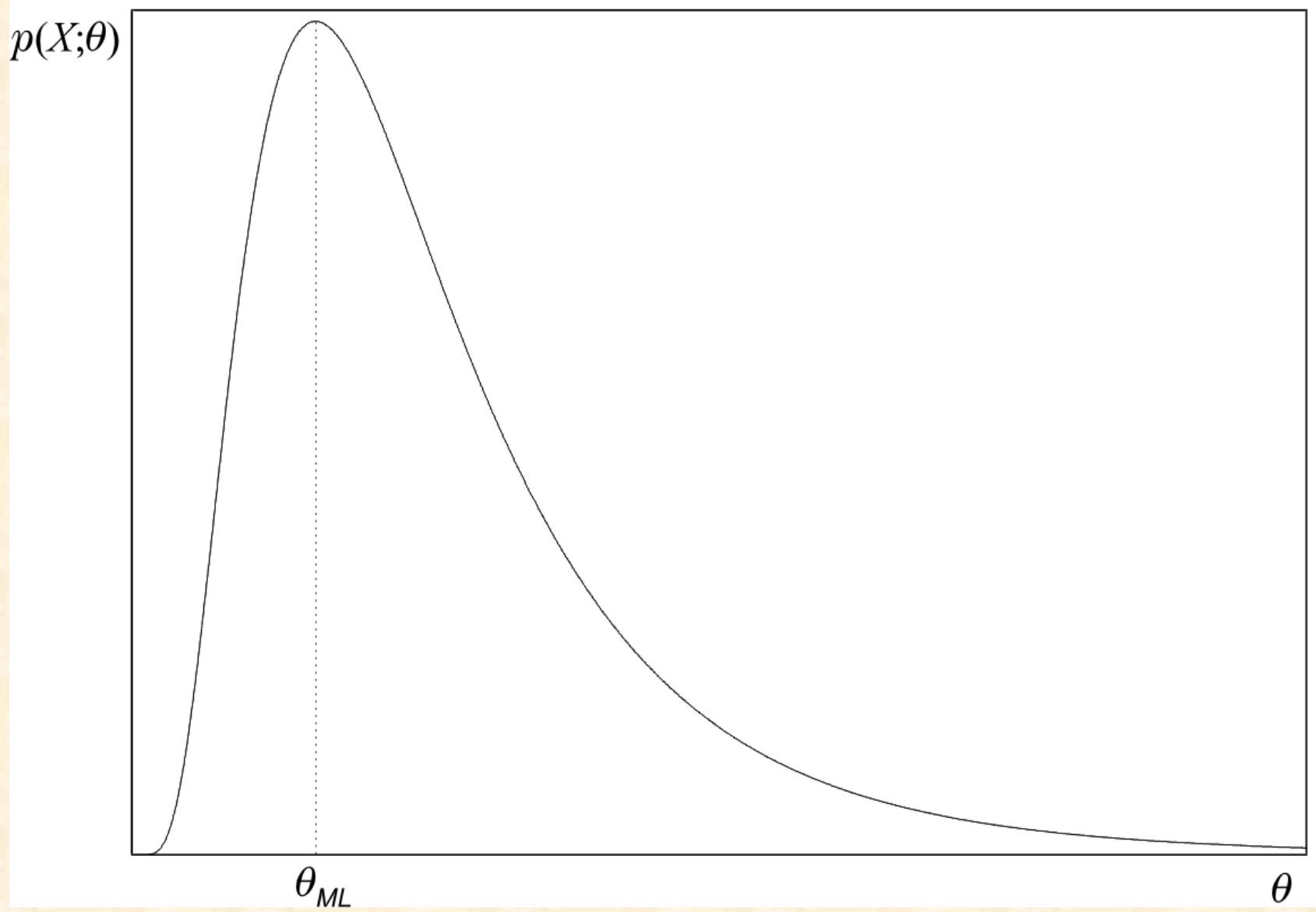
Η μέθοδος:

$$\blacktriangleright \hat{\underline{\theta}}_{\text{ML}} : \arg \max_{\underline{\theta}} \prod_{k=1}^N p(\underline{x}_k; \underline{\theta})$$

$$\blacktriangleright L(\underline{\theta}) \equiv \ln p(X; \underline{\theta}) = \sum_{k=1}^N \ln p(\underline{x}_k; \underline{\theta})$$

$$\blacktriangleright \hat{\underline{\theta}}_{\text{ML}} : \frac{\partial L(\underline{\theta})}{\partial(\underline{\theta})} = \sum_{k=1}^N \frac{1}{p(\underline{x}_k; \underline{\theta})} \frac{\partial p(\underline{x}_k; \underline{\theta})}{\partial(\underline{\theta})} = \underline{0}$$





Αν, πράγματι, υπάρχει ένα  $\underline{\theta}_0$  έτσι ώστε

$$p(\underline{x}) = p(\underline{x}; \underline{\theta}_0), \text{ τότε}$$

$$\lim_{N \rightarrow \infty} E[\underline{\theta}_{ML}] = \underline{\theta}_0$$

$$\lim_{N \rightarrow \infty} E \left\| \hat{\underline{\theta}}_{ML} - \underline{\theta}_0 \right\|^2 = 0$$

- Ασυμπτωτικά αμερόληπτος (unbiased) εκτιμητής:  
Η μέση τιμή της τυχαίας μεταβλητής  $\hat{\underline{\theta}}_{ML}$  είναι η πραγματική τιμή της άγνωστης παραμέτρου
- Ασυμπτωτικά συνεπής (consistent) εκτιμητής:  
Η διασπορά της ML εκτίμησης τείνει στο μηδέν.

## ❖ Παράδειγμα:

$$p(\underline{x}): N(\underline{\mu}, \underline{\Sigma}): \underline{\mu} \text{ άγνωστο, } \underline{x}_1, \underline{x}_2, \dots, \underline{x}_N \quad p(\underline{x}_k) \equiv p(\underline{x}_k; \underline{\mu})$$

$$L(\underline{\mu}) = \ln \prod_{k=1}^N p(\underline{x}_k; \underline{\mu}) = C - \frac{1}{2} \sum_{k=1}^N (\underline{x}_k - \underline{\mu})^T \underline{\Sigma}^{-1} (\underline{x}_k - \underline{\mu})$$

$$p(\underline{x}_k; \underline{\mu}) = \frac{1}{(2\pi)^{\frac{l}{2}} |\underline{\Sigma}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} (\underline{x}_k - \underline{\mu})^T \underline{\Sigma}^{-1} (\underline{x}_k - \underline{\mu})\right)$$

$$\frac{\partial L(\underline{\mu})}{\partial(\underline{\mu})} \equiv \begin{bmatrix} \frac{\partial L}{\partial \mu_1} \\ \cdot \\ \cdot \\ \cdot \\ \frac{\partial L}{\partial \mu_l} \end{bmatrix} = \sum_{k=1}^N \underline{\Sigma}^{-1} (\underline{x}_k - \underline{\mu}) = \underline{0} \Rightarrow \underline{\mu}_{ML} = \frac{1}{N} \sum_{k=1}^N \underline{x}_k$$

$$\text{Υπενθύμιση: Αν } A = A^T \Rightarrow \frac{\partial(\underline{\alpha}^T A \underline{\alpha})}{\partial \underline{\alpha}} = 2A \underline{\alpha}$$

❖ Εκτίμηση μέγιστης εκ των υστέρων πιθανότητας  
(Maximum A posteriori Probability Estimation)

- Στην ML μέθοδο, το  $\underline{\theta}$  λογίζοταν ως παράμετρος
- Εδώ θα θεωρήσουμε το  $\underline{\theta}$  ως τυχαίο διάνυσμα που περιγράφεται από (υποτίθεται γνωστή) pdf  $p(\underline{\theta})$ .

- Δοθέντος

$$X = \{ \underline{x}_1, \underline{x}_2, \dots, \underline{x}_N \}$$

Υπολόγισε το μέγιστο της

$$p(\underline{\theta} | X)$$

- From Bayes theorem

$$p(\underline{\theta}) p(X | \underline{\theta}) = p(X) p(\underline{\theta} | X) \text{ or}$$

$$p(\underline{\theta} | X) = \frac{p(\underline{\theta}) p(X | \underline{\theta})}{p(X)}$$

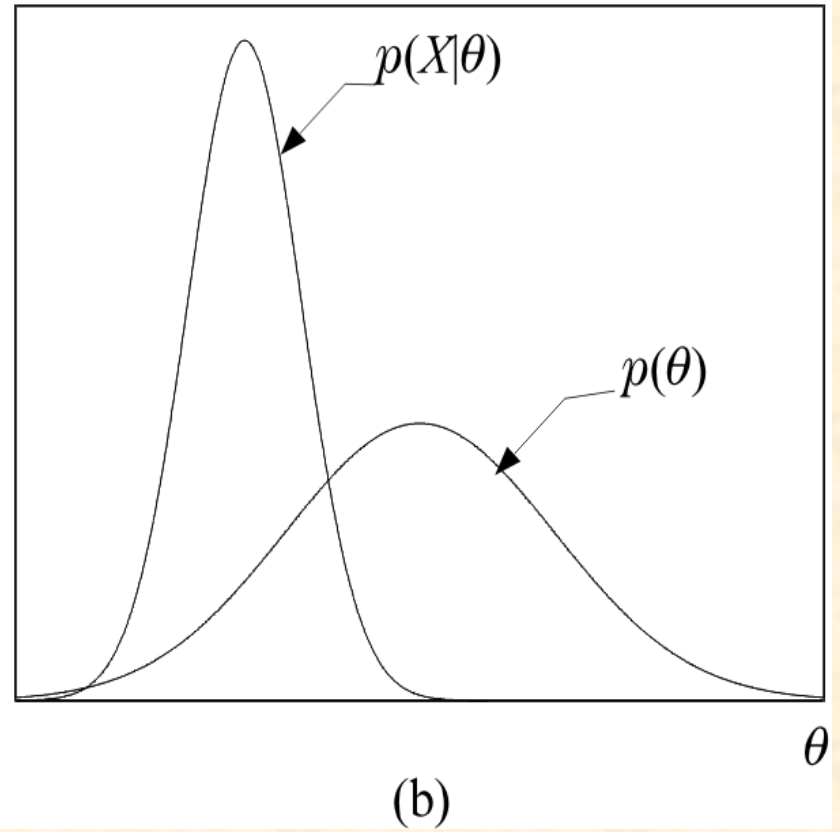
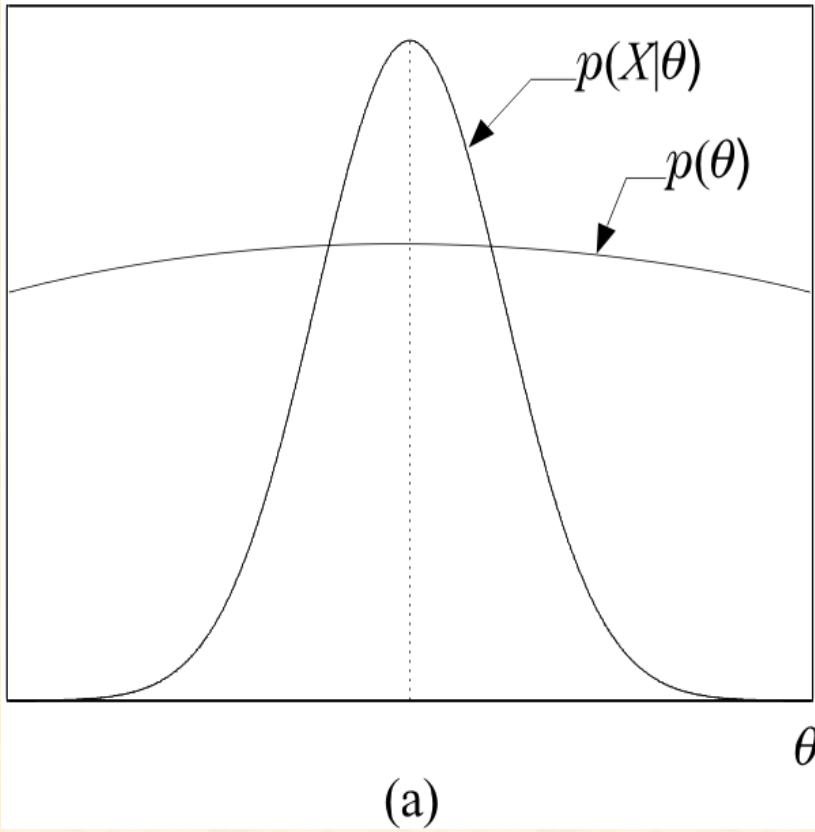
➤ Η μέθοδος:

$$\hat{\underline{\theta}}_{MAP} = \arg \max_{\underline{\theta}} p(\underline{\theta} | X) \text{ ή}$$

$$\hat{\underline{\theta}}_{MAP} : \frac{\partial}{\partial \underline{\theta}} (p(\underline{\theta}) p(X | \underline{\theta}))$$

Αν η  $p(\underline{\theta})$  είναι ομοιόμορφη ή αρκετά ευρεία:

$$\hat{\underline{\theta}}_{MAP} \cong \underline{\theta}_{ML}$$



## ❖ Παράδειγμα:

$$p(\underline{x}) : N(\underline{\mu}, \Sigma), \quad \underline{\mu} \text{ άγνωστο}, \quad X = \{\underline{x}_1, \dots, \underline{x}_N\}$$

$$p(\underline{\mu}) = \frac{1}{(2\pi)^{\frac{l}{2}} \sigma_{\mu}^l} \exp\left(-\frac{\|\underline{\mu} - \underline{\mu}_0\|^2}{2\sigma_{\mu}^2}\right)$$

$$\hat{\underline{\theta}}_{MAP} : \frac{\partial}{\partial \underline{\mu}} \ln(\prod_{k=1}^N p(\underline{x}_k | \underline{\mu}) p(\underline{\mu})) = \underline{0} \quad \text{ή} \quad \sum_{k=1}^N \frac{1}{\sigma^2} (\underline{x}_k - \hat{\underline{\mu}}) - \frac{1}{\sigma_{\mu}^2} (\hat{\underline{\mu}} - \underline{\mu}_0) = \underline{0} \Rightarrow$$

$$\hat{\underline{\mu}}_{MAP} = \frac{\underline{\mu}_0 + \frac{\sigma_{\mu}^2}{\sigma^2} \sum_{k=1}^N \underline{x}_k}{1 + \frac{\sigma_{\mu}^2}{\sigma^2} N} \quad \text{For} \quad \frac{\sigma_{\mu}^2}{\sigma^2} \gg 1, \quad \text{ή για } N \rightarrow \infty$$

$$\hat{\underline{\mu}}_{MAP} \cong \hat{\underline{\mu}}_{ML} = \frac{1}{N} \sum_{k=1}^N \underline{x}_k$$