❖ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ
❖(PATTERN RECOGNITION)

Σέργιος Θεοδωρίδης

Κωνσταντίνος Κουτρούμπας

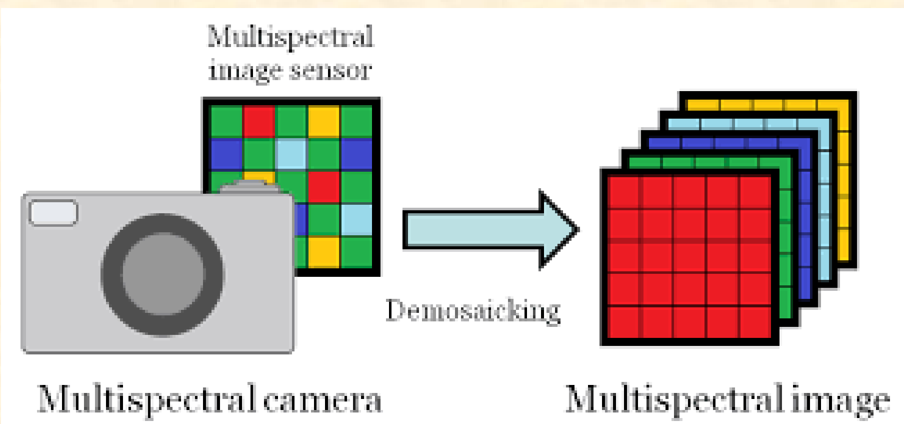
Version 2

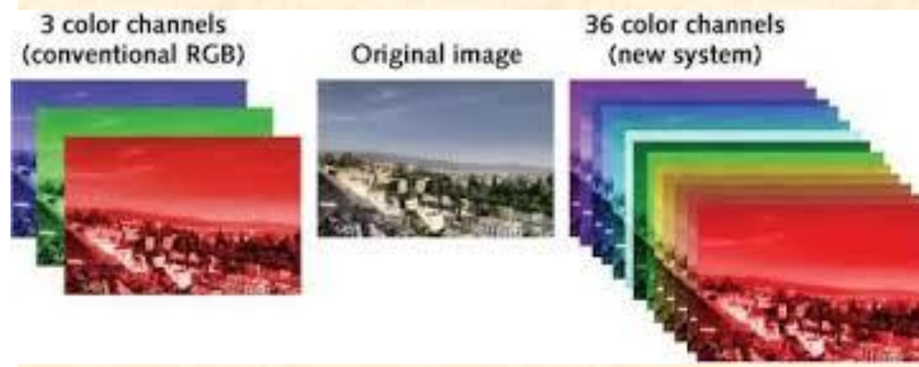
❖ΔΕΔΟΜΕΝΑ

❖Δεδομένα μπορούν να αποκτηθούν στα πλαίσια διαφόρων εφαρμογών, χρησιμοποιώντας, όπου είναι απαραίτητο, κατάλληλο εξοπλισμό.

❖Μερικά παραδείγματα είναι:

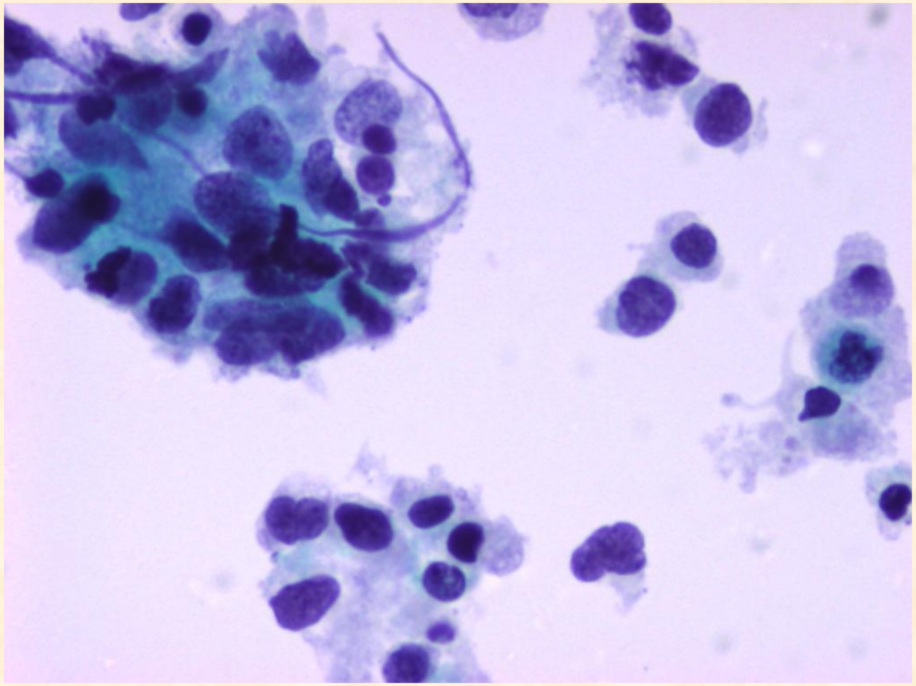
❖**(Α)** Εικόνες (εικόνες κλίμακας του γκρι (grayscale), πολυφασματικές (multispectral), υπερφασματικές (hyperspectral)) που λαμβάνονται από κατάλληλα συστήματα απεικόνισης (imagers).

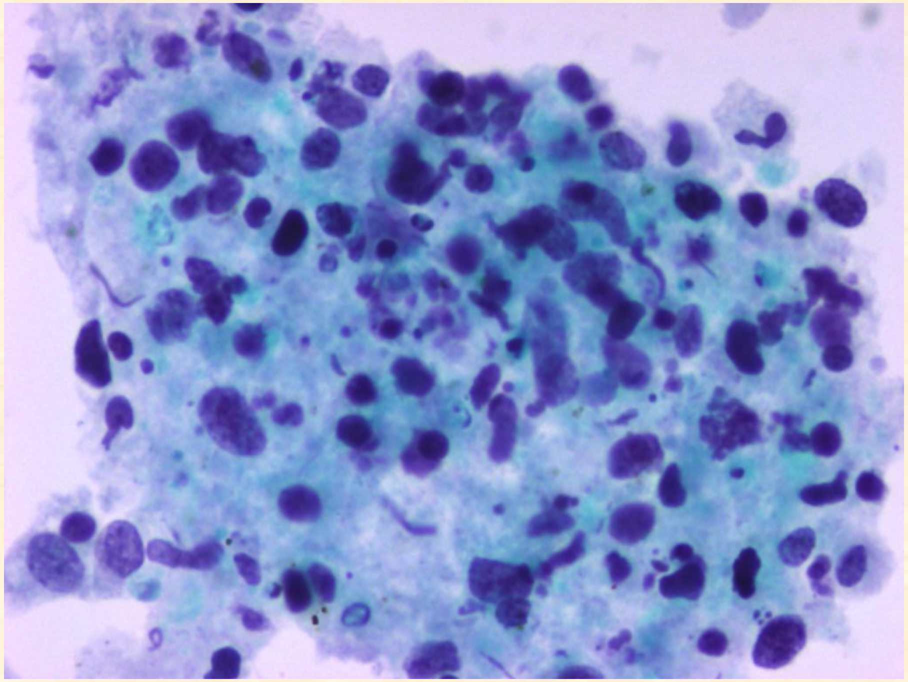




❖ΔΕΔΟΜΕΝΑ

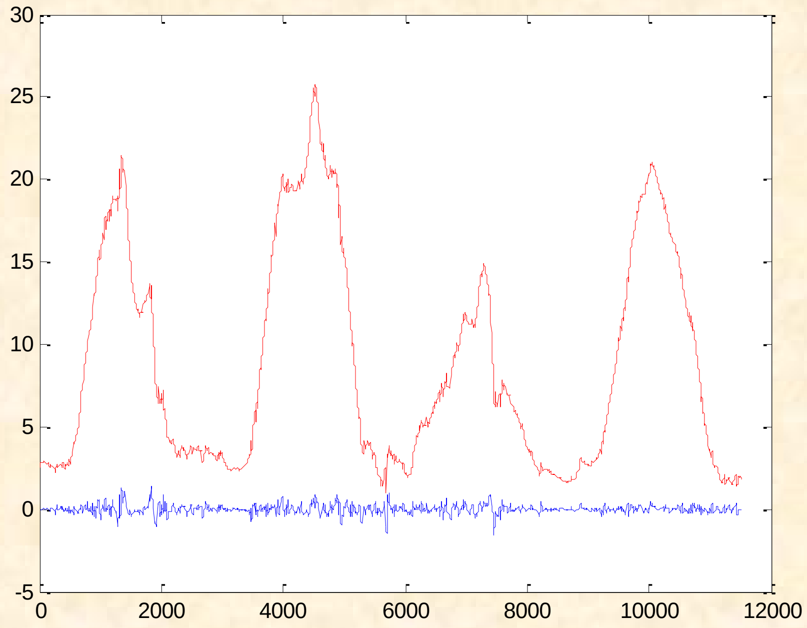
❖**(Β)** Ανθρώπινα γαστρικά κύτταρα.





❖ΔΕΔΟΜΕΝΑ

❖**(C)** Χρονοσειρές (Time series).



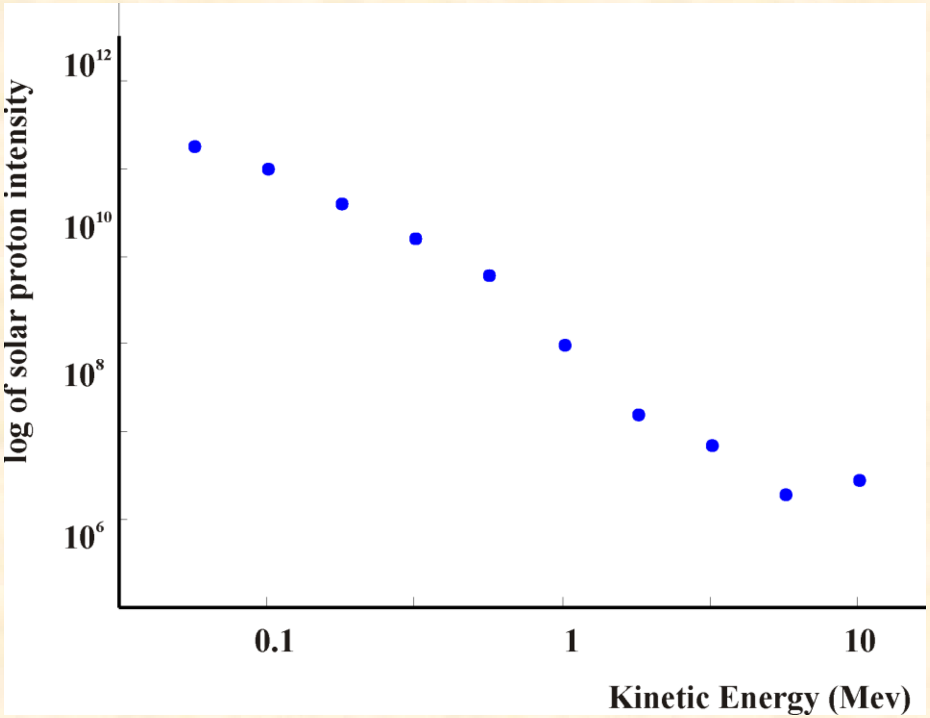
❖<u>ΔΕΔΟΜΕΝΑ</u>

❖**(D)** Κείμενο

ABCDEFGHIJKLMNO
PQRSTUVWXYZÅØüa
bcdefghijklmnop
qrstuvwxyz&1234
567890($£.,!?)

❖ΔΕΔΟΜΕΝΑ

❖**(Ε)** Ζεύγη τιμών σχετιζόμενων ποσοστήτων (κάθε σημείο αντιστοιχεί σ’ ένα ζεύγος)



❖<u>**ΕΞΑΓΩΓΗ ΓΝΩΣΗΣ (KNOWLEDGE EXTRACTION)**</u>

❖**Στόχος επεξεργασίας δεδομένω:** Εξαγωγή γνώσης (knowledge extraction) από τα δεδομένα.

❖***<u>ΣΗΜ:</u>*** Στη συνέχεια θεωρούμε μόνο αριθμητικά δεδομένα (numerical data).

❖Στο τρέχον πλαίσιο όρος ***<u>«γνώση»</u>*** (“***<u>knowledge</u>***”) είναι η απάντηση σε τουλάχιστον ένα από τα ακόλουθα ερωτήματα:

- (a) Ποιο είναι το μοντέλο που γεννά τα δεδομένα;

- (b) Πώς μπορούμε να χρησιμοποιήσουμε τα γνωστά δεδομένα προκειμένου να κάνουμε εκτιμήσεις (estimations) για άγνωστα δεδομένα;

- (c) Ποιο είναι το γενικό μοτίβο εξάπλωσης (spread pattern) των δεδομένων στο χώρο;

❖ΤΑ ΠΡΟΒΛΗΜΑΤΑ ΕΠΕΞΕΡΓΑΣΙΑΣ

❖**Παλινδρόμηση (Regression):**

❖Ένα σύνολο ανεξάρτητων μεταβλητών x σχετίζονται με ένα σύνολο εξαρτημένων μεταβλητών y μέσω μιας συνάρτησης της μορφής $y=f(x)+e$, όπου e είναι η “**αβεβαιότητα**”.

❖**Στόχος:** Δοθείσης μιας τιμής για το x **εκτίμησε** την αντίστοιχη τιμή για το y.

❖**Ταξινόμηση (Classification):**

❖ Ένας αριθμός κατηγοριών (κλάσεων) $\omega_1, \ldots \omega_M$ στις οποίες θα πρέπει να ταξινομηθούν τα στοιχεία ενός **συνόλου “οντοτήτων”**.

❖**Στόχος:** Δοθείσης μιας οντότητας **καταχώρησέ** την στην “πιο κατάλληλη” κατηγορία.

❖**Ομαδοποίηση (Clustering):**

❖Διατίθεται ένα σύνολο **οντοτήτων**.

❖**Στόχος: Ομαδοποίησε (a)** “όμοιες” οντότητες στην ίδια ομάδα και **(b)** “λιγότερο όμοιες” ποσότητες σε διαφορετικές ομάδες.

❖In the sequel we consider only **classification** and **clustering**.

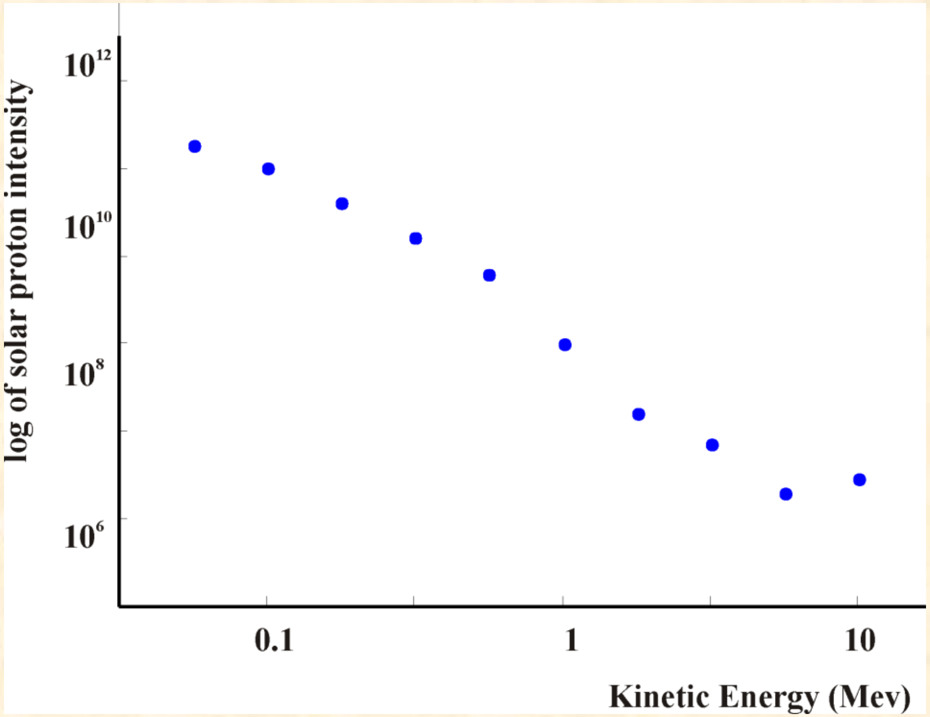
❖ΕΞΑΓΩΓΗ ΓΝΩΣΗΣ

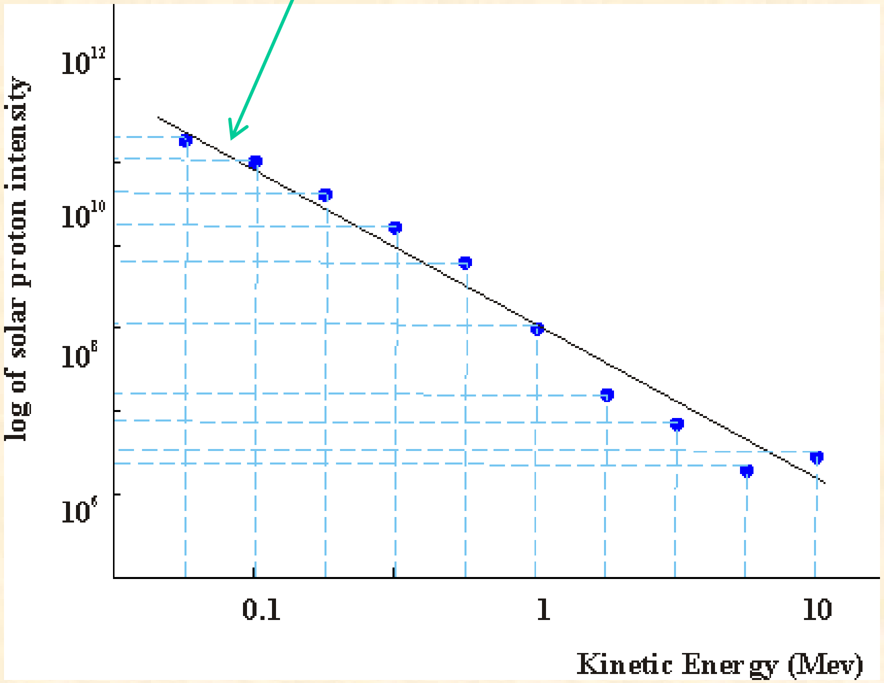
❖ (a) Ποιο είναι το μοντέλο που γεννά τα δεδομένα; (**παλινδρόμηση**)

❖Η γραμμή

❖**$y=ax+b$**

❖Αναπαριστά τη "**γνώση**" που συσχετίζει τις δύο ποσότητες

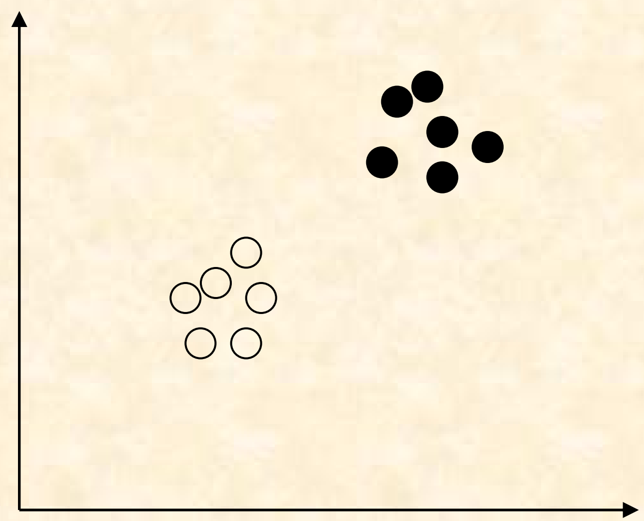




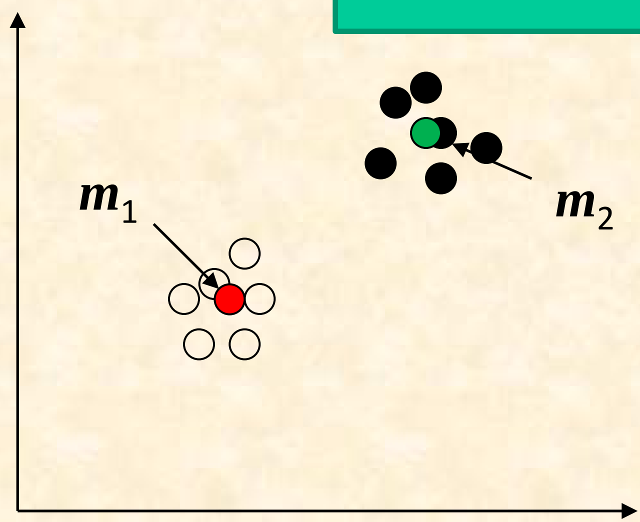
❖<u>ΕΞΑΓΩΓΗ ΓΝΩΣΗΣ</u>

❖ (a) Ποιο είναι το μοντέλο που γεννά τα δεδομένα; (**ταξινόμηση**)

❖Τα σημεία $\boldsymbol{m_1}$, $\boldsymbol{m_2}$ Αναπαριστούν τη "**γνώση**" (σημεία γύρω από τα οποία συγκεντρώνονται τα σημεία των δύο κλάσεων)



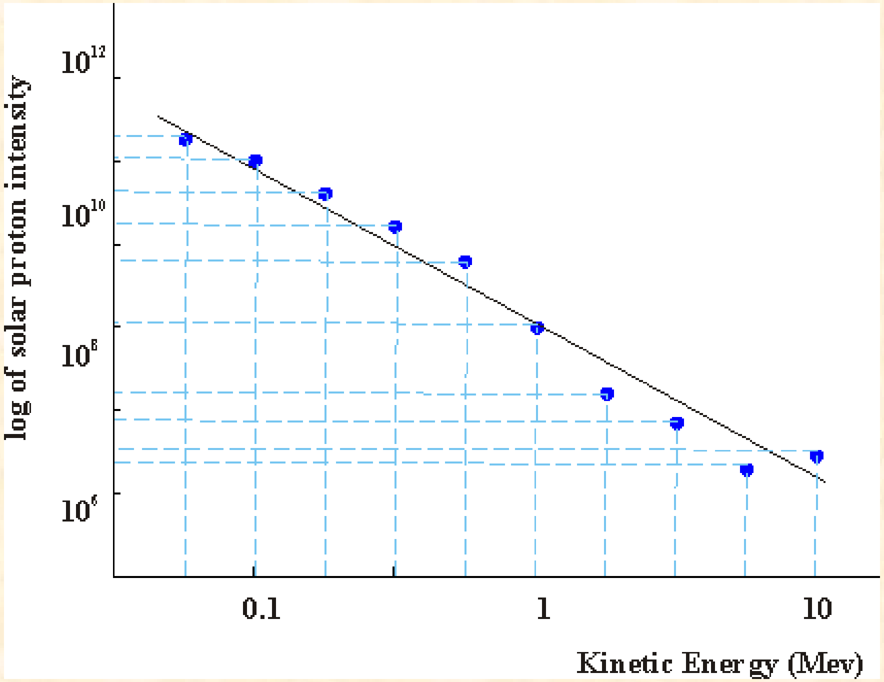
○ ❖● → 1 ❖ → 2

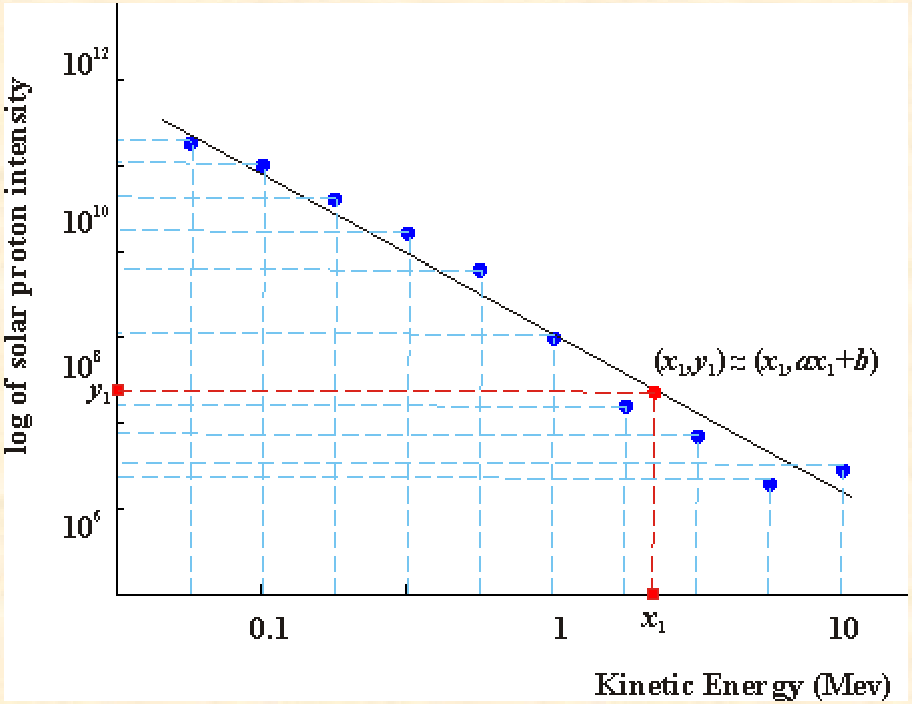


○ ❖● → 1 ❖ → 2

❖ΕΞΑΓΩΓΗ ΓΝΩΣΗΣ

❖(b) Πώς μπορούμε να χρησιμοποιήσουμε τα γνωστά δεδομένα προκειμένου να κάνουμε εκτιμήσεις (estimations) για άγνωστα δεδομένα; (**παλινδρόμηση**)





❖**Παραμετρική εκτίμηση**

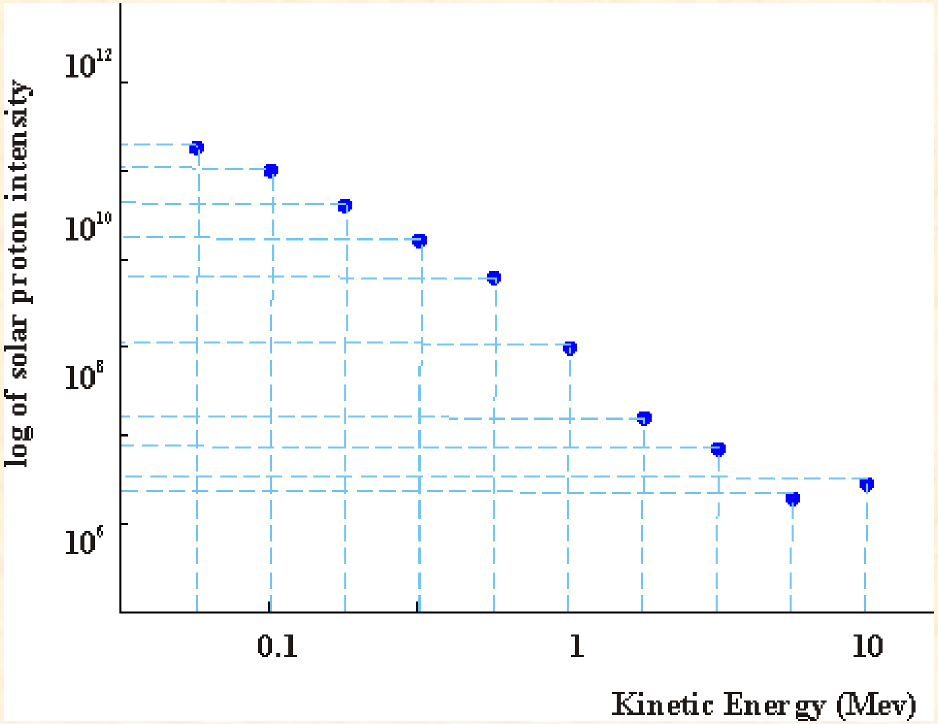
❖*Μοντέλο:* **γραμμή y=ax+b**

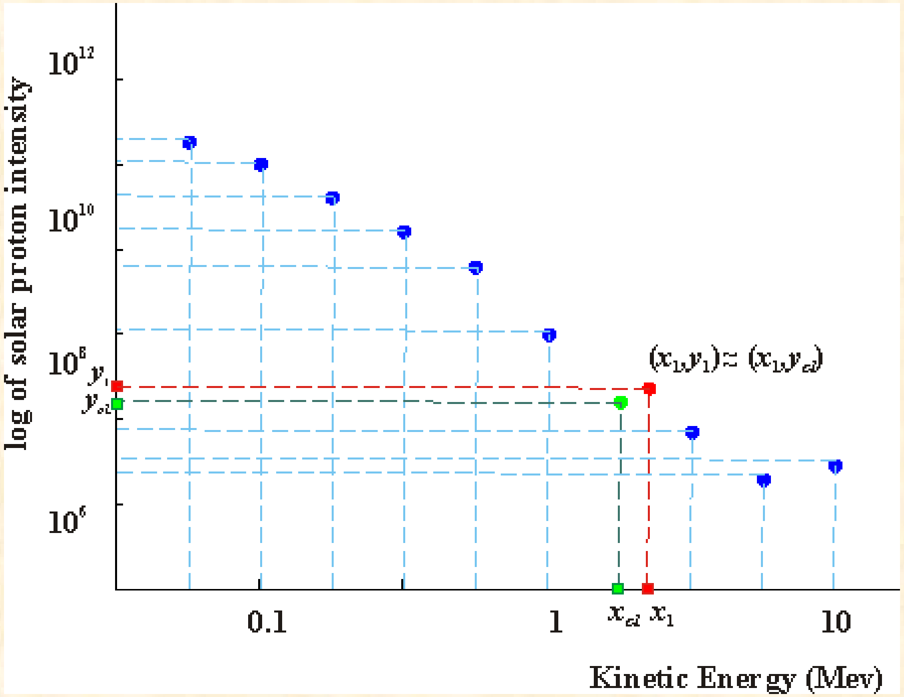
❖*Παράμετροι μοντέλου:* **a, b**

❖ΕΞΑΓΩΓΗ ΓΝΩΣΗΣ

❖(b) Πώς μπορούμε να χρησιμοποιήσουμε τα γνωστά δεδομένα προκειμένου να κάνουμε εκτιμήσεις (estimations) για άγνωστα δεδομένα;

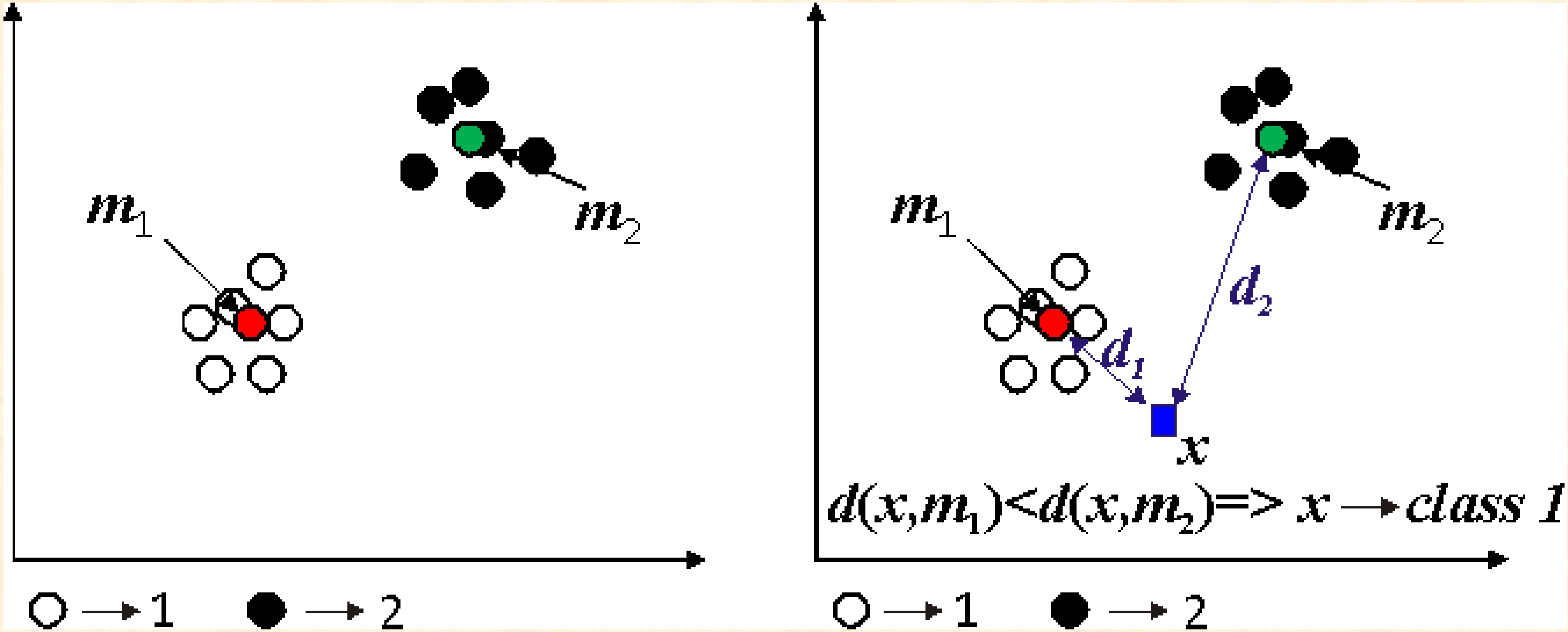
❖**Μη παραμετρική ταξινόμηση**





❖ΕΞΑΓΩΓΗ ΓΝΩΣΗΣ

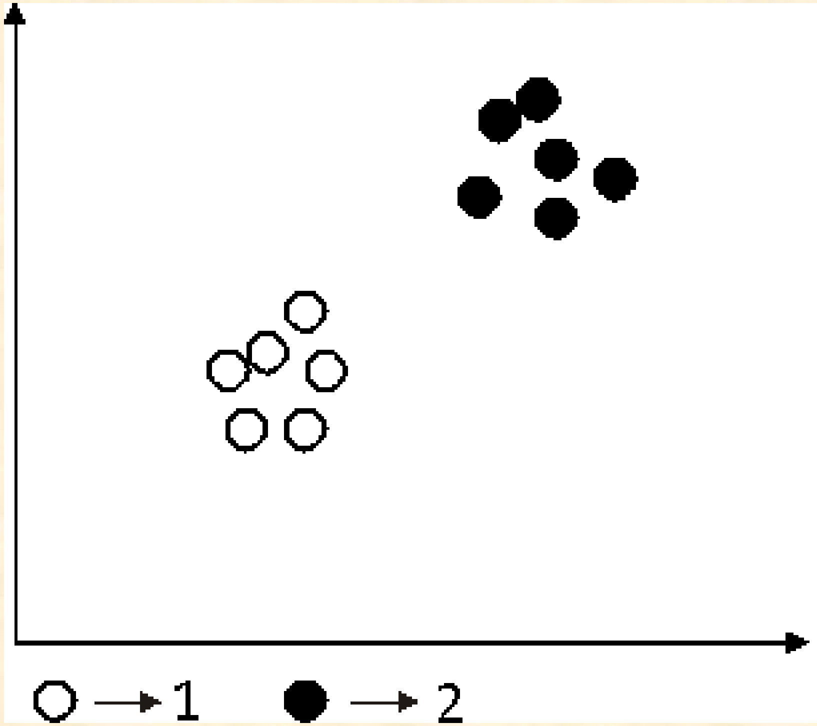
❖(b) Πώς μπορούμε να χρησιμοποιήσουμε τα γνωστά δεδομένα προκειμένου να κάνουμε εκτιμήσεις (estimations) για άγνωστα δεδομένα; (**ταξινόμηση**)

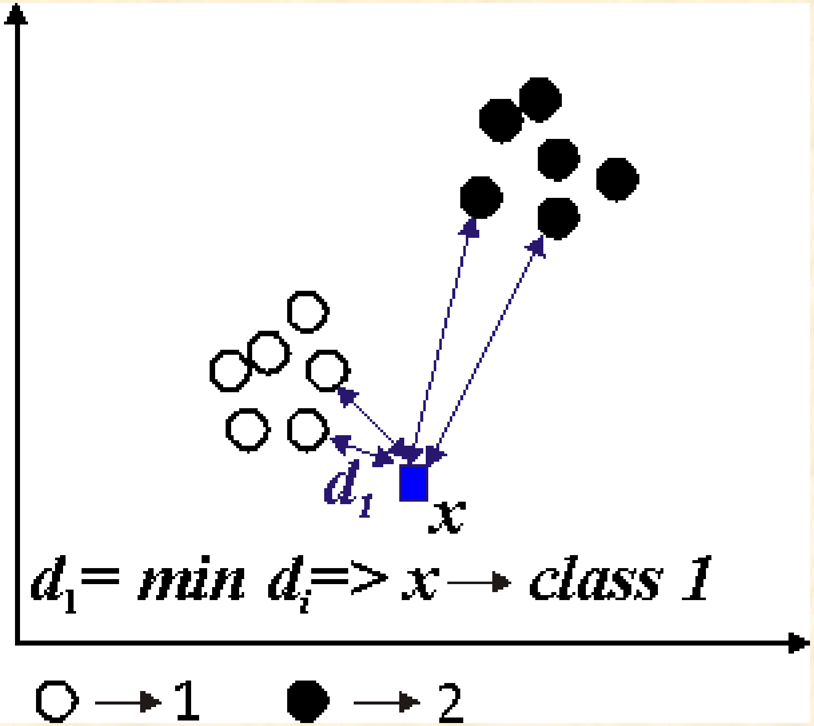


❖**Παραμετρική εκτίμηση**
❖**Παράμετροι:** m_1, m_2

❖ΕΞΑΓΩΓΗ ΓΝΩΣΗΣ

❖(b) Πώς μπορούμε να χρησιμοποιήσουμε τα γνωστά δεδομένα προκειμένου να κάνουμε εκτιμήσεις (estimations) για άγνωστα δεδομένα; (**ταξινόμηση**)

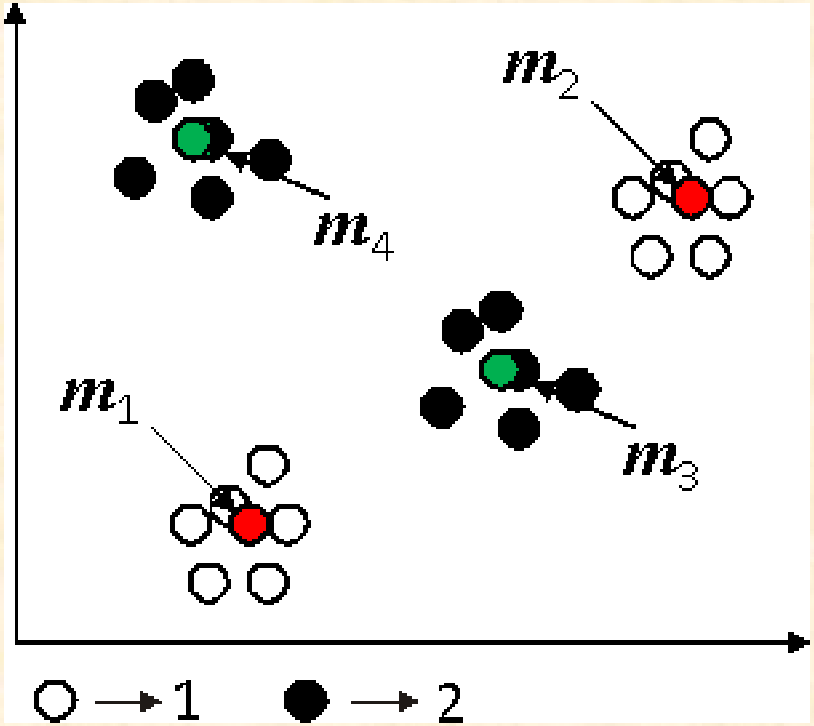




❖**Μη παραμετρική εκτίμηση**

❖ΕΞΑΓΩΓΗ ΓΝΩΣΗΣ

❖ (c) Ποιο είναι το γενικό μοτίβο εξάπλωσης (spread pattern) των δεδομένων στο χώρο; (**ταξινόμηση**)

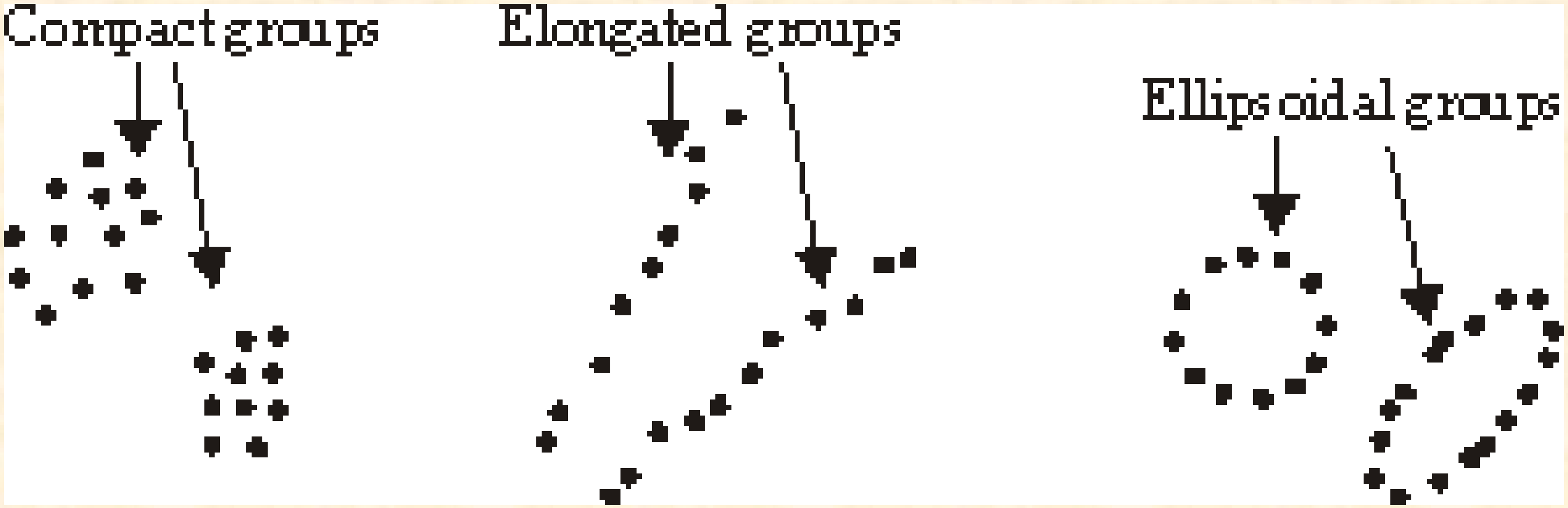


❖Η **κατηγορία 1** συγκεντρώνεται γύρω από τα $\boldsymbol{m}_1$ και $\boldsymbol{m}_2$

❖Η **κατηγορία 2** συγκεντρώνεται γύρω από τα $\boldsymbol{m}_3$ και $\boldsymbol{m}_4$

❖ΕΞΑΓΩΓΗ ΓΝΩΣΗΣ

❖ (c) Ποιο είναι το γενικό μοτίβο εξάπλωσης (spread pattern) των δεδομένων στο χώρο; (**ομαδοποίηση**)



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)

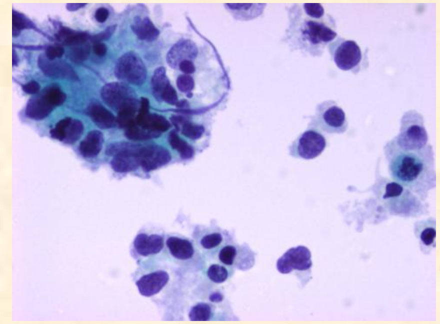
- **<u>Το πρόβλημα:</u>** Καταχώρηση άγνωστων αντικειμένων – προτύπων – στη σωστή κατηγορία (κλάση). Το πρόβλημα είναι γνωστό ως ταξινόμηση (classification).

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

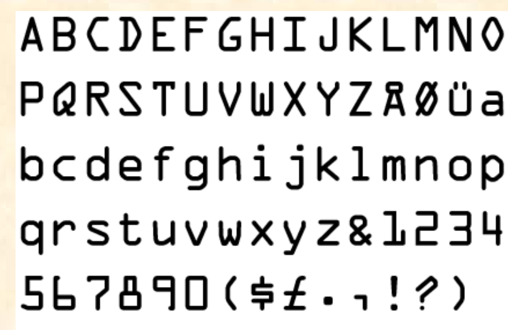
  - **Με επίβλεψη**:
    - Γνωστός αριθμός κλάσεων (κατηγοριών)
    - Διαθέσιμα αντικείμενα για τα οποία είναι γνωστή η κλάση στην οποία ανήκουν.
  - **Χωρίς επίβλεψη**:
    - Άγνωστος αριθμός κλάσεων (γενικά).
    - Διαθέσιμα αντικείμενα για τα οποία δεν είναι γνωστή οποιαδήποτε πληροφορία σχετική με κλάση.

Παραδείγματα:

- Εφαρμογή στην **κυτταρολογία**.
  - **Αντικείμενα:** πυρήνες κυττάρων
  - **Κατηγορίες:** καλοήθη, κακοήθη



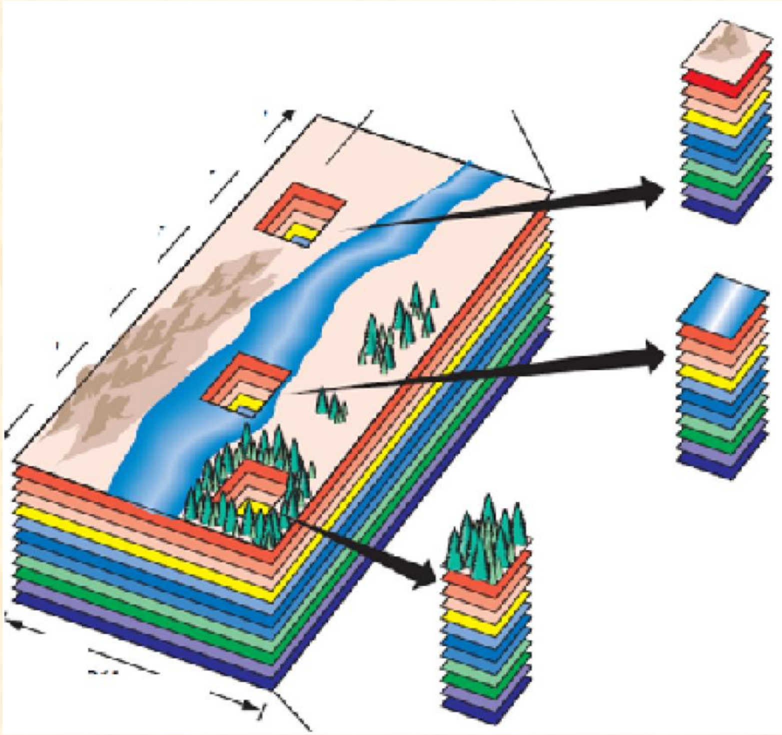
- Εφαρμογή σε αλυσίδα παραγωγής στη **βιομηχανία**.
  - **Αντικείμενα:** Π.χ. βίδες
  - **Κατηγορίες:** Ελαττωματική, μη ελαττωματική
- Εφαρμογή στην **αναγνώριση χαρακτήρων κειμένου**
  - **Αντικείμενα:** Αλφαριθμητικοί χαρακτήρες, σημεία στίξης
  - **Κατηγορίες:** 26 κατηγορίες πεζών+26 κατ. κεφαλαίων+10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



❖Εφαρμογή σε **ταξινόμηση υπερφασματικών εικόνων**

➢**Αντικείμενα:** *εικονοστοιχεία*

➢**Κατηγορίες:** *τύποι εδάφους* (π.χ., γυμνό έδαφο, νερό, βλάστηση κλπ.).



❖ 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

$$x_1, \ldots, x_l,$$

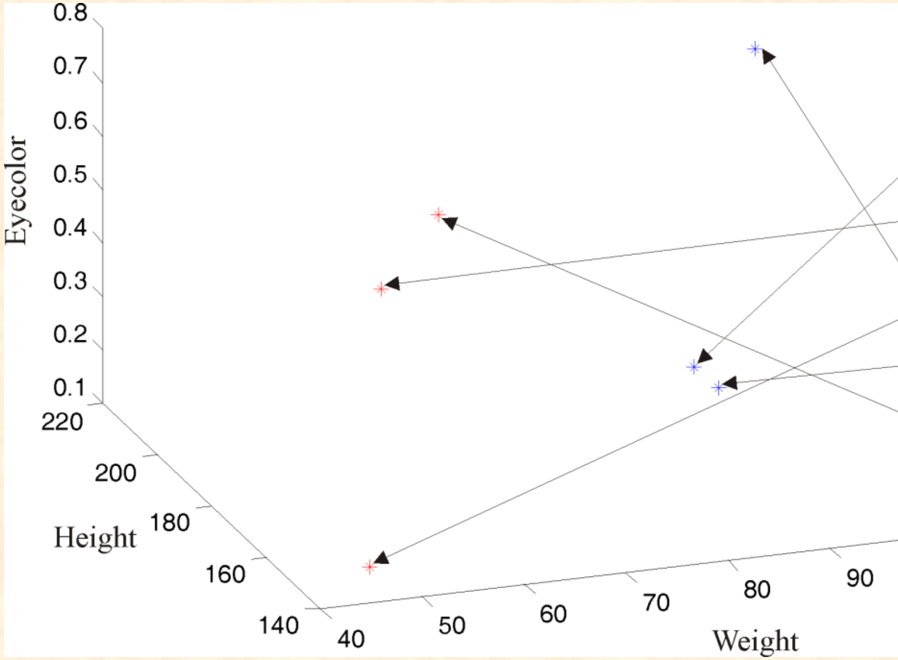
συνιστούν το διάνυσμα χαρακτηριστικών

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

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

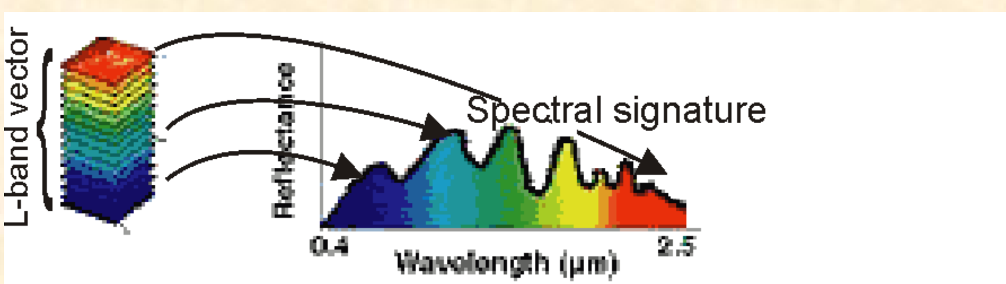
Παραδείγματα:

❖**Σύνολο αθλητών: Χαρακτηριστικά** βάρος, ύψος, χρώμα ματιών



	Weight	Height	Eyecolor
Athlete1	90	190	.2
Athlete2	50	155	.6
Athlete3	100	205	.7
Athlete4	48	152	.1
Athlete5	95	200	.1
Athlete6	57	160	.7

❖**Σύνολο εικονοστοιχείων σε υπερφασματικές εικόνες: Χαρακτηριστικά** φασματικές μετρήσεις

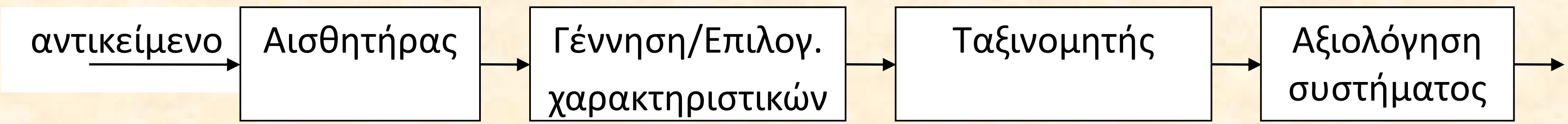


Αναπαράσταση αντικειμένων σε σύστημα αναγνώρισης προτύπων:

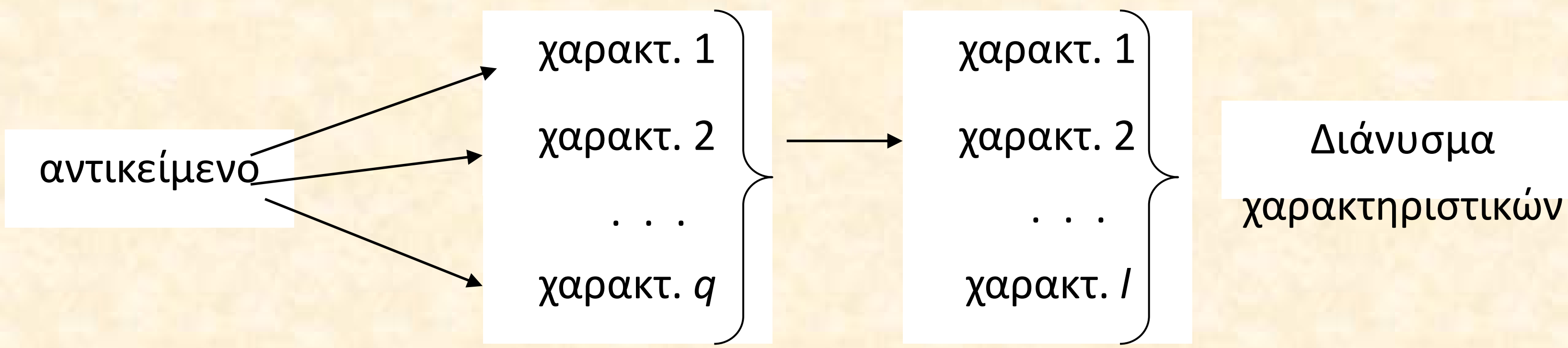
Αντικείμενο ⇐===⇒ Διάνυσμα χαρακτηριστικών μετρήσεων (feature vector)

Αντικείμενο⇐=⇒ 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:

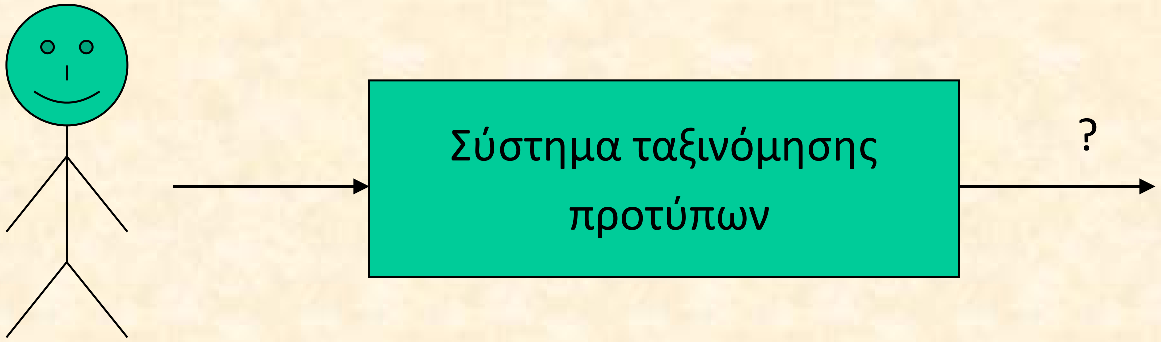
(A) Ποια χαρακτηριστικά θα επιλέξω;

(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) **<u>Χαρακτηριστικά</u>**

ύψος, βάρος.

Περιμένουμε ότι

	D	B
Βάρος	μικρό	μεγάλο
Ύψος	μικρό	μεγάλο

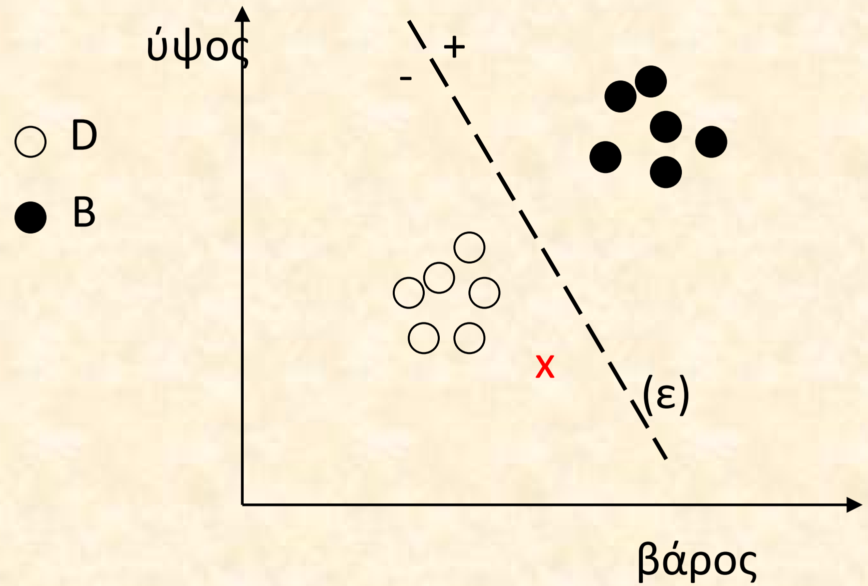
Πώς καθορίζεται ποσοτικά όμως το «**μεγάλο**» και το «**μικρό**»;

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.

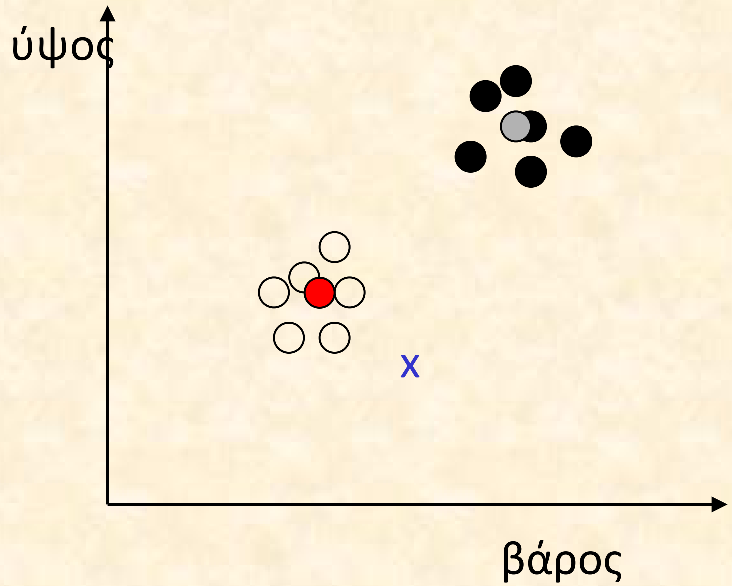


Παράδειγμα 1:

Δοθέντος ενός χαρακτηριστικού διανύσματος, ο ταξινομητής εξετάζει αν αυτό βρίσκεται

(α) στην αρνητική πλευρά της (ε), οπότε το κατατάσσει στην κατηγορία “D”

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

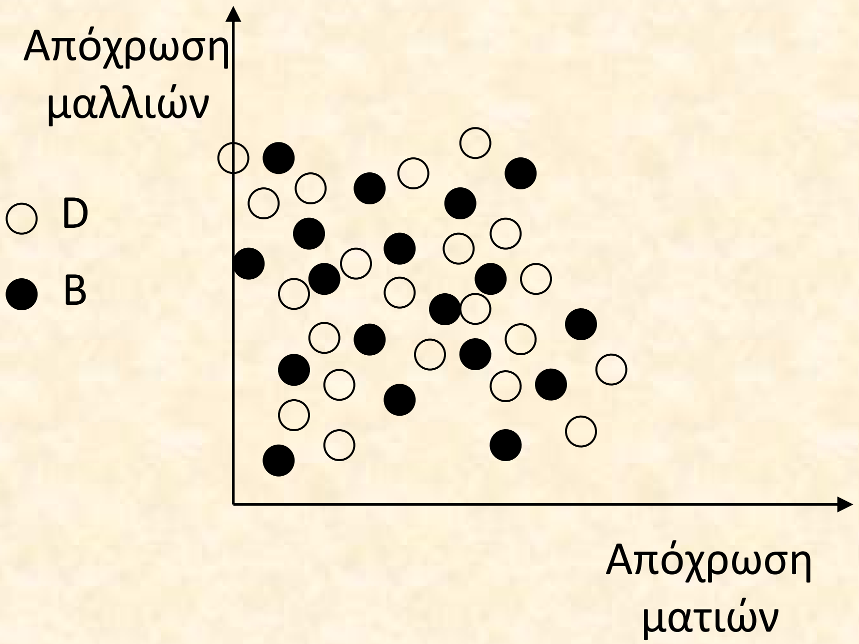


Παράδειγμα 2:

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

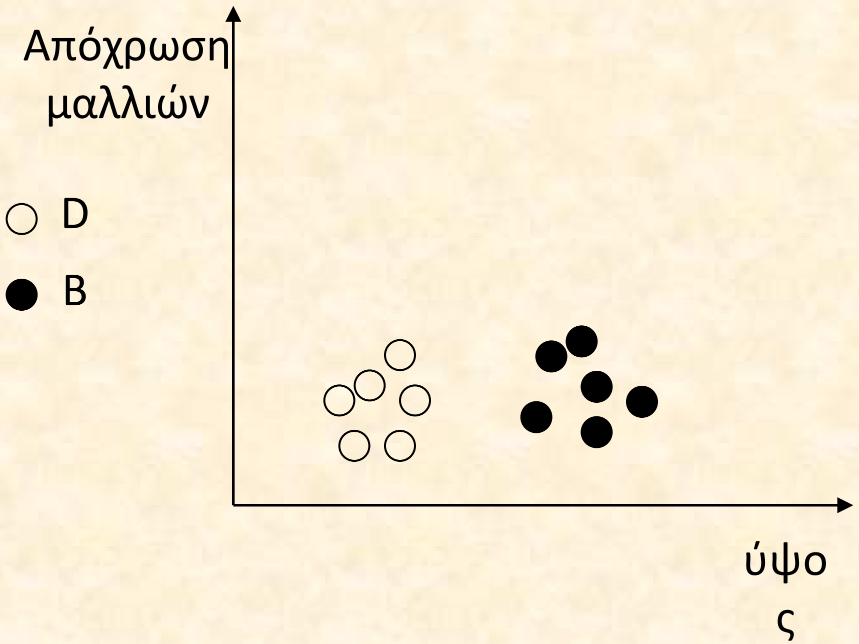
Χαρακτηριστικά (ξανά)

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 προκαταρτικά

Έστω $\mathcal{A}_i, i = 1, 2, \ldots, M$, *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 προκαταρτικά (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

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

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

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

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

A posteriori (εκ των υστέρων) πιθανότητα

Μερικά προκαταρτικά (συν.)

Αν

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

τότε

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

Αν ο *A* είναι συμμετρικός τότε

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

Αν ο *A* είναι διαγώνιος τότε

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

Μερικά προκαταρτικά (συν.)

Αν

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

Ο *A* είναι **οριστικά θετικός** (**μη αρνητικός**) ανν

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

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 $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$

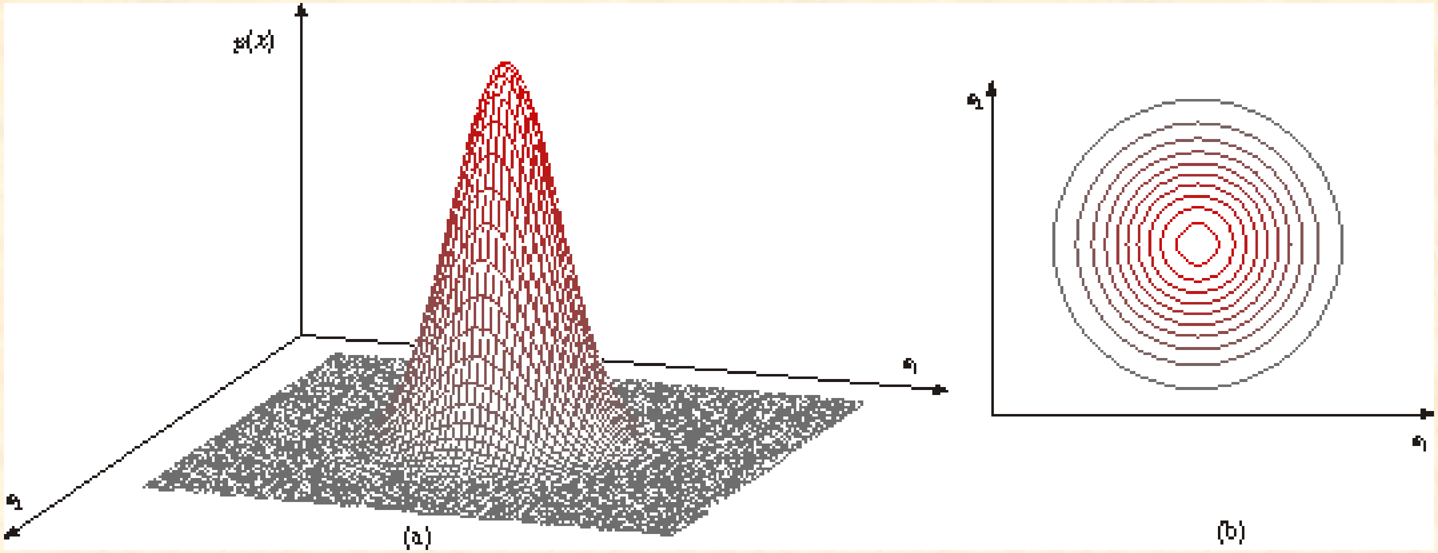
$$p(\boldsymbol{x}) = \frac{1}{(2\pi)^{l/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{\mu})\right)$$

όπου

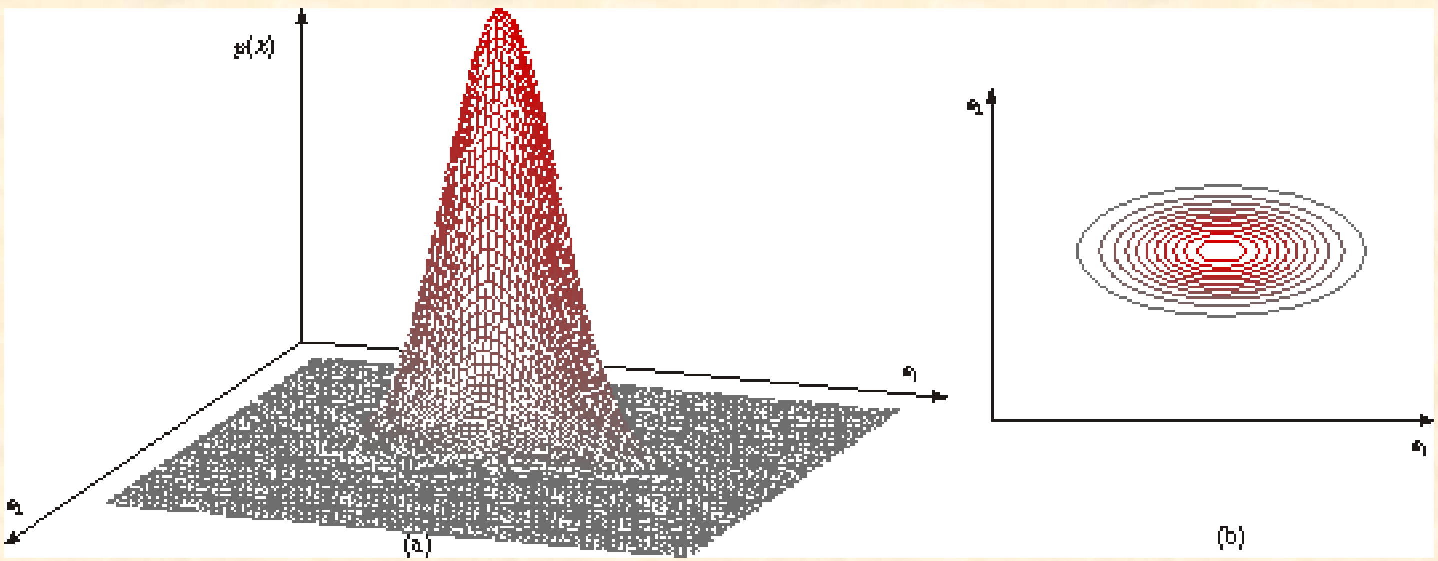
$$E(\boldsymbol{x}) \equiv E\left[[x_1, x_2, \ldots, x_l]^T\right] = [\mu_1, \mu_2, \ldots, \mu_l]^T \equiv \boldsymbol{\mu}$$

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

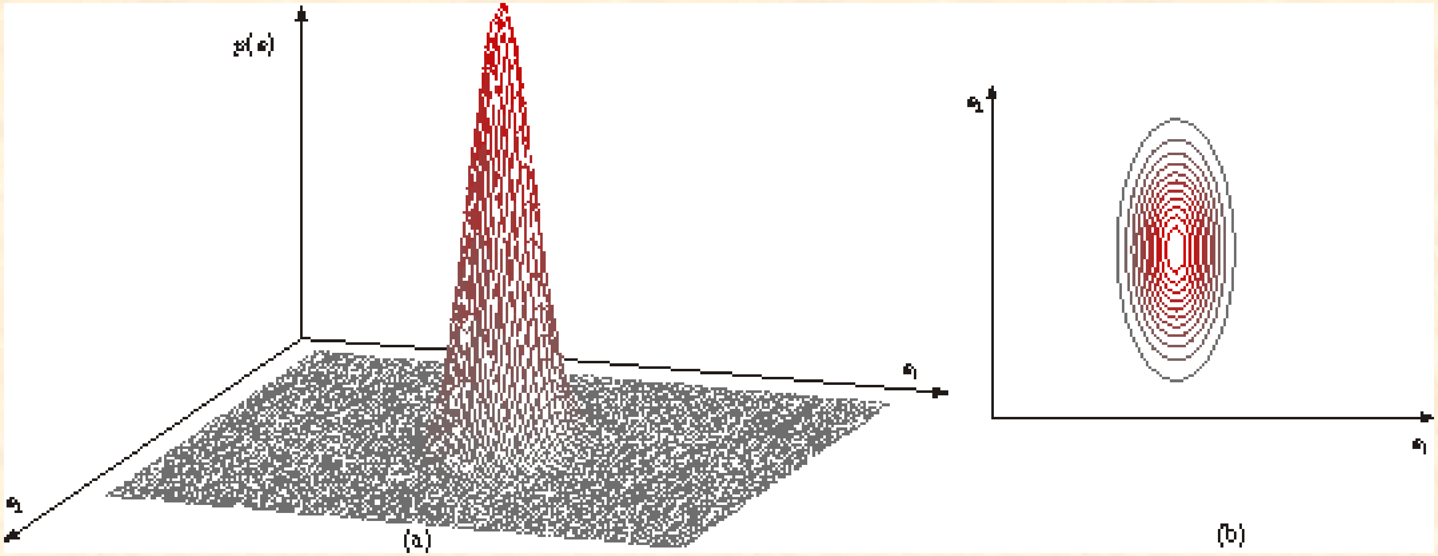
$$\sigma_i^2 = E[(x_i-\mu_i)^2], \sigma_{ij} = \sigma_{ji} = E[(x_i-\mu_i)(x_j-\mu_j)]$$



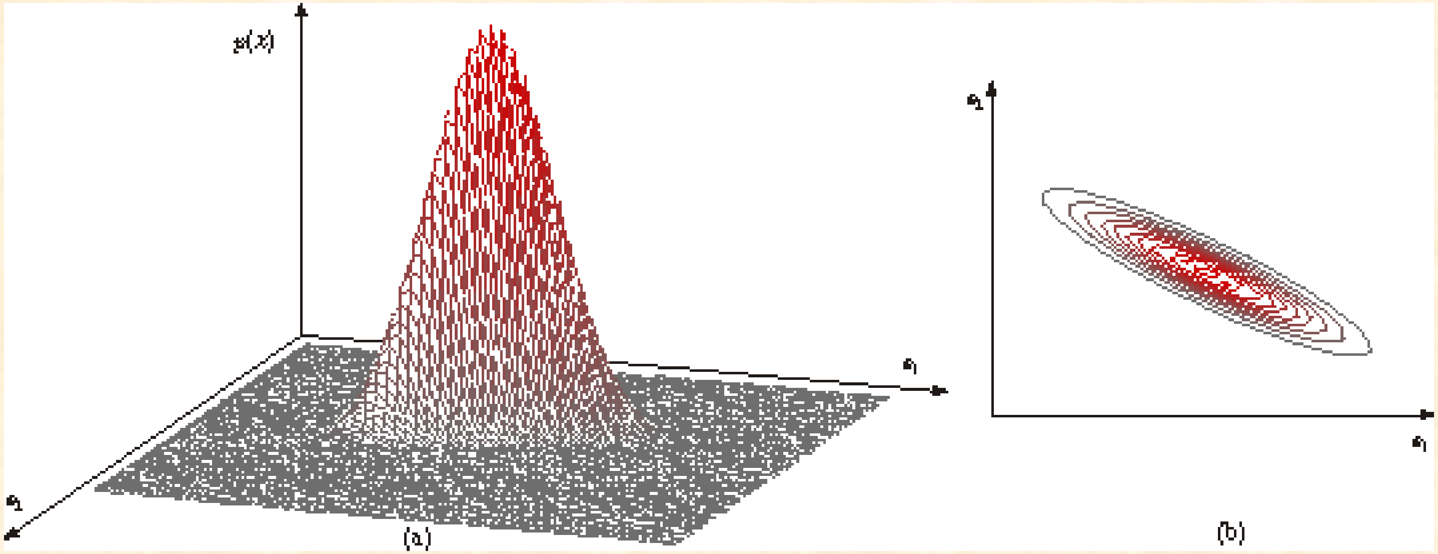
Σ: διαγώνιος
με ίσα διαγώνια
στοιχεία



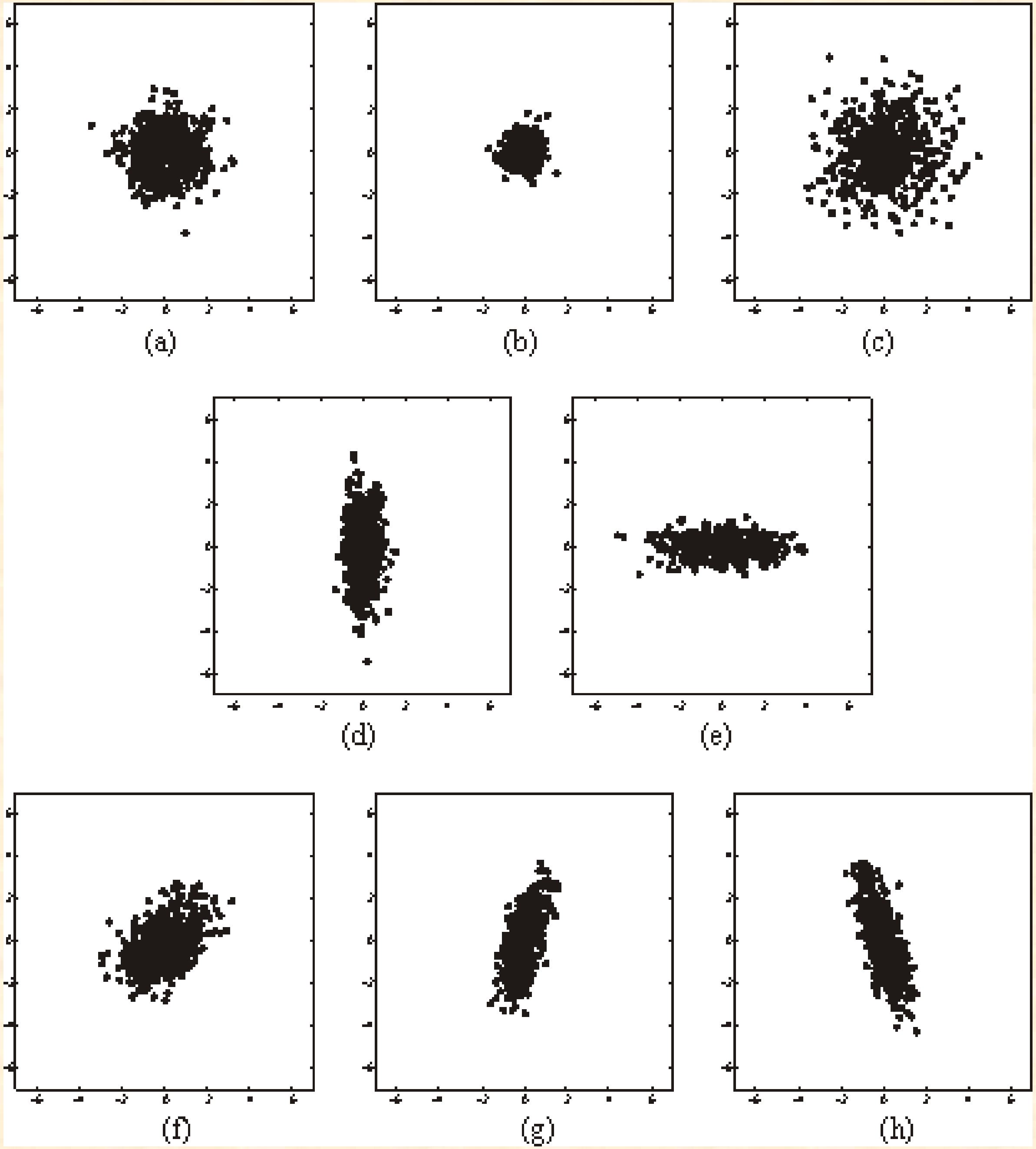
Σ: διαγώνιος
με $\sigma_1^2 >> \sigma_2^2$



Σ: διαγώνιος
με $\sigma_1^2 << \sigma_2^2$



Σ: μη διαγώνιος



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

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

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

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

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

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

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

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

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

Μερικά προκαταρτικά (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 *Σ* είναι διαγώνιος

Στην περίπτωση της πολυδιάστατης κανονικής κατανομής ισχύει:

Σ διαγώνιος ⇔ οι συνιστώσες του ***x*** είναι στατιστικά ανεξάρτητες

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

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

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

Υπόθεση: Σε μια μεγάλη αίθουσα γυμναστηρίου βρίσκονται *100* χορευτές (**D**) και *200* καλαθοσφαιριστές (**B**)

Ερώτηση 1: *«Αν επιλέξουμε κάποιον από την αίθουσα, μαντέψτε αν είναι D ή B.»*

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

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

Άρα επιλέγεται η απάντηση **B**

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

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

Ερώτηση 2: *«Αν επιλέξουμε κάποιον από την αίθουσα, μαντέψτε αν είναι D ή B.»*

Απάντηση: Ίδια με την ερώτηση 1.

Ερώτηση 3: *««Αν επιλέξουμε κάποιον από την αίθουσα, μαντέψτε αν είναι D ή B, δοθέντος ότι το ύψος του είναι 155 εκ;»*

Απάντηση: Ίδια με την ερώτηση 1.

Ερώτηση 4: *««Αν επιλέξουμε κάποιον από την αίθουσα, μαντέψτε αν είναι D ή B, δοθέντος ότι το βάρος του είναι 60 κιλά;»*

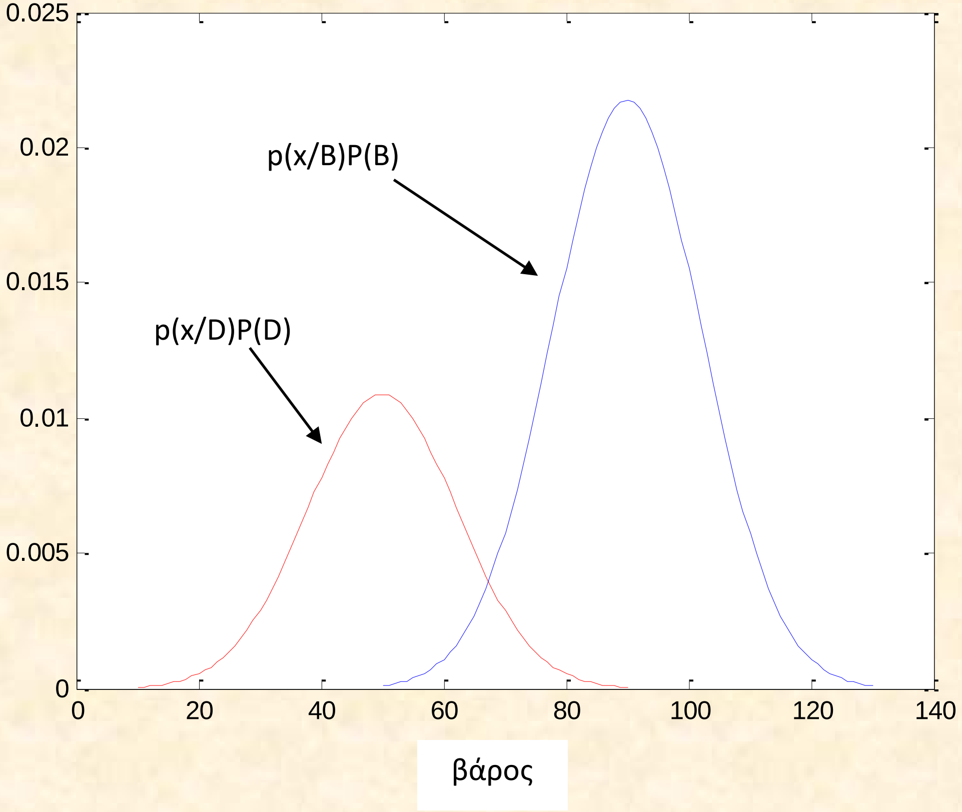
Απάντηση:

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

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

Ερώτηση 5: *««Αν επιλέξουμε κάποιον από την αίθουσα, μαντέψτε αν είναι D ή B, δοθέντος ότι η απόχρωση των μαλλιών του είναι y∈[χ1,χ2]».*

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 **D**.

Διαφορετικά καταχώρησε τον αθλητή στην κατηγορία **B**.

Λαμβανομένου υπόψη ότι

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

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

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

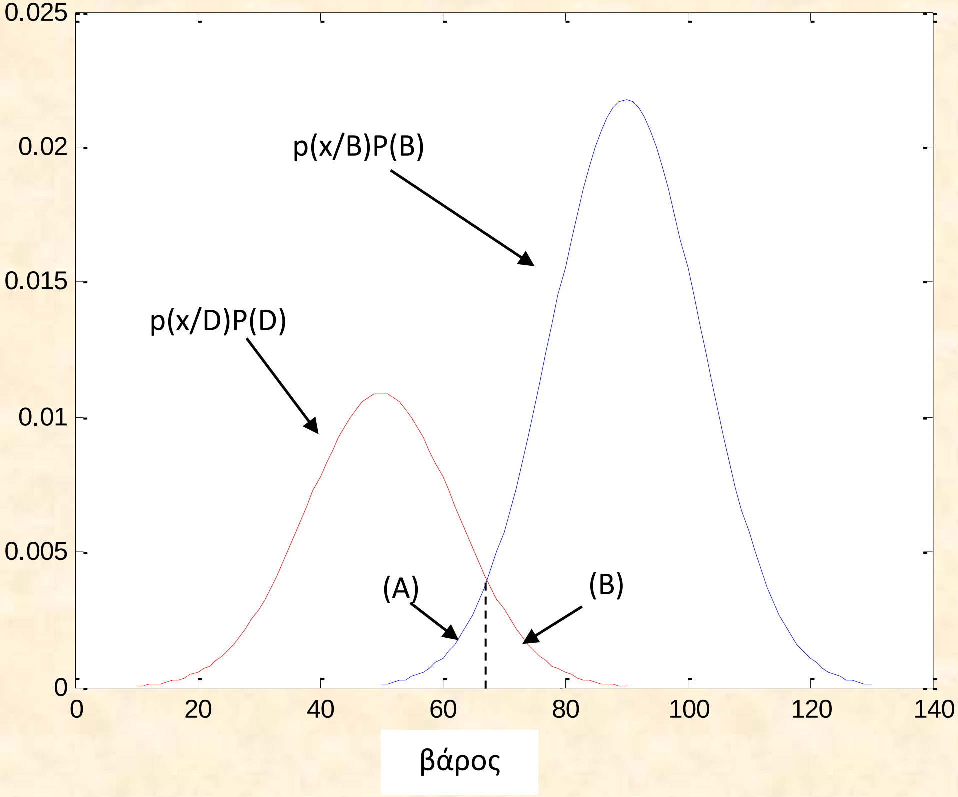
οριοθετούν τις περιοχές του χώρου των χαρακτηριστικών (βάρος) που αντιστοιχούν στις δύο κατηγορίες

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

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

Σχεδιασμός ταξινομητή

Συμβολισμός:

- $\omega_1,\omega_2,\ldots,\omega_c$ οι κατηγορίες στις οποίες θα ταξινομηθεί μια οντότητα
- l το πλήθος των χαρακτηριστικών που αναπαριστούν μια οντότητα (διάσταση του χώρου των χαρακτηριστικών)
- $\mathbf{x}=[x_1,x_2,\ldots x_l]^T$ 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 ω_i.

Ερώτηση 7: *«Δοθέντος ενός διανύσματος μετρήσεων* ***x*** *μιας υπό εξέταση οντότητας να προσδιοριστεί η κατηγορία ωi στην οποία ανήκει η οντότητα»*

Bayes 1: *«καταχώρησε την υπό εξέταση οντότητα στην κατηγορία ω_i για την οποία*

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

Δεδομένου ότι:

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

Bayes 2: *«καταχώρησε την υπό εξέταση οντότητα στην κατηγορία ω_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 ωi 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:

ω1: $P(\omega1)=1/2$, $p(x|\omega1)=(1/\sqrt{2\pi})*\exp(-x^2/2)$

ω2: $P(\omega2)=1/3$, $p(x|\omega2)=(1/\sqrt{2\pi})*\exp(-(x-1)^2/2)$

ω3: $P(\omega3)=1/6$, $p(x|\omega3)=(1/\sqrt{2\pi})*\exp(-(x-2)^2/2)$

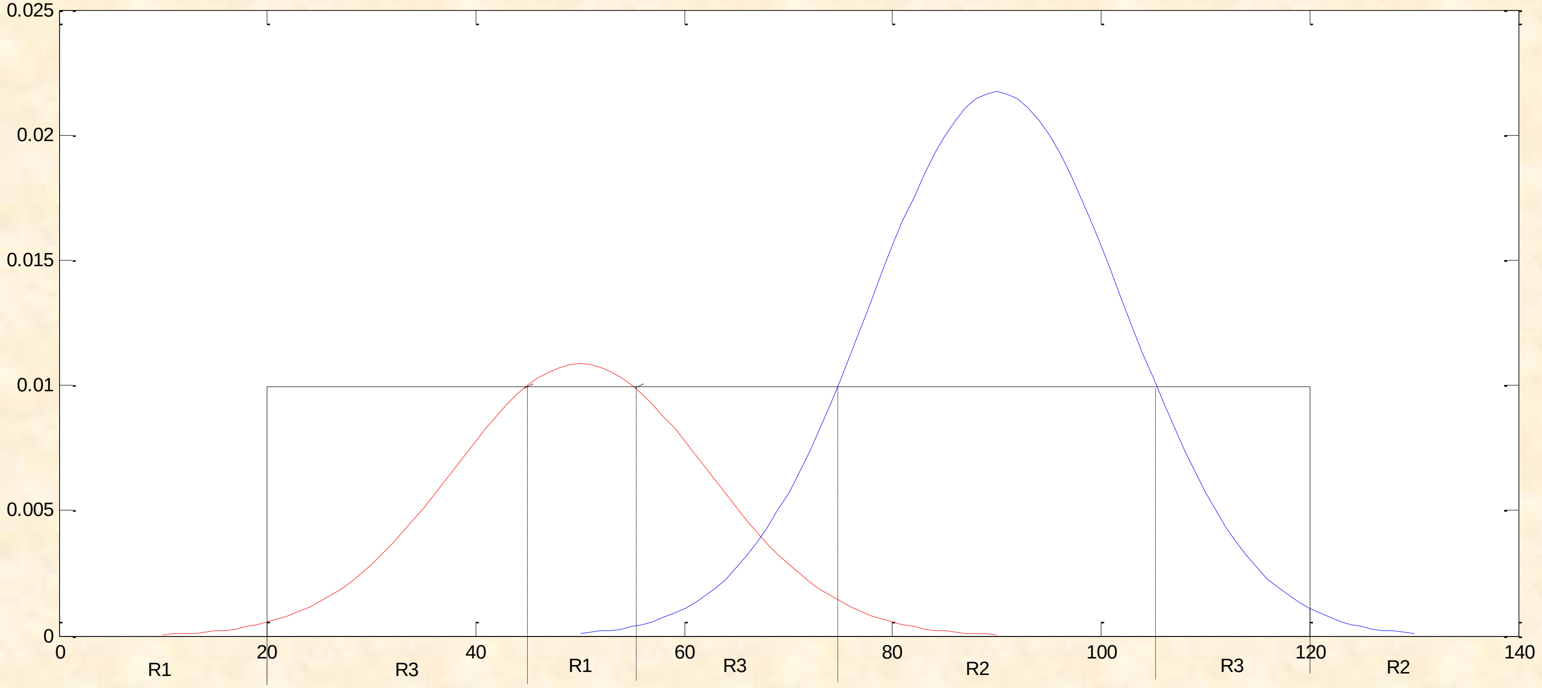
Ταξινομήστε σε μία από τις παραπάνω κατηγορίες άγνωστο δείγμα με τιμή χαρακτηριστικού **x=1.6**.

Είναι:

$P(\omega1)p(x|\ \omega1)=$ **0.0555**, $P(\omega2)p(x|\ \omega2)=$ **0.1111**, $P(\omega3)p(x|\ \omega3)=$ **0.0614**

Συνεπώς το δείγμα καταχωρείται στην κατηγορία **ω2.**

Παράδειγμα 3 τριών κατανομών



Πιθανότητα λάθους:

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

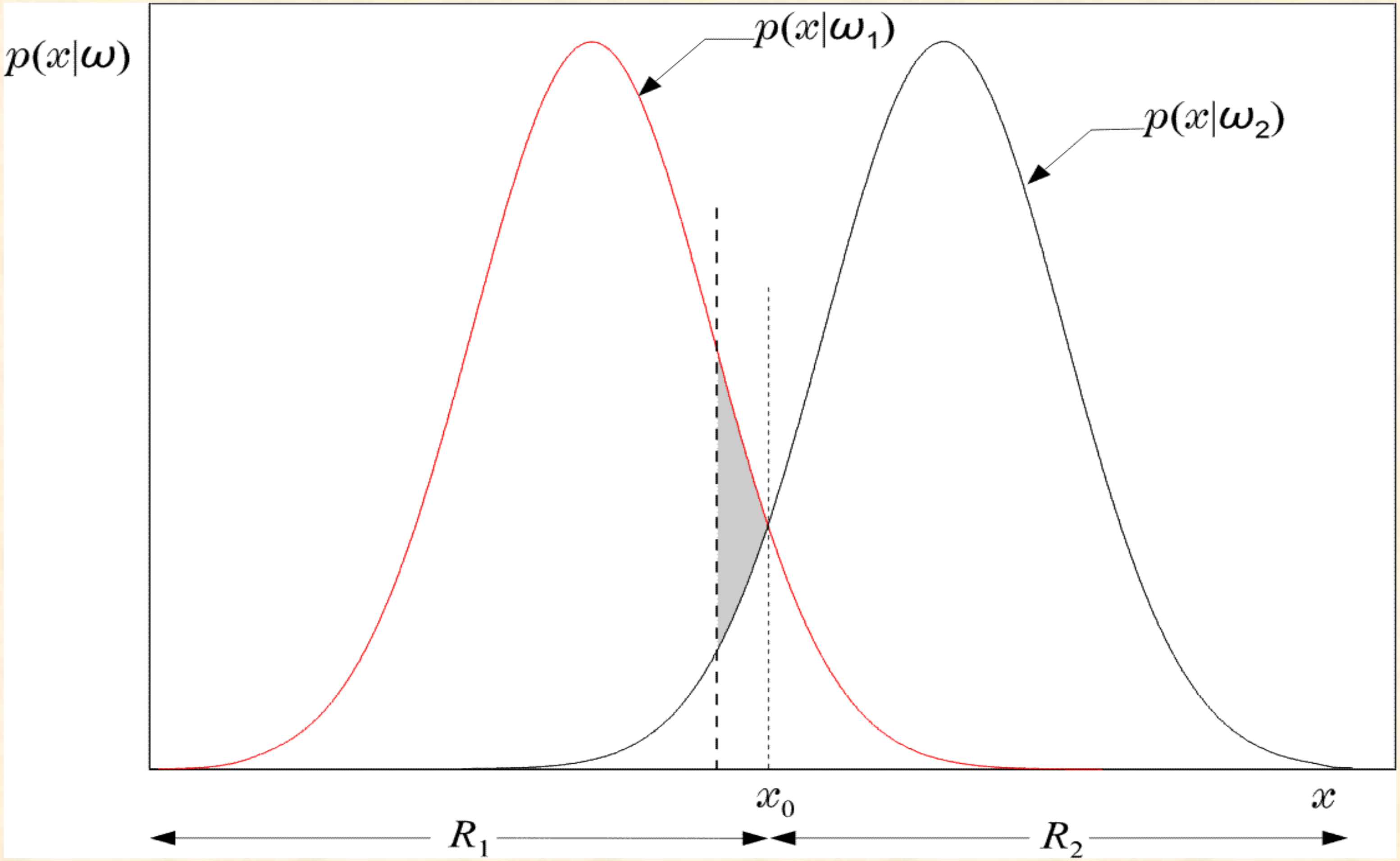
Ερώτηση: Γιατί τόση επιμονή στον ταξινομητή Bayes;

Απάντηση:

•*γιατί η λογική του δεν έρχεται σε αντίθεση με την διαίσθησή μας.*

•*Γιατί περιλαμβάνει ως ειδικές περιπτώσεις πολλούς ταξινομητές που χρησιμοποιούνται στην πράξη.*

•*Ο κανόνας ταξινόμησης κατά Bayes είναι βέλτιστος με την έννοια ότι ελαχιστοποιεί την πιθανότητα λάθους.*

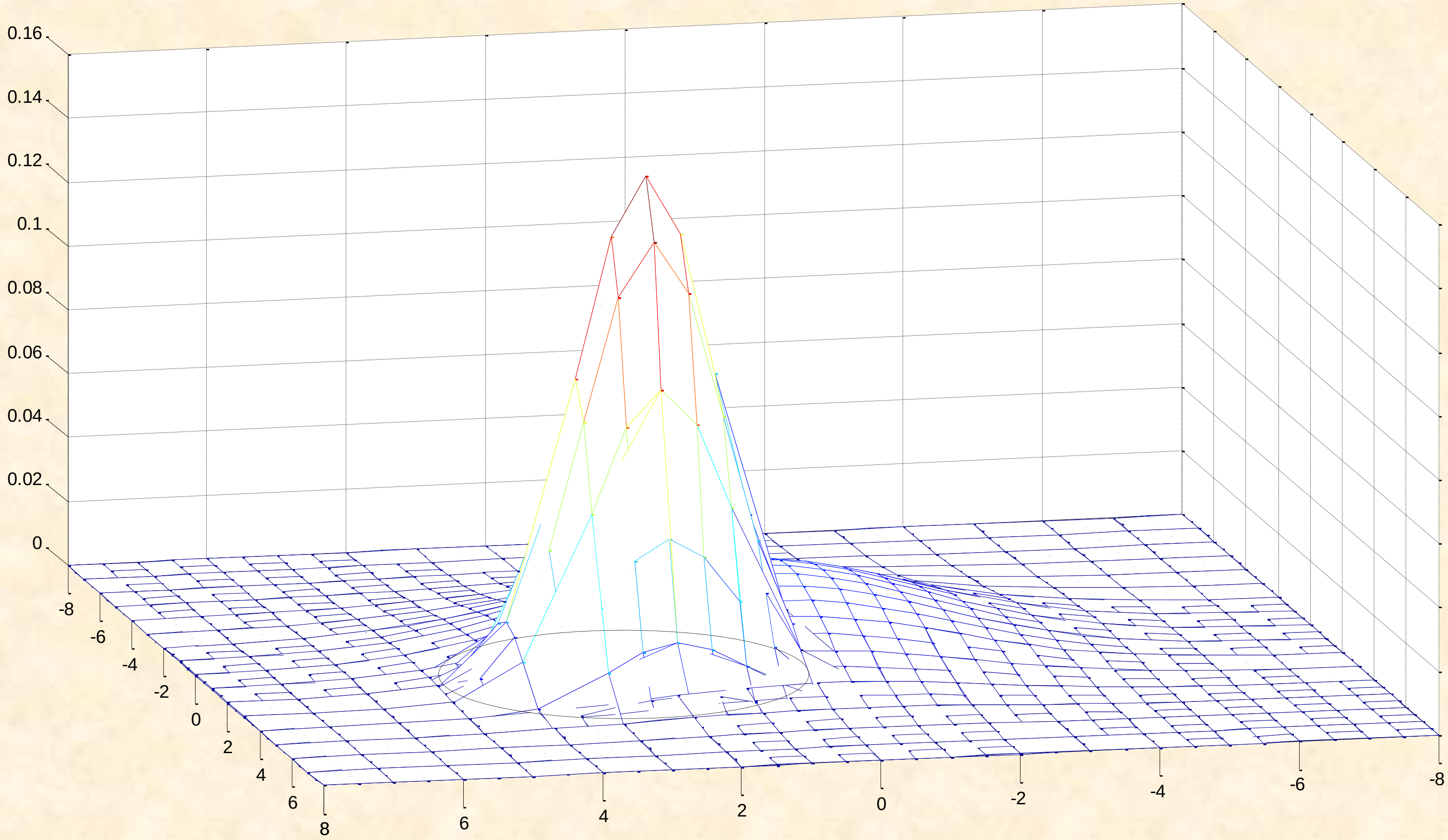


Παρατήρηση: οι πυκνότητες πιθανότητας των κατηγοριών μαζί με τις a priori πιθανότητές τους ορίζουν **μονοσήμαντα** μια διαμέριση του χώρου του *l*-διάστατου χώρου των χαρακτηριστικών, σε (όχι απαραίτητα ενιαίες) περιοχές (R_i) όπου κάθε μια αντιστοιχεί και σε μια κατηγορία (ω_i).

Έτσι, η ταξινόμηση ενός διανύσματος 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.

Εκφράζοντας τον ταξινομητή Bayes

(a) Με χρήση **συναρτήσεων διάκρισης** (**discriminant functions**)

- Έστω $g_q(\mathbf{x})=f(P(\omega_q)p(\mathbf{x}|\omega_q))$, $q=1,\dots,M$, όπου f ***γνησίως αύξουσα*** συνάρτηση.

- Ταξινόμησε δεδομένο $\mathbf{x}$ στην κλάση ω_j για την οποία $g_j(\mathbf{x}) = max_{q=1,\dots,M}\, g_q(\mathbf{x})$.

(b) Με χρήση **επιφανειών απόφασης** (**decision surfaces**)

- Ταξινόμησε δεδομένο $\mathbf{x}$ στην κλάση ω_j για την οποία $x\in R_j$, όπου

$$R_j=\{\mathbf{x} \in R^l: P(\omega_j)p(\mathbf{x}|\omega_j) = max_{q=1,\dots,M}\, P(\omega_q)p(\mathbf{x}|\omega_q) \}$$

A NOTE: Οι R_js μπορούν να ταυτοποιηθούν μέσω των συνόρων τους.

Σημαντική παρατήρηση: Στην πράξη η εύρεση της μίας έκφρασης από την άλλη είναι μια καθόλου προφανής διαδικασία. Αυτός είναι ένας βασικός λόγος για τον οποίο έχουν αναπτυχθεί ταξινομητές χρησιμοποιώντας ανεξάρτητα τις παραπάνω εκφράσεις.

ΤΑΞΙΝΟΜΗΤΗΣ BAYES ΓΙΑ ΚΑΝΟΝΙΚΕΣ ΚΑΤΑΝΟΜΕΣ

- Πολυδιάστατη Gaussian pdf

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

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

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

Πίνακας συνδιασποράς (covariance matrix)

Ο ταξινομητής Bayes: Η περίπτωση των κανονικών κατανομών

❖Ας θεωρήσουμε την περίπτωση όπου

$$p(x|\omega_j) \equiv N(\mu_j, \Sigma_j) = \frac{1}{(2\pi)^{l/2}|\Sigma_j|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu_j)^T \Sigma_j^{-1}(x-\mu_j)\right)$$

Είναι

$$g_j(x) = \ln\left(P(\omega_j)p(x|\omega_j)\right) = -\frac{1}{2}(x-\mu_j)^T \Sigma_j^{-1}(x-\mu_j) + c_j + \ln P(\omega_j) \qquad (A)$$

ή, **ισοδύναμα**,

$$g_j(x) = -\frac{1}{2}x^T \Sigma_j^{-1} x + \mu_j^T \Sigma_j^{-1} x - \frac{1}{2}\mu_j^T \Sigma_j^{-1}\mu_j + c_j + \ln P(\omega_j) \qquad (B)$$

με

$$c_j = -\frac{l}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma_j|$$

Ο ταξινομητής Bayes: Η περίπτωση των κανονικών κατανομών

❖Επίσης $g_{jk}(\boldsymbol{x})=0 \Leftrightarrow g_j(\boldsymbol{x})- g_k(\boldsymbol{x})=0$ είναι **ισοδύναμη** με

❖Εξίσωση συνόρου συνεχόμενων κλάσεων

$$\begin{cases} -\frac{1}{2}x^T\Sigma_j^{-1}x+\mu_j^T\Sigma_j^{-1}x-\frac{1}{2}\mu_j^T\Sigma_j^{-1}\mu_j+c_j+\ln P(\omega_j)= \\ -\frac{1}{2}x^T\Sigma_k^{-1}x+\mu_k^T\Sigma_k^{-1}x-\frac{1}{2}\mu_k^T\Sigma_k^{-1}\mu_k+c_k+\ln P(\omega_k) \end{cases}$$

(i) $\Sigma_i \neq \Sigma_k$:

Σ’ αυτή την περίπτωση η $g_{jk}(\boldsymbol{x})=0$ αναπαριστά μια καμπύλη δευτέρου βαθμού (π.χ. υπερ-έλλειψη, υπερ-παραβολή) (τετραγωνικός ταξινομητής - quadratic discriminant analysis)

❖ (ii) $\Sigma_i=\Sigma_k$:

Σ’ αυτή την περίπτωση η $g_{jk}(\boldsymbol{x})=0$ αναπαριστά μια καμπύλη πρώτου βαθμού (υπερεπίπεδο) (γραμμικός ταξινομητής - linear discriminant analysis)

Ο ταξινομητής Bayes: Η περίπτωση των κανονικών κατανομών

(ii)-a $\Sigma_i=\Sigma_k$ ***και*** $P(\omega_j)=P(\omega_k)$:

Αγνοώντας τις κοινές ποσότητες που εμφανίζονται σε όλες τις $g_j(\boldsymbol{x})$s, η $g_j(\boldsymbol{x})$ γίνεται

$$g_j(x) = -\frac{1}{2}(x-\mu_j)^T \Sigma^{-1}(x-\mu_j)$$

Θυμίζουμε ότι

$$d_M(x,\mu_j) = \left(\frac{1}{2}(x-\mu_j)^T \Sigma^{-1}(x-\mu_j)\right)^{1/2}$$

είναι η γνωστή **Mahalanobis απόσταση.**

Σ' αυτή την περίπτωση, ο ταξινομητής Bayes μπορεί να γραφτεί **ισοδύναμα** ως

- Καταχώρησε το $\boldsymbol{x}$ στην ω_i με:
- $d_M(\boldsymbol{x},\boldsymbol{\mu}_j) = \boldsymbol{min}_{q=1,\ldots M}\, d_M(\boldsymbol{x},\boldsymbol{\mu}_q)$

Ταξινομητής ελάχιστης Mahalanobis απόστασης

Ο ταξινομητής Bayes: Η περίπτωση των κανονικών κατανομών

(ii)-b $\Sigma_i=\Sigma_k=\sigma^2 I$ ***και*** $P(\omega_j)=P(\omega_k)$:

Αγνοώντας τις κοινές ποσότητες που εμφανίζονται σε όλες τις $g_j(\boldsymbol{x})$s, η $g_j(\boldsymbol{x})$ γίνεται

$$g_j(x) = -\frac{1}{2}(x-\mu_j)^T(x-\mu_j) = -\frac{1}{2}\| x-\mu_j \|^2$$

όπου $||\boldsymbol{x}-\boldsymbol{\mu}_j||$ είναι η γνωστή **Ευκλείδεια απόσταση.**

Σ' αυτή την περίπτωση, ο ταξινομητής Bayes μπορεί να διατυπωθεί **ισοδύναμα** ως εξής

- Καταχώρησε το $\boldsymbol{x}$ στην ω_i με:
- $||\boldsymbol{x} - \boldsymbol{\mu}_j|| = \boldsymbol{min}_{q=1,\ldots M} ||\boldsymbol{x} - \boldsymbol{\mu}_q||$

Ταξινομητής ελάχιστης Ευκλείδειας απόστασης

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

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

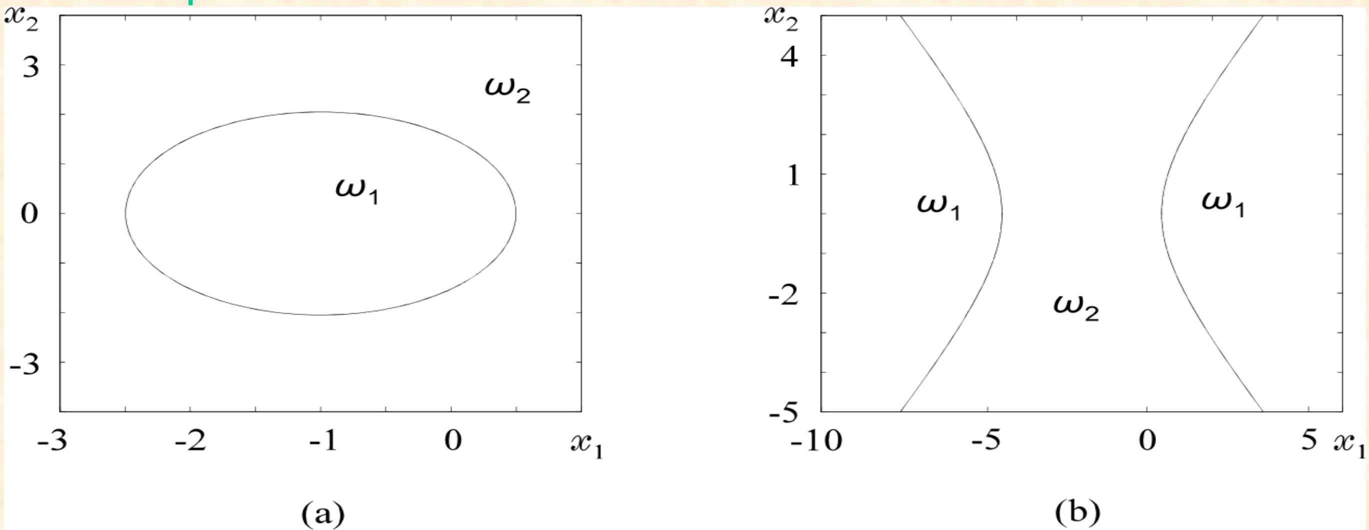
$$-\frac{1}{2\sigma^2}(\mu_{i1}^2 + \mu_{i2}^2) + \ln P(\omega_i) + C_i$$

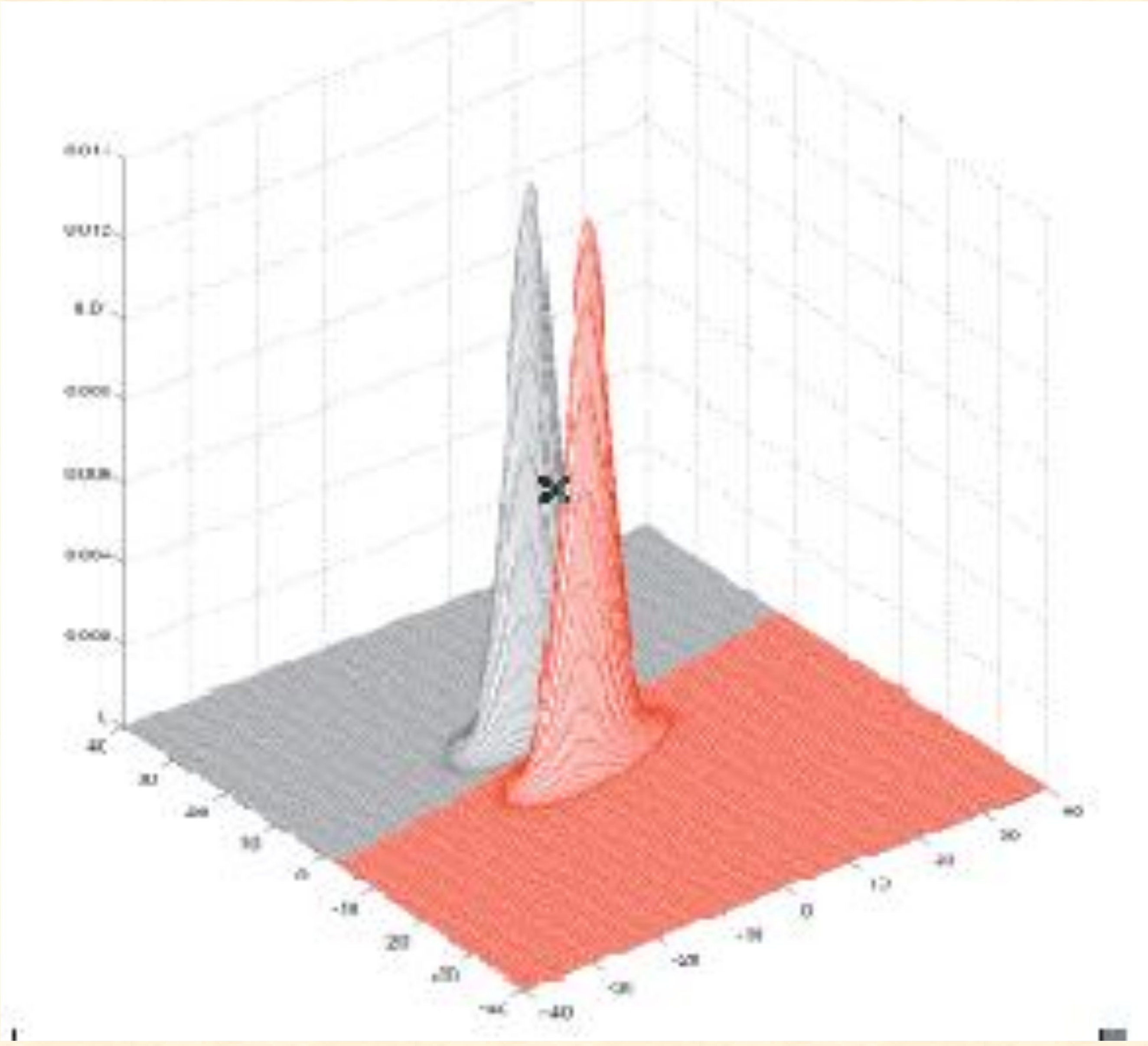
Δηλαδή, η $g_i(x)$ είναι τετραγωνική. Σημειώστε ότι οι επιφάνειες

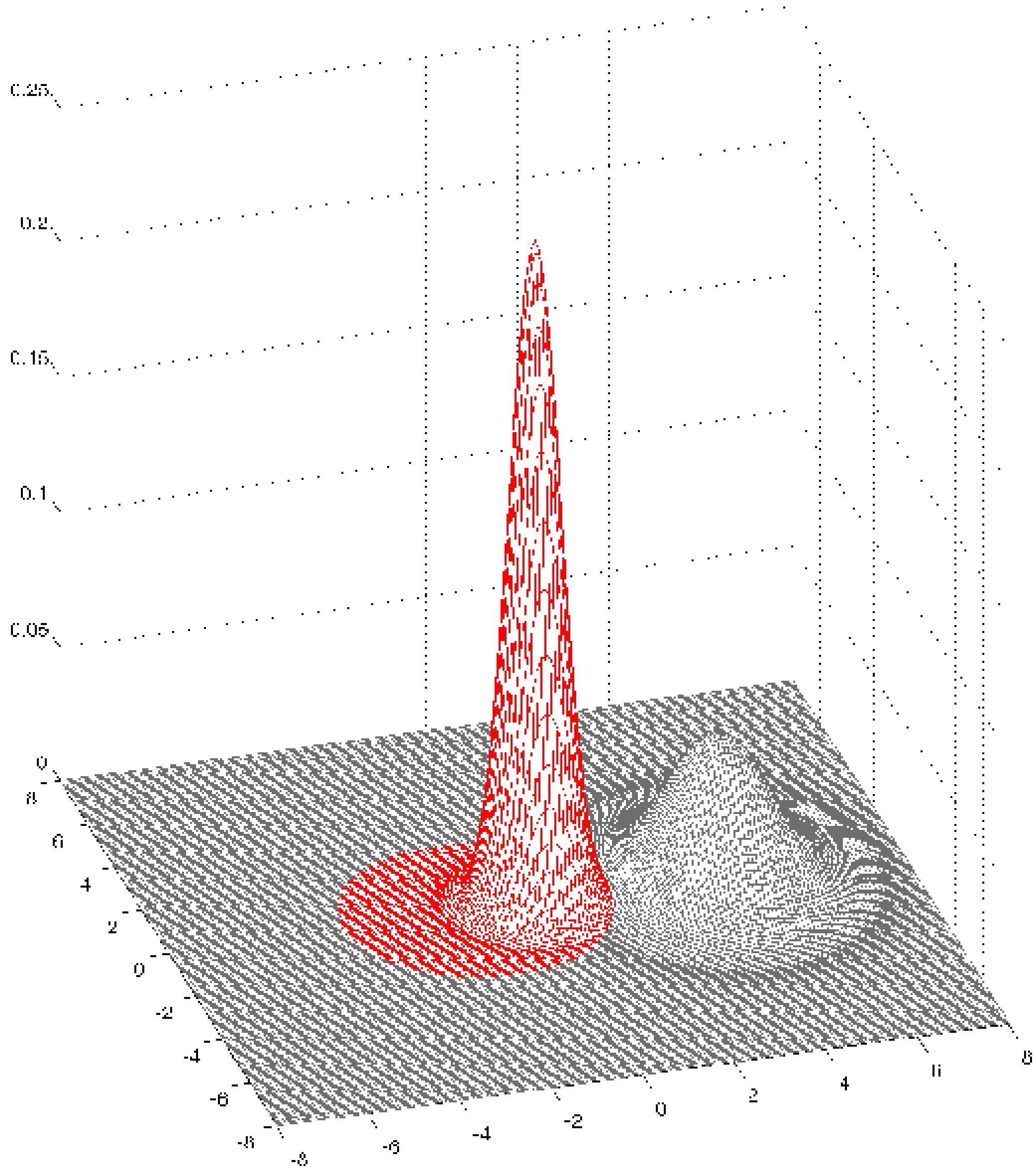
$$g_i(\underline{x}) - g_j(\underline{x}) = 0$$

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

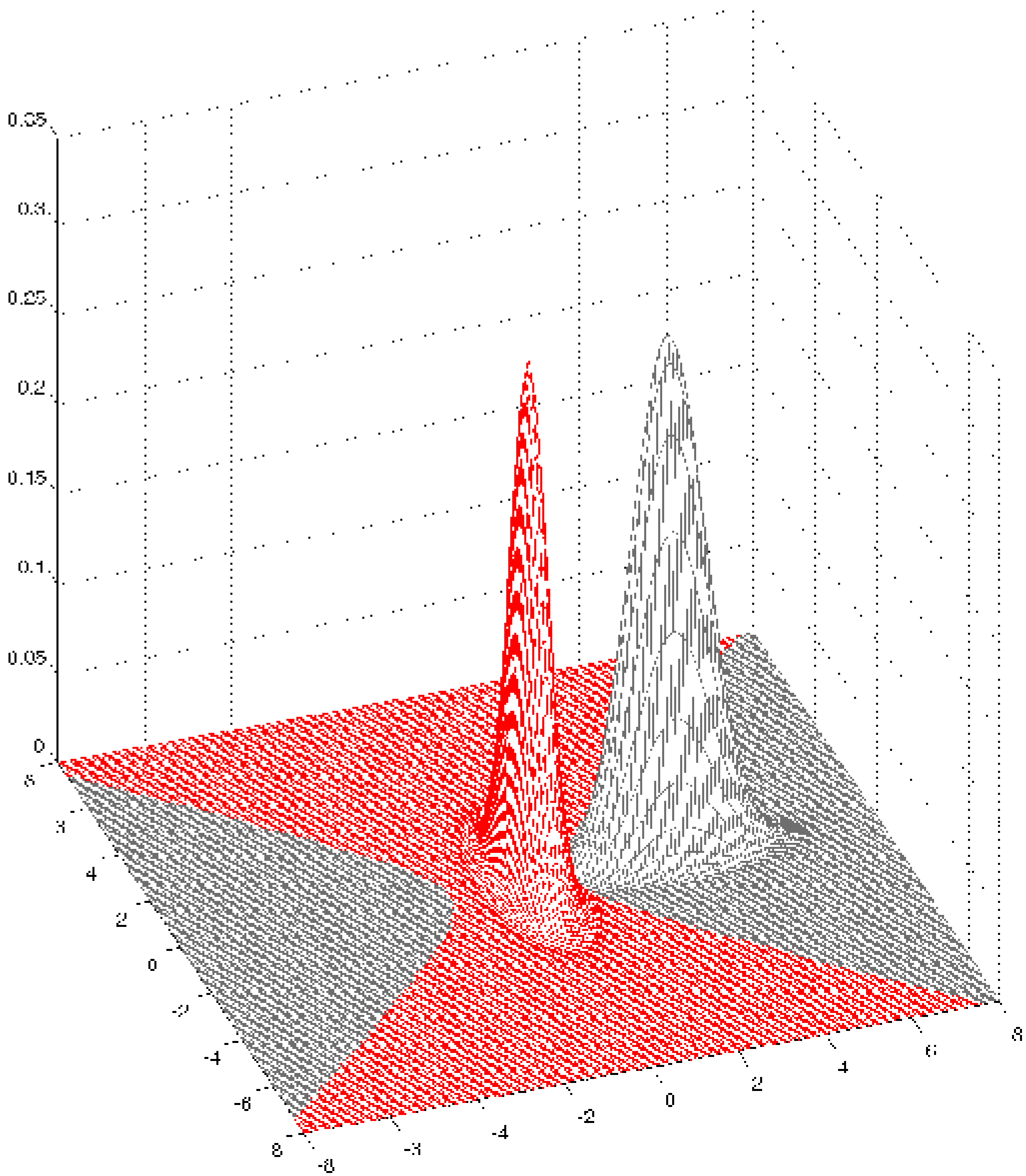
For example:







0.25
0.2
0.15
0.1
0.05
0
8
6
4
2
0
-2
-4
-6
-8
-8
-6
-4
-2
0
2
4
6
8



0.35
0.3
0.25
0.2
0.15
0.1
0.05
0

❖ Έκφραση γραμμικών ταξινομητών μέσω συναρτήσεων διάκρισης

- 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:

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

Οι συναρτήσεις διάκρισης είναι ΓΡΑΜΜΙΚΕΣ

❖Έκφραση γραμμικών ταξινομητών μέσω συναρτήσεων διάκρισης

➢ Έστω επιπλέον ότι:

- $$\Sigma = \sigma^2 I. \text{ Τότε}$$

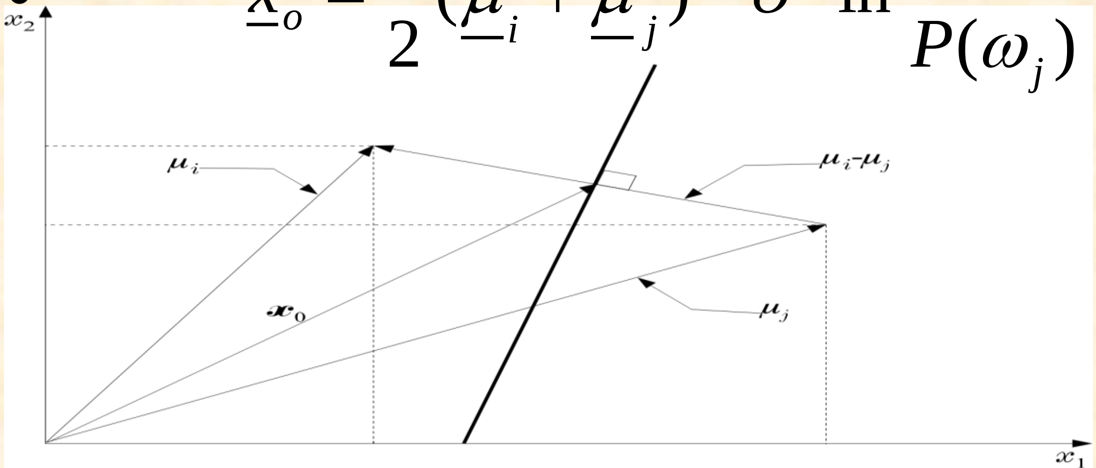
$$g_i(\underline{x}) = \frac{1}{\sigma^2} \underline{\mu}_i^T \underline{x} + w_{i0}$$

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

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

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

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



❖Έκφραση γραμμικών ταξινομητών μέσω συναρτήσεων διάκρισης

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

- $$g_{ij}(\underline{x}) = \underline{w}^T(\underline{x} - \underline{x}_0) = 0$$
- $$\underline{w} = \Sigma^{-1}(\underline{\mu}_i - \underline{\mu}_j)$$
- $$\underline{x}_0 = \frac{1}{2}(\underline{\mu}_i + \underline{\mu}_j) - \ell\text{n}\left(\frac{P(\omega_i)}{P(\omega_j)}\right)\frac{\underline{\mu}_i - \underline{\mu}_j}{\left\|\underline{\mu}_i - \underline{\mu}_j\right\|^2_{\Sigma^{-1}}}$$

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

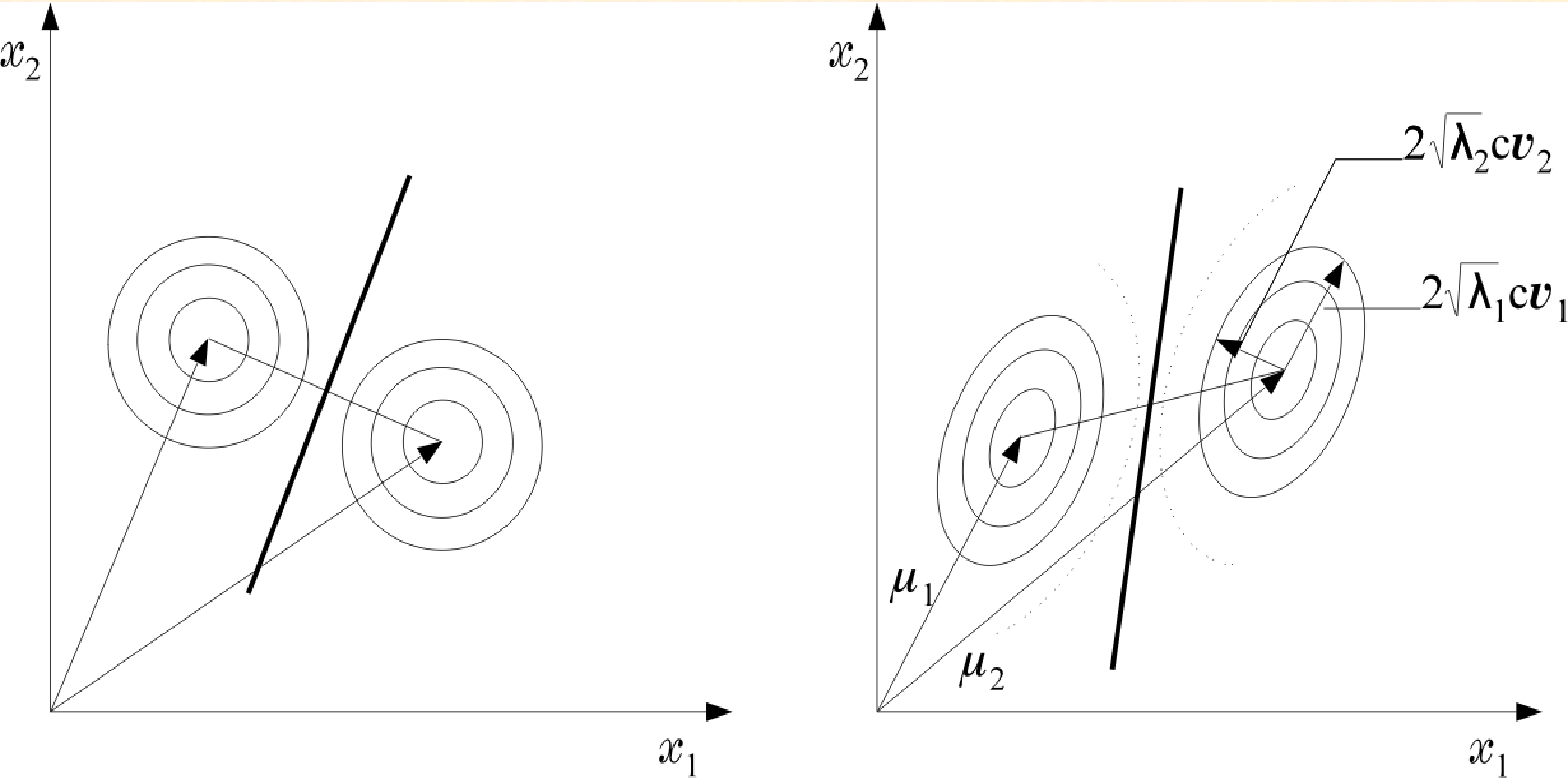
➢ Υπερεπίπεδο απόφασης

όχι κάθετο στο $\underline{\mu}_i - \underline{\mu}_j$

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

❖ Ταξινομητές ελάχιστης απόστασης – Μία διευκρίνιση:

Οι ταξινομητές ελάχιστης Ευκλείδειας και ελάχιστης Mahalanobis απόστασης μπορούν να χρησιμοποιηθούν **ΚΑΙ** στην περίπτωση όπου δεν συντρέχουν οι υποθέσεις κάτω από τις οποίες είναι ισοδύναμοι με τον ταξινομητή Bayes (και κατά συνέπεια βέλτιστοι ως προς το κριτήριο της πιθανότητας λάθους). Βέβαια στην περίπτωση αυτή θα οδηγήσουν σε υποβέλτιστες λύσεις, οι οποίες όμως, πολλές φορές είναι αποδεκτές σε πραγματικές εφαρμογές.


x_2
x_1
x_2
$2\sqrt{\lambda_2}c\boldsymbol{v}_2$
$2\sqrt{\lambda_1}c\boldsymbol{v}_1$
μ_1
μ_2
x_1

(a) (b)

❖ Παράδειγμα:

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

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

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 $\mu_1,\ \mu_2$: $d^2_{m,1} = \begin{bmatrix} 1.0, & 2.2 \end{bmatrix}$

  $$\Sigma^{-1} \begin{bmatrix} 1.0 \\ 2.2 \end{bmatrix} = 2.952,\ d^2_{m,2} = \begin{bmatrix} -2.0, & -0.8 \end{bmatrix} \Sigma^{-1} \begin{bmatrix} -2.0 \\ -0.8 \end{bmatrix} = 3.672$$

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

Ο ταξινομητής Bayes: Η περίπτωση των κανονικών κατανομών

Ερώτηση: Κάτω από ποιες συνθήκες ο ταξινομητής ελάχιστης Ευκλείδειας απόστασης είναι βέλτιστος ως προς το κριτήριο της πιθανότητας λάθους;

Απάντηση: Όταν οι κατηγορίες

(a) είναι ισοπίθανες,

(b) Μοντελοποιούνται από κανονικές κατανομές με πίνακες συνδιασποράς της μορφής $\sigma^2 I$.

Ερώτηση: Κάτω από ποιες συνθήκες ένας γραμμικός ταξινομητής **μπορεί** να είναι βέλτιστος ως προς το κριτήριο της πιθανότητας λάθους;

Απάντηση: Όταν οι κατηγορίες μοντελοποιούνται από κανονικές κατανομές με κοινό μητρώο συνδιασποράς.

Ερώτηση: Κάτω από ποιες συνθήκες ένας τετραγωνικός ταξινομητής **μπορεί** να είναι βέλτιστος ως προς το κριτήριο της πιθανότητας λάθους;

Απάντηση: Όταν οι κατηγορίες μοντελοποιούνται από κανονικές κατανομές με διαφορετικές κατανομές συνδιασποράς.

Σχεδιαζόντας ταξινομητές: Τα δεδομένα

Στην πράξη η γνώση σχετικά διαδικασία γέννεσης των δεδομένων είναι πολύ **σπάνια γνωστή**. Το μόνο που έχουμε στη διάθεσή μας είναι ένα σύνολο δεδομένων (data set).

$$Y = \{(x_i, d_i), x_i \in R^l, d_i \in \{1,2,\ldots,m\}, i = 1,\ldots D\}$$

όπου

$\mathbf{x}_i$ είναι η *l*-διάστατη αναπαράσταση του *i*-στου από τις *N* οντότητες (διάνυσμα δεδομένων – data vector)

d_i είναι η ετικέτα της κλάσης όπου ανήκει το $\mathbf{x}_i$ (1 για την ω_1, 2 για την ω_2,…).

Ο βασικός στόχος ενός ταξινομητή

Δοθέντος ενός ***x*** καταχώρησέ το στην πιο κατάλληλη κατηγορία.

Ανάλογα με τον τρόπο που σχεδιάζεται ένας ταξινομητής, έχουμε δύο κύριες κατηγορίες:

- **Παραμετρικοί ταξινομητές (Parametric classifiers)**
- **Μη παραμετρικοί ταξινομητές (Nonparametric classifiers)**

Σχεδιάζοντας ταξινομητές: Η παραμετρική περίπτωση

Στην περίπτωση αυτή:

• Υποθέτουμε ότι η **μορφή** του παραμετρικού μοντέλου παράγει τα διανύσματα δεδομένων είναι γνωστή, ενώ οι (**άγνωστες**) παράμετροί του εκτιμώνται με βάση το διαθέσιμο σύνολο δεδομένων.

• Η ταξινόμηση ενός νέου διανύσματος δεδομένων ***x*** βασίζεται στο μοντέλο που ορίστηκε παραπάνω.

➢ **ΣΗΜ:** Η ταξινόμηση του ***x*** εξαρτάται **έμμεσα** από το διαθέσιμο σύνολο δεδομένων.

Σχεδιάζοντας ταξινομητές: Η μη παραμετρική περίπτωση

Στην περίπτωση αυτή:

➢ Η ταξινόμηση του ***x*** εξαρτάται **άμεσα** από το διαθέσιμο σύνολο δεδομένων.

Σχεδιάζοντας ταξινομητές: Εκτιμώντας την απόδοσή τους

Ένα σημαντικό ζήτημα: Πώς γνωρίζουμε το **βαθμό αξιοπιστίας** του ταξινομητή που σχεδιάστηκε;

Είναι σαφές ότι, αν χρησιμοποιήσουμε όλα τα διανύσματα δεδομένων του διαθέσιμου συνόλου δεδομένων *Y*, δεν μπορούμε να χρησιμοποιήσουμε τα ίδια δεδομένα για να αξιολογήσουμε αντικειμενικά την απόδοσή του, καθώς ή γνώση που περιέχεται σ’ αυτά έχει ήδη χρησιμοποιηθεί για το σχεδιασμό του ταξινομητή.

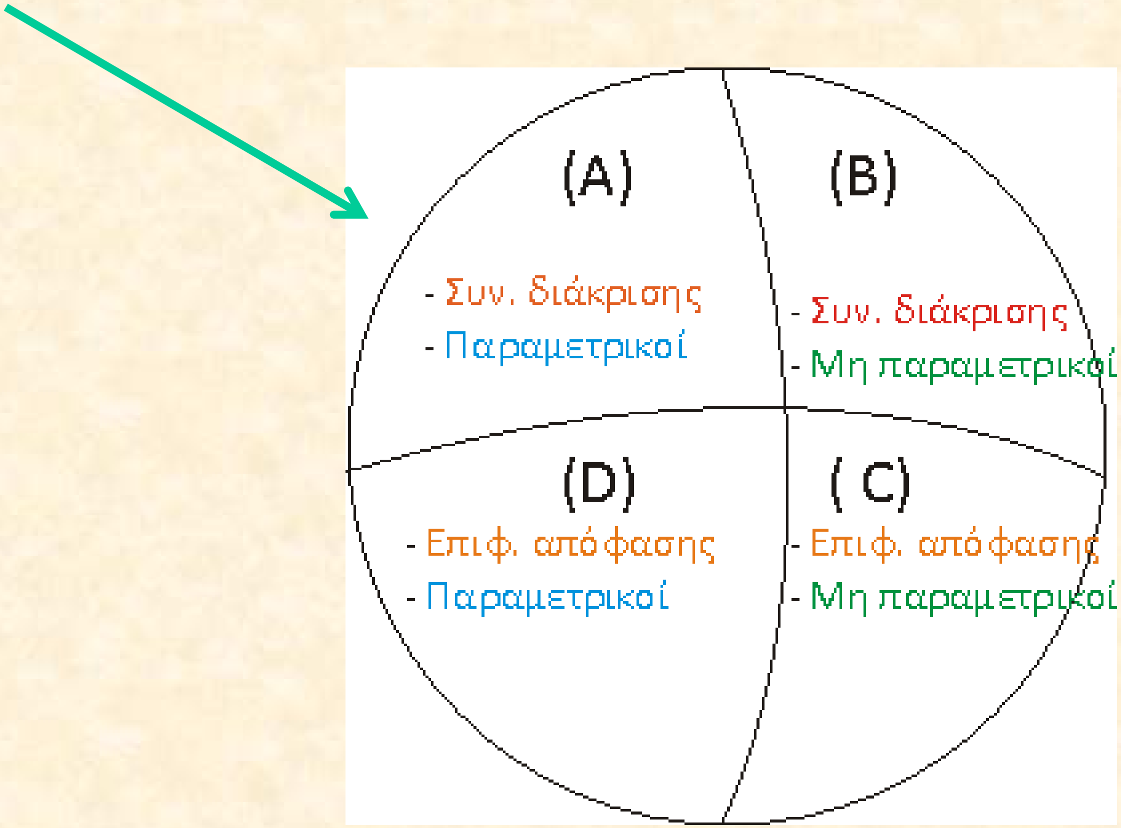
(*) Άλλη σχετική τεχνική είναι το k-fold cross validation.

Μια συνηθισμένη (και απλή) στρατηγική ταξινόμησης(*):

- Διαίρεσε το *Y* σε δύο σύνολα, **X** και **X’** N και N’ στοιχείων, αντιστοίχως.
- Χρησιμοποίησε το **X** (**σύνολο εκπαίδευσης**) για τη σχεδίαση του ταξινομητή
- Χρησιμοποίησε το **X’** (**σύνολο δοκιμής**) για την αξιολόγηση της απόδοσης του σχεδιασθέντος ταξινομητή
- Αν η απόδοση είναι ικανοποιητική, ο ταξινομητής χρησιμοποιείται στη συνέχεια σε ‘‘πραγματικό περιβάλλον‘‘. Διαφορετικά, θα πρέπει να αναθεωρηθούν κάποια βήματα της διαδικασίας σχεδιασμού.

Με βάση ποιο κριτήριο αξιολογείται η απόδοση του ταξινομητή;

Συνήθως με βάση το ποσοστό των εσφαλμένα ταξινομημένων διανυσμάτων του συνόλου δοκιμής X’ (που είναι μια εκτίμηση της πιθανότητας σφάλματος)


(A)
- Συν. διάκρισης
- Παραμετρικοί
(B)
- Συν. διάκρισης
- Μη παραμετρικοί
(D)
- Επιφ. απόφασης
- Παραμετρικοί
(C)
- Επιφ. απόφασης
- Μη παραμετρικοί

(Α) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Για **ΟΛΟΥΣ τους παραμετρικούς ταξινομητές**, $y=f(\boldsymbol{x};\boldsymbol{w})$, υπάρχουν δύο κύριες φάσεις

-> Καθορισμός του παραμετρικού μοντέλου

-> Εκτίμησητων παραμέτρων

Ο ταξινομητής Bayes με χρήση παραμετρικών μοντέλων

Ποιες είναι οι παράμετροι που εμπλέκονται στον ταξινομητή Bayes στην περίπτωση αυτή;

- Οι εκ των προτέρων πιθανότητες $P(\omega_j)$, j=1,…,M.
- Οι **άγνωστες** παράμετροι $\boldsymbol{\vartheta}_j$ των pdf των κλάσεων $p(\boldsymbol{x}|\omega_j;\boldsymbol{\vartheta}_j)$, j=1,…,$M$.

Παράδειγμα 1: Αν η $p(\boldsymbol{x}|\omega_j)$ **μοντελοποιείται** από μία κανονική κατανομή, $\mathrm{N}(\mu_j,\Sigma_j)$, και οι άγνωστες παράμετροι είναι το μέσο διάνυσμα και το μητρώο συνδιασποράς τότε $\boldsymbol{\vartheta}_j=\{\boldsymbol{\mu}_j, \Sigma_j\}$

Παράδειγμα 2: Αν η $p(\boldsymbol{x}|\omega_j)$ **μοντελοποιείται** από μία κανονική κατανομή, $\mathrm{N}(\mu_j,\Sigma_j)$, και η άγνωστη παράμετρος είναι το μέσο διάνυσμα, τότε $\boldsymbol{\vartheta}_j=\boldsymbol{\mu}_j$.

(A) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η περίπτωση του ταξινομητή Bayes

Στην πράξη

- οι τιμές των παραμέτρων είναι πολύ σπάνια γνωστές.
- Έτσι, χρειάζεται να τις **εκτιμήσουμε**.

Στην πράξη έχουμε στην διάθεσή μας ένα σύνολο δεδομένων (σύνολο εκπαίδευσης)

$$X = \{(x_i, d_i), x_i \in R^l, d_i \in \{1,2,\ldots,M\}, i = 1,\ldots N\}$$

όπου

$\boldsymbol{x}_i$ είναι το *i*-στό διάνυσμα εκπαίδευσης (training vector)

d_i είναι η ετικέτα της κλάσης όπου ανήκει το $\boldsymbol{x}_i$ (1 για την ω_1, 2 για την ω_2,…).

(Α) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η περίπτωση του ταξινομητή Bayes

Έστω

$$X_j = \{x_i \in R^l : d_i = j, i = 1,\ldots N_j\}, \qquad j = 1,\ldots,M$$

$$\boxed{N_1 + \cdots + N_M = N}$$

Το σύνολο που περιέχει τα διανύσματα δεδομένων της *j*-στης κλάσης.

<u>Εκτίμηση παραμέτρων</u> (βασίζεται στο σύνολο X)**:**

- Για κάθε $P(\omega_j)$:

  Χρησιμοποίησε την απλή προσέγγιση $P(\omega_j) \approx N_j / N$.
- Για κάθε $p(\boldsymbol{x}|\omega_j)$:
- Υπάρχουν αρκετές μέθοδοι
  - **Εκτίμηση μέγιστης πιθανοφάνειας** (**Maximum Likelihood** (ML) estimation),
  - **Εκτίμησης μέγιστης εκ των υστέρων πιθανότητας** (**Maximum a posteriori** (MAP) estimation),
  - Εκτίμηση μέγιστης εντροπίας (Maximum entropy (ME) estimation),
  - **Expectation-Maximization** (EM) estimation (mixture models) etc

Στη συνέχεια περιγράφουμε τις μεθόδους ML, MAP και (σύντομα) την μέθοδο EM.

(Α) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η περίπτωση του ταξινομητή Bayes

Η μέθοδος εκτίμησης μέγιστης πιθανοφάνειας (ML)

- Έστω $Y=\{ \boldsymbol{x}_1,\boldsymbol{x}_2,\ldots,\boldsymbol{x}_N\}$ ένα σύνολο από ανεξάρτητα διανύσματα δεδομένων
- Έστω $p(\boldsymbol{x})$ μια pdf γνωστής παραμετρικής μορφής αλλά με άγνωστο το διάνυσμα παράμετρων του $\boldsymbol{\vartheta}$; $p(\boldsymbol{x})=p(\boldsymbol{x};\boldsymbol{\vartheta})$.

  Παραδείγματα:

  Αν η $p(\boldsymbol{x};\boldsymbol{\vartheta})$ είναι κανονική με άγνωστο μέσο διάνυσμα $\boldsymbol{\mu}$, το $\boldsymbol{\vartheta}$ είναι απλά το $\boldsymbol{\mu}$.

  Αν η $p(\boldsymbol{x};\boldsymbol{\vartheta})$ είναι κανονική με άγνωστο μέσο διάνυσμα $\boldsymbol{\mu}$ και άγνωστο μητρώο συνδιασποράς Σ, το $\boldsymbol{\vartheta}$ περιέχει τις συνιστώσες τόσο του $\boldsymbol{\mu}$ όσο και του $\boldsymbol{\Sigma}$.

Το πρόβλημα: Εκτίμησε το $\boldsymbol{\vartheta}$ ώστε η $p(\boldsymbol{x};\boldsymbol{\vartheta})$ να είναι το **καλύτερο δυνατό ταίριασμα** για το σύνολο δεδομένων Y.

(Α) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η περίπτωση του ταξινομητή Bayes

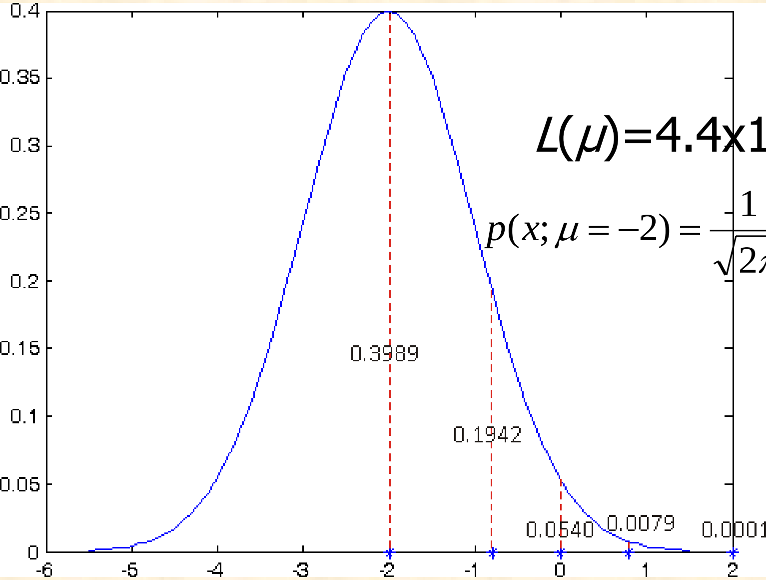
Η μέθοδος εκτίμησης μέγιστης πιθανοφάνειας (ML)

- Έστω η **συνάρτηση πιθανοφάνειας (likelihood function)** του $\boldsymbol{\vartheta}$ ως προς το *Y*:

$$p(Y;\vartheta) = p(x_1,\ldots,x_N;\vartheta) = \prod_{i=1}^{N} p(x_i;\vartheta)$$

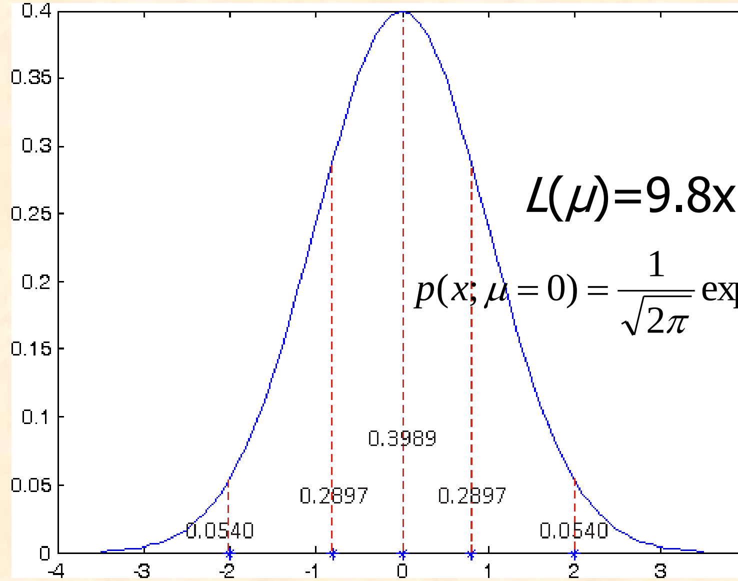
- ***Στην πράξη***, είναι πιο βολικό να μεγιστοποιήσουμε την συνάρτηση του **λογάριθμου της πιθανοφάνειας** (**log-likelihood function**) ***L*(*ϑ*)** του $\boldsymbol{\vartheta}$ ως προς το *Y*, η οποία ορίζεται ως

$$L(\vartheta) = \ln p(Y;\vartheta) = \ln\, p(x_1,\ldots,x_N;\vartheta) = \sum_{i=1}^{N} \ln p(x_i;\vartheta)$$



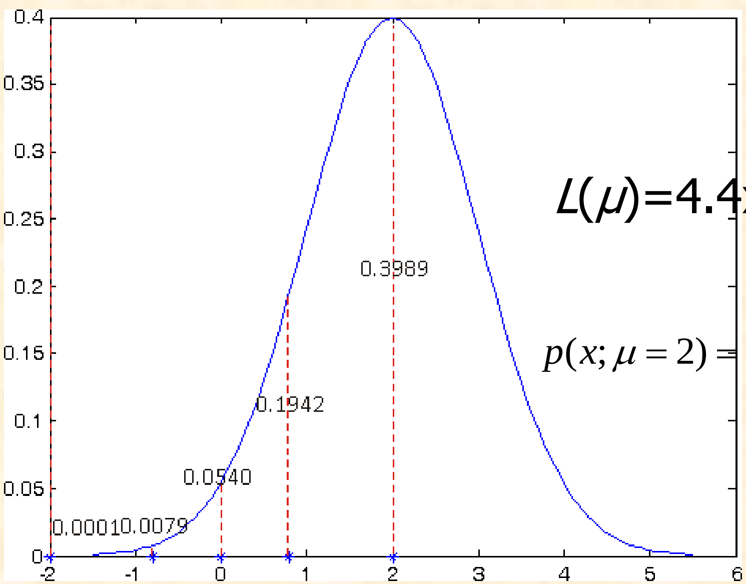
$L(\mu)=4.4\text{x}10^{-9}$

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



$L(\mu)=9.8\text{x}10^{-5}$

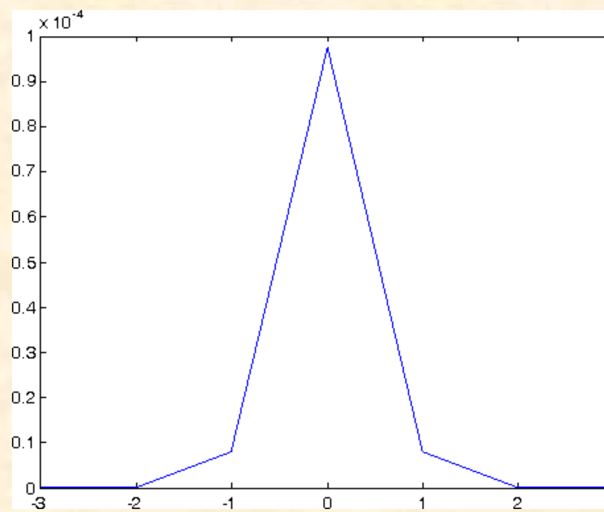
$$p(x;\mu=0)=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)$$



$L(\mu)=4.4\text{x}10^{-9}$

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

$L(\mu)$

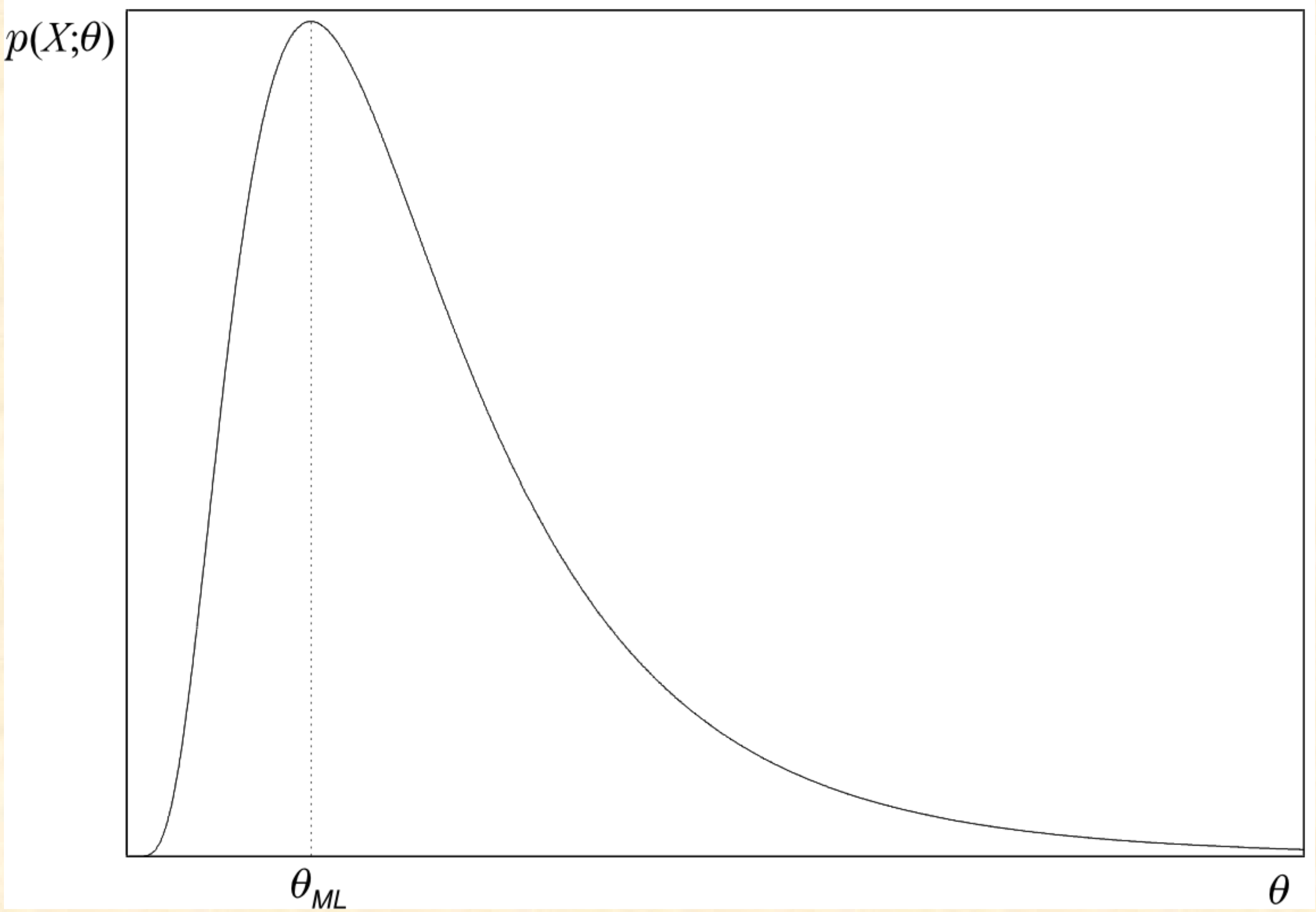


μ

(Α) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η περίπτωση του ταξινομητή Bayes

Η μέθοδος εκτίμησης μέγιστης πιθανοφάνειας (ML)



(Α) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η περίπτωση του ταξινομητή Bayes

Η μέθοδος εκτίμησης μέγιστης πιθανοφάνειας: **Μεγιστοποίησε** την $p(Y;\boldsymbol{\vartheta})$ (ή την $L(\boldsymbol{\vartheta})$) ως προς το $\boldsymbol{\vartheta}$.

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

(A) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η περίπτωση του ταξινομητή Bayes

Η μέθοδος εκτίμησης μέγιστης πιθανοφάνειας (ML) - Ένα παράδειγμα

-Έστω *Y* ένα σύνολο *N* διανυσμάτων δεδομένων, $\boldsymbol{x}_i$, i=1,…,*N*.

-Ποια είναι η κανονική κατανομή $p(\boldsymbol{x};\boldsymbol{\vartheta})$ γνωστού μητρώου συνδιασποράς που μοντελοποιεί καλύτερα τα διανύσματα του συνόλου *Y*?

Λύση:

-Το άγνωστο διάνυσμα παραμέτρων στην περίπτωση αυτή είναι το μέσο διάνυσμα $\boldsymbol{\mu}$, δηλ. $\boldsymbol{\vartheta}=\boldsymbol{\mu}$.

-Είναι

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

$$\ln p(x;\mu) = \ln\left(\frac{1}{(2\pi)^{l/2}|\Sigma|^{1/2}}\right) - \frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu) = C - \frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)$$

$$L(\mu) = \sum_{i=1}^{N} \ln p(x_i;\mu) = NC - \frac{1}{2}\sum_{i=1}^{N}(x_i-\mu)^T \Sigma^{-1}(x_i-\mu)$$

(Α) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η περίπτωση του ταξινομητή Bayes

Η μέθοδος εκτίμησης μέγιστης πιθανοφάνειας (ML) - Ένα παράδειγμα

-Θέτοντας την παράγωγο της $L(\boldsymbol{\mu})$ ως προς το $\boldsymbol{\mu}$ ίση με **0** έχουμε

$$\frac{\partial L(\mu)}{\partial \mu} = \frac{\partial}{\partial \mu}\left(NC - \sum_{i=1}^{N} \frac{1}{2}(x_i - \mu)^T \Sigma^{-1}(x_i - \mu) \right) = 0$$

$$\sum_{i=1}^{N} \Sigma^{-1}(x_i - \mu) = 0 \Leftrightarrow \sum_{i=1}^{N}(x_i - \mu) = 0 \Leftrightarrow \sum_{i=1}^{N} x_i - N\mu = 0$$

$$\mu_{ML} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

(Α) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η περίπτωση του ταξινομητή Bayes

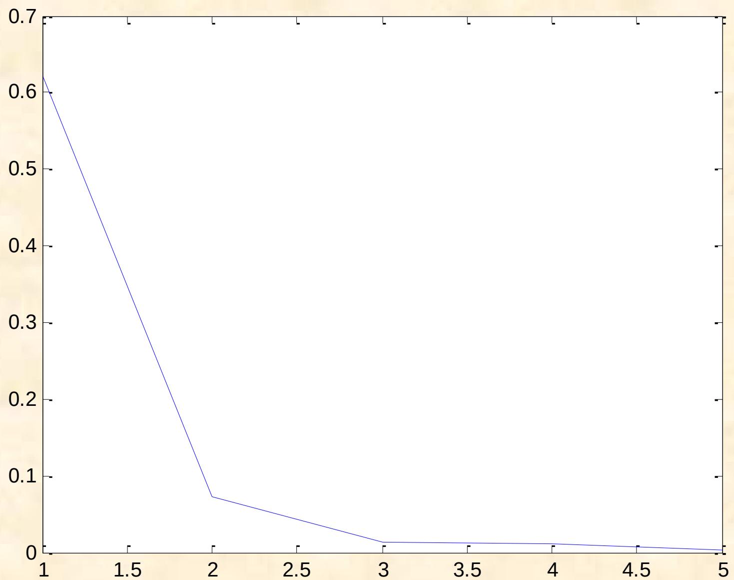
- Έστω N=10 σημεία που παρήχθησαν από τη μονοδιάστατη κανονική κατανομή μηδενικής μέσης τιμής και μοναδιαίας συνδασποράς, *N*(0,1).

- Ας **προσποιηθούμε** ότι έχουμε στη διάθεσή μας μόνο τα ακόλουθα:
  - Τη γνώση ότι η κατανομή που παρήγαγε τα σημεία είναι κανονική με διασπορά ίση με 1 και άγνωστο μέσο διάνυσμα μ.
  - τα 10 σημεία.

➢ Η **ML εκτίμηση** του μέσου διανύσματος, $\hat{\mu}$, της κατανομής είναι 0.6210 (η πραγματική τιμή είναι 0).

Όσο περισσότερα σημεία είναι διαθέσιμα, τόσο πιο ακριβείς εκτιμήσεις θα έχουμε για το μέσο διάνυσμα.

N	Σφάλμα εκτίμησης
10	0.6210
100	0.0727
1000	0.0138
10000	0.0118
100000	0.0034



(Α) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η περίπτωση του ταξινομητή Bayes

Η μέθοδος εκτίμησης μέγιστης πιθανοφάνειας (ML)

Αν πράγματι υπάρχει $\boldsymbol{\vartheta}_0$ έτσι ώστε $p(\boldsymbol{x})=p(\boldsymbol{x};\boldsymbol{\vartheta}_0)$ τότε

$$\lim_{N\to\infty} E[\vartheta_{ML}] = \vartheta_0$$

$$\lim_{N\to\infty} E\|\vartheta_{ML} - \vartheta_0\|^2 = 0$$

- Ασυμπτωτικά αμερόληπτος (Asymptotically unbiased) εκτιμητής:

Η μέση τιμή του $\boldsymbol{\vartheta}$, $\boldsymbol{\vartheta}_{ML}$, **τείνει προς την πραγματική τιμή** της άγνωστης παραμέτρου ($\boldsymbol{\vartheta}_0$), καθώς $N\to\infty$.

- Ασυμπτωτικά συνεπής (Asymptotically consistent) εκτιμητής:

Η διασπορά της ML εκτίμησης τείνει στο 0, καθώς $N\to\infty$.

(Α) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η περίπτωση του ταξινομητή Bayes

Η μέθοδος εκτίμησης μέγιστης πιθανοφάνειας (ML)

Άσκηση: Η τυχαία μεταβλητή x ακολουθεί την κατανομή Erlang

$$p(x;\vartheta) = \vartheta^2 x \exp(-\vartheta x) u(x)$$

με $u(x)=1$ αν $x>0$ και 0, διαφορετικά.

Δεδομένου ενός συνόλου δειγμάτων $x_1,\ldots,x_N$, του x, να δείξετε ότι η εκτίμηση ML του ϑ είναι

$$\vartheta_{ML} = \frac{2N}{\sum_{k=1}^{N} x_k}$$

(Α) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η περίπτωση του ταξινομητή Bayes

Εκτίμηση μέγιστης εκ των υστέρων πιθανότητας (Maximum Aposteriori Probability (MAP) Estimation)

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

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

Υπολόγισε το μέγιστο της

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

- From Bayes theorem

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

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

Εκτίμηση μέγιστης εκ των υστέρων πιθανότητας (MAP)

- Η μέθοδος:

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

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

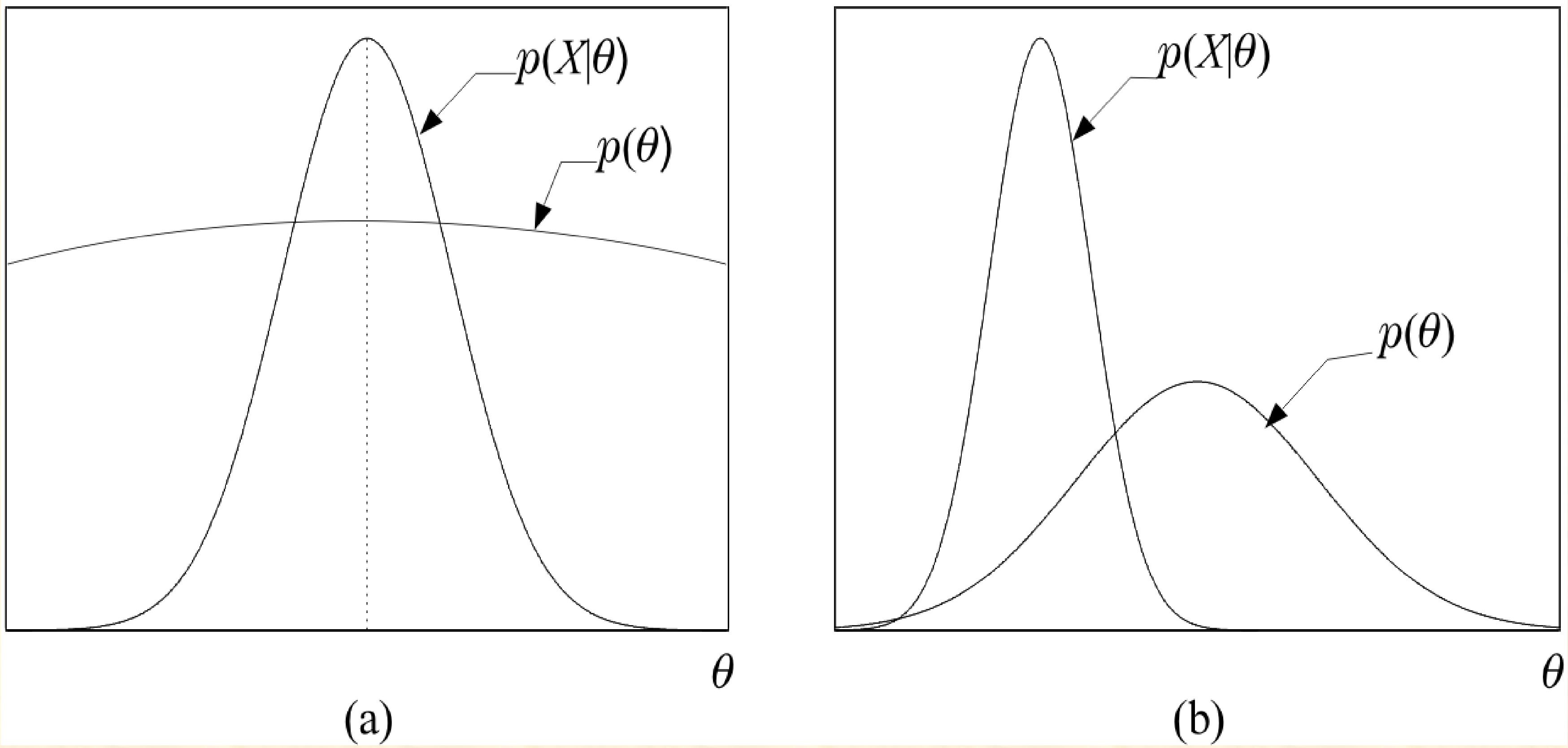
Αν η $p(\underline{\theta})$ είναι ομοιόμορφη ή αρκετά ευρεία:

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

(Α) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η περίπτωση του ταξινομητή Bayes

Εκτίμηση μέγιστης εκ των υστέρων πιθανότητας (MAP)



(Α) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η περίπτωση του ταξινομητή Bayes

Εκτίμηση μέγιστης εκ των υστέρων πιθανότητας (MAP) - Παράδειγμα

$$p(\underline{x}): \; N(\underline{\mu}, \Sigma), \quad \underline{\mu} \;\; \acute{\alpha}\gamma\nu\omega\sigma\tau o, \quad X = \{\underline{x}_1, \ldots, \underline{x}_N\}$$

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

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

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

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

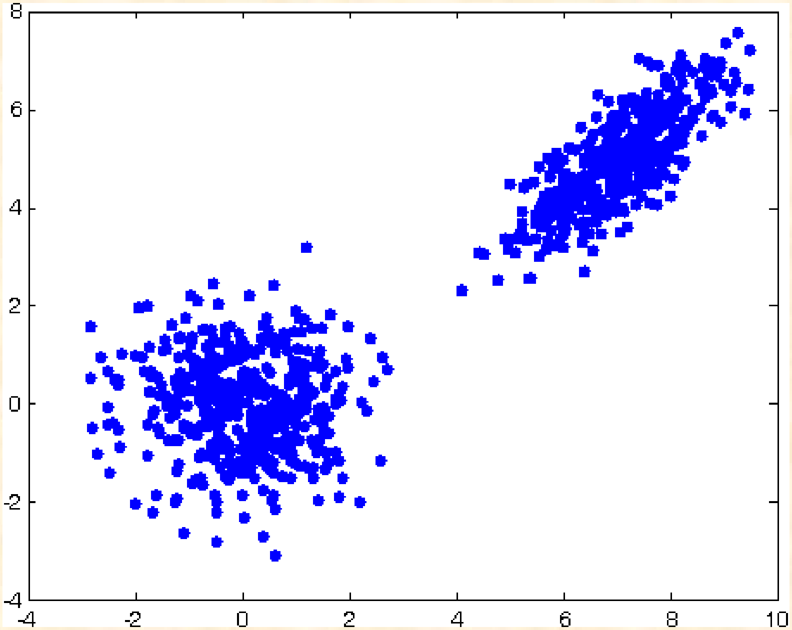
(Α) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η περίπτωση του ταξινομητή Bayes

Η μέθοδος αναμονής-μεγιστοποίησης (expectation-maximization – EM) (μοντέλα μίξης - mixture models)

Μοντέλο μίξης: Σταθμισμένο άθροισμα pdfs γνωστής παραμετρικής μορφής

$$p(x)=\sum_{k=1}^{K} P_k\, p(x\,|\,k), \qquad \sum_{k=1}^{K} P_k = 1, \qquad \int_{-\infty}^{+\infty} p(x\,|\,k) = 1$$



❖Η μέθοδος ML δεν μπορεί να αξιοποιηθεί εδώ εξαιτίας των *ετικετών k*, που είναι επίσης άγνωστες. **Λύση:** Ο αλγόριθμος **EM**.

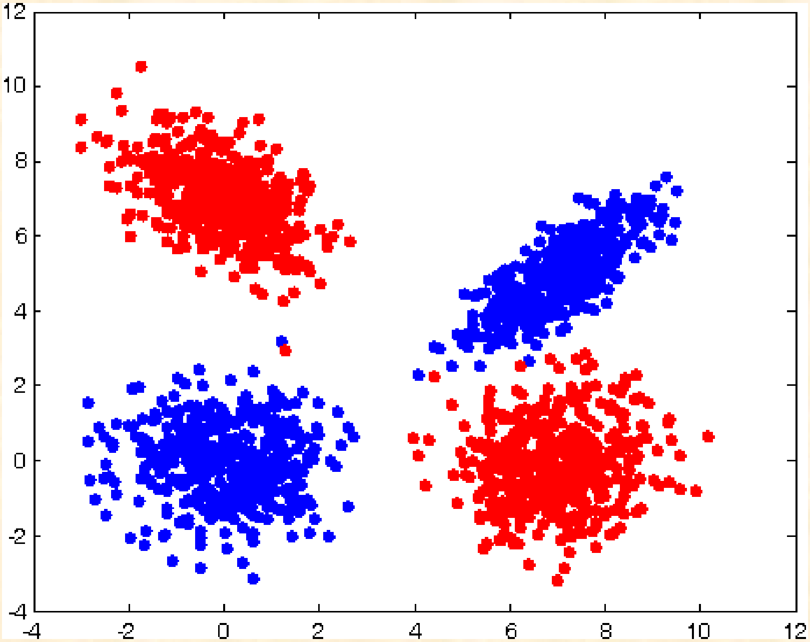
Συμβολισμοί:

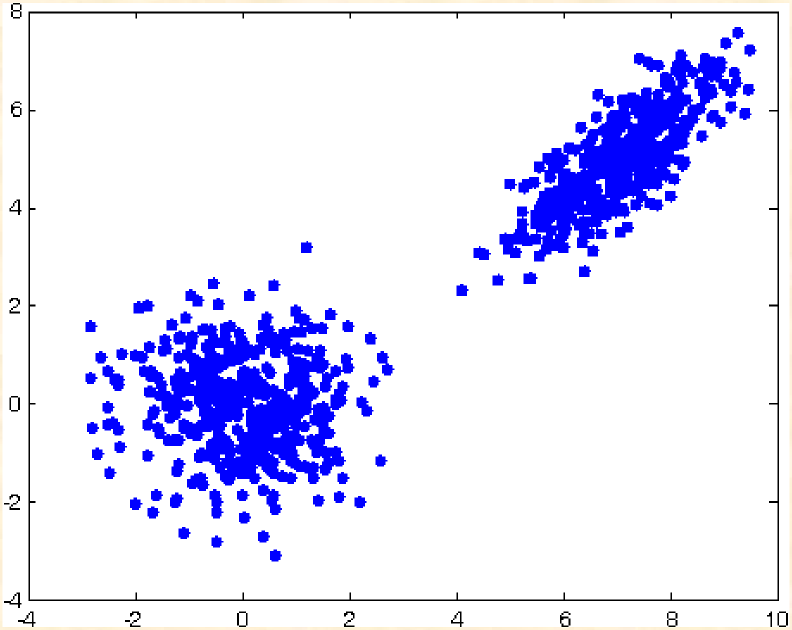
- $p(\boldsymbol{x}|k) = p(\boldsymbol{x}|k;\boldsymbol{\vartheta}_k)$
- $\boldsymbol{\Theta}=(\boldsymbol{\vartheta}_1,\ldots,\boldsymbol{\vartheta}_K)$, $\boldsymbol{P}=[P_1,\ldots,P_k]^T$, $\boldsymbol{\Xi}=(\boldsymbol{\Theta},\boldsymbol{P})$.
- $X=\{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N\}$: μη πλήρες (incomplete) (παρατηρούμενο) σύνολο δεδομένων.
- $X^c=\{(\boldsymbol{x}_1,k_1),\ldots,(\boldsymbol{x}_N,k_N)\}$: πλήρες (complete) σύνολο δεδομένων

(A) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η περίπτωση του ταξινομητή Bayes

Η μέθοδος EM (μοντέλα μίξης)





(Α) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η περίπτωση του ταξινομητή Bayes

Η μέθοδος EM (μοντέλα μίξης)

Στόχος: Εκτίμηση των ***Θ*** και ***P*** μέσω της ελαχιστοποίησης της λογαριθμικής συνάρτησης πιθανοφάνειας (log-likelihood) του πλήρους συν. δεδομένων.

$$\ln p(X^c;\Theta,P)=\sum_{n=1}^{N}\ln p(x_n,k_n;\vartheta_{k_n})=\sum_{n=1}^{N}\ln\Big(p(x_n\,|\,k_n;\vartheta_{k_n})P_{k_n}\Big)$$

Πρόβλημα: Για κάθε ***x*** είναι άγνωστη η επιμέρους κατανομή της μίξης από την οποία προήλθε.

Λύση: Μεγιστοποίηση της μέσης τιμής της log-likelihood ως προς *P*(k_n|***x***$_n$;*Ξ*)

$$E\left[\sum_{n=1}^{N}\ln\Big(p(x_n\,|\,k_n;\vartheta_{k_n})P_{k_n}\Big)\right]_{P(k_n|x_n;\Xi)}=\sum_{n=1}^{N}E\Big(\ln\Big(p(x_n\,|\,k_n;\vartheta_{k_n})P_{k_n}\Big)\Big)_{P(k_n|x_n;\Xi)}$$

$$\sum_{n=1}^{N}\sum_{k_n=1}^{K}\ln\Big(p(x_n\,|\,k_n;\vartheta_{k_n})P_{k_n}\Big)P(k_n\,|\,x_n;\Xi)=\sum_{n=1}^{N}\sum_{k=1}^{K}\ln\Big(p(x_n\,|\,k;\vartheta_k)P_k\Big)P(k\,|\,x_n;\Xi$$

(A) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η περίπτωση του ταξινομητή Bayes - Η μέθοδος EM (μοντέλα μίξης)

$$\sum_{n=1}^{N}\sum_{k=1}^{k}\ln\left(p(x_n \mid k_n;\vartheta_{k_n})P_{k_n}\right)P(k_n \mid x_n;\Xi)=\sum_{n=1}^{N}\sum_{k=1}^{K}\ln\left(p(x_n \mid k;\vartheta_k)P_k\right)P(k \mid x_n;\Xi)$$

Επιπλέον πρόβλημα: Οι ποσότητες $P(k|\mathbf{x}_n;\Xi)$ είναι άγνωστες.

Λύση: Δεδομένου ότι από τον κανόνα του Bayes είναι

$$P(k \mid x_n;\Xi)=\frac{p(x_n \mid k;\vartheta_k)P_k}{\sum_{j=1}^{k}p(x_n \mid j;\vartheta_j)P_j} \qquad \textbf{(A)}$$

Η λύση είναι ένας αναδρομικός αλγόριθμος.

Αρχικοποιώντας το $\boldsymbol{\Xi}=(\boldsymbol{\Theta},\boldsymbol{P})$ στις τιμές $\boldsymbol{\Xi}(0)=(\boldsymbol{\Theta}(0),\boldsymbol{P}(0))$

Εκτιμούμε τις $P(k|\mathbf{x}_k;\Xi(0))$ από την (A)

(**E-step**) Υπολογίζουμε τη μέση τιμή της *expected log-likelihood* με βάση τις $P(k|\mathbf{x}_k;\Xi(0))$

$$Q(\Xi \mid \Xi(0))=\sum_{n=1}^{N}\sum_{k=1}^{K}\ln\left(p(x_n \mid k;\vartheta_k)P_k\right)P(k \mid x_n;\Xi(0))$$

(**M-step**) Μεγιστοποιούμε την $Q(\Xi|\Xi(0))$ ως προς τις παραμέτρους Ξ.

$$\partial Q(\Xi \mid \Xi(0))\big/\partial\vartheta_k=0,\qquad \partial Q(\Xi \mid \Xi(0))/\partial P_k=0$$

(Α) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η περίπτωση του ταξινομητή Bayes - Η μέθοδος ΕΜ (μοντέλα μίξης)

Ο αλγόριθμος ΕΜ

- Αρχικοποίησε το $\boldsymbol{\Xi}=(\boldsymbol{\Theta},\boldsymbol{P})$ στις τιμές $\boldsymbol{\Xi}(0)=(\boldsymbol{\Theta}(0),\boldsymbol{P}(0))$
- t=0
- Επανέλαβε
  - Εκτίμησε τις $P(k|\mathbf{x}_k;\Xi(\mathrm{t}))$ από την (Α)
  - (**E-step**) Υπολόγισε τη μέση τιμή της *expected log-likelihood* με βάση τις $P(k|\mathbf{x}_k;\Xi(\mathrm{t}))$

$$Q(\Xi\,|\,\Xi(t))=\sum_{n=1}^{N}\sum_{k=1}^{K}\ln\left(p(x_n\,|\,k;\vartheta_k)P_k\right)P(k\,|\,x_n;\Xi(t))$$

  - (**M-step**) Εκτίμησε το $\Xi(t+1)$ μεγιστοποιώντας την $Q(\Xi|\Xi(t))$ ως προς τις παραμέτρους $\boldsymbol{\Xi}=(\boldsymbol{\Theta},\boldsymbol{P}) = ((\boldsymbol{\vartheta}_1,\ldots\boldsymbol{\vartheta}_k), (P_1,\ldots,P_k))$.

$$\partial Q(\Xi\,|\,\Xi(t))\Big/\partial\vartheta_k=0,\qquad \partial Q(\Xi\,|\,\Xi(t))/\partial P_k=0$$

  - t=t+1
- Έως ότου επιτευχθεί σύγκλιση.

(Α) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η περίπτωση του ταξινομητή Bayes

Η μέθοδος ΕΜ (μοντέλα μίξης)

Μίξη κανονικών κατανομών (Mixtures of Gaussians).

$$p(x) = \sum_{k=1}^{K} P_k \, p(x \mid k; \mu_k, \Sigma_k),$$

$$p(x \mid k; \mu_k, \Sigma_k) = \frac{1}{(2\pi)^{l/2} \mid \Sigma_k \mid^{1/2}} \exp\left(-0.5 \cdot (x-\mu_k)^T \Sigma_k^{-1} (x-\mu_k)\right)$$

Ο αλγόριθμος **ΕΜ** για την περίπτωση **μίξης κανονικών κατανομών**.

$$Q(\Xi \mid \Xi(t)) =$$

$$\sum_{n=1}^{N} \sum_{k=1}^{K} P(k \mid x_n; \Xi(t)) \left(-0.5 \ln \mid \Sigma_k \mid -0.5 (x-\mu_k)^T \Sigma_k^{-1} (x-\mu_k) + \ln P_k\right)$$

(Α) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η περίπτωση του ταξινομητή Bayes

❖Ο αλγόριθμος **EM** για την περίπτωση **μίξης κανονικών κατανομών**.

- Αρχικοποίησε $\boldsymbol{\mu}_k=\boldsymbol{\mu}_k^{(0)}$, $\Sigma_k=\Sigma_k^{(0)}$, $\boldsymbol{P}=\boldsymbol{P}^{(0)}$
- t=0
- Επανέλαβε

$$P\left(k \mid x_n; \Theta^{(t)}, P^{(t)}\right) = \frac{p\left(x_n \mid k; \vartheta_k^{(t)}\right) P_k^{(t)}}{\sum_{q=1}^{K} p\left(x_n \mid q; \vartheta_q^{(t)}\right) P_q^{(t)}} \equiv \gamma_{kn}^{(t)}$$

$$\mu_k^{(t+1)} = \sum_{n=1}^{N} \gamma_{kn}^{(t)} x_n \Big/ \sum_{n=1}^{N} \gamma_{kn}^{(t)}$$

$$\Sigma_k^{(t+1)} = \sum_{n=1}^{N} \gamma_{kn}^{(t)} (x_n - \mu_k^{(t+1)})(x_n - \mu_k^{(t+1)})^T \Big/ \sum_{n=1}^{N} \gamma_{kn}^{(t)}$$

$$P_k^{(t+1)} = \frac{1}{N} \sum_{n=1}^{N} \gamma_{kn}^{(t)}$$

$t=t+1$

- Έως ότου ικανοποιηθεί ένα κατάλληλο κριτήριο τερματισμού

(Α) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Ανασκόπηση του (προσεγγιστικού) ταξινομητή Bayes

- Available data: $X = X_1 \cup X_2 \cup \ldots X_M$. Each X_j corresponds to class ω_j.

Σχεδιασμός ταξινομητή

- **Υιοθέτησε** ένα μοντέλο pdf για κάθε ω_j, με άγνωστες παραμέτρους $\boldsymbol{\vartheta}_j$.
- **Εφάρμοσε** την ML (ή τις MAP, EM) μεθόδους ***M* φορές** (μία για κάθε κλάση) για την εκτίμηση των $\boldsymbol{\vartheta}_j$'s, με βάση τα αντίστοιχα σύνολα X_j,

$$\hat{p}(x \mid \omega_j) \equiv p(x \mid \omega_j; \hat{\vartheta}_j)$$

-**Προσέγγισε** τις $P(\omega_j)$ ως εξής

$$\hat{P}(\omega_j) = N_i / N$$

-**Όρισε**

$$g_j(x) = \hat{P}(\omega_j)\hat{p}(x \mid \omega_j)$$

Χρήση ταξινομητή

-Για δεδομένο $\boldsymbol{x}$,

-Υπολόγισε τις τιμές των $g_j(\boldsymbol{x})$, $j=1,\ldots,M$.

-Καταχώρησε το $\boldsymbol{x}$ στην κλάση k με $g_k(\boldsymbol{x}) = \max_{j=1,\ldots,M} g_j(\boldsymbol{x})$.

Ένα σημαντικό ζήτημα – Η κατάρα της διαστατικότητας (Curse of dimensionality)

– Σε όλες τις μεθόδους που εξετάσαμε μέχρι στιγμής είδαμε ότι όσο μεγαλύτερος είναι ο αριθμός των σημείων, N, τόσο καλύτερες είναι και οι προκύπτουσες εκτιμήσεις.

– Αν στο μονοδιάστατο χώρο για ένα διάστημα μήκους L, N σημεία είναι αρκετά για να δώσουν μια καλή εκτίμηση των εμπλεκόμενων παραμέτρων, στο διδιάστατο χώρο για μια επιφάνεια διαστάσεων LxL απαιτούνται N2 σημεία για να πάρουμε καλές εκτιμήσεις και, γενικά, στον l-διάστατο χώρο για έναν υπερκύβο διαστάσεων Ll απαιτούνται Nl σημέια.

– Η εκθετική αύξηση του αριθμού των σημείων που απαιτούνται για μια καλή εκτίμηση των εμπλεκόμενων παραμέτρων με τη διάσταση του χώρου είναι γνωστή ως κατάρα της διαστατικότητας (curse of dimensionality). Πρόκειται για ένα σημαντικό πρόβλημα που απαντάται σε προβλήματα που απεικονίζονται σε χώρους υψηλής διαστατικότητας.

(Α) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Ο απλοϊκός ταξινομητής Bayes (naive Bayes classifier)

- Διαθέσιμα δεδομένα: $X=X_1 \cup X_2 \cup \ldots X_M$. Κάθε X_j αντιστοιχεί στην κλάση ω_j. Έστω $X_j=X_{j1}$ x…x X_{jl}

-**Υπόθεση:** Όλα τα **χαρακτηριστικά (features)** είναι μεταξύ τους **στατιστικώς ανεξάρτητα**. Έτσι

$$p(x \mid \omega_j) = \prod_{k=1}^{l} p(x_k \mid \omega_j), \qquad j = 1, \ldots M$$

-Ως εκ τούτου, μπορούμε για κάθε κλάση να εργαστούμε για κάθε $p(\boldsymbol{x}_k \mid \omega_j)$ ανεξάρτητα από τις υπόλοιπες $p(\boldsymbol{x}_q \mid \omega_j)$, q=1,…,l, q ≠k.

(Α) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Ο απλοϊκός ταξινομητής Bayes (naive Bayes classifier)

Σχεδιασμός ταξινομητή

- **Υιοθέτησε** ένα μοντέλο pdf για κάθε ω_j, με άγνωστες παραμέτρους $\boldsymbol{\vartheta}_j$.
- **Εφάρμοσε** την ML μέθοδο ***M l* φορές** (***l*** φορές για κάθε κλάση) για την εκτίμηση των $\boldsymbol{\vartheta}_{jk}$'s των μονοδιάστατων $p(\boldsymbol{x}_k|\omega_j)$, με βάση τα αντίστοιχα X_{jk}

-**Προσέγγισε** τις $P(\omega_j)$ ως ακολούθως

$$\hat{P}(\omega_j) = N_i / N$$

-**Όρισε**

$$g_j(x) = \hat{P}(\omega_j)\hat{p}(x|\omega_j) = \hat{P}(\omega_j)\prod_{k=1}^{l}\hat{p}(x_k|\omega_j)$$

Χρήση ταξινομητή

-Για δεδομένο $\boldsymbol{x}$,

-Υπολόγισε τις τιμές των $g_j(\boldsymbol{x})$, j=1,…,M.

-Καταχώρησε το $\boldsymbol{x}$ στην κλάση k με $g_k(\boldsymbol{x})=\max_{j=1,\dots,M} g_j(\boldsymbol{x})$.

(Α) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Γραμμικοί ταξινομητές – Ο ταξινομητής ελάχιστης Ευκλείδειας απόστασης

- Διαθέσιμα δεδομένα: $X=X_1 \cup X_2 \cup \ldots X_M$. Κάθε X_j αντιστοιχεί στην κλάση ω_j.

Σχεδιασμός ταξινομητή

- **Εκτίμησε** το **μέσο διάνυσμα** , m_j, κάθε κλάσης ω_j, με βάση το X_j , j=1,…,N_j.

(συνήθως, η μέθοδος ML χρησιμοποιείται για κάθε κλάση, με την **υπόθεση** ότι κάθε κλάση μοντελοποιείται από μια κανονική κατανομή **αγνώστου μέσου διανύσματος** και το m_j τίθεται ίσο με την προκύπτουσα εκτίμηση του μέσου διανύσματος)

-**Όρισε**

$$g_j(\boldsymbol{x})=-||\mathbf{x}-\boldsymbol{m}_j||^2$$

Χρήση ταξινομητή

-Για δεδομένο $\boldsymbol{x}$,

-Υπολόγισε τις τιμές των $g_j(\boldsymbol{x})$, j=1,…,M.

-Καταχώρησε το $\boldsymbol{x}$ στην κλάση k με $g_k(\boldsymbol{x})=\max_{j=1,\ldots,M} g_j(\boldsymbol{x})$.

ΣΗΜ: Γενικά, ο ταξινομητής ελάχιστης Ευκλείδειας απόστασης δεν είναι βέλτιστος ως προς το κριτήριο της πιθανότητας λάθους (κάτω από ποιές συνθήκες ο ταξινομητής αυτός θα μπορούσε να είναι βέλτιστος;)

(A) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Γραμμικοί ταξινομητές – Ο ταξινομητής ελάχιστης Mahalanobis απόστασης

- Διαθέσιμα δεδομένα: $X=X_1 \cup X_2 \cup \ldots X_M$. Κάθε X_j αντιστοιχεί στην κλάση ω_j.

Σχεδιασμός ταξινομητή

- **Εκτίμησε** το **μέσο διάνυσμα** , m_j, και το μητρώο συνδιασποράς, S_j, κάθε κλάσης ω_j, με βάση το X_j , j=1,…,N_j.

(συνήθως, η μέθοδος ML χρησιμοποιείται για κάθε κλάση, με την **υπόθεση** ότι κάθε κλάση μοντελοποιείται από μια κανονική κατανομή **αγνώστου μέσου διανύσματος** και **μητρώου συνδιασποράς** και τα m_j και S_j, τίθενται ίσα με τις προκύπτουσες εκτιμήσεις (το κοινό μητρώο συνδιασποράς μπορεί να τεθεί ίσο με το μέσο των εκτιμήσεων για τα S_j)

-**Όρισε**

$$g_j(\boldsymbol{x})=-(\boldsymbol{x}-\boldsymbol{m}_j)^T S^{-1}(\boldsymbol{x}-\boldsymbol{m}_j)$$

Χρήση ταξινομητή

-Για δεδομένο $\boldsymbol{x}$,

-Υπολόγισε τις τιμές των $g_j(\boldsymbol{x})$, j=1,…,M.

-Καταχώρησε το $\boldsymbol{x}$ στην κλάση k με $g_k(\boldsymbol{x})=\max_{j=1,\ldots,M} g_j(\boldsymbol{x})$.

ΣΗΜ: Γενικά, ο ταξινομητής ελάχιστης Mahalanobis απόστασης δεν είναι βέλτιστος ως προς το κριτήριο της πιθανότητας λάθους (κάτω από ποιές συνθήκες ο ταξινομητής αυτός θα μπορούσε να είναι βέλτιστος;)

(Α) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Τετραγωνικοί ταξινομητές

- Διαθέσιμα δεδομένα: $X = X_1 \cup X_2 \cup \ldots X_M$. Κάθε X_j αντιστοιχεί στην κλάση ω_j.

Σχεδιασμός ταξινομητή

- **Εκτίμησε** το **μέσο διάνυσμα** , m_j, και το μητρώο συνδιασποράς, S_j, κάθε κλάσης ω_j, με βάση το X_j , j=1,…,N_j.

(συνήθως, η μέθοδος ML χρησιμοποιείται για κάθε κλάση, με την **υπόθεση** ότι κάθε κλάση μοντελοποιείται από μια κανονική κατανομή **αγνώστου μέσου διανύσματος** και **μητρώου συνδιασποράς** και τα m_j και S_j, τίθενται ίσα με τις προκύπτουσες εκτιμήσεις

-**Όρισε**

$$g_j(\boldsymbol{x}) = -(\boldsymbol{x}-\boldsymbol{m}_j)^T S_j^{-1}(\boldsymbol{x}-\boldsymbol{m}_j)$$

Χρήση ταξινομητή

-Για δεδομένο $\boldsymbol{x}$,

-Υπολόγισε τις τιμές των $g_j(\boldsymbol{x})$, j=1,…,M.

-Καταχώρησε το $\boldsymbol{x}$ στην κλάση k με $g_k(\boldsymbol{x}) = \max_{j=1,\ldots,M} g_j(\boldsymbol{x})$.

ΣΗΜ: Γενικά, ο τετραγωνικός ταξινομητής δεν είναι βέλτιστος ως προς το κριτήριο της πιθανότητας λάθους (κάτω από ποιές συνθήκες ο ταξινομητής αυτός θα μπορούσε να είναι βέλτιστος;)