❖ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ
❖(PATTERN RECOGNITION)

Σέργιος Θεοδωρίδης

Κωνσταντίνος Κουτρούμπας

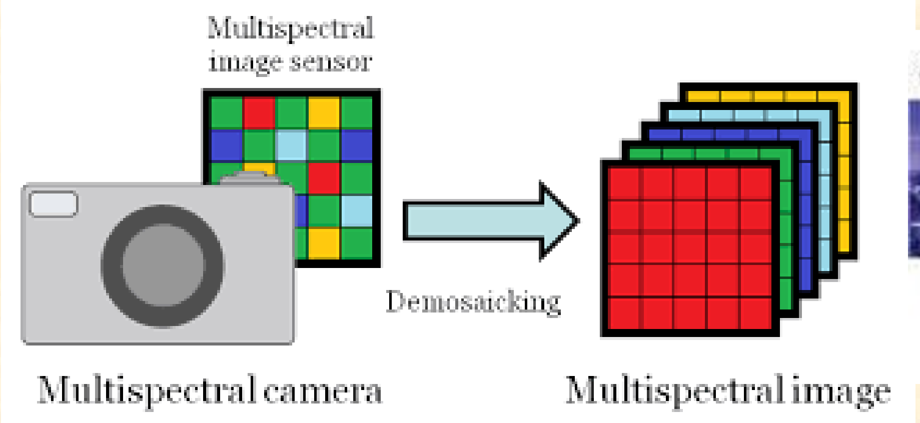
Version 2

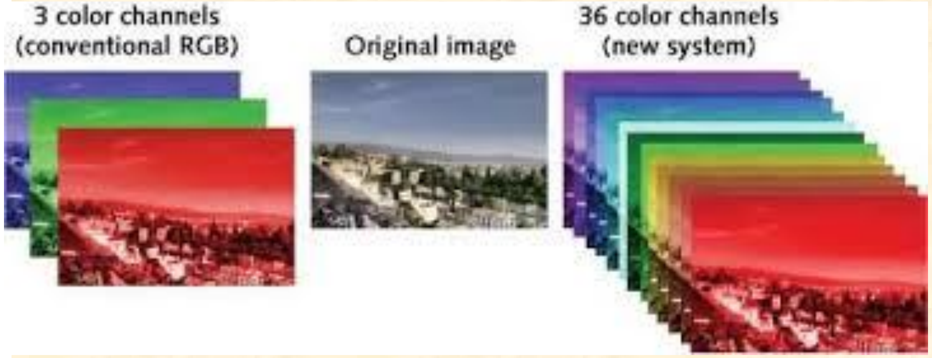
❖ΔΕΔΟΜΕΝΑ

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

❖Μερικά παραδείγματα είναι:

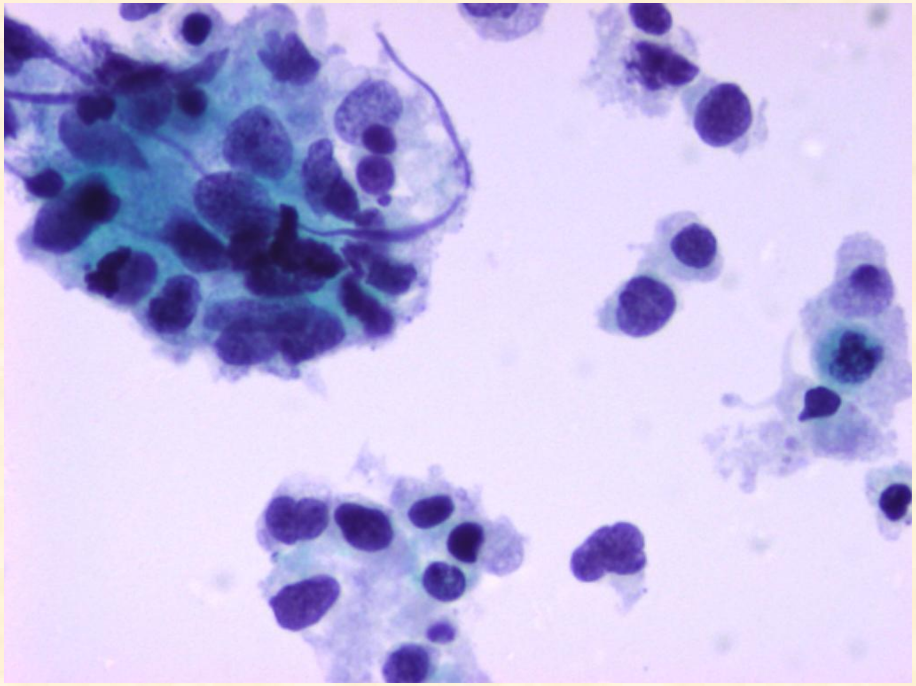
❖**(A)** Εικόνες (εικόνες κλίμακας του γκρι (grayscale), πολυφασματικές (multispectral), υπερφασματικές (hyperspectral)) που λαμβάνονται από κατάλληλα συστήματα απεικόνισης (imagers).

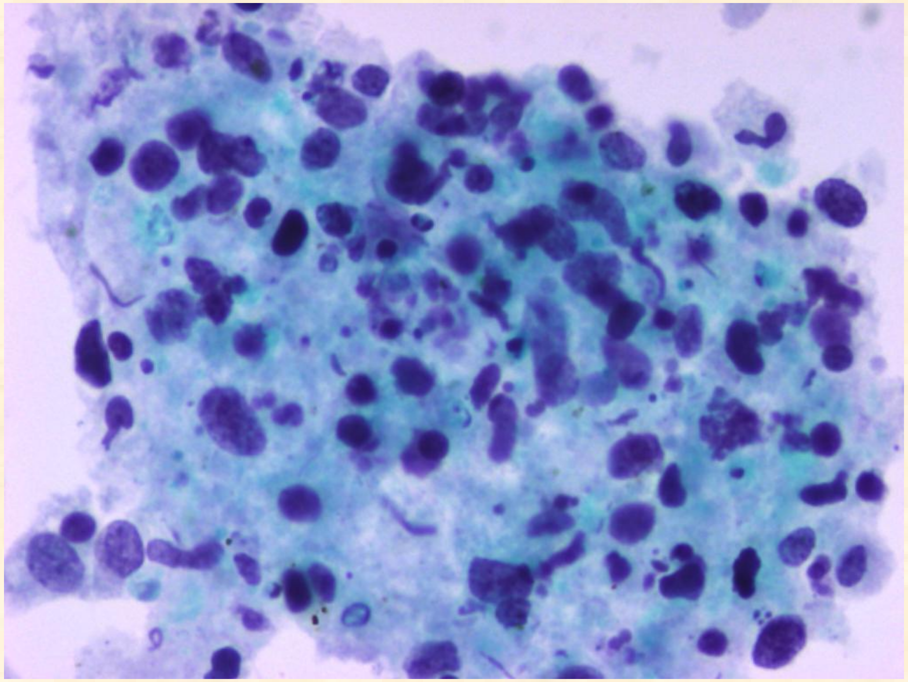




❖ΔΕΔΟΜΕΝΑ

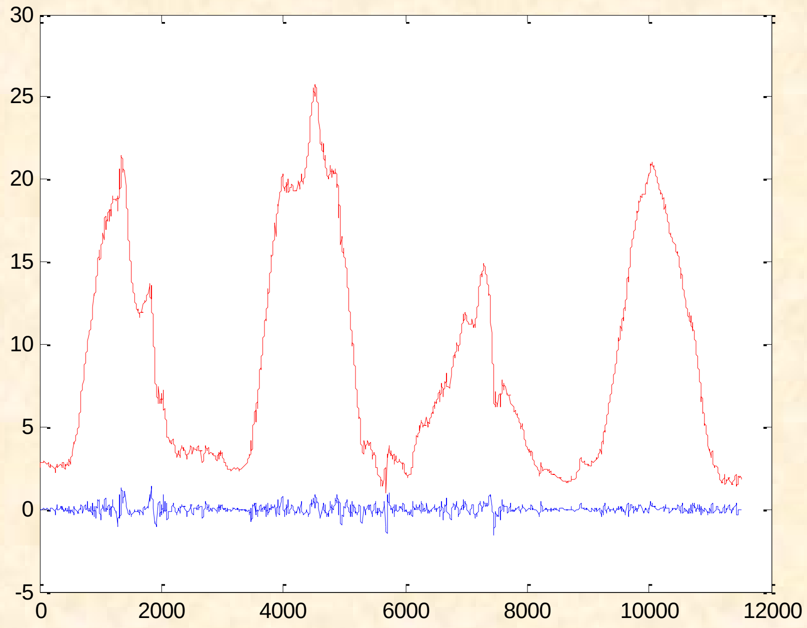
❖**(Β)** Ανθρώπινα γαστρικά κύτταρα.





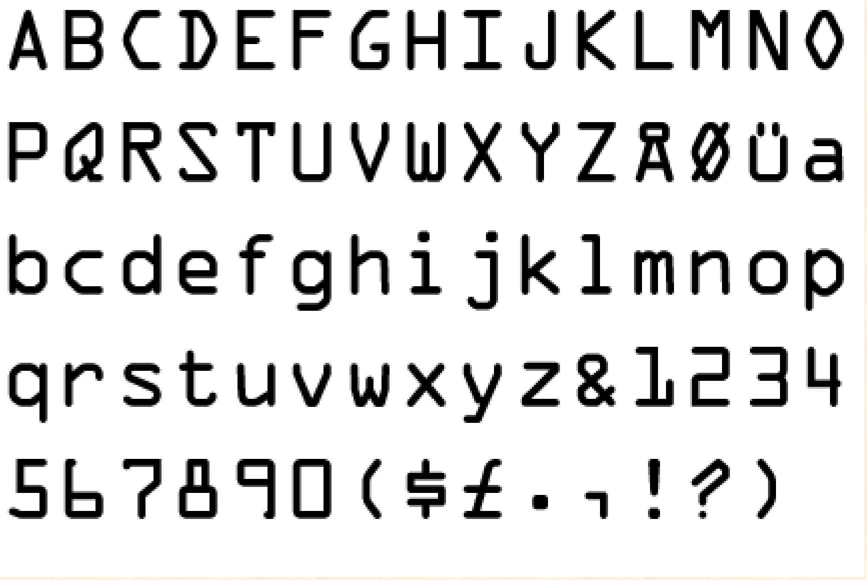
❖ΔΕΔΟΜΕΝΑ

❖**(C)** Χρονοσειρές (Time series).



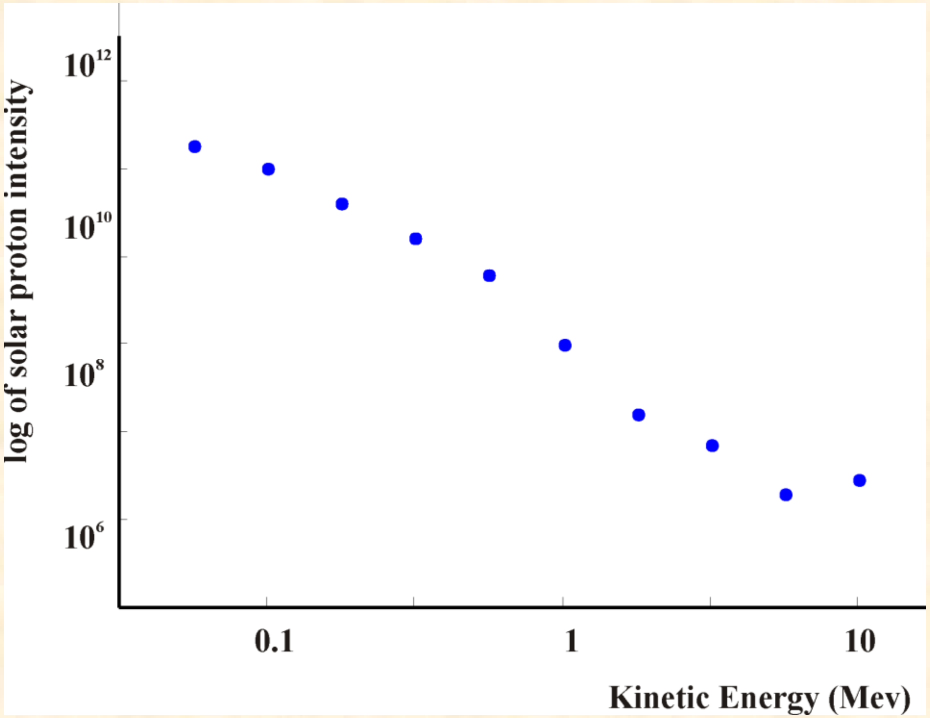
❖ΔΕΔΟΜΕΝΑ

❖**(D)** Κείμενο



❖ΔΕΔΟΜΕΝΑ

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



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

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

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

❖Στο τρέχον πλαίσιο ο 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.

❖Στη συνέχεια ασχολούμαστε μόνο με **ταξινόμηση** και **ομαδοποίηση**.

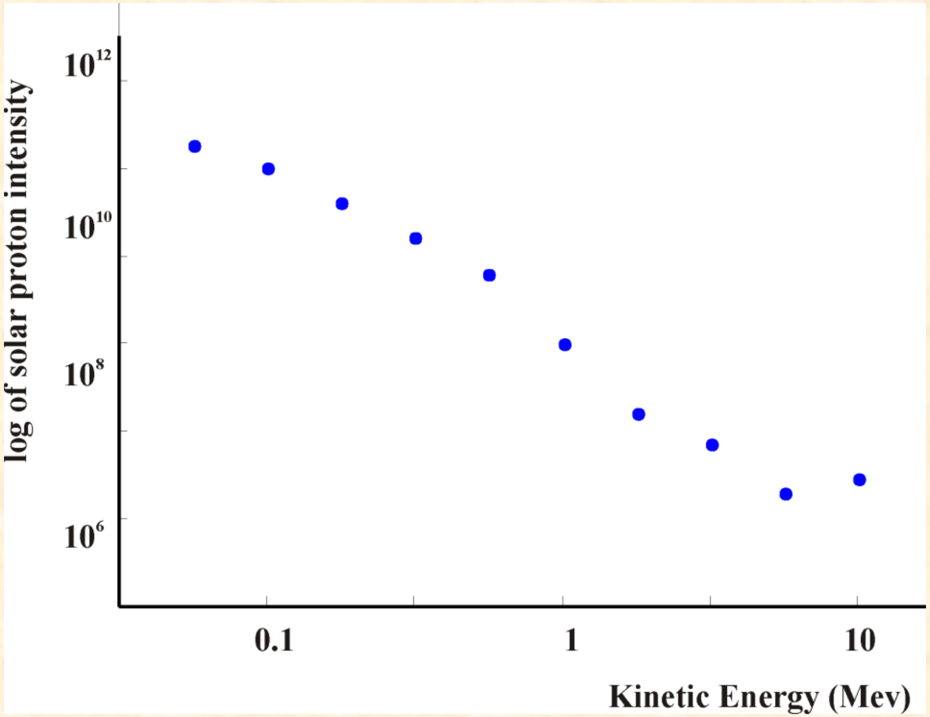
❖ΕΞΑΓΩΓΗ ΓΝΩΣΗΣ

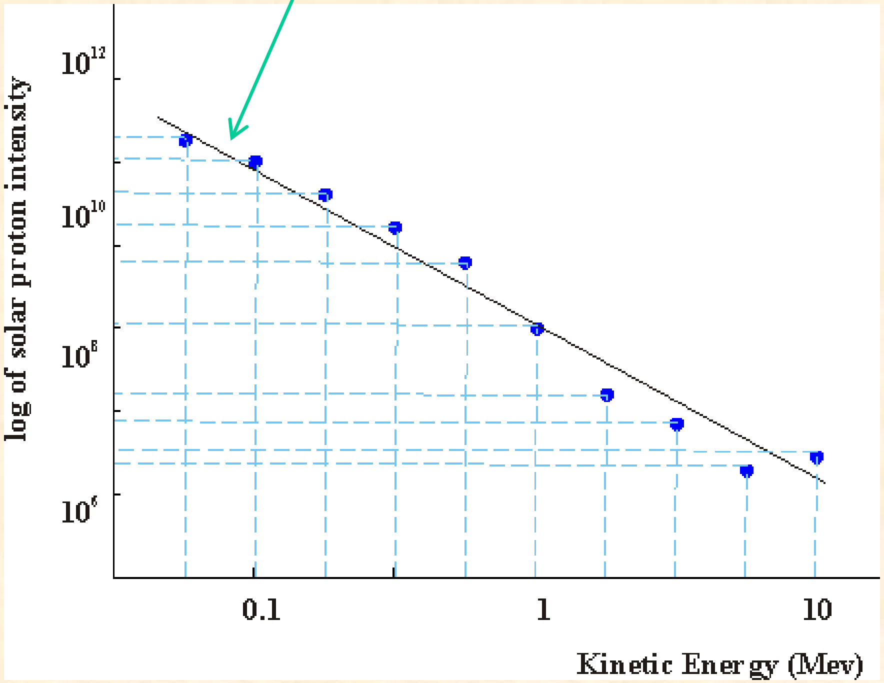
❖ (a) Ποιο είναι το μοντέλο που γεννά τα δεδομένα; (**παλινδρόμηση**)

❖Η γραμμή

❖**y=ax+b**

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

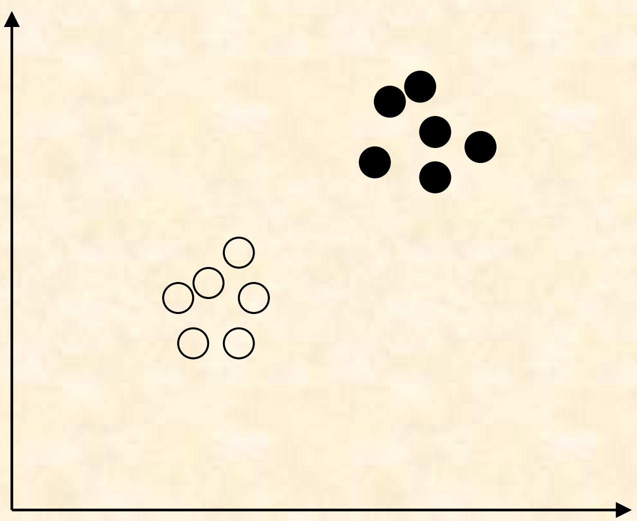


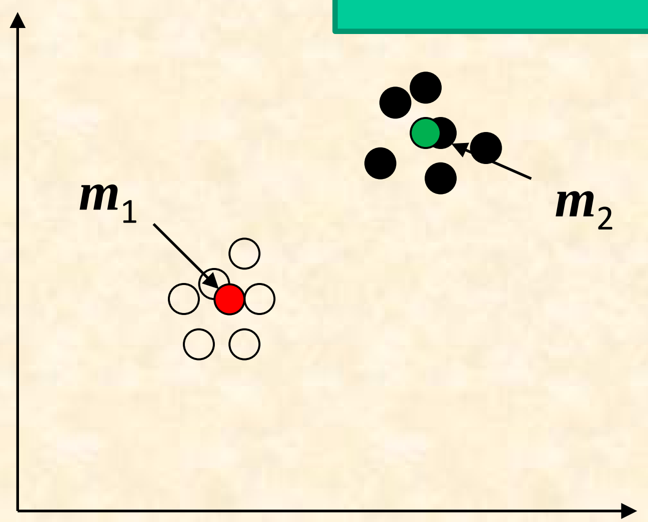


❖ΕΞΑΓΩΓΗ ΓΝΩΣΗΣ

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

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



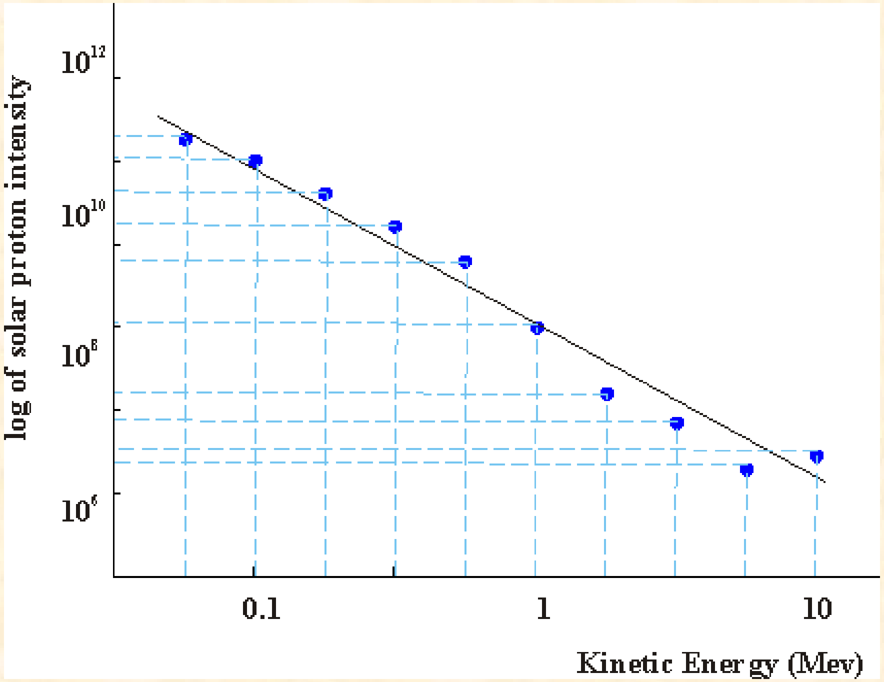


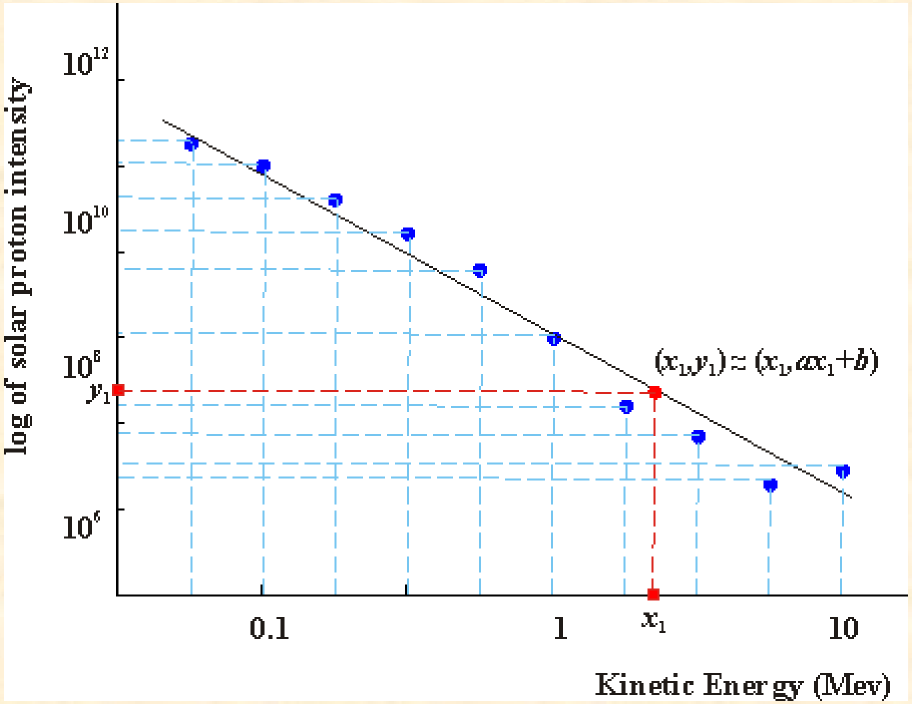
○ ❖● → 1 ❖→ 2

○ ❖● → 1 ❖→ 2

❖ΕΞΑΓΩΓΗ ΓΝΩΣΗΣ

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





❖**Παραμετρική εκτίμηση**

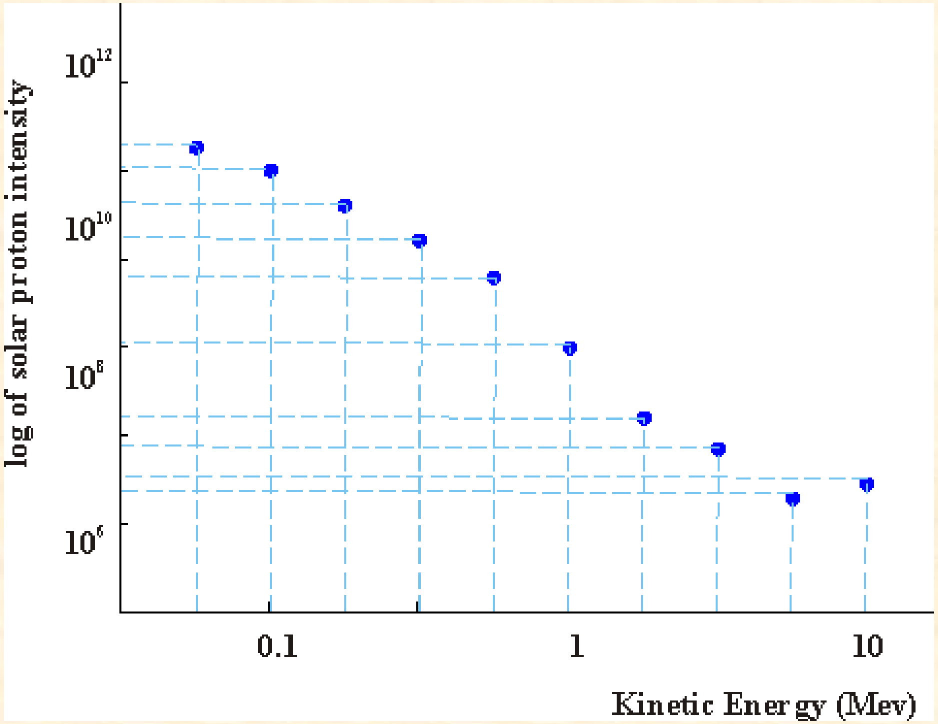
❖***Μοντέλο:*** **γραμμή y=ax+b**

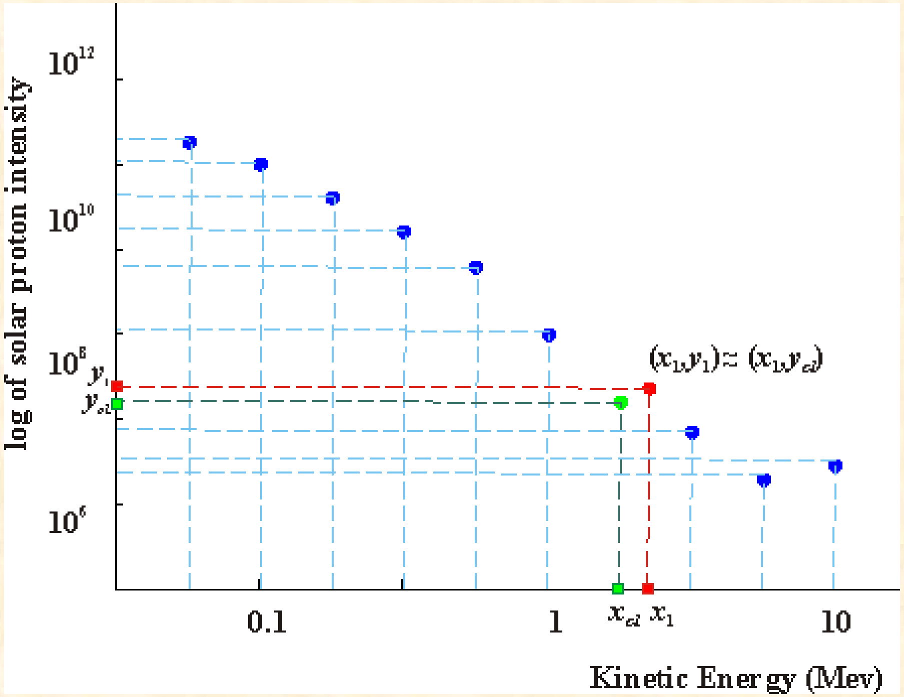
❖***Παράμετροι μοντέλου:*** **a, b**

❖ΕΞΑΓΩΓΗ ΓΝΩΣΗΣ

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

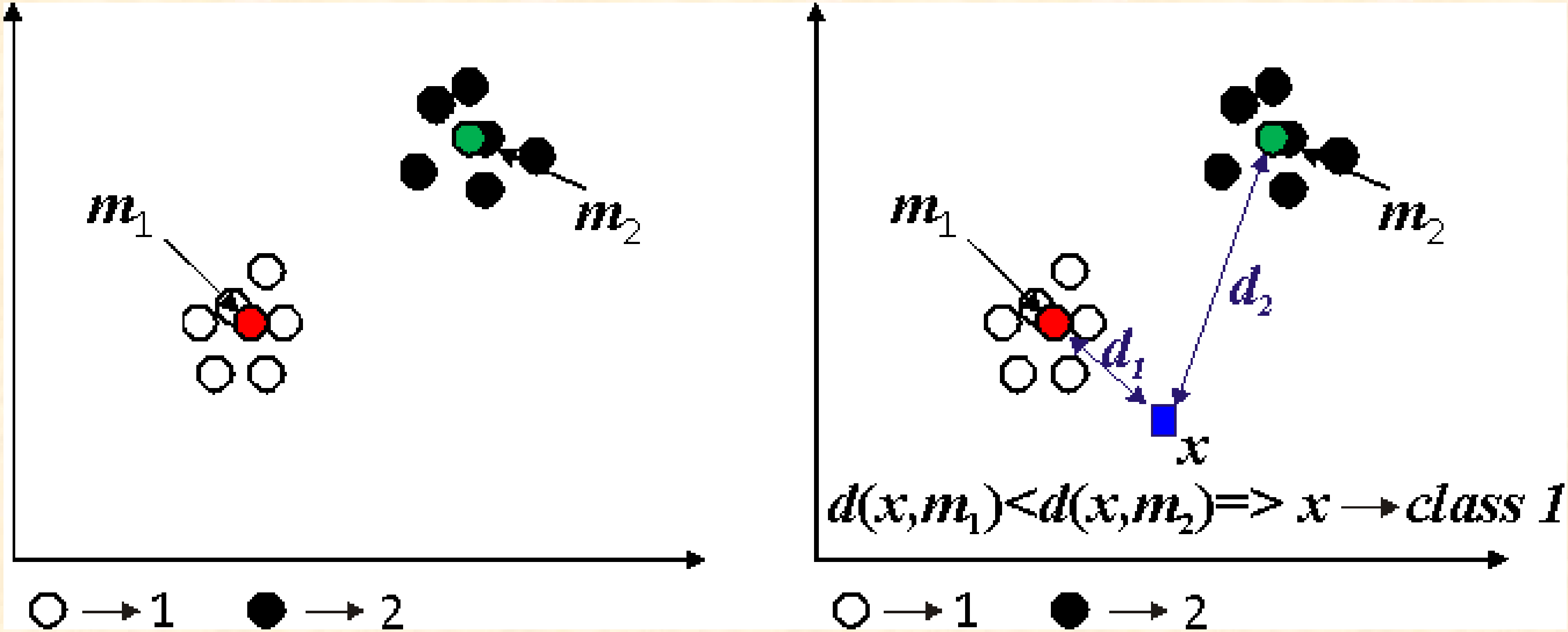
❖**Μη παραμετρική ταξινόμηση**





❖ΕΞΑΓΩΓΗ ΓΝΩΣΗΣ

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

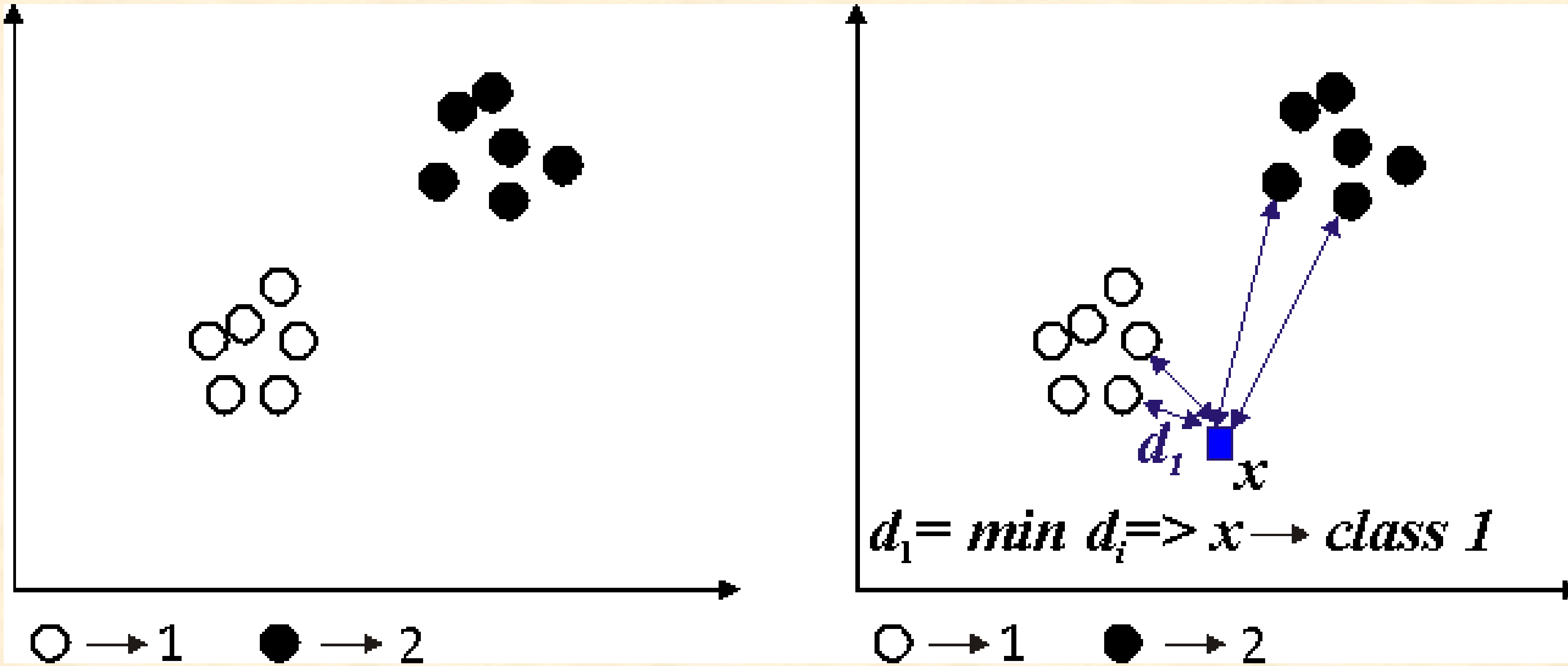


❖**Παραμετρική εκτίμηση**

❖**Παράμετροι: m_1, m_2**

❖ΕΞΑΓΩΓΗ ΓΝΩΣΗΣ

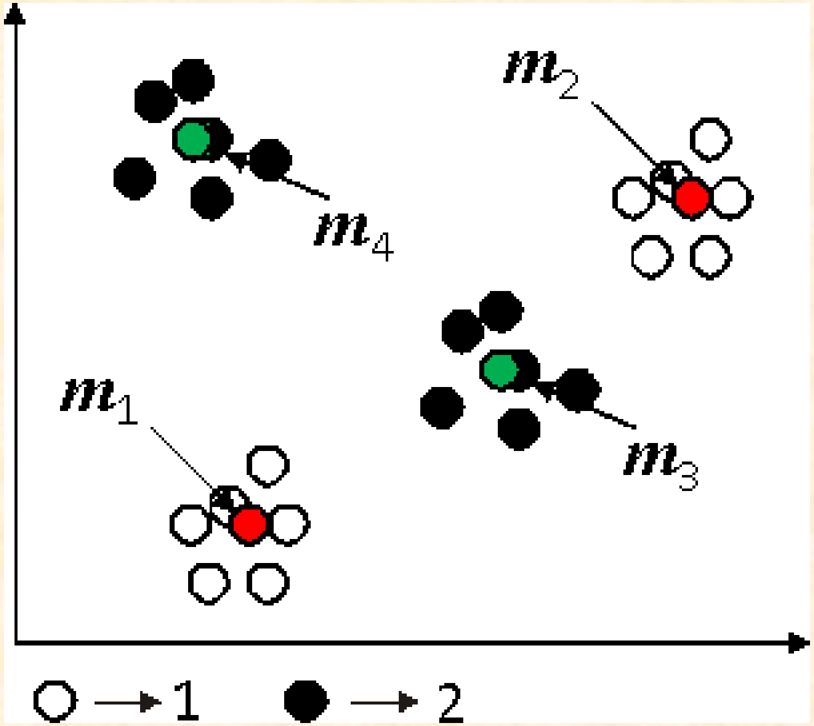
❖(b) Πώς μπορούμε να χρησιμοποιήσουμε τα γνωστά δεδομένα προκειμένου να κάνουμε εκτιμήσεις (estimations) για άγνωστα δεδομένα; (**ταξινόμηση**)



❖**Μη παραμετρική εκτίμηση**

❖ΕΞΑΓΩΓΗ ΓΝΩΣΗΣ

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

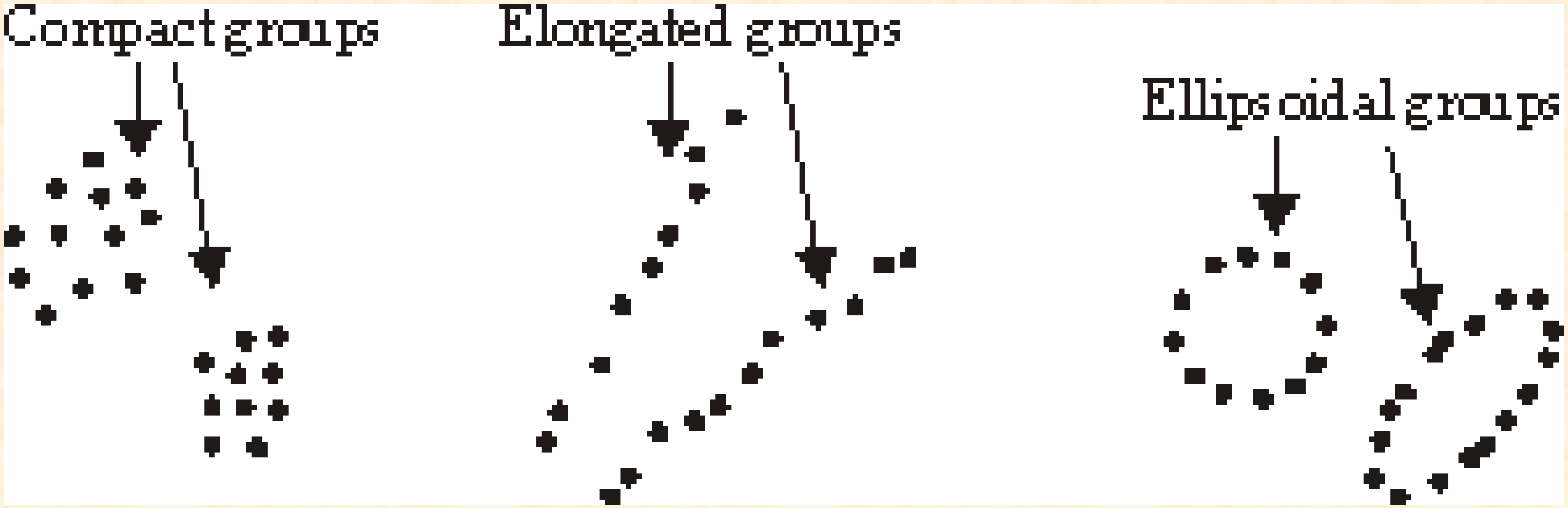


❖Η **κατηγορία 1** συγκεντρώνεται γύρω από τα m_1 και m_2

❖Η **κατηγορία 2** συγκεντρώνεται γύρω από τα m_3 και 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)

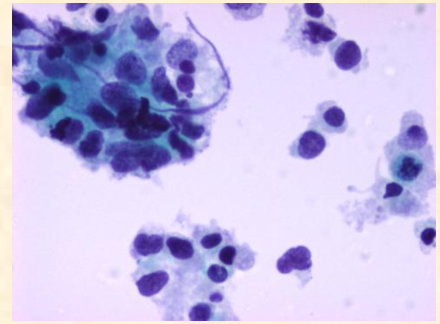
- **Το πρόβλημα:** Καταχώρηση άγνωστων αντικειμένων – προτύπων – στη σωστή κατηγορία (κλάση). Το πρόβλημα είναι γνωστό ως ταξινόμηση (classification).

- 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**.



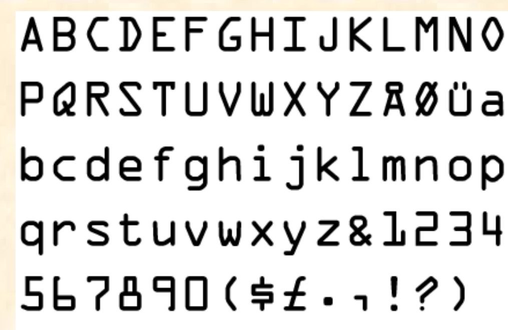
- **Αντικείμενα:** πυρήνες κυττάρων
- **Κατηγορίες:** καλοήθη, κακοήθη

❖ Εφαρμογή σε αλυσίδα παραγωγής στη **βιομηχανία**.

- **Αντικείμενα:** Π.χ. βίδες
- **Κατηγορίες:** Ελαττωματική, μη ελαττωματική

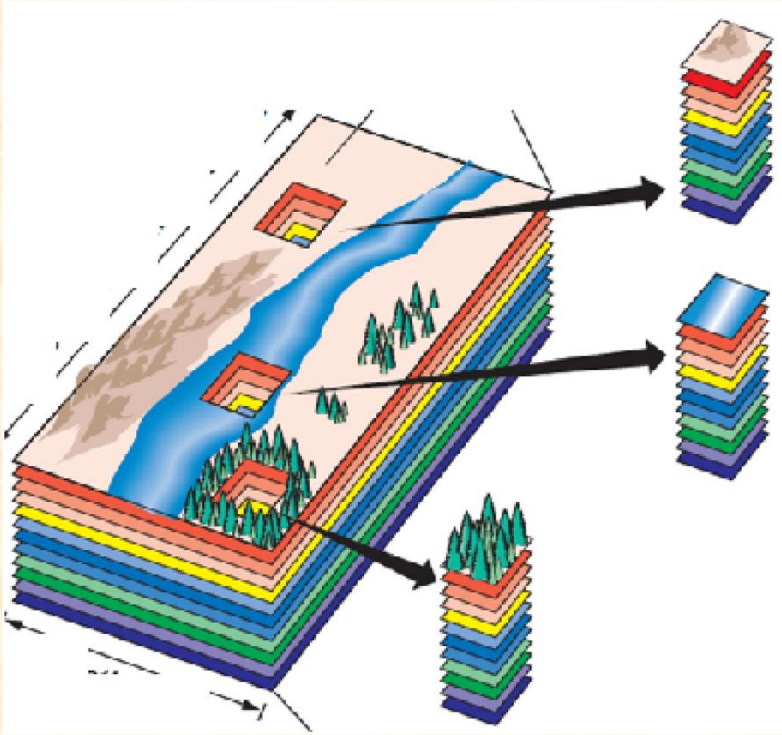
❖ Εφαρμογή στην **αναγνώριση χαρακτήρων κειμένου**

- **Αντικείμενα:** Αλφαριθμητικοί χαρακτήρες, σημεία στίξης
- **Κατηγορίες:** 26 κατηγορίες πεζών+26 κατ. κεφαλαίων+10 κατ. ψηφίων +κατηγορίες σημείων στίξης.



❖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

$$x_1, \ldots, x_l,$$

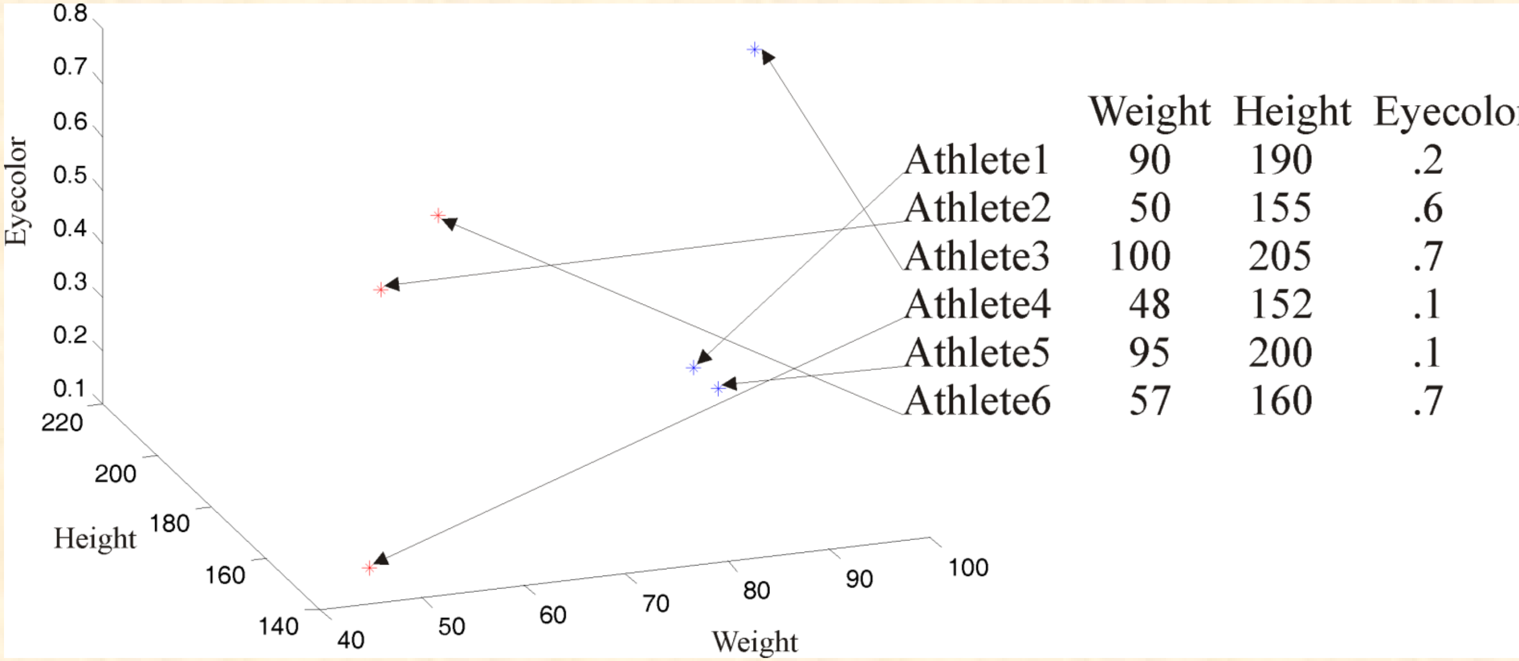
συνιστούν το διάνυσμα χαρακτηριστικών

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

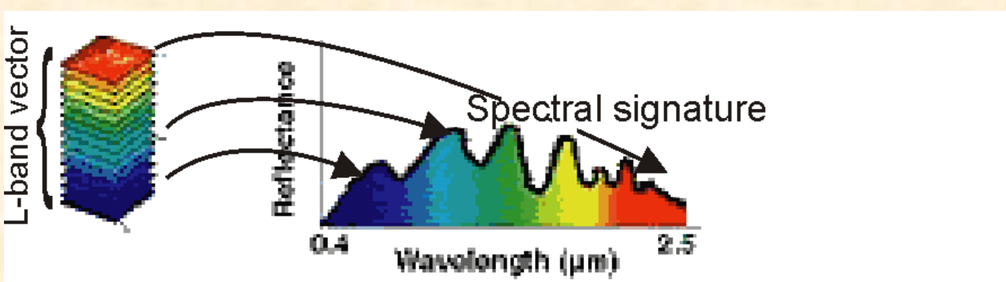
Τα διανύσματα χαρακτηριστικών θεωρούνται ως τυχαίες μεταβλητές (random vectors).

Παραδείγματα:

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



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

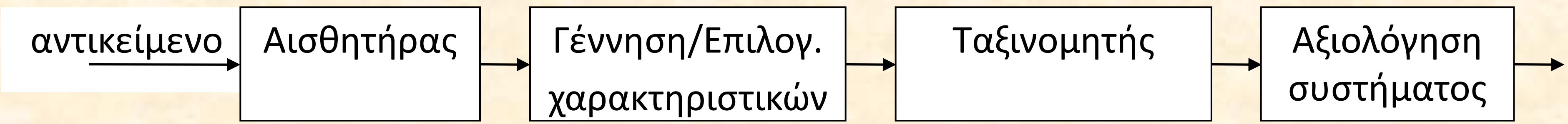


Αναπαράσταση αντικειμένων σε σύστημα αναγνώρισης προτύπων:

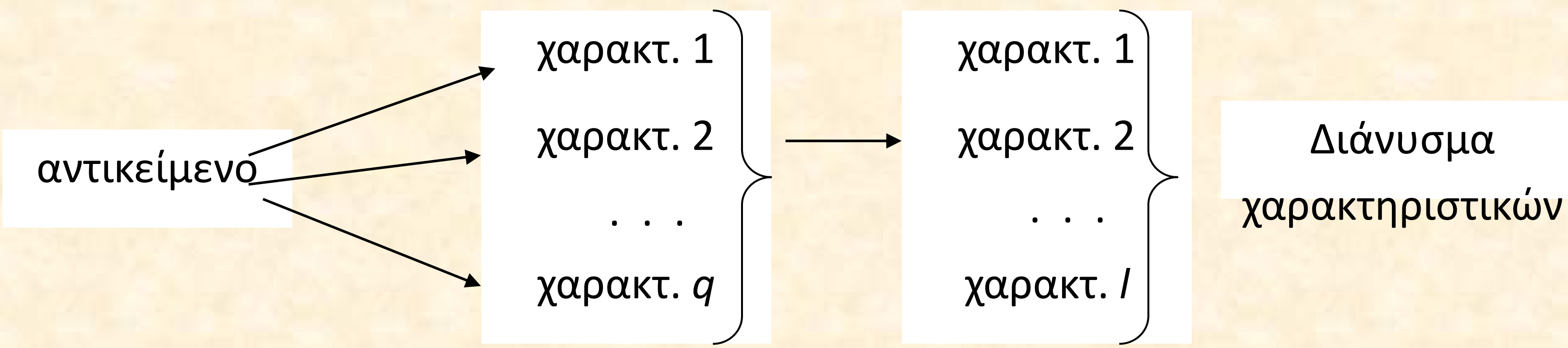
Αντικείμενο ←===→ Διάνυσμα χαρακτηριστικών μετρήσεων (feature vector)

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.

❖**Παράδειγμα:** Να σχεδιασθεί ταξινομητής που να αναγνωρίζει αν ένα αντικείμενο είναι ντομάτα ή αγγούρι.

Αντικείμενα: ντομάτες, αγγούρια

Κλάσεις: «ντομάτα» , «αγγούρι»






<u>Σχεδιασμός:</u>

Αισθητήρας: Π.χ. μια *φωτογραφική μηχανή*. Κάθε αντικείμενο ταυτοποιείται από τη RGB φωτογραφία του.

Γέννηση χαρακτηριστικών: Π.χ. (α) χρώμα (η τιμή στο **R** (red) κανάλι) και (β) σχήμα (**λ**= λόγος κύριου προς δευτερεύοντα άξονα της έλλειψης που προσεγγίζει το σχήμα του αντικειμένου στη φωτογραφία κυκλικό (λ≅1), ελλειψοειδές (λ>>1)).

(Αντιπροσώπευση του αρχικού αντικειμένου διάνυσμα δύο αριθμών [R, λ])

Ταξινόμηση: Ταξινόμηση αντικειμένου βάσει του αντίστοιχου διανύσματος [R, λ], με βάση τον κανόνα: **Αν** R>> **ΚΑΙ** λ ≅1 **τότε** «ντομάτα», **διαφορετικά** «αγγούρι».

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 τη στατιστική των κλάσεων).

- Στη συνέχεια ακολουθεί μελέτη των προτύπων ώστε να προσδιορίσουμε τα χαρακτηριστικά που διαφοροποιούν πιο πολύ τα πρότυπα των δύο κλάσεων.

Προβλήματα:

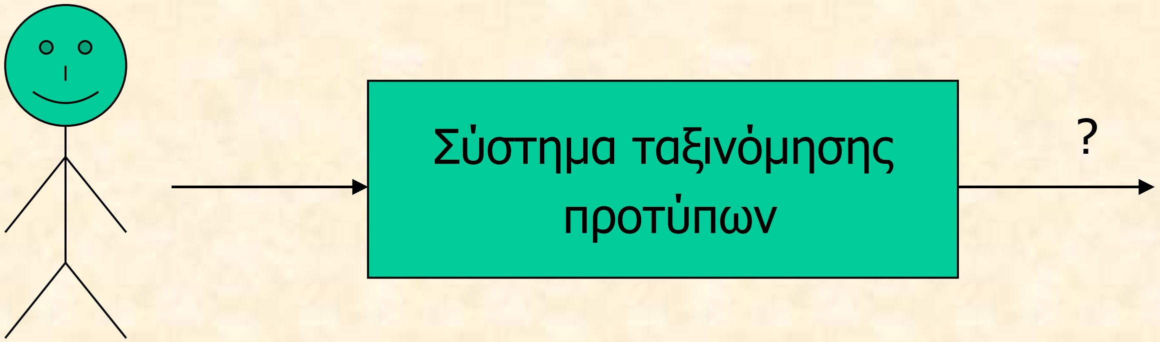
(A) Ποια χαρακτηριστικά θα επιλέξω;

(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

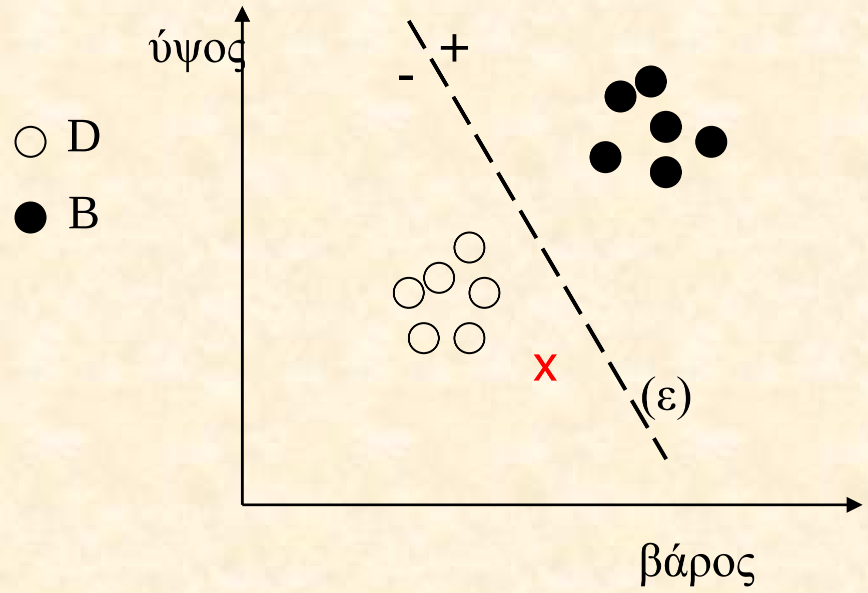
Πώς καθορίζεται ποσοτικά όμως το «**μεγάλο**» και το «**μικρό**»;

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) **<u>Ταξινομητής</u>**

Σύνολο συναρτήσεων/κανόνων βάσει των οποίων γίνεται η ταξινόμηση.

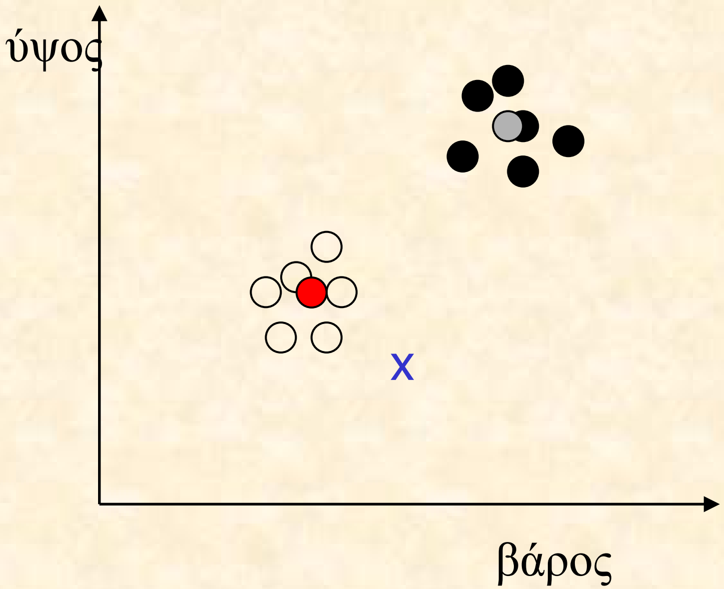


Παράδειγμα 1:

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”

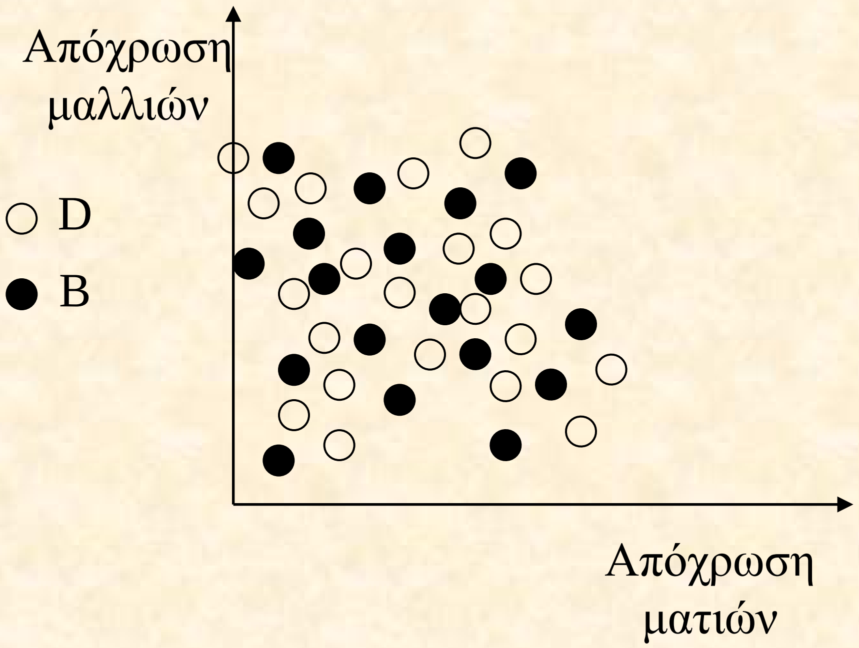


Παράδειγμα 2:

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

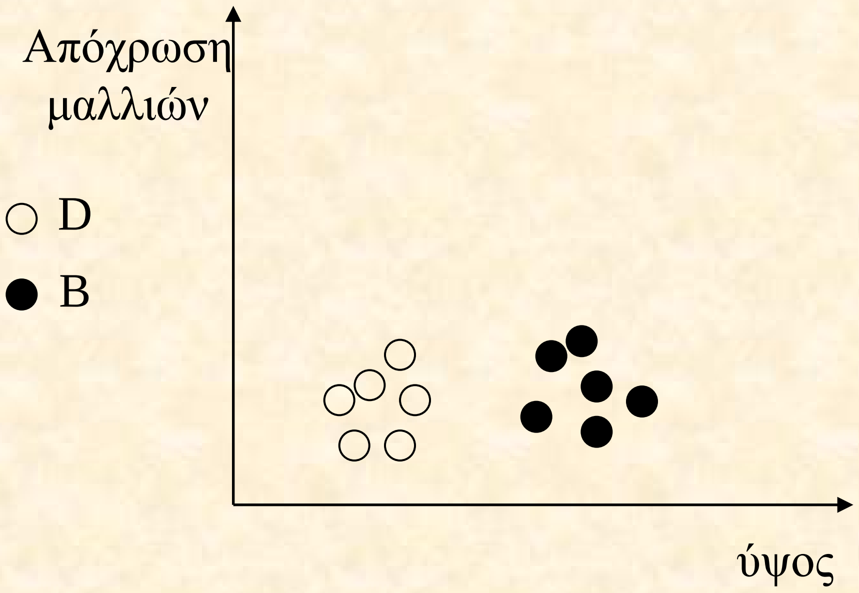
Χαρακτηριστικά (ξανά)

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

Ακόμη είναι:

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

Μερικά προκαταρτικά (συν.)

Τυχαία μεταβλητή (**random variable)** *x* ονομάζεται μια μεταβλητή που η τιμή της είναι το αποτέλεσμα ενός στατιστικού πειράματος. Με άλλα λόγια, η *x* **μοντελοποιεί** το αποτέλεσμα ενός στατιστικού πειράματος.

Τυχαίο διανυσμα (random vector) ***x*** είναι ένα διάνυσμα τυχ. μεταβλητών.

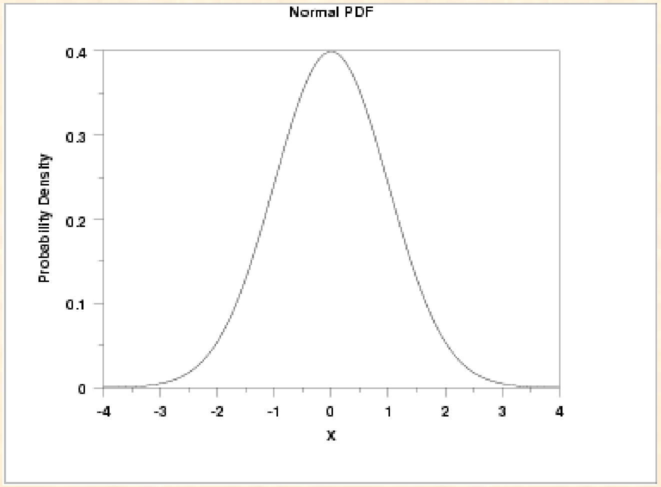
Ένα **συνεχές (continuous) τυχαίο διάνυσμα** ***x*** παίρνει τιμές σε ένα υποσύνολο *S* του R^l.

Η **συνάρτηση πυκνότητας πιθανότητας** (**probability density function** - **pdf**) του τυχαίου διανύσματος ***x***, $p(\boldsymbol{x})$, είναι μια μη αρνητική συνάρτηση πάνω στο *S*, με το ολοκλήρωμά της πάνω στο S ναι είναι μονάδα. Η $p(\boldsymbol{x})$ παρουσιάζει ``μεγάλες'' τιμές σε περιοχές του *S* που αντιστοιχούν στα πιο πιθανά ενδεχόμενα για το υπό μελέτη πείραμα.

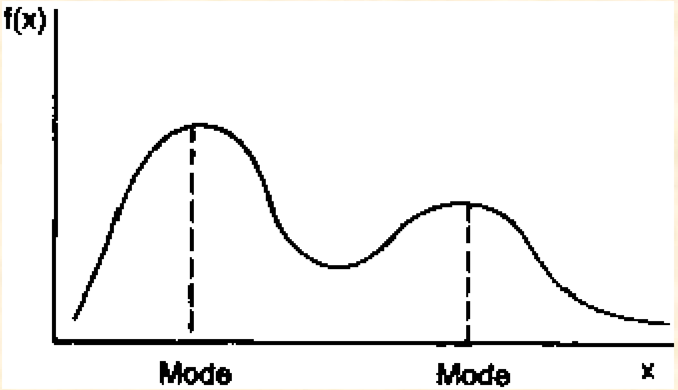
Στο πλαίσιο της **ταξινόμησης**: Ένα τυχαίο διάνυσμα ***x*** **μοντελοποιεί** την κατανομή των διανυσμάτων αντιπροσώπευσης των αντικειμένων στο χώρο.

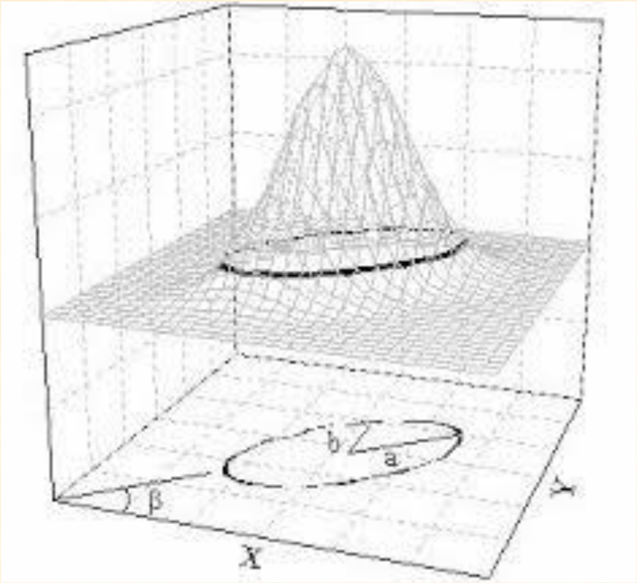
Μερικά προκαταρτικά (συν.)

<u>Παραδείγματα συναρτήσεων πυκνότητας πιθανότητας (pdf)</u>

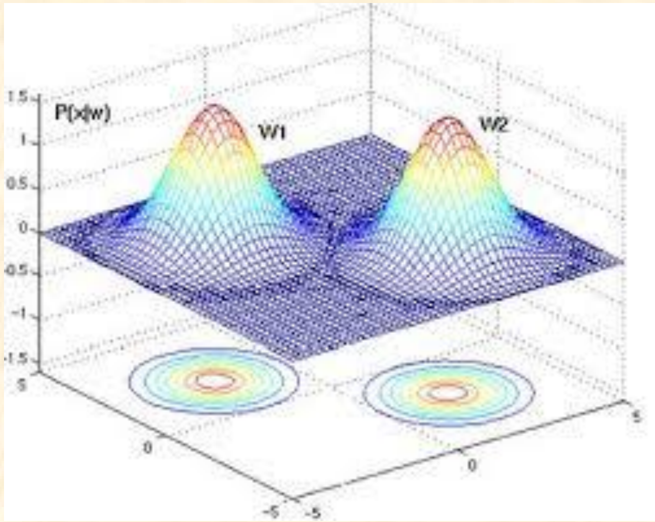


Μονοδιάστατες pdf





Δισδιάστατες pdf



Μερικά προκαταρτικά (συν.)

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 **ορισιτκά θετικός** (**μη αρνητικός**) ανν

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

Ο **αντίστροφος** του *A* (αν υπάρχει) συμβολίζεται με A^{-1} και είναι

$$A A^{-1} = A^{-1} A = I$$

Αν ο A είναι **οριστικά θετικός** τότε είναι **αντιστρέψιμος.**

Αν ο A είναι **διαγώνιος** με $\alpha_{ii} \neq 0$, **υπάρχει ο αντίστροφός του**, ο οποίος είναι επίσης διαγώνιος με το (i,i) στοιχείο του να ισούται με $1/a_{ii}$.

Μερικά προκαταρτικά (συν.)

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

Το **ίχνος (trace) του** A είναι

$$trace\left(\begin{bmatrix} a_{11} & \cdots & a_{1l} \\ \vdots & \ddots & \vdots \\ a_{l1} & \cdots & a_{ll} \end{bmatrix}\right) = a_{11} + \cdots + a_{ll}$$

Είναι $x^T x = x_1^2 + \cdots + x_l^2$, $\quad xx^T = \begin{bmatrix} x_1^2 & \cdots & x_1 x_l \\ \vdots & \ddots & \vdots \\ x_l x_1 & \cdots & x_l^2 \end{bmatrix}$

και $trace(xx^T) = x^T x$

Μερικά προκαταρτικά (συν.)

Έστω *A* ένας *l*x*l* πίνακας. Οι *l* αριθμοί λ και τα αντίστοιχα διανύσματα ***x*** που ικανοποιούν την εξίσωση

$$A\boldsymbol{x}=\lambda\boldsymbol{x}$$

ονομάζονται αντίστοιχα **ιδιοτιμές** και **ιδιοδιανύσματα** του πίνακα Α.

-> Ο *A* είναι οριστικά θετικός, τότε και μόνον τότε όταν έχει θετικές ιδιοτιμές.

-> Αν ο Α είναι συμμετρικός τότε τα ιδιοδιανύσματά του είναι ορθογώνια (κάθετα) μεταξύ τους.

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

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

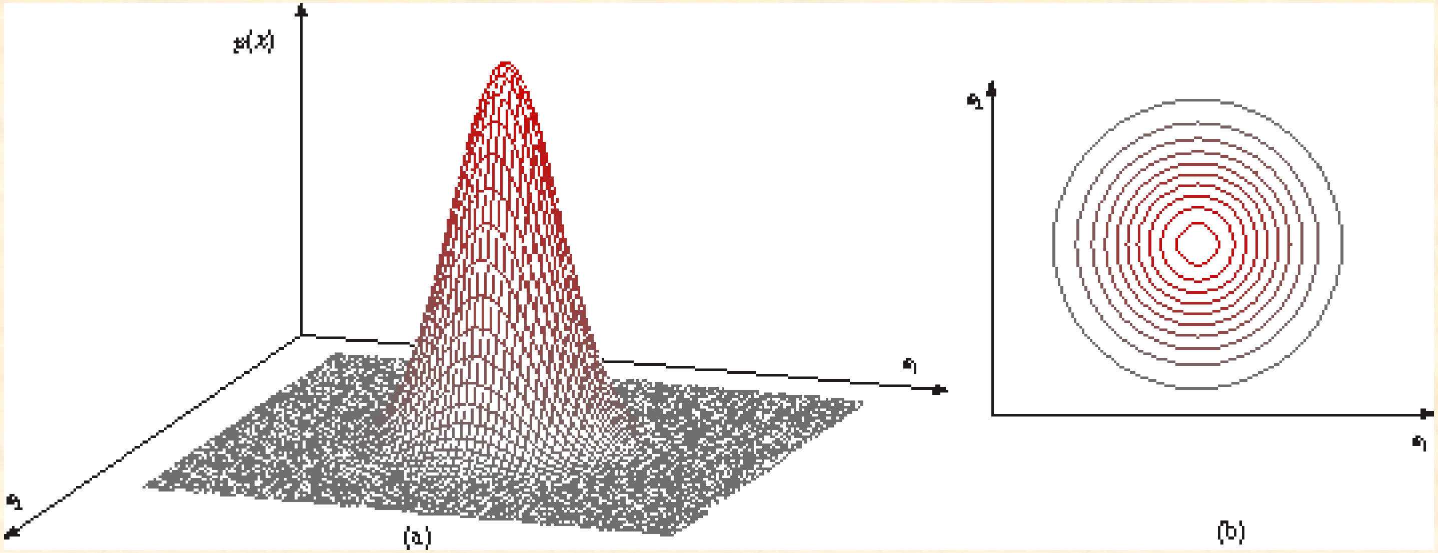
Σ, *Σ*$^{-1}$ είναι **θετικά (ήμι) ορισμένοι.**

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

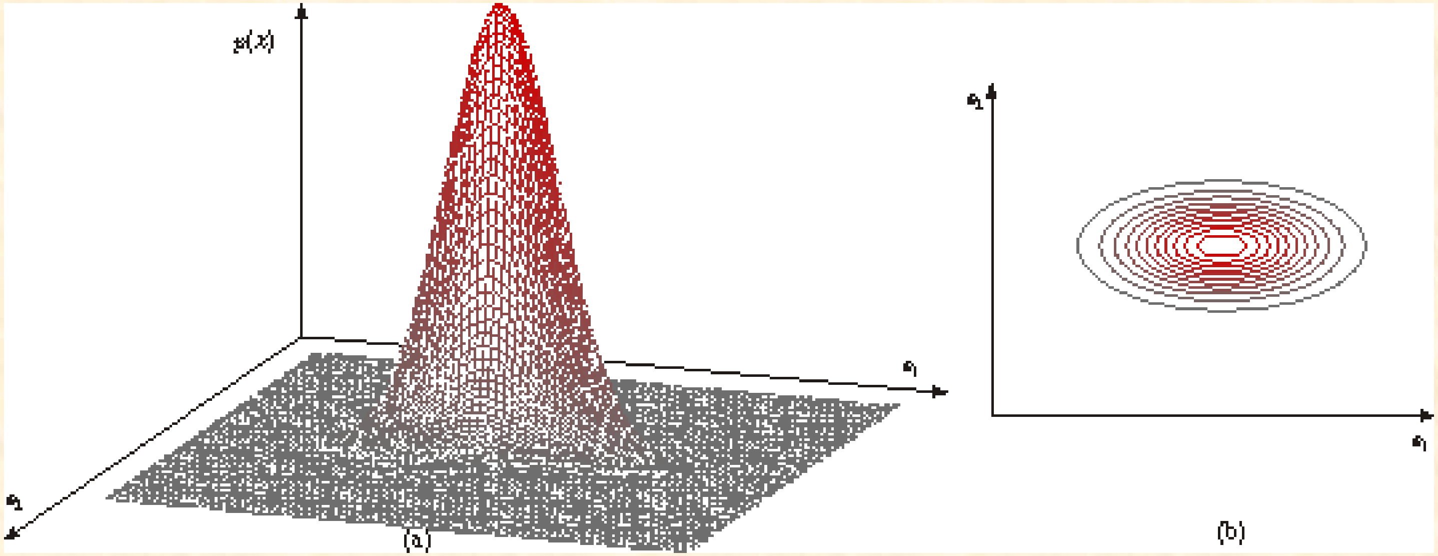
$$\sigma_i^2 = E[(x_i - \mu_i)^2], \sigma_{ij} = \sigma_{ji} = E[(x_i - \mu_i)(x_j - \mu_j)]$$

Μερικά προκαταρτικά (συν.)

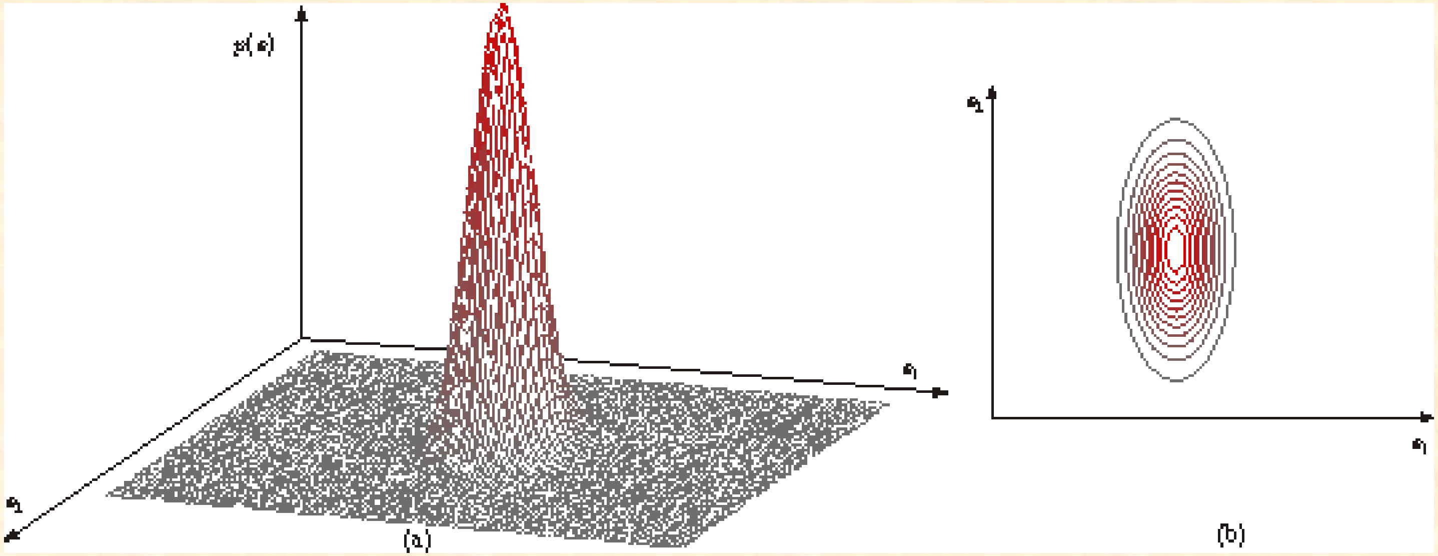
-> Το ιδιοδιάνυσμα που αντιστοιχεί στη μεγαλύτερη ιδιοτιμή του πίνακα συνδιασποράς, ορίζει την κατεύθυνση κατά την οποία το σύνολο δεδομένων παρουσιάζει τη μέγιστη διασπορά, το ιδιοδιάνυσμα που αντιστοιχεί στην αμέσως επόμενη ιδιοτιμή είναι κάθετο στην πρώτη και αντιστοιχεί στη διεύθυνση κατά την οποία το σύνολο δεδομένων παρουσιάζει την αμέσως επόμενη μέγιστη διασπορά κλπ.



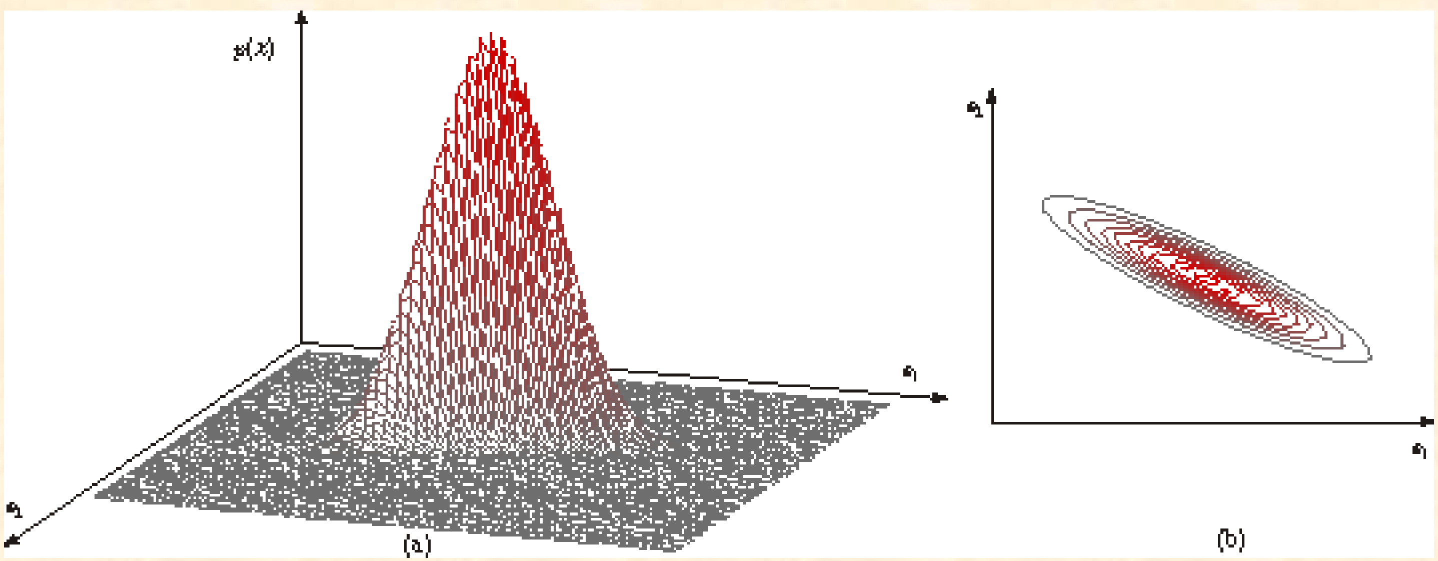
Σ: διαγώνιος με ίσα διαγώνια στοιχεία



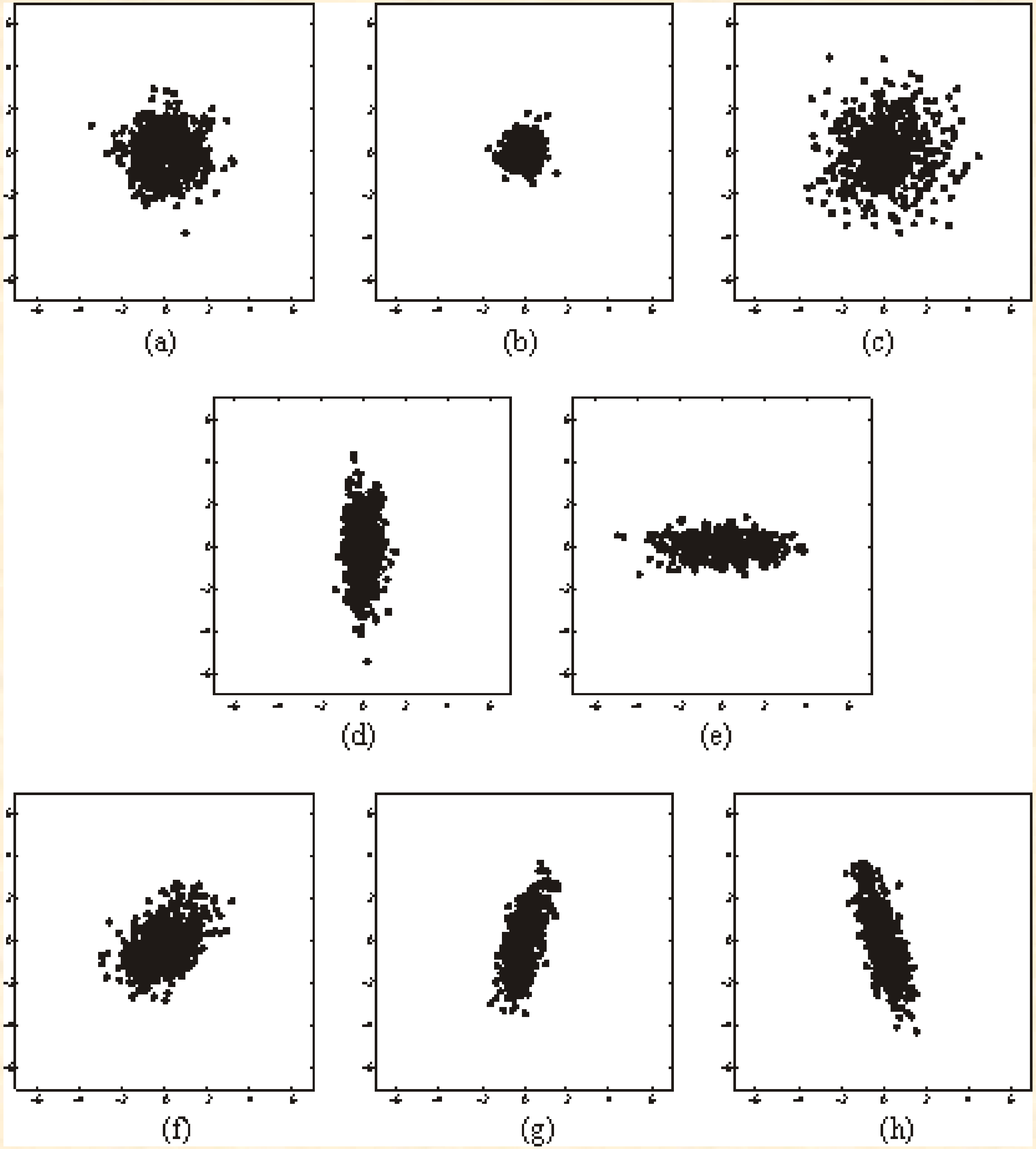
Σ: διαγώνιος με $\sigma_1^2 >> \sigma_2^2$



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 $\Leftrightarrow$ οι συνιστώσες του ***x*** είναι στατιστικά ανεξάρτητες

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

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

Μερικά προκαταρτικά (συν.)

Έστω $x = [x_1, \ldots, x_l]^T$ $b = [b_1, \ldots, b_l]^T$

$$A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1l} \\ a_{21} & a_{22} & \cdots & a_{2l} \\ \vdots & \vdots & \ddots & \vdots \\ a_{l1} & a_{l2} & \cdots & a_{ll} \end{bmatrix}$$

συμμετρικός, *c* βαθμωτό μέγεθος

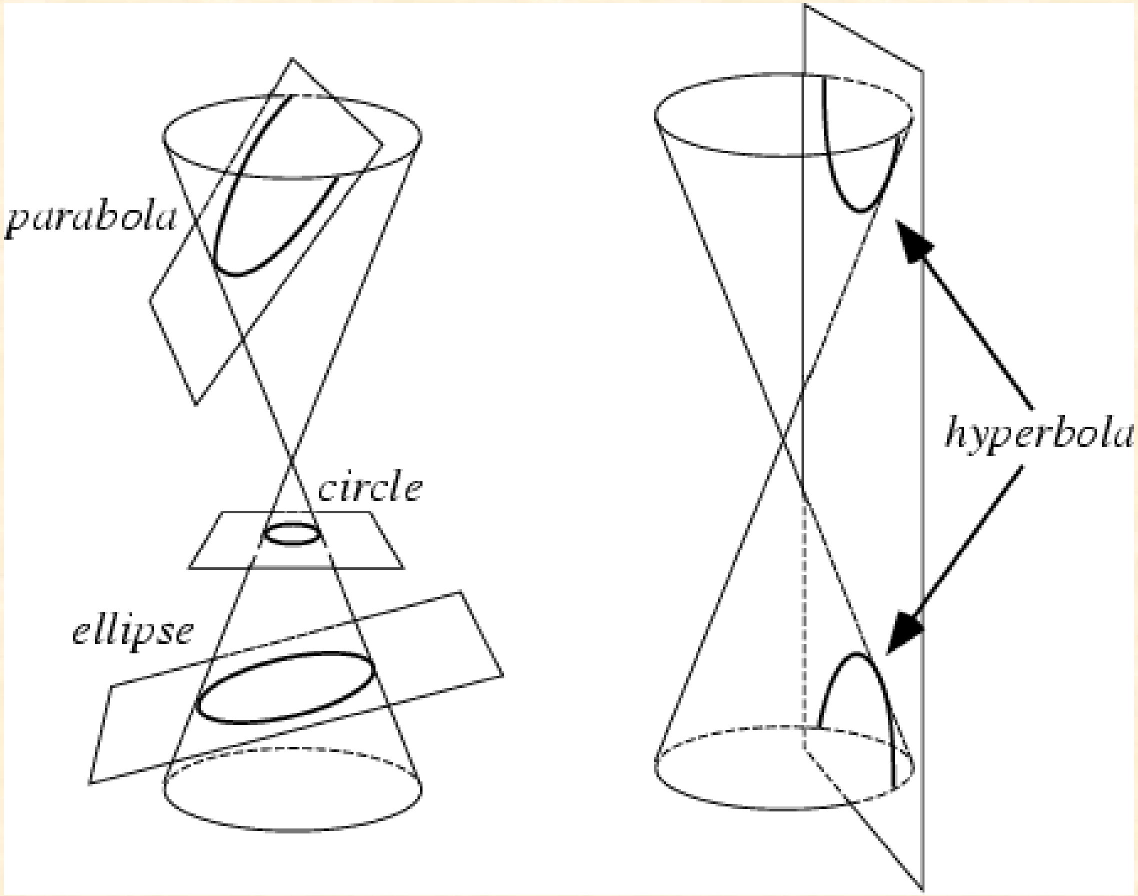
Η εξίσωση $$x^T A x + b^T x + c = 0$$

Παριστάνει μια **υπερεπιφάνεια (hypersurface)** στο χώρο R^l (υπερ-έλλειψη, υπερ-υπερβολή, υπέρ-παραβολή, ζεύγος υπερεπιπέδων).

-> Αν Α είναι **οριστικά θετικός**, η υπερεπιφάνεια είναι **υπερέλλειψη**.

-> Αν *A=O*, η υπερεπιφάνεια είναι **υπερεπίπεδο**.

Μερικά προκαταρτικά (συν.)



Μερικά προκαταρτικά (συν.)

Ισχύουν τα παρακάτω:

$$\frac{\partial}{\partial x}\left(c^T x\right) = c^T$$

$$\frac{\partial}{\partial x}\left(x^T A x\right) = x^T (A^T + A)$$

$$\frac{\partial}{\partial x}\left(x^T A x\right) = 2x^T A, \quad \alpha\nu \quad A \quad \varepsilon\acute{\iota}\nu\alpha\iota \quad \sigma\upsilon\mu\mu\varepsilon\tau\rho\iota\kappa\acute{o}\varsigma$$

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;»*

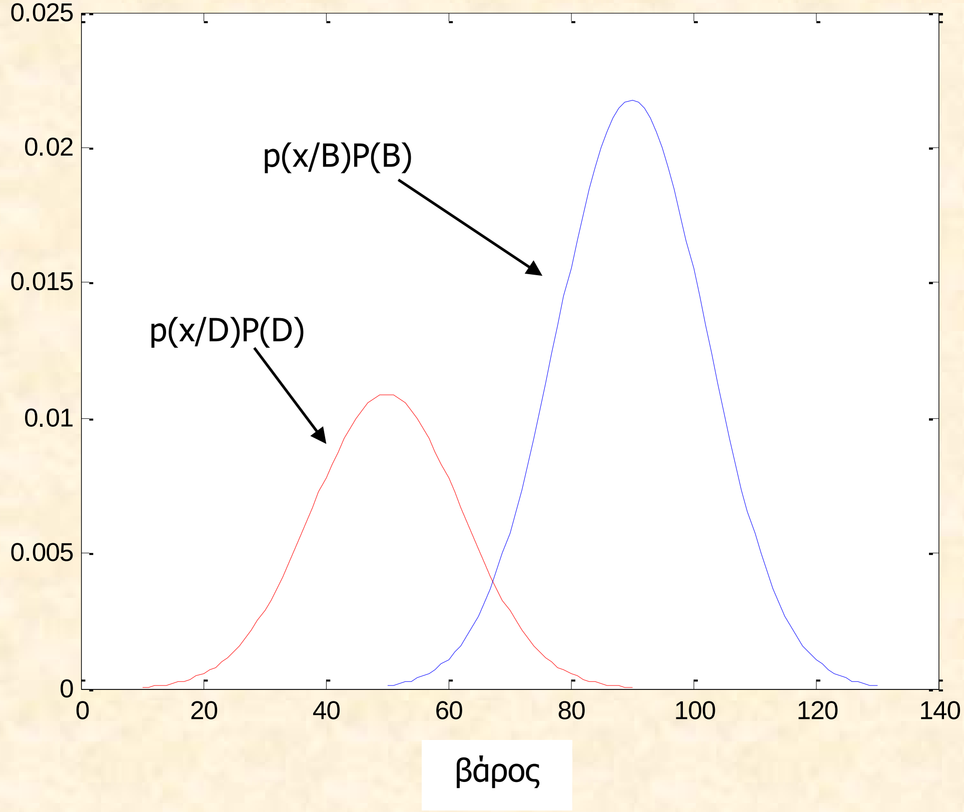
Απάντηση:

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

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

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

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

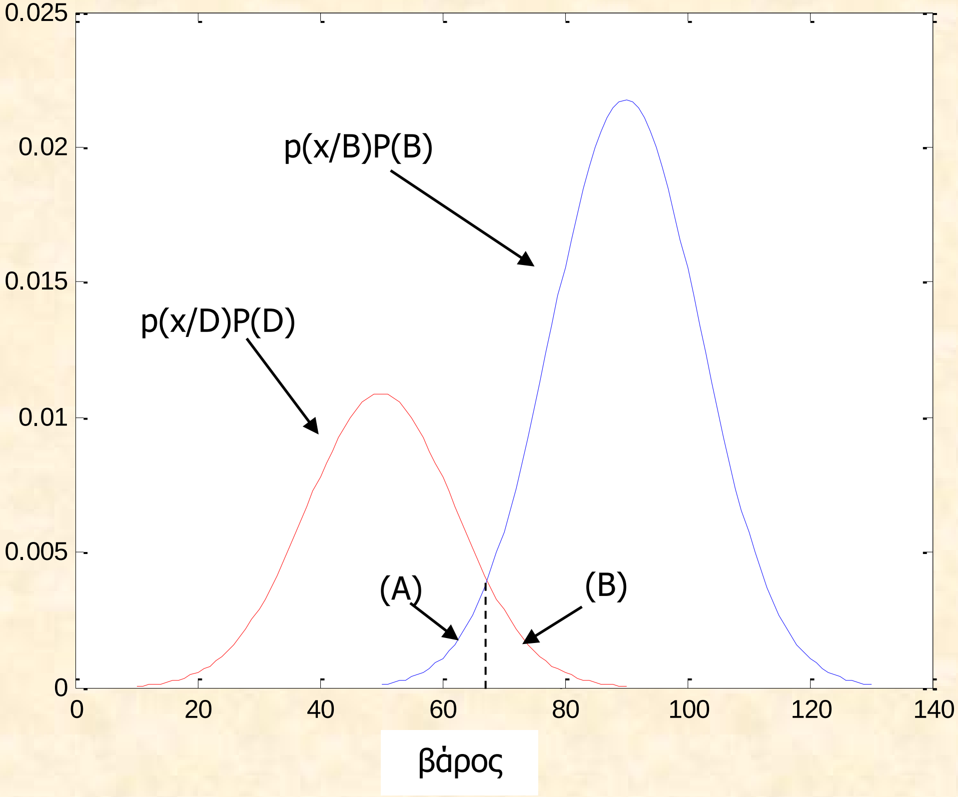
οριοθετούν τις περιοχές του χώρου των χαρακτηριστικών (βάρος) που αντιστοιχούν στις δύο κατηγορίες

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

Ορισμός περιοχών: Λύνοντας την εξίσωση

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

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



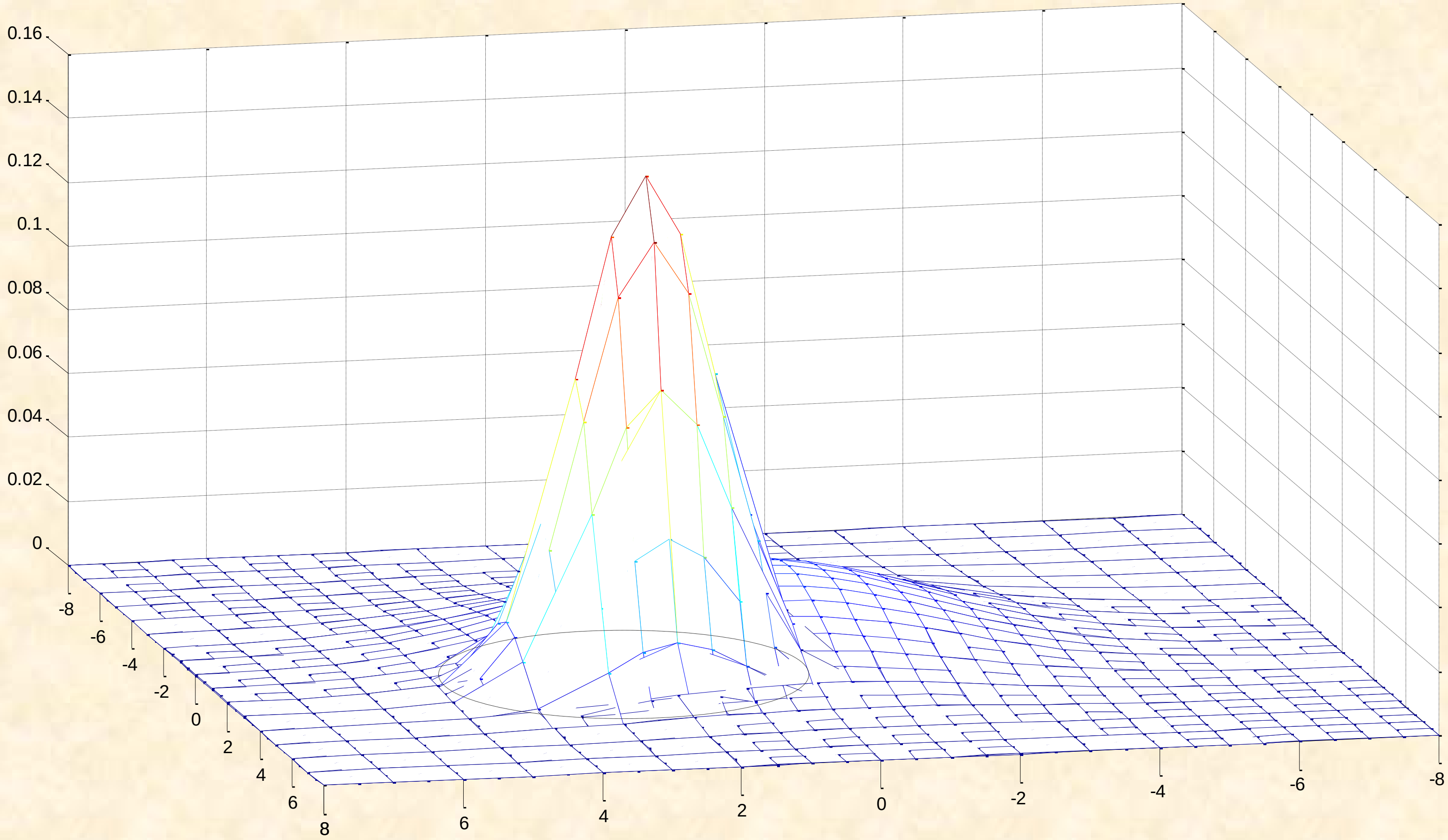
Χορευτής: $R_D=\{x: x<64.7\}$

Καλαθοσφ.: $R_B=\{x: x>64.7\}$

Ερώτηση: Ποιά είναι η πιθανότητα λάθους;

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

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



Παράδειγμα δύο κατηγοριών στον 2-διάστατο χώρο.

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*

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

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*

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

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

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

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

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

ω3: $P(\omega 3)=1/6$, $p(x|\omega 3)=(1/\sqrt{2\pi})*\exp(-(x-2)^2/2)$

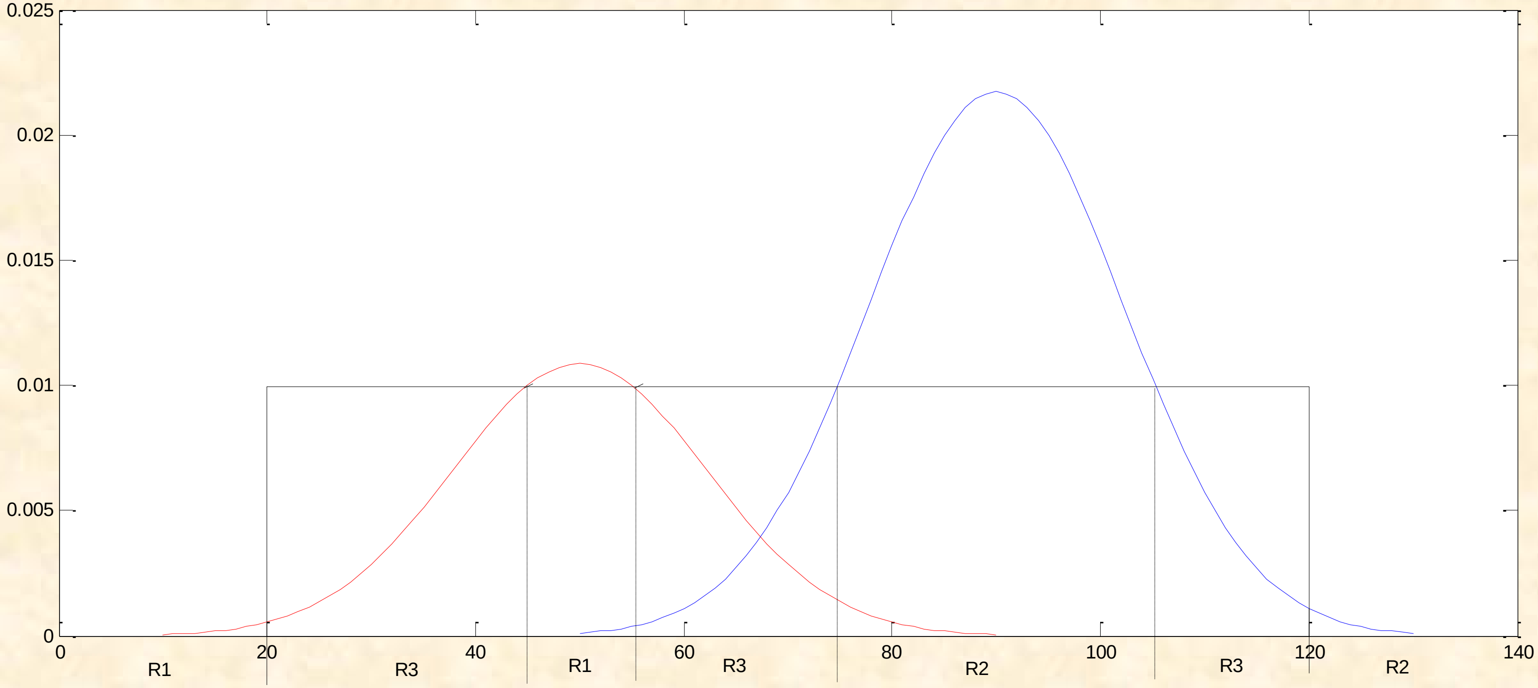
Ταξινομήστε σε μία από τις παραπάνω κατηγορίες άγνωστο δείγμα με τιμή χαρακτηριστικού **x=1.6**.

Είναι:

$P(\omega 1)p(x|\omega 1)=$ **0.0555**, $P(\omega 2)p(x|\omega 2)=$ **0.1111**, $P(\omega 3)p(x|\omega 3)=$ **0.0614**

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

Παράδειγμα 3 τριών κατανομών



Πιθανότητα λάθους:

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

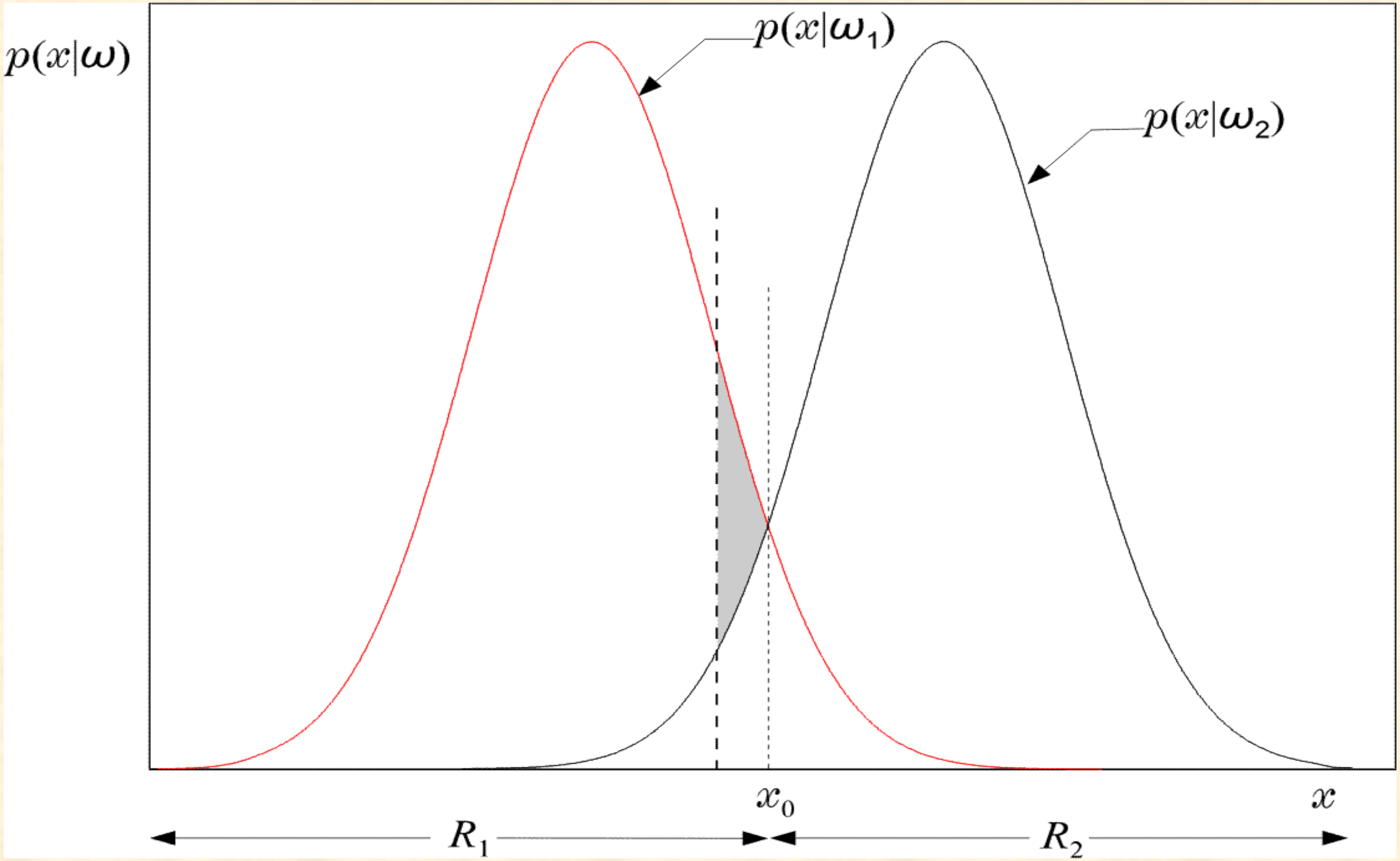
Ερώτηση: Γιατί τόση επιμονή στον ταξινομητή Bayes;

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.*

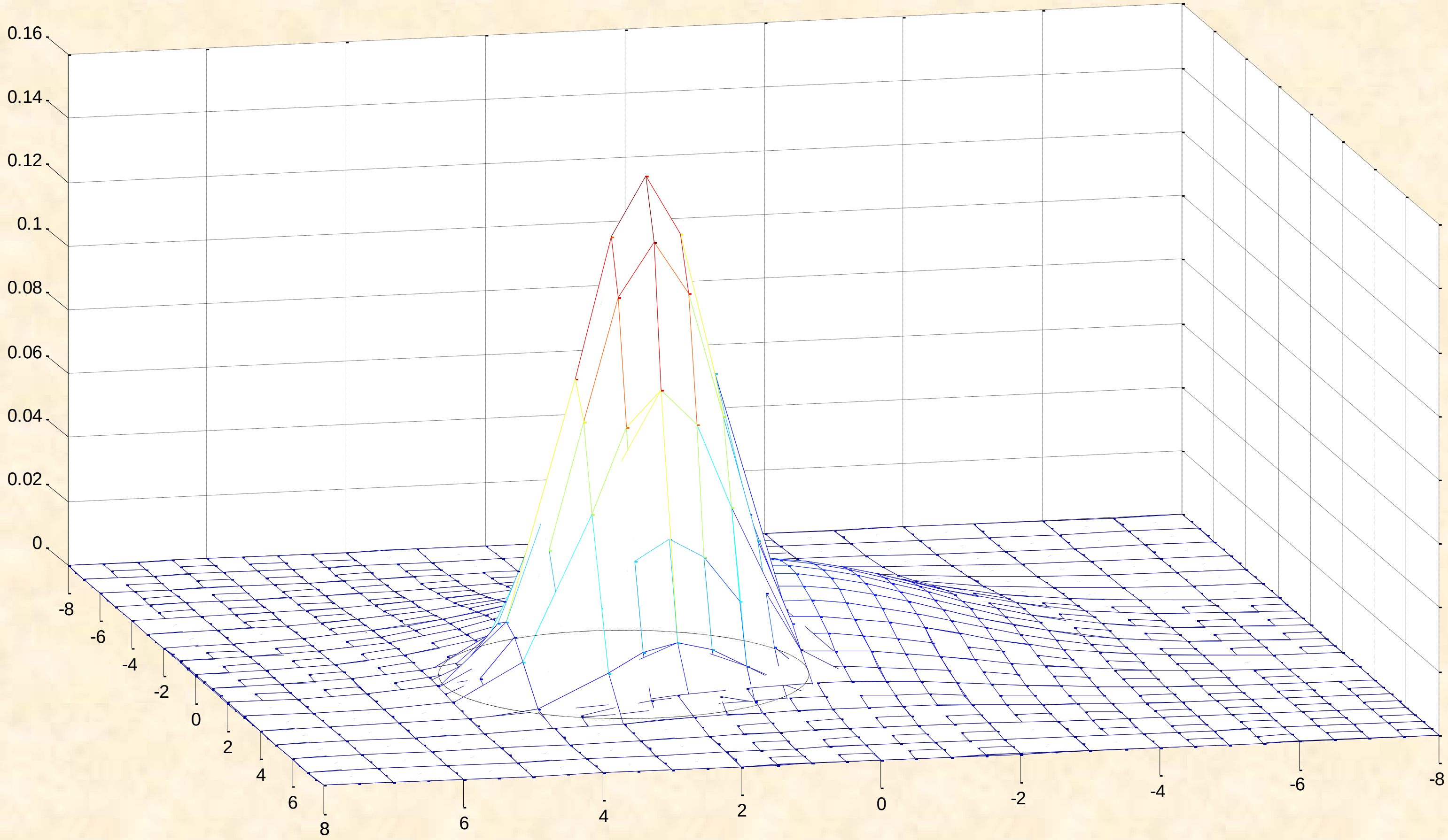


Παρατήρηση: οι πυκνότητες πιθανότητας των κατηγοριών μαζί με τις a priori πιθανότητές τους ορίζουν **μονοσήμαντα** μια διαμέριση του χώρου του *l*-διάστατου χώρου των χαρακτηριστικών, σε (όχι απαραίτητα ενιαίες) περιοχές (R_i) όπου κάθε μια αντιστοιχεί και σε μια κατηγορία (ω_i).

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.