Ψηφιακή Εικόνα – *Αναγνώριση Προτύπων*

(DIP: **C**omputer **V**ision – **P**attern **R**ecognition)

Χειμερινό εξάμηνο

2019-2020

Κλασσική θεωρία ταξινόμησης.

Θεωρία αποφάσεων κατά Bayes

Bayesian Decision Theory

- Η ικανότητα των μηχανών να αντιλαμβάνονται το περιβάλλον τους είναι σχετικά περιορισμένη.

- Με την βοήθεια αισθητήρων (transducers) φυσικά μεγέθη όπως το φως, ο ήχος η θερμοκρασία κ.λ.π, μπορούν να μετατραπούν σε ηλεκτρικά σήματα.

- Στην συνέχεια με κατάλληλη επεξεργασία από ηλεκτρονικό σύστημα θα πρέπει να αναλυθούν και να κατανοηθούν.

- Η δυνατότητα αποθήκευσης και επεξεργασίας μεγάλου όγκου δεδομένων έχει προσδώσει ανεκτίμητες δυνατότητες στην προσπάθεια εξομοίωσης του ανθρώπινου τρόπου σκέψης από τον ηλεκτρονικό υπολογιστή.

Ένα Παράδειγμα

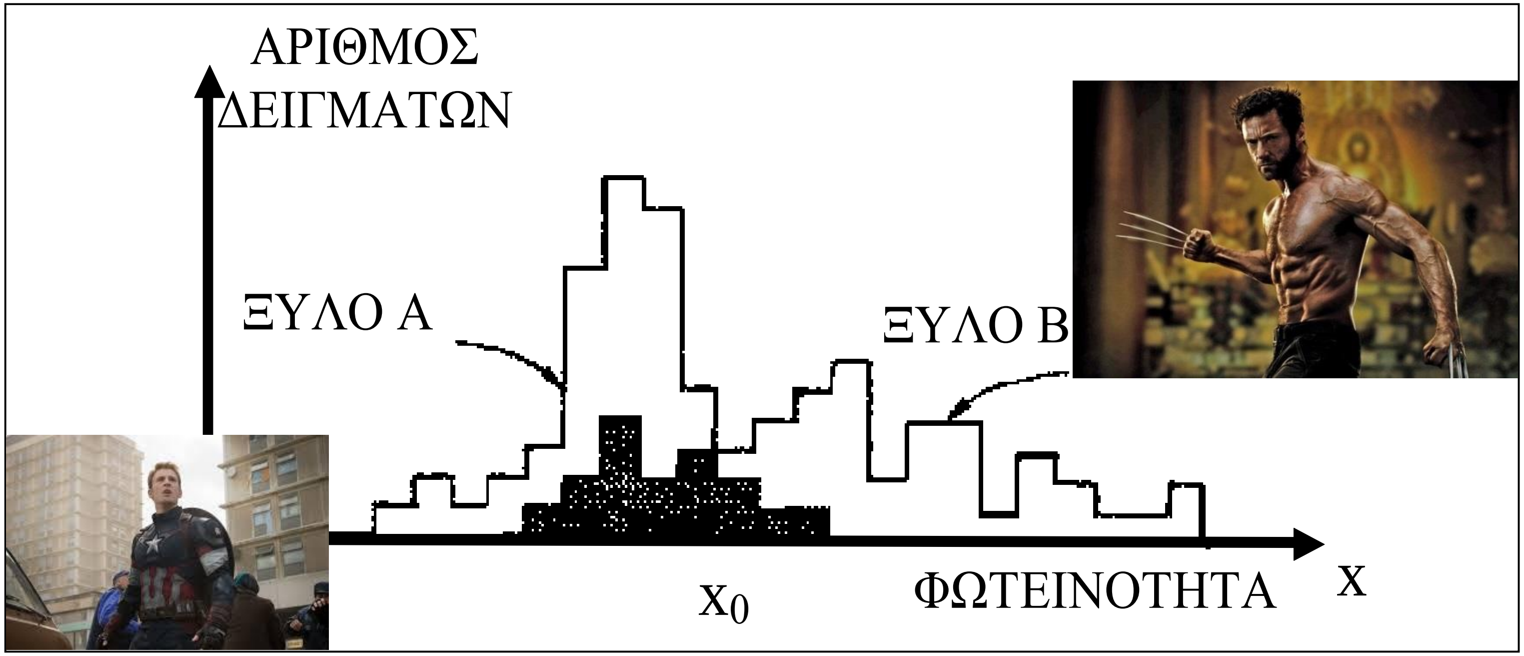
- Έστω ότι ένα πριονιστήριο που παράγει προϊόντα ξυλείας θέλει να αυτοματοποιήσει την διαδικασία της ταξινόμησης των ξύλων σύμφωνα με τον τύπο τους (διαχωρισμός του ξύλου Α από το ξύλο Β).

- Ένα σύστημα που θα μπορούσε να εκτελέσει αυτό τον διαχωρισμό φαίνεται στο παρακάτω σχήμα.
  - Η κάμερα λαμβάνει την εικόνα του ξύλου και την μεταβιβάζει σε ένα εξαγωγέα χαρακτηριστικών (**feature extractor**), που σκοπό έχει να ελαττώσει την διαθέσιμη πληροφορία μετασχηματίζοντας τα δεδομένα σε μια πιο συνοπτική μορφή (άνυσμα).
  - Τα χαρακτηριστικά εισέρχονται σε κατάλληλο ταξινομητή (**classifier**) που αποφασίζει σχετικά με τον τύπο του ξύλου.

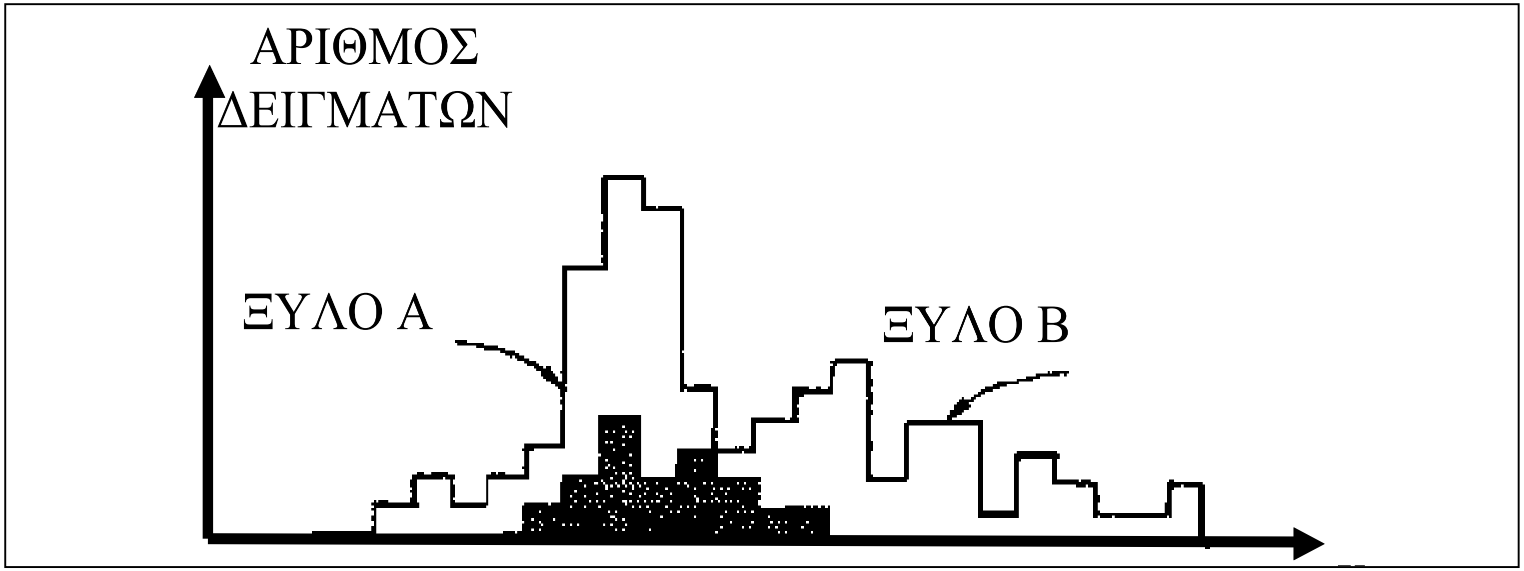
Εξαγωγή Χαρακτηριστικών

- Έστω ότι *το ξύλο A είναι πιο φωτεινόχρωμο από το ξύλο B*
- Τότε είναι προφανές ότι η φωτεινότητα, οι τιμές της οποίας αντιστοιχούν στη μεταβλητή x, αποτελεί ένα χαρακτηριστικό για τον διαχωρισμό του τύπου των ξύλων.
- Αρχικά θα πρέπει να μετρήσουμε μερικά δείγματα της φωτεινότητας x και να δούμε τα αποτελέσματα.
  - *Κατώφλι απόφασης:* x_0 *(τιμή της μεταβλητής* x*)*

Εξαγωγή Χαρακτηριστικών



- Το παραπάνω σχήμα δείχνει το ιστόγραμμα της φωτεινότητας ή την κατανομή του κάθε τύπου ξύλου ως συνάρτηση της φωτεινότητας (υπό συνθήκη πυκνότητα πιθανότητας).
- Θα ταξινομήσουμε τα ξύλα με βάση το αν η τιμή της φωτεινότητας x ξεπερνάει μια τιμή κατωφλίου x_0



- Από την μορφή των ιστογραμμάτων συμπεραίνουμε ότι το ξύλο Β είναι πιο φωτεινό από το ξύλο Α.

- Συμπεραίνουμε επίσης, ότι το χαρακτηριστικό αυτό από μόνο του δεν αρκεί για έναν χωρίς λάθη διαχωρισμό των τύπων των ξύλων αφού υπάρχουν περιοχές στις οποίες δεν μπορούμε να αποφασίσουμε καθώς οι τιμές της φωτεινότητας και για τα δύο ξύλα συμπίπτουν.

Εξαγωγή Χαρακτηριστικών

- Στην προσπάθεια για αναζήτηση άλλων χαρακτηριστικών μπορούμε να δώσουμε έμφαση στην προεξέχουσα υφή των ξύλων (τα γνωστά νερά).

- Το χαρακτηριστικό αυτό είναι πολύ πιο δύσκολο να μετρηθεί από την μέση φωτεινότητα.

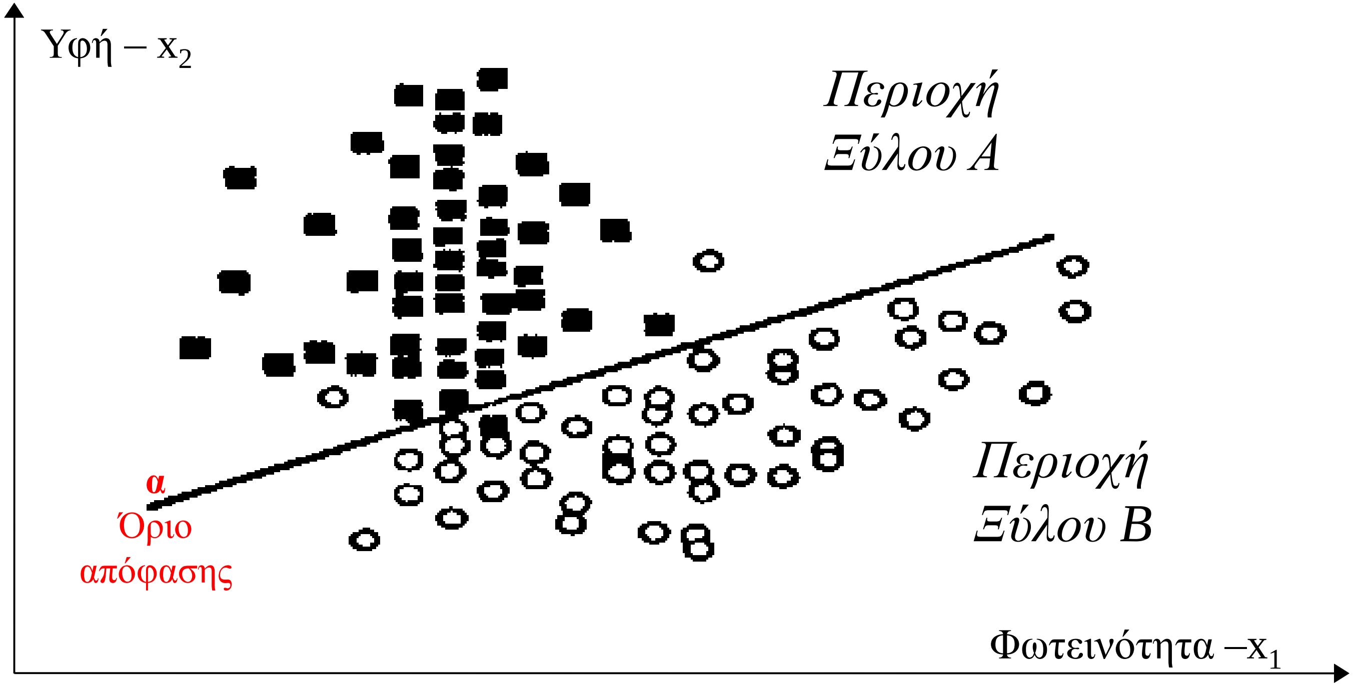
  – Είναι όμως λογικό να υποθέσουμε ότι μπορούμε να έχουμε ένα μέτρο σύγκρισης της υφής εξετάζοντας τις μεταβάσεις από φωτεινή σε σκοτεινή περιοχή.

Εξαγωγή Χαρακτηριστικών

- **Τώρα λοιπόν έχουμε δύο χαρακτηριστικά για την ταξινόμηση του τύπου του ξύλου.**
  - Έτσι ο εξαγωγέας χαρακτηριστικών έχει μετατρέψει κάθε εικόνα σε ένα σημείο ή διάνυσμα χαρακτηριστικών **x** σε ένα χώρο δυο διαστάσεων όπου:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \text{Φωτεινότητα} \\ \text{Υφή} \end{bmatrix}$$

- Επαναλαμβάνουμε τις μετρήσεις του αρχικού πληθυσμού για το διάνυσμα **x**:
  - Ας δούμε τα αποτελέσματα:



- Η μορφή του σχήματος μπορεί να δώσει τον επόμενο κανόνα για την ταξινόμηση οποιουδήποτε τύπου ξύλου που μπορεί να εμφανιστεί:
  - Ταξινόμησε το ξύλο ως τύπου Α εάν το διάνυσμα χαρακτηριστικών βρίσκεται πάνω από την γραμμή α ή ως τύπου Β σε αντίθετη περίπτωση.
- Το πρόβλημα της ταξινόμησης είναι βασικά ο διαχωρισμός του χώρου των χαρακτηριστικών σε περιοχές όπου η κάθε μια χαρακτηρίζει διαφορετική τάξη.
  - Σε αυτήν την περίπτωση το πρόβλημα της ταξινόμησης γίνεται πρόβλημα στατιστικής θεωρίας αποφάσεων.
  - Είναι φανερό ότι το πρόβλημα της εξαγωγής χαρακτηριστικών εξαρτάται από την εφαρμογή.

Θεωρία Αποφάσεων κατά Bayes

- Η θεωρία αποφάσεων του Bayes αποτελεί τη συμβατική στατιστική προσέγγιση στο πρόβλημα της θεωρίας ταξινόμησης.
- Ας θυμηθούμε το προηγούμενο παράδειγμα.
  - Εκεί είχαμε σκοπό τον σχεδιασμό ενός συστήματος που να διαχωρίζει δυο είδη ξύλων.

Θεωρία Αποφάσεων κατά Bayes

Χρησιμοποιώντας την ορολογία της στατιστικής λέμε ότι: *καθώς ένα* ***καινούργιο ξύλο μπαίνει στο πριονιστήριο*** *είναι δυνατόν να επιλεγεί μια από τις δύο δυνατές κατηγορίες ξύλου*:

- Ας ονομάσουμε $\boldsymbol{\omega}$ την μεταβλητή που απεικονίζει την κατηγορία ξύλου, με $\boldsymbol{\omega}=\boldsymbol{\omega}_1$ να αναπαριστά το είδος του ξύλου Α και $\boldsymbol{\omega}=\boldsymbol{\omega}_2$ να αναπαριστά το είδος του ξύλου Β.

Λόγω του ότι η φύση είναι απρόβλεπτη στην εκλογή της λέμε ότι η μεταβλητή $\boldsymbol{\omega}$ είναι μια:

Τυχαία Μεταβλητή (ΤΜ - RV).

Εκ των προτέρων πιθανότητα (A-priori)

- Υποθέτουμε ότι υπάρχει μια εκ των προτέρων πιθανότητα (a-*priori*) P($\boldsymbol{\omega}_1$) το επόμενο κομμάτι ξύλου να είναι τύπου **A** και μια εκ των προτέρων πιθανότητα P($\boldsymbol{\omega}_2$) το επόμενο κομμάτι ξύλου να είναι τύπου **B**.
  – *Εάν το πριονιστήριο παρήγαγε τόση ξυλεία από το πρώτο είδος όση και από το δεύτερο είδος, θα μπορούσαμε να πούμε ότι το επόμενο κομμάτι που θα εισέλθει στο πριονιστήριο θα είναι είτε του τύπου A είτε του τύπου B με πιθανότητα ίδια και για τα δύο ξύλα*.

Εκ των προτέρων πιθανότητα (A-priori)

- Οι a-priori πιθανότητες **αντανακλούν την προηγούμενη γνώση μας** ότι θα δούμε ένα από τα δύο είδη πριν την εμφάνιση του ξύλου στο πριονιστήριο.
- Είναι προφανές ότι $P(\boldsymbol{\omega}_1)$ και $P(\boldsymbol{\omega}_2)$ είναι μη αρνητικές ποσότητες και το άθροισμά τους ισούται με την μονάδα.

$$P(\boldsymbol{\omega}_1)+ P(\boldsymbol{\omega}_2)=1$$

Απόφαση για το Άγνωστο

- Ας υποθέσουμε ότι είμαστε υποχρεωμένοι να πάρουμε μια απόφαση για τον τύπο του ξύλου που θα εμφανιστεί αμέσως μετά.
- **<u>Αν η μόνη πληροφορία που έχουμε</u>** αφορά τις εκ των προτέρων πιθανότητες P($\boldsymbol{\omega}_1$) και P($\boldsymbol{\omega}_2$) τότε θα ήτανε αρκετά λογικό να διατυπώσουμε τον ακόλουθο κανόνα αποφάσεως:

Απόφαση για το Άγνωστο

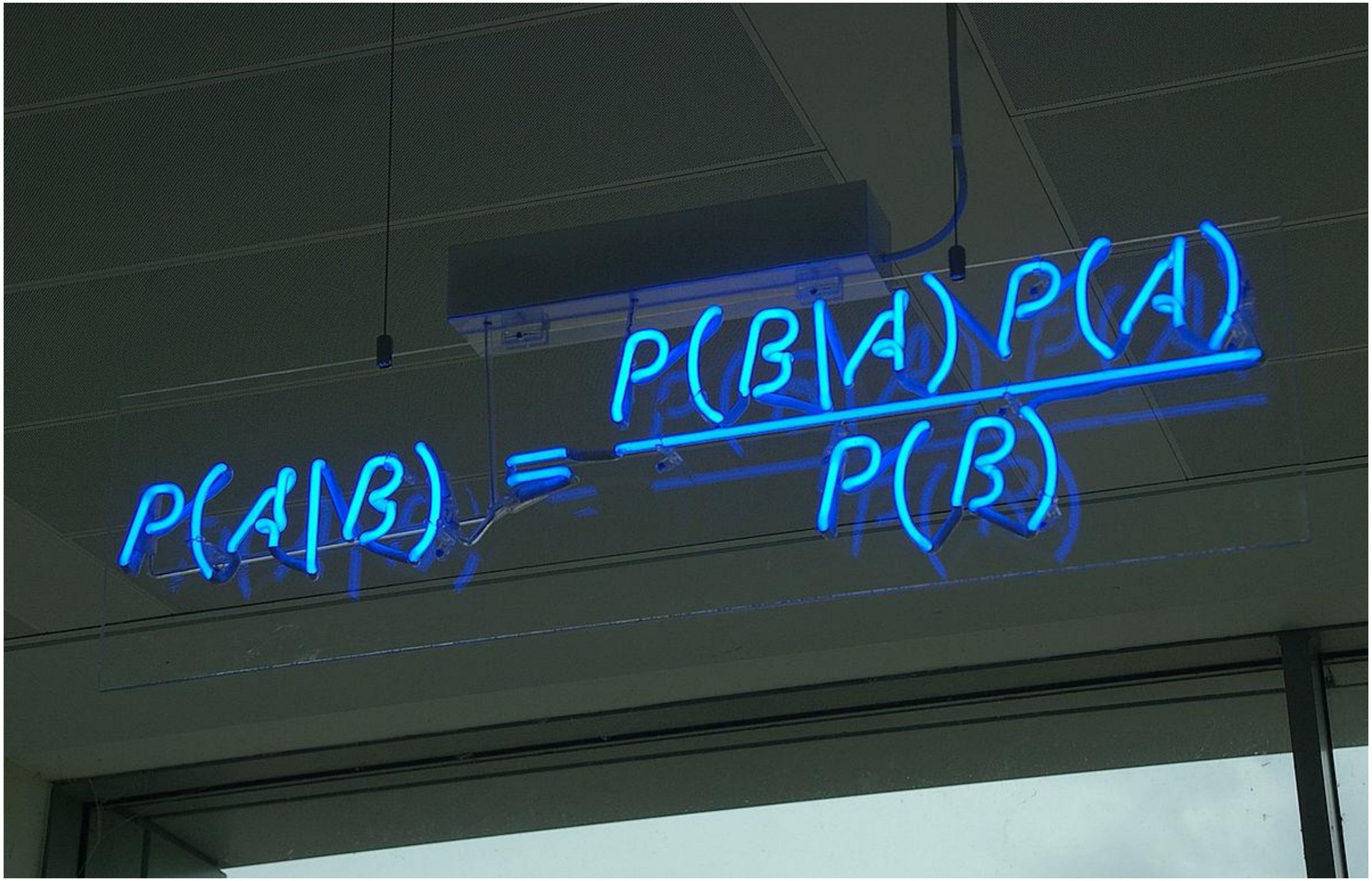
Αποφάσισε $\boldsymbol{\omega}_1$ εάν $P(\boldsymbol{\omega}_1) > P(\boldsymbol{\omega}_2)$ αλλιώς αποφάσισε $\boldsymbol{\omega}_2$

- Θα κάνουμε πάντα την ίδια εκτίμηση έστω και αν γνωρίζουμε ότι υπάρχουν δύο τύποι ξύλων.
- Το πόσο καλά αυτή η απόφαση θα δουλέψει εξαρτάται από τις τιμές των a-priori πιθανοτήτων.
  - *Εάν η τιμή του $P(\boldsymbol{\omega}_1)$ είναι πολύ μεγαλύτερη από την τιμή του $P(\boldsymbol{\omega}_2)$ η απόφασή μας υπέρ του $\boldsymbol{\omega}_1$ θα είναι σωστή τις περισσότερες φορές.*
  - *Εάν $P(\boldsymbol{\omega}_1) = P(\boldsymbol{\omega}_2)$ έχουμε πιθανότητα 50% να είμαστε σωστοί.*

Η πιθανότητα του να πάρω την λανθασμένη απόφαση είναι:

<u>Η μικρότερη</u> από τις $\{P(\omega_1), P(\omega_2)\}$.

Κανόνας του Bayes



Κανόνας του Bayes

$$posterior = \frac{likelihood \times prior}{evidence}$$

Κανόνας του Bayes

$$P\left(\mathrm{A}\middle|\mathrm{B}\right) = \frac{P(B|\mathrm{A}) \cdot P(A)}{P(B)}$$

- **P(A), P(B):** A priori πιθανότητες των γεγονότων A, B.
- **P(A | B):** Υπό συνθήκη πιθανότητα της παρατήρησης του γεγονότος A δεδομένου ότι το γεγονός B συνέβη.
- **P(A | B):** Εκ των υστέρων (a posteriori) πιθανότητα του A δεδομένου του B.

Πιθανότητα υπό Συνθήκη

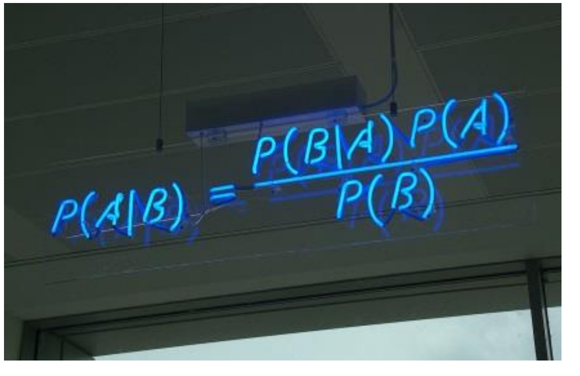
Ορισμός: $$P\left(\mathrm{A}\middle|\mathrm{B}\right) = \frac{P(A \cap B)}{P(B)}$$

$$\mathrm{P}(\mathrm{A} \cap B) = \mathrm{P}\left(\mathrm{A}\middle|\mathrm{B}\right) \cdot P(\mathrm{B})$$

$$P(A \cap B) = P(B \cap A)$$

$$P(B \cap A) = P\left(\mathrm{A}\middle|\mathrm{B}\right) \cdot P(A)$$

$$P\left(\mathrm{A}\middle|\mathrm{B}\right) \cdot P(B) = P\left(B\middle|\mathrm{A}\right) \cdot P(A)$$



Τυχαία μεταβλητή (RV)

- Ας θεωρήσουμε x μια τυχαία μεταβλητή.
  - *Θεωρούμε ότι η κατανομή της εξαρτάται από την κατάσταση της φύσης ($\boldsymbol{\omega_1}$, $\boldsymbol{\omega_2}$)*

$$p(x/\boldsymbol{\omega_1}),\ p(x/\boldsymbol{\omega_2})$$

- $p(x/\boldsymbol{\omega_j})$: Υπό συνθήκη πυκνότητα πιθανότητας της τυχαίας μεταβλητής x δεδομένης της κατάστασης $\boldsymbol{\omega_j}$

$$p(x/\boldsymbol{\omega_1})\text{: πιο σωστά } p_x(x/\boldsymbol{\omega_1})$$

Πιθανότητα και Κατανομή τυχαίας μεταβλητής

- Έστω X μια τυχαία μεταβλητή.
  - p(x) ή $p_X(x)$ ή $f_X(x)$: Η πυκνότητα πιθανότητας **(ή κατανομή)** της τυχαίας μεταβλητής X.

$$\Pr(a \le X \le b) = \mathrm{P}(a \le X \le b) = \int_a^b p_X(x)\, dx$$

- **Διαισθητικά:** $p_X(x)dx$ είναι η πιθανότητα της τυχαίας μεταβλητής X να λάβει τιμή μεταξύ [x, x+dx]

$$P_X(x) = \int_{-\infty}^{x} p_X(\mathrm{u})\, du$$

$$p_X(\mathrm{x}) = \frac{d}{dx} P_X(\mathrm{x}) = \frac{d}{dx}\left(\int_{-\infty}^{x} p_X(u)\, du\right)$$

Τυχαία μεταβλητή

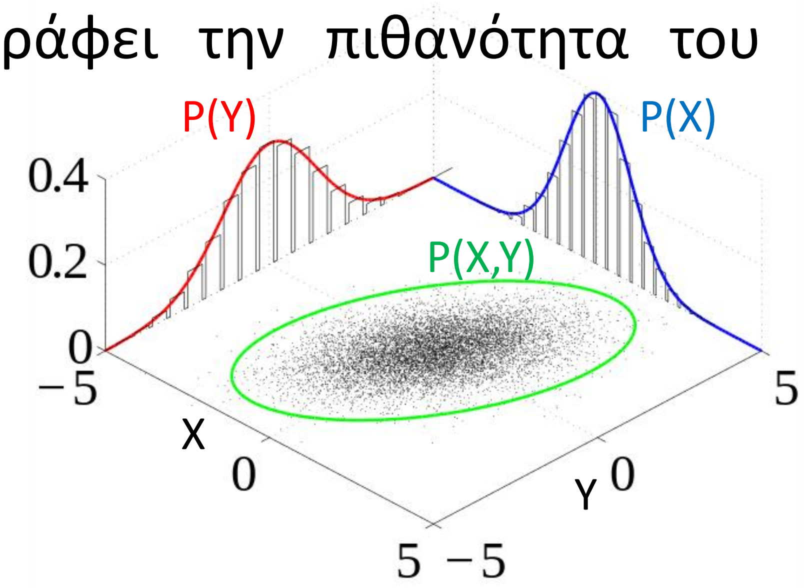
$p_X(x/\boldsymbol{\omega_1})$ - $p_X(x/\boldsymbol{\omega_2})$

Περιγραφή της διαφοράς του χαρακτηριστικού x (π.χ: Φωτεινότητα)

μεταξύ των πληθυσμών που αντιπροσωπεύονται από τις τάξεις $\boldsymbol{\omega_1}$, $\boldsymbol{\omega_2}$

Από κοινού κατανομή

- Δεδομένου **δύο** τυχαίων μεταβλητών X, Y, η **από κοινού κατανομή** (joint probability distribution) για τις X, Y, είναι μια κατανομή (probability distribution) που περιγράφει την πιθανότητα του συνδυασμού των γεγονότων:



- X λαμβάνει τιμές μεταξύ:
  - Είτε κάποιων διακριτών τιμών
  - Είτε σε ένα διάστημα τιμών.
- Y λαμβάνει τιμές μεταξύ:
  - Είτε κάποιων διακριτών τιμών
  - Είτε σε ένα διάστημα τιμών.
- Αν έχω 2 τυχαίες μεταβλητές (**bivariate distribution**)
- Αν έχω πολλές τυχαίες μεταβλητές (**multivariate distribution**)

Ιδιότητες

$$P(\mathrm{X}=x \textbf{ και } Y=y) @ P(\mathrm{X}=x, Y=y)=$$

$$P(Y=y \big| \mathrm{X}=x)\cdot P(\mathrm{X}=x)=P(X=x \big| Y=y)\cdot P(Y=y)$$

$$\sum_i \sum_j \mathrm{P}(X=x_i \text{ and } Y=y_j)=1,$$

$$\begin{aligned}\mathrm{P}(X_1=x_1,\ldots,X_n=x_n) &= \mathrm{P}(X_1=x_1)\times\mathrm{P}(X_2=x_2\mid X_1=x_1)\\ &\times\mathrm{P}(X_3=x_3\mid X_1=x_1,X_2=x_2)\\ &\ldots\\ &\times P(X_n=x_n\mid X_1=x_1,X_2=x_2,\ldots,X_{n-1}=x_{n-1})\end{aligned}$$

$$\sum_i \sum_j \ldots \sum_k \mathrm{P}(X_1=x_{1i},X_2=x_{2j},\ldots,X_n=x_{nk})=1.$$

Απόφαση κατά Bayes

- Μπορούμε να θεωρήσουμε την φωτεινότητα *x* ως μια συνεχή τυχαία μεταβλητή, η κατανομή της οποίας εξαρτάται από την εκλογή της κατηγορίας του ξύλου.
  - Η πυκνότητα πιθανότητας $p_X(x/\boldsymbol{\omega}j)$ αφορά την κατανομή της μεταβλητής *X* υπό την συνθήκη ότι έχουμε την κατηγορία $\boldsymbol{\omega}_j$ (conditional density).

- Αν οι $P(\boldsymbol{\omega}_1)$ και $P(\boldsymbol{\omega}_2)$ και οι $p(x/\boldsymbol{\omega}_1)$ και $p(x/\boldsymbol{\omega}_2)$ είναι γνωστές, τότε η μέτρηση της φωτεινότητας *X* του ξύλου επηρεάζει την γνώμη μας για το πραγματικό είδος του ξύλου σύμφωνα με τον κανόνα του Bayes.

Συνεχής μεταβλητή

Κατανομή

$$p(\omega_j, x) = P(\omega_j \mid x)\, p(x) = p(x \mid \omega_j)\, P(\omega_j)$$

Διακριτή μεταβλητή

Πιθανότητα

Κατανομή

$$p(x) = \sum_{j=1}^{2} p\left(x \middle| \omega_j\right) P\left(\omega_j\right)$$

$$P\left(\omega_j \middle| x\right) = \frac{p\left(x \middle| \omega_j\right) P\left(\omega_j\right)}{p(x)} = \frac{p\left(x \middle| \omega_j\right) P\left(\omega_j\right)}{\sum_{j=1}^{2} p\left(x \middle| \omega_j\right) P\left(\omega_j\right)}$$

Απόφαση για το Άγνωστο

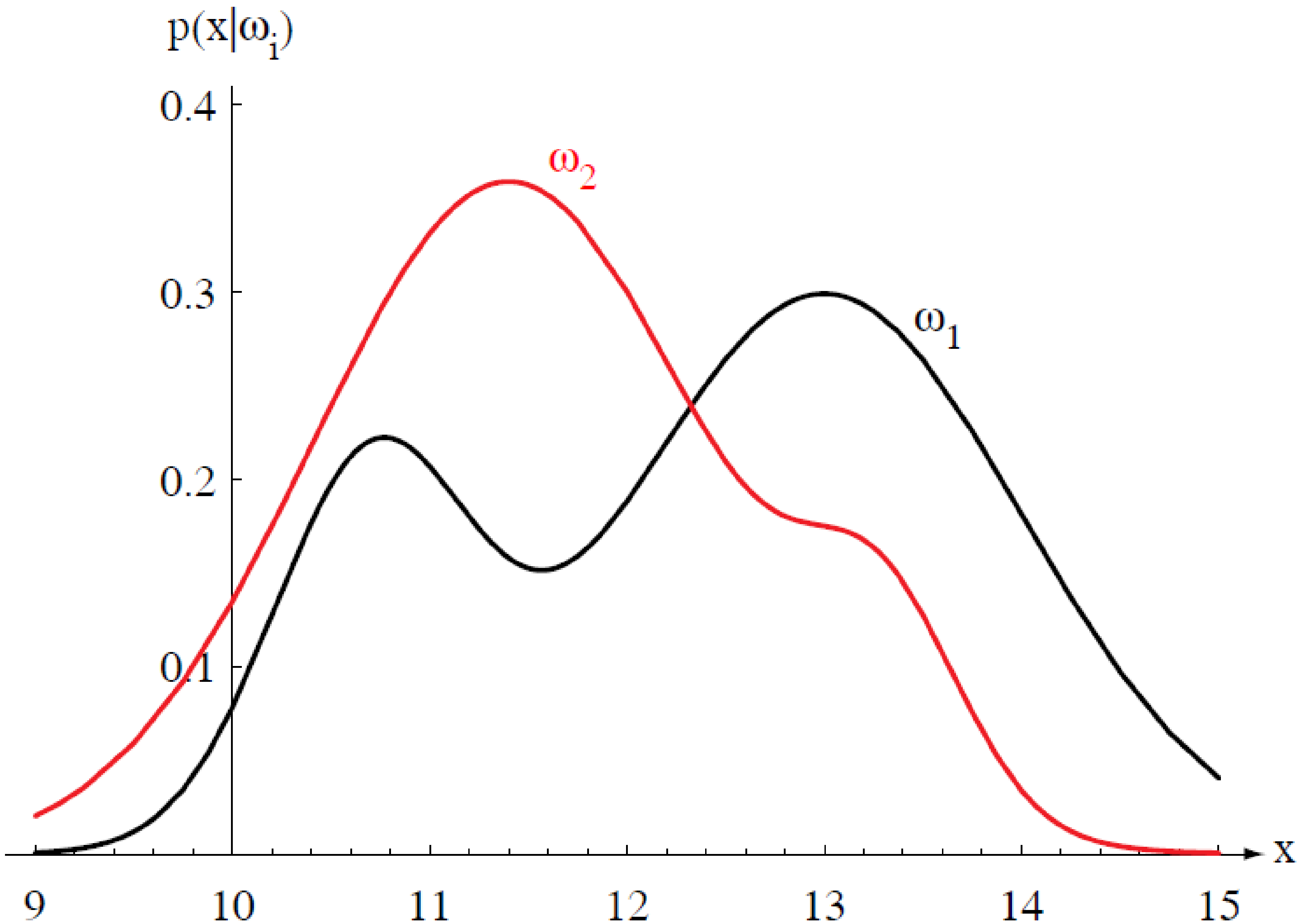


Figure 2.1: Hypothetical class-conditional probability density functions show the probability density of measuring a particular feature value x given the pattern is in category ω_i. If x represents the length of a fish, the two curves might describe the difference in length of populations of two types of fish. Density functions are normalized, and thus the area under each curve is 1.0.

Απόφαση για το Άγνωστο (2)

Εάν $P(\omega_1 | x) > P(\omega_2 | x)$
τότε: επέλεξε ω_1
αλλιώς
επέλεξε ω_2

- Ο κανόνας του Bayes δείχνει πως παρατηρώντας την τιμή της μεταβλητής *x* έχουμε αποδώσει σε κάθε μια από τις δύο κατηγορίες μια πιθανότητα να συμβεί (a-posteriori propability)

Ισοδύναμα,
Επέλεξε ω_1
εάν $p(x|\omega_1)P(\omega_1) > p(x|\omega_2)P(\omega_2)$,
αλλιώς επέλεξε ω_2.

Απόφαση για το Άγνωστο (3)

- Ο τύπος του Bayes δείχνει το πώς η μέτρηση μια τιμής x της τυχαίας μεταβλητής X, μετατρέπει την a-priori γνώση-πιθανότητα $P(\omega_j)$ στην εκ των υστέρων πιθανότητα $P(\omega_j/x)$.

- $P(\omega_j/x)$: η πιθανότητα η φύση να είναι στην κατάσταση ω_j όταν μετράμε το χαρακτηριστικό X να έχει τιμή x.
- $p(x/\omega_j)$ είναι η πιθανοφάνεια (likelihood) της κατάστασης ω_j *σε σχέση με την τιμή* x της μεταβλητής – χαρακτηριστικό X.

- **Φυσική ερμηνεία** $p(x/\omega_j)$: Η κατηγορία ω_j για την οποία η $p(x/\omega_j)$ είναι μεγάλη είναι πιθανό να είναι η σωστή κατηγορία.

- Το γινόμενο $p(x/\omega_j)$ x $P(\omega_j)$ είναι το πιο σημαντικό. Ο παράγοντας τεκμηρίωσης **evidence** , $p(x)$, είναι απλά μια κλίμακα ώστε το άθροισμα των εκ των υστέρων πιθανοτήτων = 1

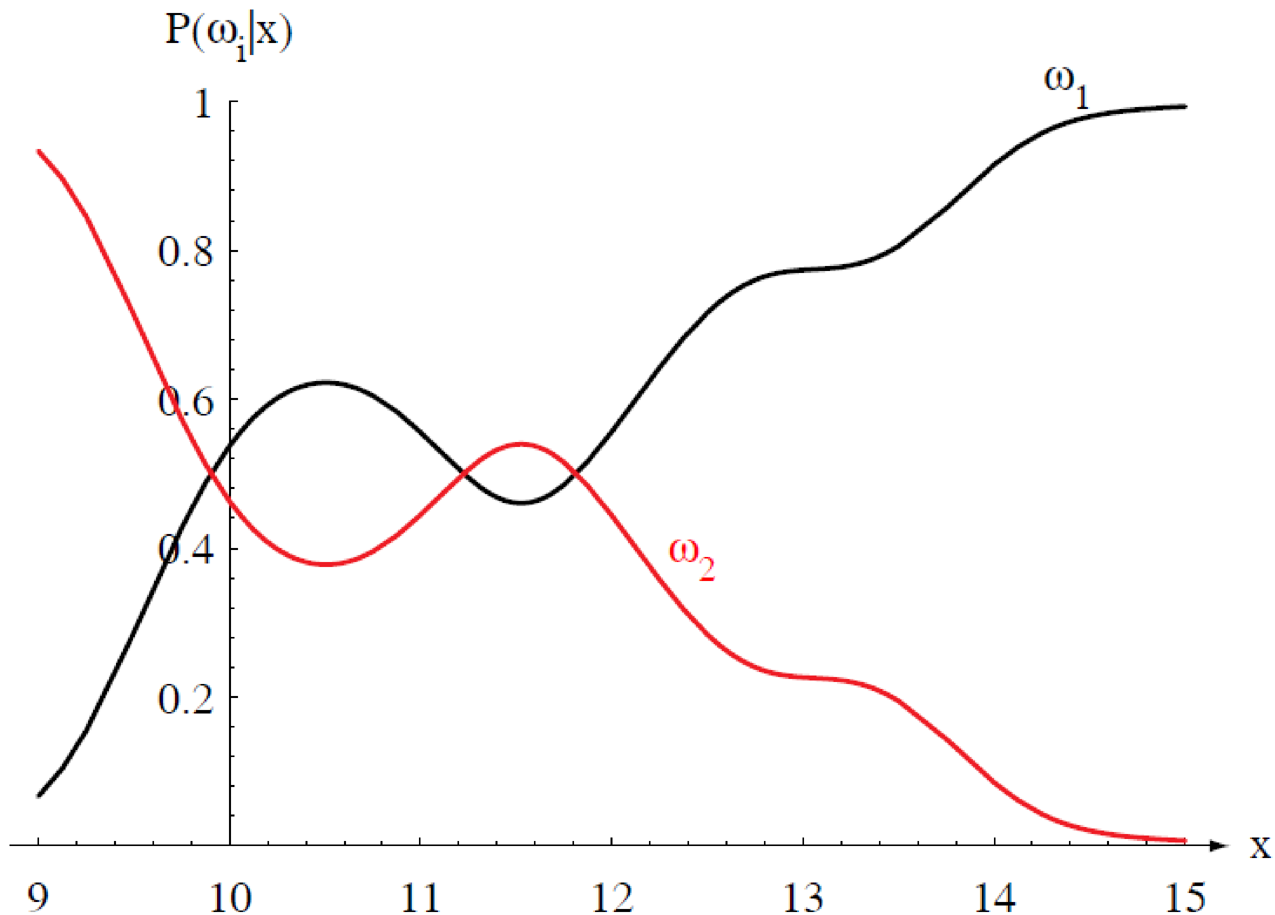


Figure 2.2: Posterior probabilities for the particular priors $\mathrm{P}(\omega_1) = 2/3$ and $\mathrm{P}(\omega_2) = 1/3$ for the class-conditional probability densities shown in Fig. 2.1. Thus in this case, given that a pattern is measured to have feature value $x = 14$, the probability it is in category ω_2 is roughly 0.08, and that it is in ω_1 is 0.92. At every x, the posteriors sum to 1.0.

Είδη σφαλμάτων

Εάν $P(\omega_1 | x) > P(\omega_2 | x)$

τότε: επέλεξε ω_1

αλλιώς

επέλεξε ω_2

- Μπορεί να αποφασίσουμε $\boldsymbol{\omega_1}$ αλλά η πραγματική κατάσταση της φύσης να είναι $\boldsymbol{\omega_2}$.

ή

- Μπορεί να αποφασίσουμε $\boldsymbol{\omega_2}$ αλλά η πραγματική κατάσταση της φύσης να είναι $\boldsymbol{\omega_1}$.

Πιθανότητα σφάλματος

$$P(Error|x) = \begin{cases} P(\omega_1|x), \text{ αν αποφασίσουμε } \omega_2 \\ P(\omega_2|x), \text{ αν αποφασίσουμε } \omega_1 \end{cases}$$

Για δεδομένο x, μπορούμε να ελαχιστοποιήσουμε την πιθανότητα του σφάλματος αν:

Επέλεξε ω_1
Εάν $P(\omega_1|x) > P(\omega_2|x)$
αλλιώς επέλεξε ω_2

Πιθανότητα σφάλματος

- Ο προηγούμενος κανόνας ελαχιστοποιεί την μέση πιθανότητα του σφάλματος.

$$P(Error) = \int_{-\infty}^{+\infty} P(Error, x)dx = \int_{-\infty}^{+\infty} P(Error \mid x)p(x)dx$$

Επέλεξε ω_1
Εάν $P(\omega_1 | x) > P(\omega_2 | x)$
αλλιώς επέλεξε ω_2

$$\mathrm{P}(Error \mid x) = \begin{cases} \mathrm{P}(\omega_1 \mid x) \\ \mathrm{P}(\omega_2 \mid x) \end{cases}$$

$$P(Error \mid x) = \min[P(\omega_1 \mid x), P(\omega_2 \mid x)]$$

Πιθανότητα σφάλματος

- Στην περίπτωση της ταξινόμησης σε δυο κατηγορίες ο ταξινομητής έχει διαιρέσει τον χώρο σε δυο περιοχές τις οποίες ονομάζουμε R_1 και R_2 αντίστοιχα.
- Υπάρχουν δυο τρόποι για να συμβεί ένα λάθος στην ταξινόμηση ενός αντικειμένου.
  - Είτε μια μέτρηση **x** να εμπίπτει στην περιοχή R_2 ενώ η αληθινή κατηγορία να είναι η ω_1.
  - Είτε να εμπίπτει στην περιοχή R_1 ενώ η αληθινή κατηγορία να είναι η ω_2.

Πιθανότητα σφάλματος

- Τα δυο αυτά γεγονότα είναι αμοιβαία αποκλειόμενα και διαμερίζουν πλήρως τον χώρο πιθανοτήτων, άρα:

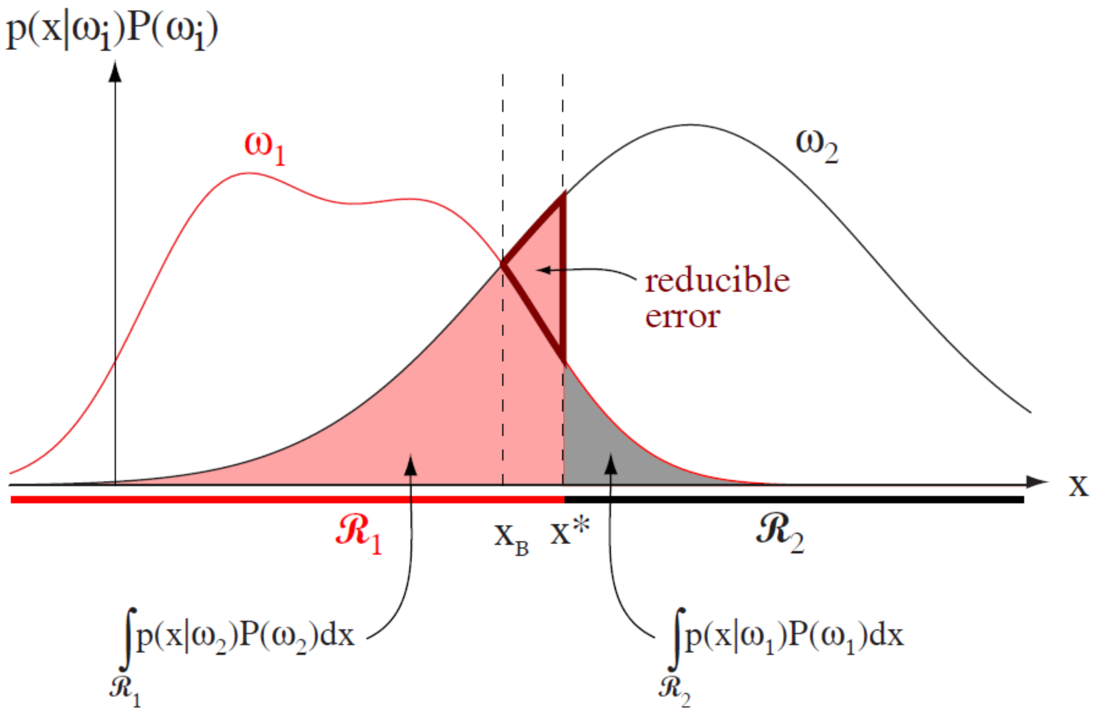
$$P(\text{σφάλματος}) = P(x \in R_2, \omega_1) + P(x \in R_1, \omega_2) =$$

$$P(x \in R_2 \mid \omega_1)P(\omega_1) + P(x \in R_1 \mid \omega_2)P(\omega_2) =$$

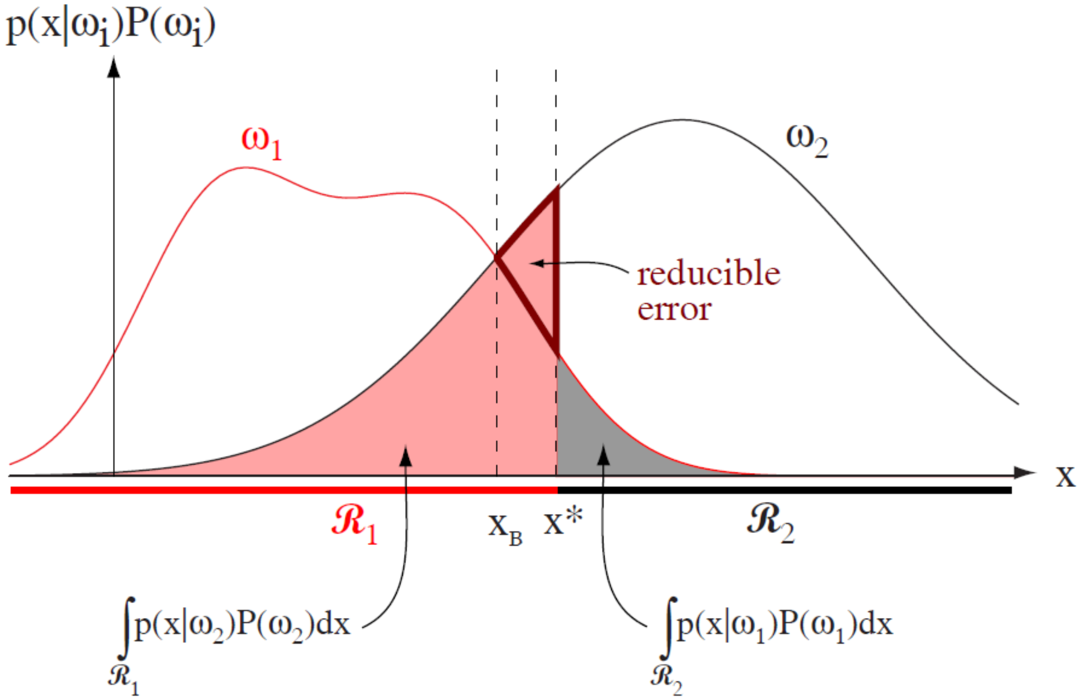
$$\int_{R_2} p(x \in R_2 \mid \omega_1)P(\omega_1)dx + \int_{R_1} p(x \in R_1 \mid \omega_2)P(\omega_2)dx$$

Πιθανότητα σφάλματος

- Λόγω της αυθαίρετης επιλογής των περιοχών R_1 και R_2 η πιθανότητα του λάθους δεν είναι όσο μικρή θα έπρεπε.
- Ακολούθως, μετακινώντας το όριο που σηματοδοτεί την λήψη απόφασης προς την μια ή την άλλη μεριά μπορούμε να εξαλείψουμε την άνω έντονα γραμμοσκιασμένη περιοχή και να λάβουμε την μικρότερη πιθανότητα του σφάλματος.

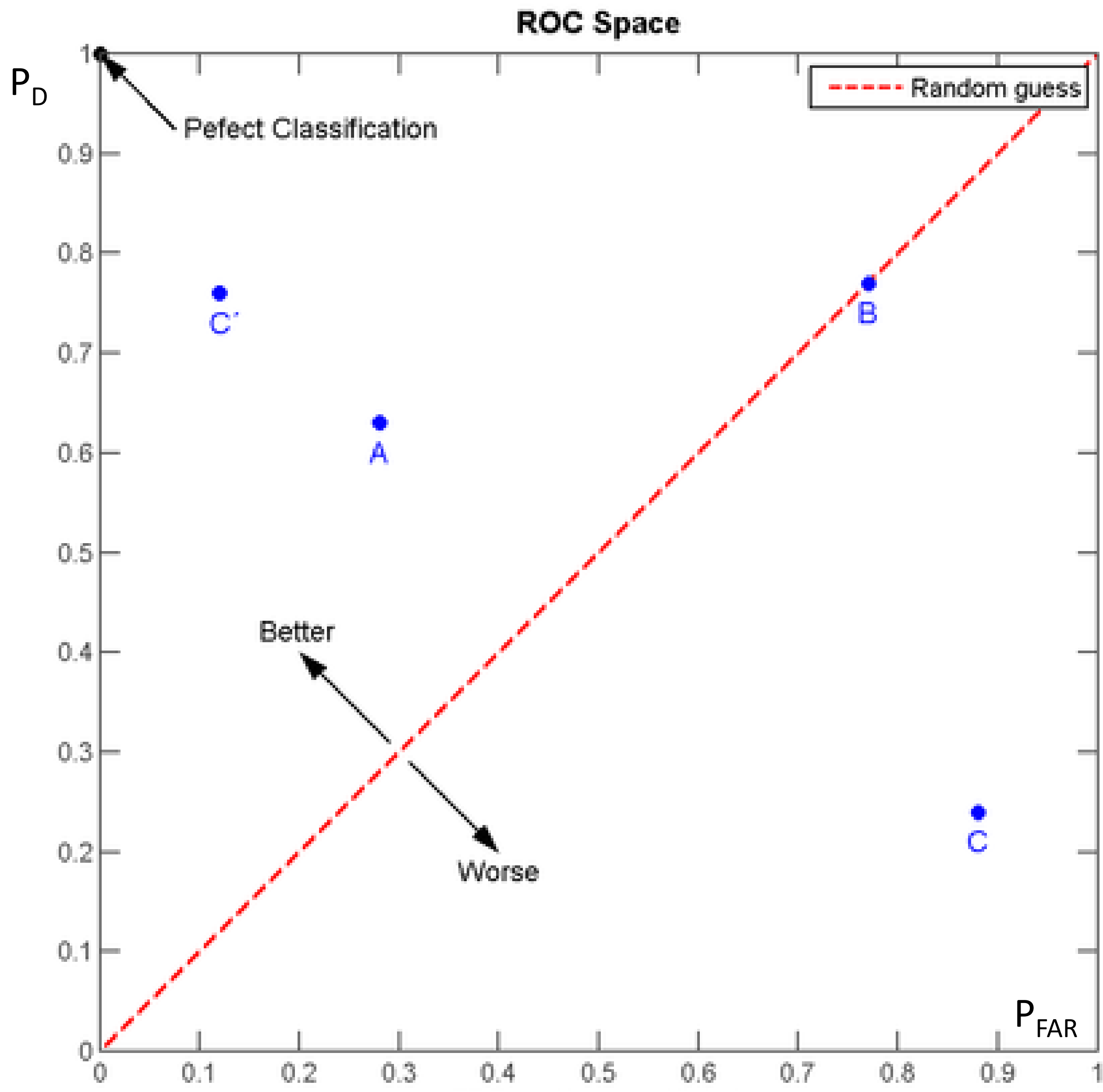


Μέτρηση απόδοσης



- Εσφαλμένη αποδοχή. (False acceptance) FAR - P_{FAR}
- Εσφαλμένη απόρριψη. (False rejection) FRR - P_{FRR}
- P_D=1- P_{FRR} = Πιθανότητα σωστής αποδοχής

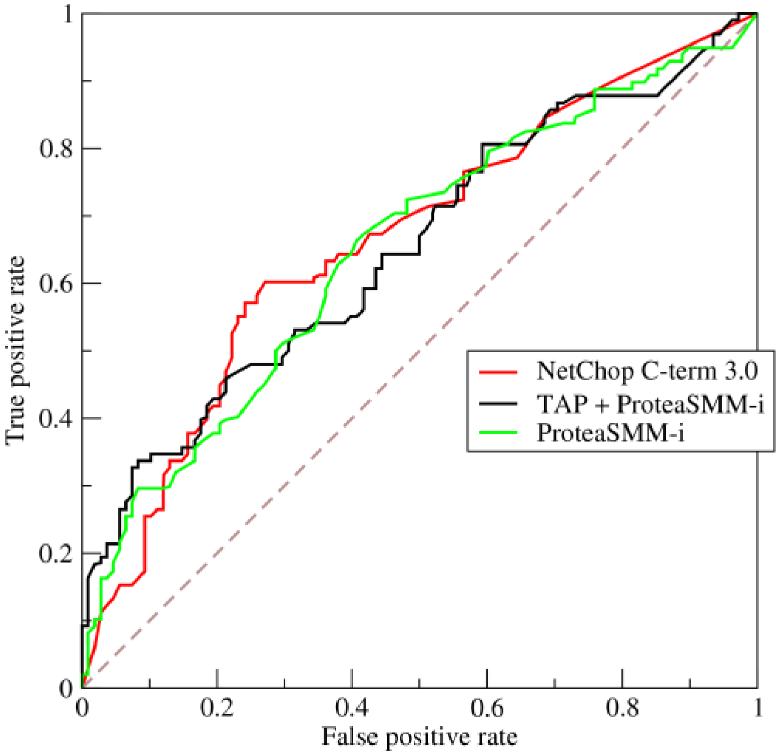
Receiver operating characteristic (ROC)



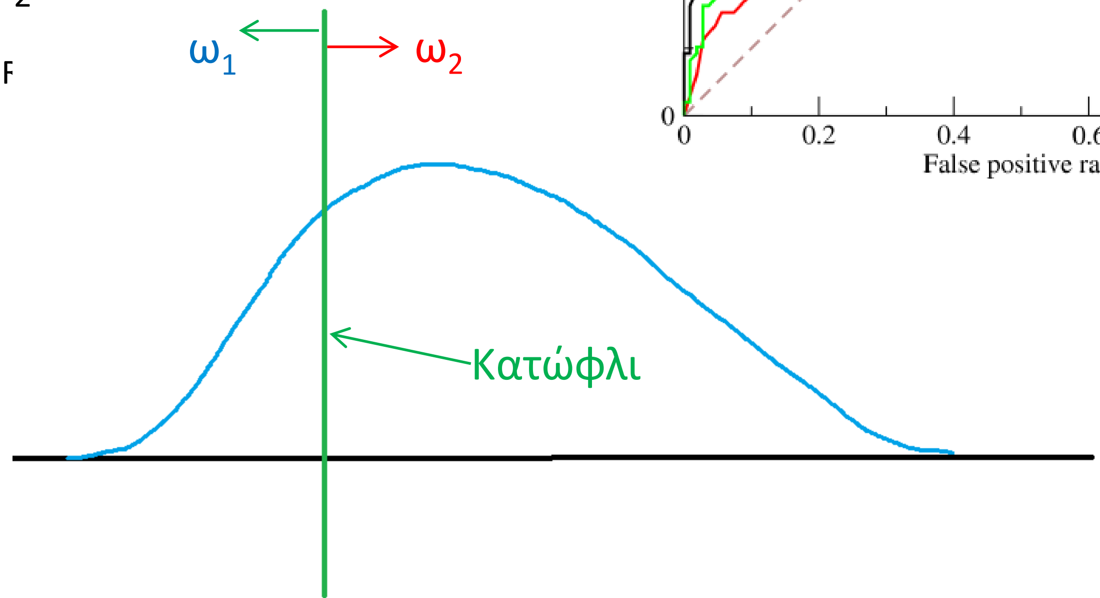
Πίνακας σύγχυσης

		predicted condition	
	total population	prediction positive	prediction negative
true condition	condition positive	**True Positive (TP)**	**False Negative (FN)** (type II error)
	condition negative	**False Positive (FP)** (Type I error)	**True Negative (TN)**

Receiver operating characteristic (ROC)



- Κατώφλι τέρμα αριστερά.
  - Όλα ω_2 κανένα ω_1
  - $\mathbf{P_{FAR}=1, P_D=0,}$ $P_{FRR}=1$
- Κατώφλι τέρμα δεξιά.
  - Όλα ω_1 κανένα ω_2
  - $\mathbf{P_{FAR}=0, P_D=1,}$ P_{FRF}



Τεκμηρίωση - Evidence – $p_X(x)$

$$p(\mathrm{x}) = \sum_{\mathrm{j}=1}^{2} \mathrm{p}\left(x \middle| \omega_{\mathrm{j}}\right) \mathrm{P}\left(\omega_{\mathrm{j}}\right)$$

- Είναι μια συνάρτηση
- Δεν είναι απαραίτητη για να λάβουμε μια απόφαση.
- Είναι απλά ένας παράγοντας κλιμάκωσης
- Η παρουσία της εξασφαλίζει ότι:

$$P(\omega_1 | x) + P(\omega_2 | x) = 1$$

Επέλεξε ω_1

Εάν:

$$p(x \mid \omega_1)P(\omega_1) > p(x \mid \omega_2)P(\omega_2)$$

αλλιώς

επέλεξε ω_2

Εάν:

$$p(x \mid \omega_1) = p(x \mid \omega_2)$$

Τότε όλα εξαρτώνται από τα $P(\omega_1)$, $P(\omega_2)$

Εάν:

$$P(\omega_1) = P(\omega_2)$$

Τότε όλα εξαρτώνται από τα $p(x|\omega_1)$, $p(x|\omega_2)$

Κανόνας Απόφασης Bayes - Προπόνηση

- Αν οι υπό συνθήκη πιθανότητες για τις τάξεις ω_1, ω_2 δίνονται από τις παρακάτω σχέσεις:
- Αν επίσης $P(\omega_1) = P(\omega_2)$ εξάγετε κανόνα απόφασης.

$$P(x \mid \omega_1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-4)^2}$$

$$P(x \mid \omega_2) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-10)^2}$$

Κανόνας Απόφασης Bayes - Λύση

$$P(x\,|\,\omega_1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-4)^2} \qquad P(x\,|\,\omega_2) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-10)^2}$$

- Αντικαθιστώντας στον κανόνα:

$$\boxed{\Lambda(x) = \frac{P(x\,|\,\omega_1)}{P(x\,|\,\omega_2)} \underset{\omega_2}{\overset{\omega_1}{\gtrless}} \frac{P(\omega_2)}{P(\omega_1)}} \quad \Rightarrow \quad \Lambda(x) = \frac{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-4)^2}}{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-10)^2}} \underset{\omega_2}{\overset{\omega_1}{\gtrless}} \frac{1}{1}$$

- Απλοποιώντας

$$\Lambda(x) = \frac{e^{-\frac{1}{2}(x-4)^2}}{e^{-\frac{1}{2}(x-10)^2}} \underset{\omega_2}{\overset{\omega_1}{\gtrless}} 1,$$

- Λογαριθμώντας

$$(x-4)^2 - (x-10)^2 \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} 0 \quad \Rightarrow \quad x \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} 7$$

<u>Γενικευμένη Περίπτωση</u>

- Οι ιδέες που αναφέραμε παραπάνω γενικεύονται με τις εξής προσθήκες.
  A. Θα επιτρέψουμε περισσότερα του ενός χαρακτηριστικά (το **x** γίνεται άνυσμα).

Επιτρέποντας παραπάνω από ένα χαρακτηριστικό απλά αντικαθιστούμε την μεταβλητή *x* με ένα διάνυσμα χαρακτηριστικών (feature vector) **x (x: συμβολισμός με έντονα γράμματα).**

Γενικευμένη Περίπτωση

- Οι ιδέες που αναφέραμε παραπάνω γενικεύονται με τις εξής προσθήκες.
  B. Θα επιτρέψουμε περισσότερες από δυο δυνατές κατηγορίες.

Ας ονομάσουμε $\Omega=\{\omega_1, . . . , \omega s\}$ το πεπερασμένο σύνολο των δυνατών κατηγοριών…

Γενικευμένη Περίπτωση

- Οι ιδέες που αναφέραμε παραπάνω γενικεύονται με τις εξής προσθήκες.
  C. Θα επιτρέψουμε παραπάνω δράσεις από το να δηλώσουμε απλά την κατάσταση της φύσης.

Π.χ. Μπορούμε να επιτρέψουμε την δράση της απόρριψης (την άρνηση του να πάρουμε μια απόφαση) , ειδικά σε καταστάσεις που το κόστος είναι μικρό.

Γενικευμένη Περίπτωση

- Οι ιδέες που αναφέραμε παραπάνω γενικεύονται με τις εξής προσθήκες.
    D. Θα επιτρέψουμε και άλλα μέτρα κόστους (loss functions) σε σχέση με την πιθανότητα του σφάλματος.
- Συνάρτηση κόστους (loss function): Πόσο πολύ κοστίζει μια απόφαση.
- Συνάρτηση κόστους (loss function): Μετατρέπει την πιθανότητα σε απόφαση.
- Αντιμετωπίζει καταστάσεις κατά τις οποίες **μερικές αποφάσεις κοστίζουν περισσότερο από κάποιες άλλες**.
    - **Απλοποίηση:** τα κόστη είναι ίδια.

Γενικευμένη Περίπτωση

- Η εκ των υστέρων πιθανότητα $P(\omega_j|\mathbf{x})$ μπορεί να υπολογιστεί από την $p(\mathbf{x}|\omega_j)$ με χρήση του κανόνα του Bayes:

$$P(\omega_j|\mathbf{x}) = \frac{p(\mathbf{x}|\omega_j)P(\omega_j)}{\sum_{j=1}^{S} p(\mathbf{x}|\omega_j)P(\omega_j)}$$

- Οι πιθανότητες $P(\omega_j|\mathbf{x})$ έχουν άθροισμα μονάδα.

$$\sum_{j=1}^{S} p(\omega_j|\mathbf{x}) = 1$$

Γενικευμένη Περίπτωση

- Θα αποφασίσουμε για την πιθανότερη κατηγορία από την μεγαλύτερη τιμή της:

$$P\left(\omega_j \middle| \mathbf{x}\right), j = 1, \ldots, S$$

- Άρα:

$$\arg\max_j (P\left(\omega_j \middle| x\right))$$

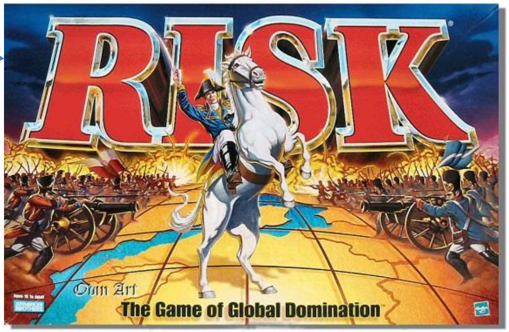
με j=1,..,S.

Γενικευμένη Περίπτωση

- Ας υποθέσουμε ότι παρατηρώντας **x** εμείς επιλέγουμε την δράση – πράξη α_i.
- Εάν η πραγματική κατάσταση της φύσης είναι ω_j τότε έχουμε την συνάρτηση κόστους $\lambda(\alpha_i|\omega_j)$.
- Εφόσον $P(\omega_j|x)$είναι η πιθανότητα της ύπαρξης της κατάστασης ω_j τότε το εκτιμώμενο κόστος (expected loss) της δράσης α_i είναι:

Υπό συνθήκας ρίσκο

$$R\left(\alpha_i \middle| \mathbf{x}\right) = \sum_{j=1}^{S} \lambda(\alpha_i|\omega_j) P\left(\omega_j \middle| \mathbf{x}\right)$$



Γενικευμένος Κανόνας Απόφασης

- (General Decision Rule - ΓΚΑ): είναι η συνάρτηση $\alpha(\mathbf{x})$ που μας υποδεικνύει την δράση που πρέπει να αναλάβουμε για κάθε **x.**
- $\alpha(\mathbf{x}) \in \{\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_a\}$
- Μερικό ρίσκο $R(\alpha_i|\mathbf{x})$ (συνυφασμένο με την (α_i)
- Ολικό ρίσκο: R (συνυφασμένο με τον ΓΚΑ)

$$R = \int R\left(\alpha(\mathbf{x}) \,|\, \mathbf{x}\right) p(\mathbf{x})\, \mathrm{d}\,\mathbf{x},\ \ \mathbf{x} \in \Re^d$$

- Αν επιλέξω $\alpha(\mathbf{x})$ να είναι μικρό για κάθε **x** τότε το R είναι το ελάχιστο.

Απόφαση κατά 

- Για να ελαχιστοποιήσουμε το R:
  - Υπολογίσατε το ρίσκο υπό συνθήκη:

$$R\left(\alpha_i(\mathbf{x})\middle|\mathbf{x}\right) = R\left(\alpha_i\middle|\mathbf{x}\right) = \sum_{j=1}^{S} \lambda(\alpha_i \mid \omega_j) P(\omega_j \mid \mathbf{x})$$

για κάθε i=1,…α και επιλέξτε την δράση α_i για την οποία το $R(a_i|\boldsymbol{x})$ είναι ελάχιστο.

- Το ελάχιστο αυτό ολικό ρίσκο καλείται:

 - ρίσκο (R*).

Είναι το ρίσκο με την καλύτερη απόδοση

Παράδειγμα: Ταξινόμηση σε δύο (2) τάξεις

- Η δράση a_1 αντιστοιχεί στο να δηλώσουμε την κατάσταση της φύσης ως ω_1.
- Η δράση a_2 αντιστοιχεί στο να δηλώσουμε την κατάσταση της φύσης ως ω_2.
- Συνάρτηση κόστους: $\lambda_{ij} = \lambda(a_i|\omega_j)$ αντιστοιχεί στο κόστος του να επιλέξω a_i , όταν η πραγματική τιμή της φύσης είναι ω_j.

Παράδειγμα: Ταξινόμηση σε δύο (2) τάξεις

- Τότε η:

$R(\alpha_i(\boldsymbol{x})|\boldsymbol{x}) = \sum_{j=1}^{S} \lambda(\alpha_i|\omega_j)P(\omega_j\,|\boldsymbol{x})$ γίνεται:

$$\begin{aligned} R(\alpha_1|\mathbf{x}) &= \lambda_{11}P(\omega_1|\mathbf{x}) + \lambda_{12}P(\omega_2|\mathbf{x}) \\ R(\alpha_2|\mathbf{x}) &= \lambda_{21}P(\omega_1|\mathbf{x}) + \lambda_{22}P(\omega_2|\mathbf{x}) \end{aligned}$$

Αποφάσισε ω_1

Εάν:

$R(\alpha_1|x) < R(\alpha_2|x)$ αλλιώς

αποφάσισε ω_2

Παράδειγμα:
Ταξινόμηση σε δύο (2) τάξεις

Αποφάσισε ω_1

Εάν:

$$(\lambda_{21}-\lambda_{11})P(\omega_1|x) > (\lambda_{12}-\lambda_{22})P(\omega_2|x)$$

αλλιώς

αποφάσισε ω_2

Συνήθως το κόστος του να κάνω λάθος είναι μεγαλύτερο από το κόστος του να πράξω σωστά.
Άρα η διαφορά > 0

Συνήθως το κόστος του να κάνω λάθος είναι μεγαλύτερο από το κόστος του να πράξω σωστά.
Άρα η διαφορά > 0

Αποφάσισε ω_1

Εάν:

$$(\lambda_{21}-\lambda_{11})p(\mathbf{x}|\omega_1)P(\omega_1) > (\lambda_{12}-\lambda_{22})\, p(\mathbf{x}|\omega_2)P(\omega_2)$$

αλλιώς αποφάσισε ω_2

Παράδειγμα: Ταξινόμηση σε δύο (2) τάξεις

Αποφάσισε ω_1 εάν:

Λόγος πιθανοφάνειας:
Likelihood Ratio (LR)

$$\mathrm{LR} \equiv \frac{p(\mathbf{x}|\omega_1)}{p(\mathbf{x}|\omega_2)} > \frac{(\lambda_{12}-\lambda_{22})P(\omega_2)}{(\lambda_{21}-\lambda_{11})P(\omega_1)}$$

Ανεξάρτητο του **x**

αλλιώς αποφάσισε ω_2

- Μπορούμε να θεωρήσουμε τα $p(\mathbf{x}|\omega_j)$ ως συνάρτηση του ω_j .
- Μετά, δημιουργώ την ποσότητα **LR**, και αποφασίζω ω_1 αν το **LR** υπερβαίνει μια τιμή κατωφλίου που δεν εξαρτάται από την παρατήρηση **x.**

Minimum Error Rate

- Συμμετρική συνάρτηση κόστους (ή zero−one):

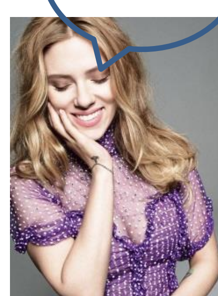
$$\lambda(\alpha_i | \omega_j) = \begin{cases} 0, & i = l \\ 1, & i \neq j \end{cases}, \ (i,j)=1,\ldots,S \qquad L = \begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{pmatrix}$$

- Όλα τα λάθη κοστίζουν το ίδιο (=1).
- Ρίσκο υπό συνθήκη:

$$\begin{aligned} R(\alpha_i|\mathbf{x}) &= \sum_{j=1}^{S} \lambda(\alpha_i|\omega_j) P(\omega_j|\mathbf{x}) \\ &= \sum_{j \neq i} P(\omega_j|\mathbf{x}) \\ &= 1 - P(\omega_i|\mathbf{x}) \end{aligned}$$

Επέλεξε την δράση που ελαχιστοποιεί το ρίσκο υπό συνθήκη.

Μ’ αρέσει



Πουχού

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

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

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

αν $\lambda_{21} = \lambda_{12} \Rightarrow$ Minimum classification error probability

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

Πουχού

- Συνέχεια:
  - $p(x|\omega_1) = \frac{1}{\sqrt{\pi}} \exp(-x^2)$
  - $p(x|\omega_2) = \frac{1}{\sqrt{\pi}} \exp(-(x-1)^2)$
  - $P(\omega_1) = P(\omega_2) = \frac{1}{2}$
  - $L = \begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{pmatrix} = \begin{pmatrix} 0 & 0.5 \\ 1.0 & 0 \end{pmatrix}$

– Η τιμή κατωφλίου για ελάχιστο P_e: (δηλαδή $\lambda_{21}=\lambda_{12}$)

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

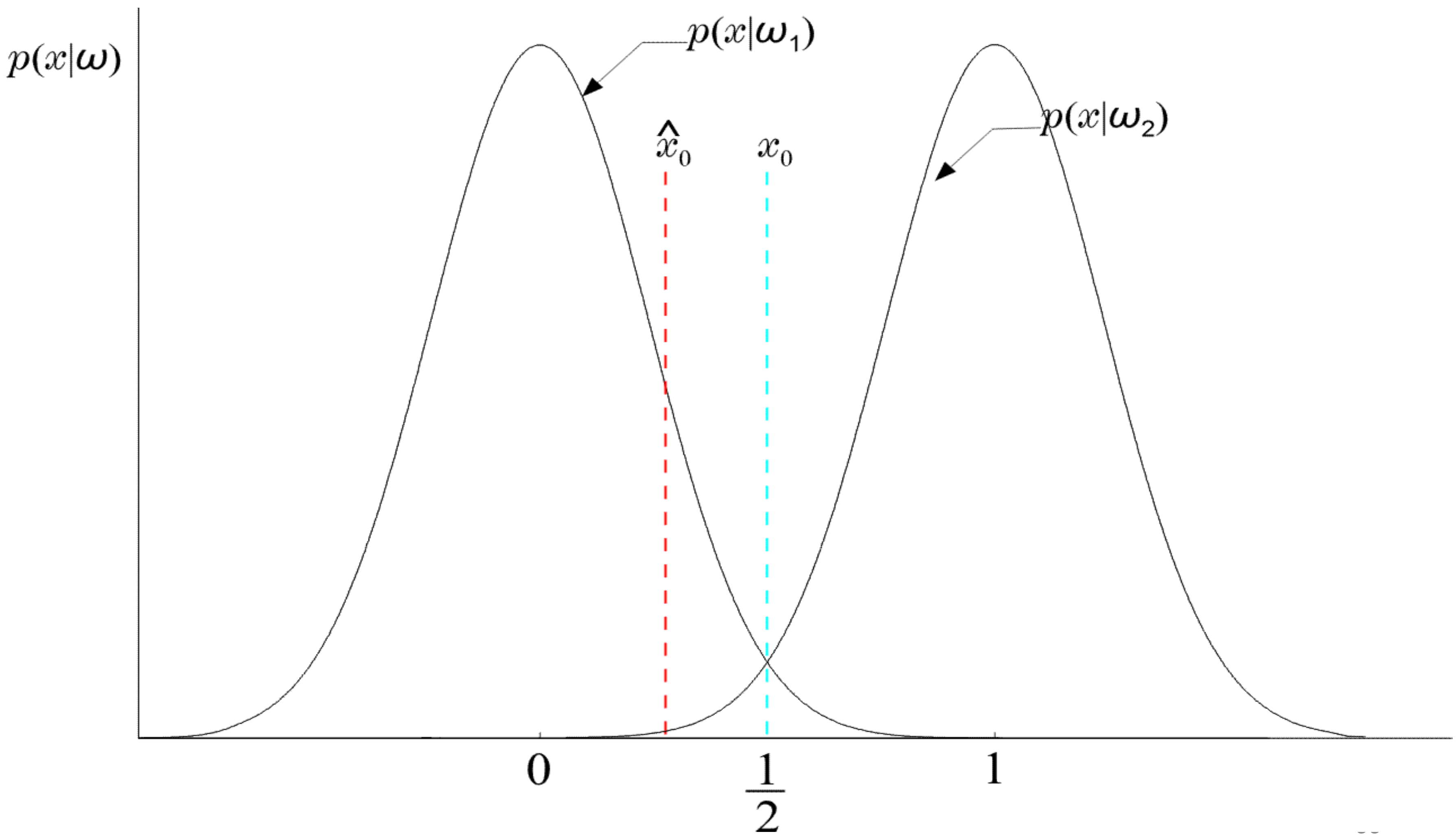
$$x_0 = \frac{1}{2}$$

– Η τιμή κατωφλίου $\hat{x}_0$ για ελάχιστο ρίσκο R

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

$$\hat{x}_0 = \frac{(1-\lambda n 2)}{2} < \frac{1}{2}$$

Από ότι φαίνεται το $\hat{x}_0$ βρίσκεται στα αριστερά του $x_0 = \dfrac{1}{2}$ (Γιατί;)



Minimum Error Rate

- Συμπέρασμα: Για ελάχιστο σφάλμα

Αποφάσισε ω_1

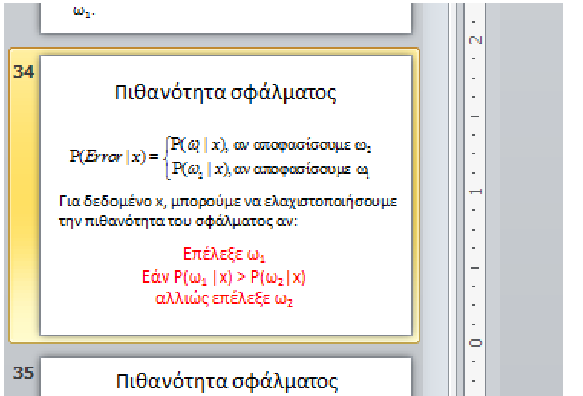
Εάν:

$P(\omega_i|\mathbf{x}) > P(\omega_j|\mathbf{x})$, για κάθε $j \neq i$

αλλιώς

αποφάσισε ω_2

Είναι η ίδια σχέση με πριν.



Για να τα δούμε πάλι μία...

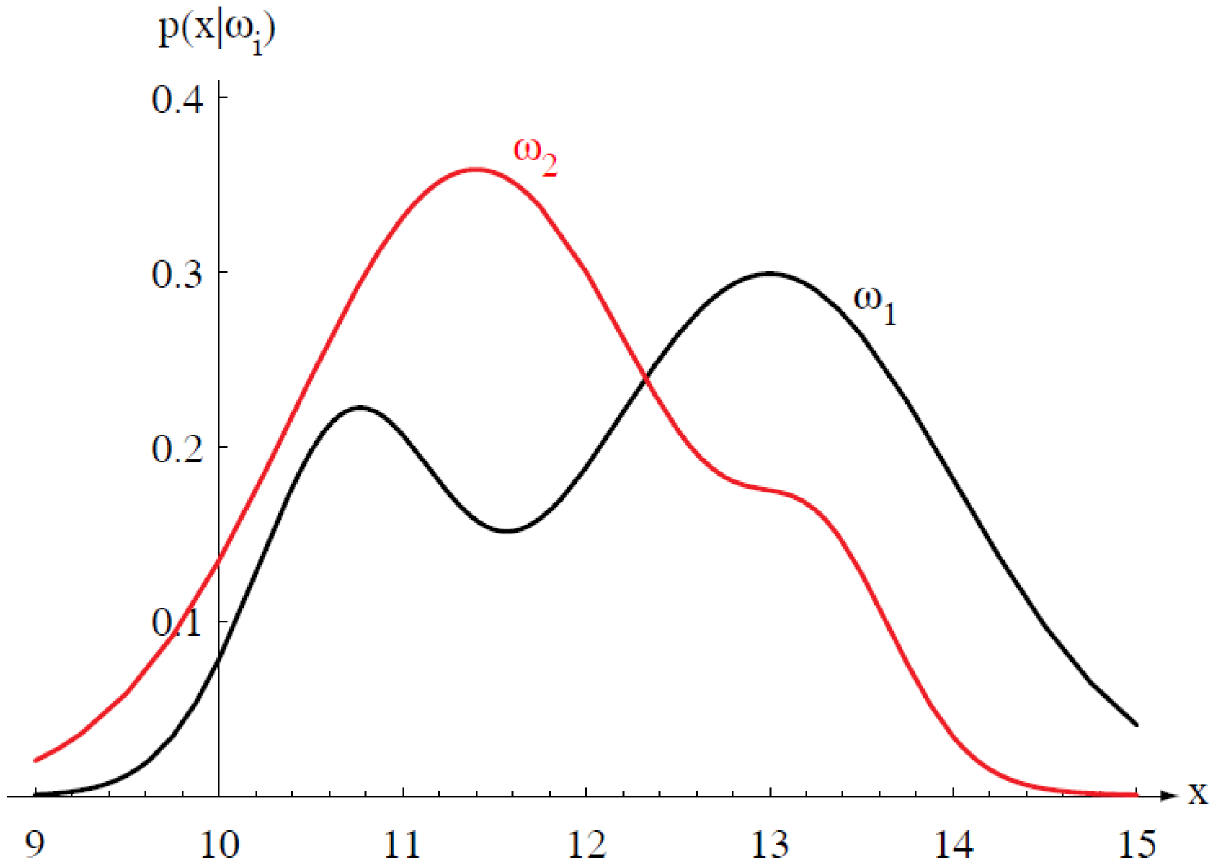


Figure 2.1: Hypothetical class-conditional probability density functions show the probability density of measuring a particular feature value x given the pattern is in category ω_i. If x represents the length of a fish, the two curves might describe the difference in length of populations of two types of fish. Density functions are normalized, and thus the area under each curve is 1.0.

Για να τα δούμε πάλι άλλη μία…

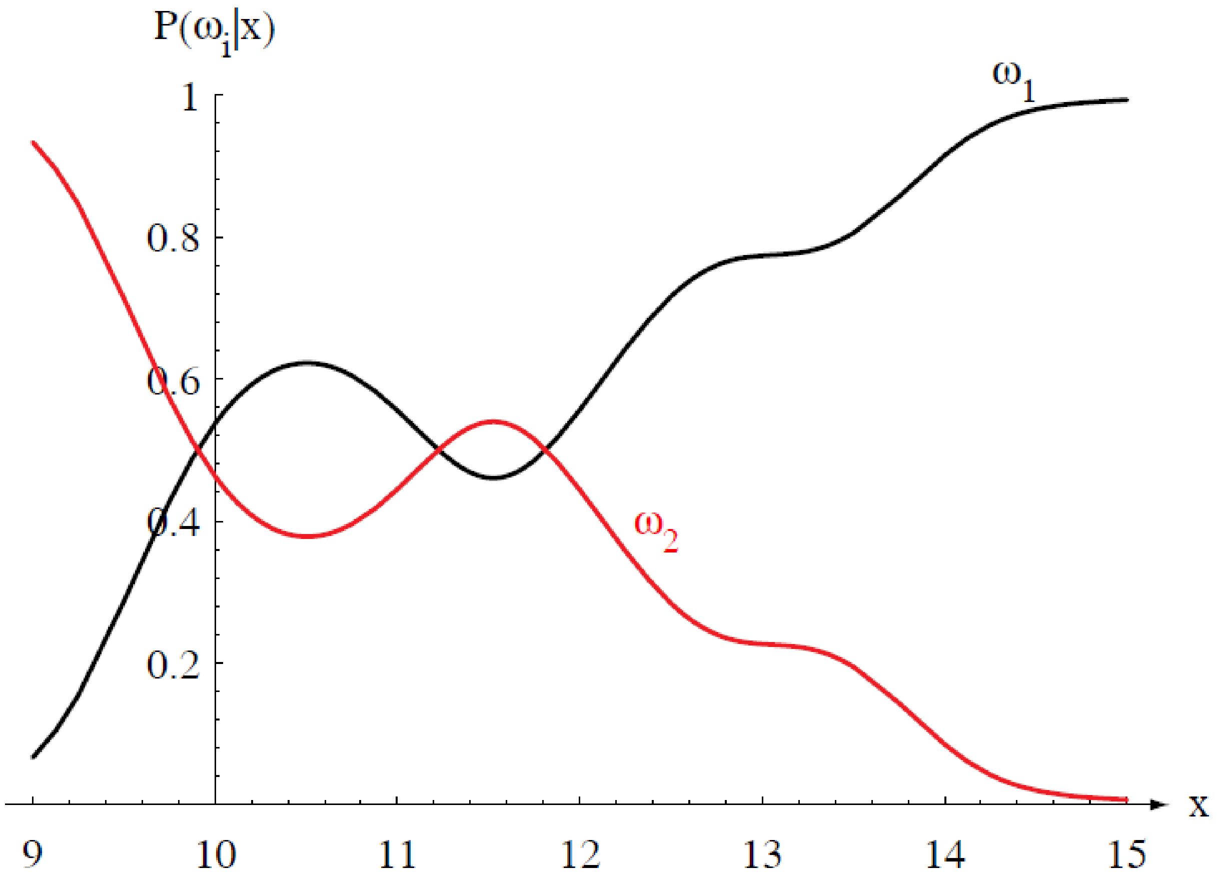


Figure 2.2: Posterior probabilities for the particular priors $P(\omega_1) = 2/3$ and $P(\omega_2) = 1/3$ for the class-conditional probability densities shown in Fig. 2.1. Thus in this case, given that a pattern is measured to have feature value $x = 14$, the probability it is in category ω_2 is roughly 0.08, and that it is in ω_1 is 0.92. At every x, the posteriors sum to 1.0.

Για να τα δούμε πάλι άλλη μία...

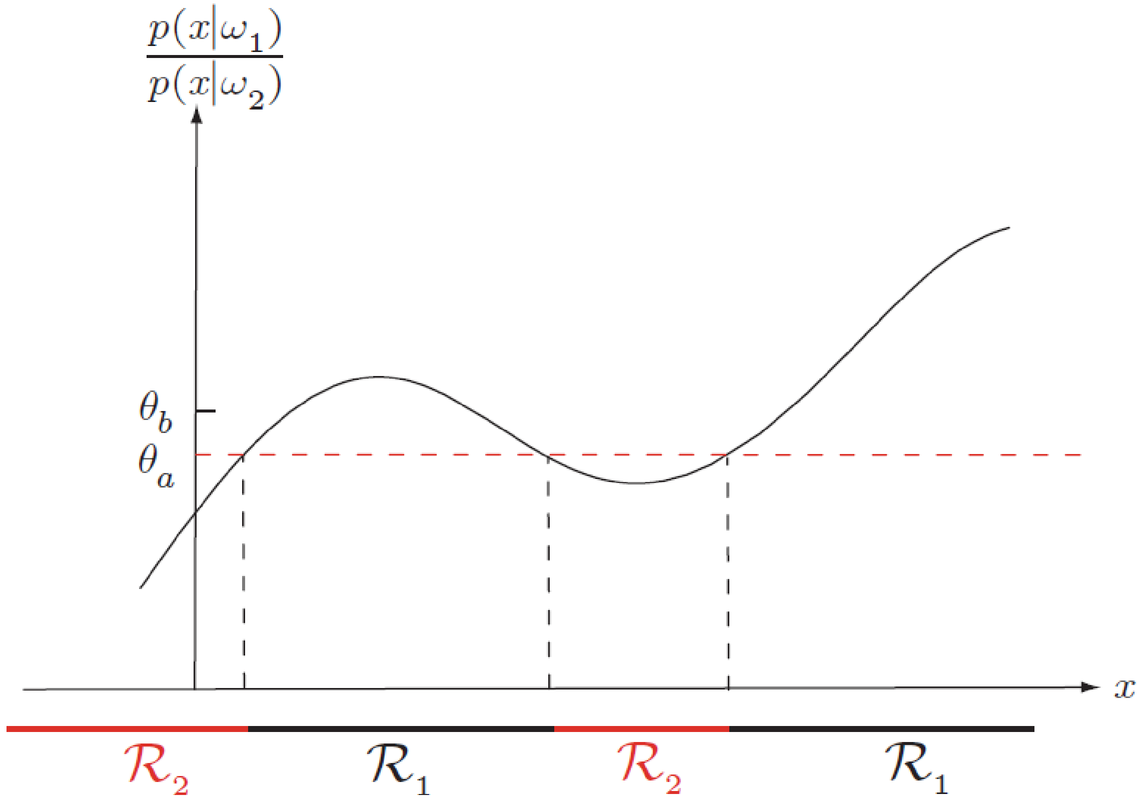
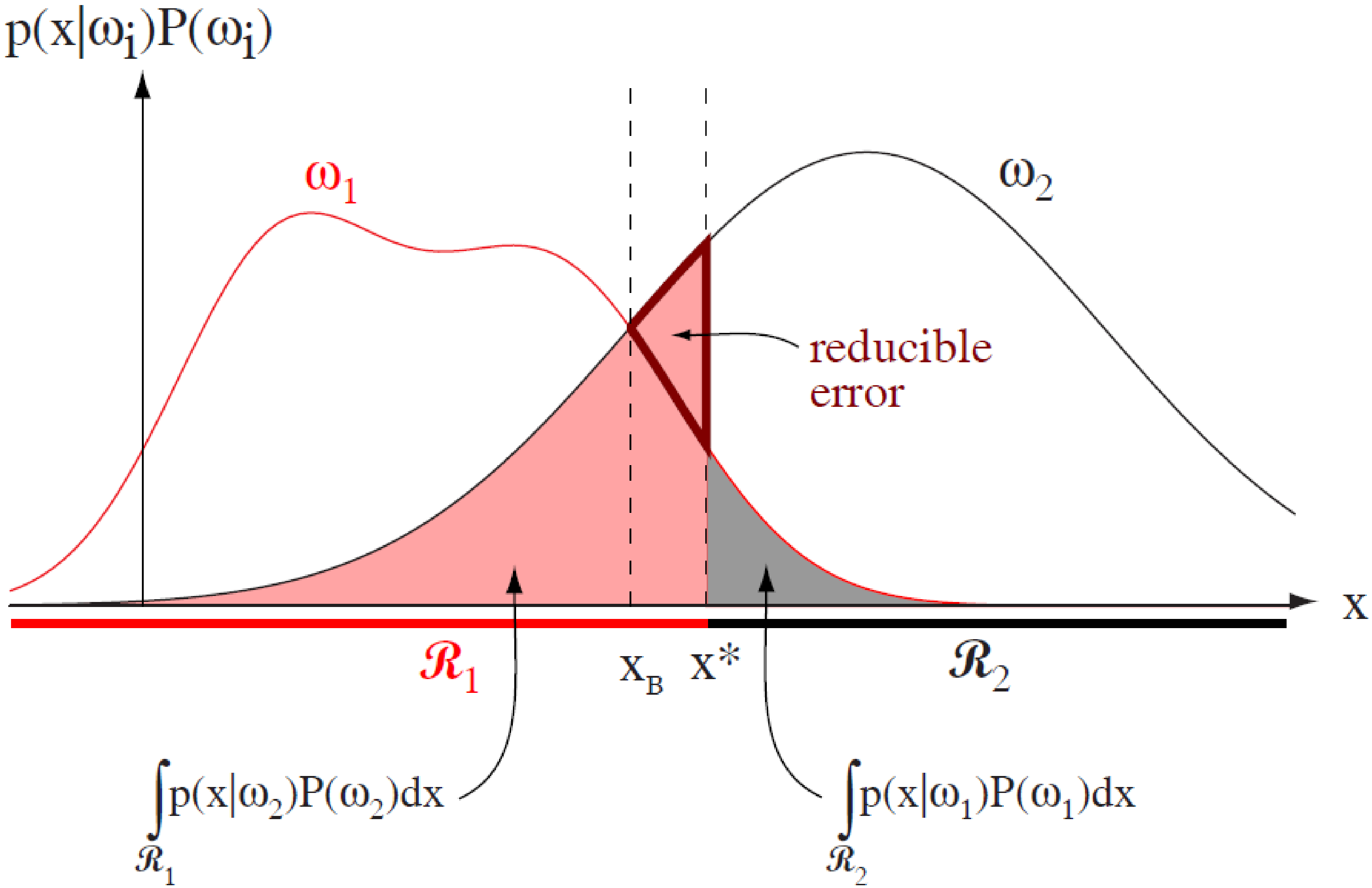


Figure 2.3: The likelihood ratio $p(x|\omega_1)/p(x|\omega_2)$ for the distributions shown in Fig. 2.1. If we employ a zero-one or classification loss, our decision boundaries are determined by the threshold θ_a. If our loss function penalizes miscategorizing ω_2 as ω_1 patterns more than the converse, (i.e., $\lambda_{12} > \lambda_{21}$), we get the larger threshold θ_b, and hence $\mathcal{R}_1$ becomes smaller.

Κριτήριο  MAX



Π.χ. Κριτήριο Minimax

- Υπάρχουν φορές που πρέπει να σχεδιάζουμε τον ταξινομητή μας να αποδίδει καλά σε ένα σύνολο διαφορετικών a-priori.
  - Π.χ: Στην κατηγοριοποίηση των ψαριών μπορούμε να φανταστούμε ότι οι τιμές των χαρακτηριστικών **x** παραμένουν σταθερές.
  - Παρόλα αυτά οι a-priori πιθανότητες μπορεί να ποικίλουν ευρέως και με απρόβλεπτο τρόπο.
  - Εναλλακτικά μπορεί να θέλουμε να χρησιμοποιήσουμε τον ταξινομητή σε άλλα εργοστάσια όπου δεν γνωρίζουμε τις a-priori πιθανότητες.
- Μια λογική προσέγγιση είναι τότε να σχεδιάσουμε τον ταξινομητή μας έτσι ώστε ο χειρότερος (**max**) συνολικός κίνδυνος για οποιαδήποτε τιμή των a-priori να είναι όσο το δυνατόν μικρότερος –
  - να ελαχιστοποιήσει (**min**) τον μέγιστο δυνατό συνολικό κίνδυνο.

Κριτήριο Minimax

- Θεωρούμε έστω την περιοχή R_1 στον χώρο των χαρακτηριστικών όπου ο ταξινομητής αποφασίζει ω_1 και παρομοίως για την $\{R_2, \omega_2\}$. Τότε το ολικό ρίσκο:

$$\begin{aligned} R &= \int_{\mathcal{R}_1} [\lambda_{11} P(\omega_1)\, p(\mathbf{x}|\omega_1) + \lambda_{12} P(\omega_2)\, p(\mathbf{x}|\omega_2)]\; d\mathbf{x} \\ &+ \int_{\mathcal{R}_2} [\lambda_{21} P(\omega_1)\, p(\mathbf{x}|\omega_1) + \lambda_{22} P(\omega_2)\, p(\mathbf{x}|\omega_2)]\; d\mathbf{x}. \end{aligned}$$

Δεδομένου: $P(\omega_2)=1- P(\omega_1)$ και:

$$\int_{\mathcal{R}_1} p(\mathbf{x}|\omega_1)\, d\mathbf{x} = 1 - \int_{\mathcal{R}_2} p(\mathbf{x}|\omega_1)\, d\mathbf{x}$$

έχουμε:

$$R(P(\omega_1)) = \overbrace{\lambda_{22} + (\lambda_{12} - \lambda_{22}) \int_{\mathcal{R}_1} p(\mathbf{x}|\omega_2)\, d\mathbf{x}}^{= R_{mm},\ \text{minimax risk}}$$

$$+ \underbrace{P(\omega_1) \left[(\lambda_{11} - \lambda_{22}) - (\lambda_{21} - \lambda_{11}) \int_{\mathcal{R}_2} p(\mathbf{x}|\omega_1)\, d\mathbf{x} - (\lambda_{12} - \lambda_{22}) \int_{\mathcal{R}_1} p(\mathbf{x}|\omega_2)\, d\mathbf{x} \right]}_{= 0 \text{ for minimax solution}}.$$

- Η εξίσωση δείχνει ότι αν τεθεί το όριο της απόφασης και κατά συνέπεια καθορίζουμε τα $\{R_1, R_2\}$ το ολικό ρίσκο είναι γραμμικό με το $P(\omega_1)$.
- Αν μπορούμε να βρούμε ένα όριο για το οποίο η σταθερά αναλογίας είναι 0, τότε το ολικό ρίσκο θα είναι ανεξάρτητο των a-priori.
- Αυτή είναι η minimax λύση και το αντίστοιχο minimax ρίσκο:

$$R_{mm} = \lambda_{22} + (\lambda_{12} - \lambda_{22}) \int_{R_1} p(\mathbf{x}|\omega_2)\, \mathrm{d}\mathbf{x} =$$

$$= \lambda_{11} + (\lambda_{21} - \lambda_{11}) \int_{R_2} p(\mathbf{x}|\omega_1)\, \mathrm{d}\mathbf{x}$$

$$R_{mm} = \lambda_{22} + (\lambda_{12} - \lambda_{22}) \int_{R_1} p(\mathbf{x}|\omega_2)\,\mathrm{d}\mathbf{x} =$$

$$= \lambda_{11} + (\lambda_{21} - \lambda_{11}) \int_{R_2} p(\mathbf{x}|\omega_1)\,\mathrm{d}\mathbf{x}$$

- Το κριτήριο minimax βρίσκει μεγάλη χρήση στη θεωρία παιγνίων απ 'ότι στην κλασσική αναγνώριση προτύπων.
- Στη θεωρία παιγνίων, υπάρχει ένας εχθρικός αντίπαλος, ο οποίος αναμένεται να λάβει μια ενέργεια που είναι πολύ επιζήμια για εμάς.
- Επομένως, έχει νόημα να προβούμε σε κάποια ενέργεια (π.χ. μια ταξινόμηση) όπου το κόστος μας - λόγω των επακόλουθων ενεργειών του αντιπάλου σας - ελαχιστοποιείται.

- Καθώς η $P(\omega_1)$ μεταβάλλεται, οι περιοχές αποφάσεων αλλάζουν.

- Με την σειρά τους αλλάζει και το μέσο ρίσκο το οποίο είναι μεγαλύτερο από το Bayes's.
- Τα δύο άκρα για την $P(\omega_1)$ είναι 0 ή 1.

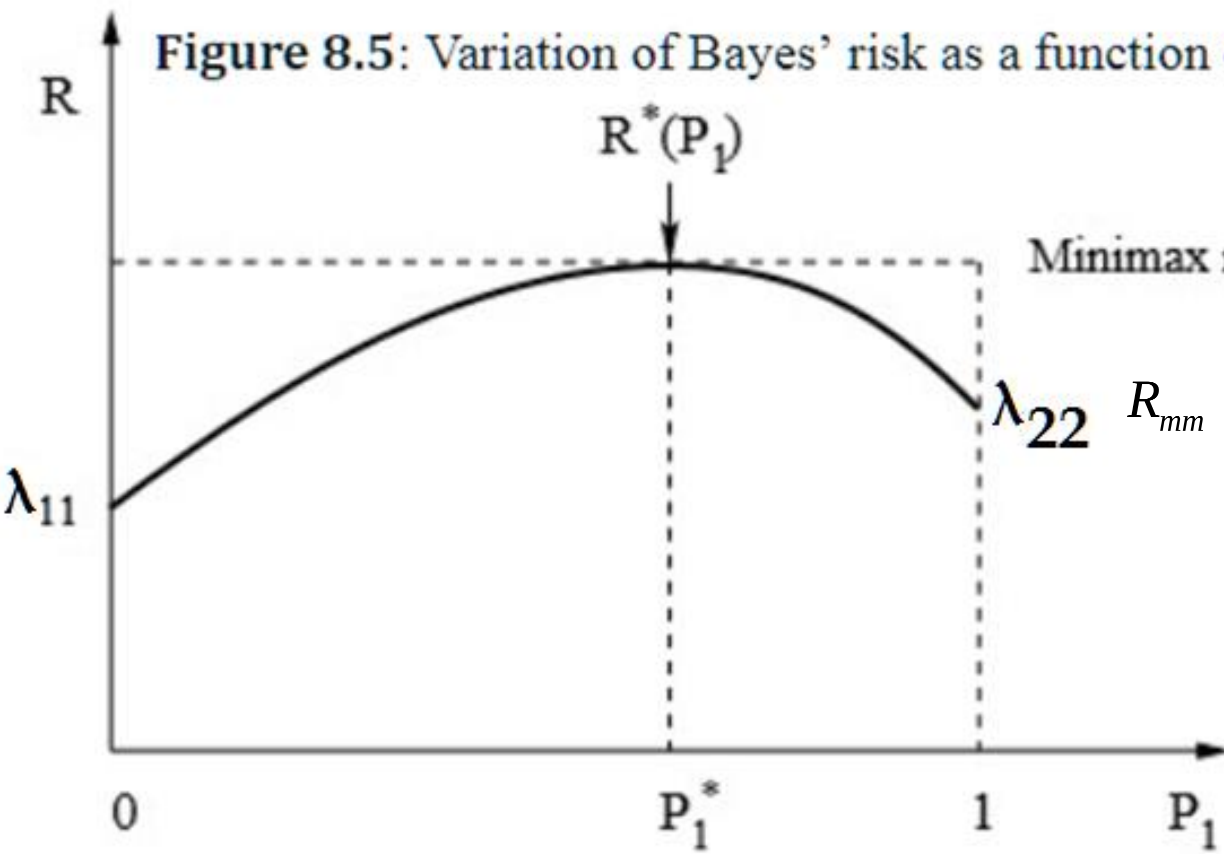


Figure 8.5: Variation of Bayes' risk as a function of P_1

$$R_{mm} = \lambda_{22} + (\lambda_{12} - \lambda_{22}) \int_{R_1} p(\mathbf{x}|\omega_2)\,\mathrm{d}\,\mathbf{x}$$

$$= \lambda_{11} + (\lambda_{21} - \lambda_{11}) \int_{R_2} p(\mathbf{x}|\omega_1)\,\mathrm{d}\,\mathbf{x}$$

- Το ρίσκο R είναι γραμμικό ως προς την a-priori $P(\omega_1)$
- Ο κανόνας του Bayes για $P(\omega_1) = P(\omega_1)^*$ παρέχει την μικρότερη τιμή ρίσκου R_{min}.
- Η εφαπτομένη στο R_{min} είναι οριζόντια και $R^*(P(\omega_1))$ στο σημείο $P(\omega_1) = P(\omega_1)^*$ αντιπροσωπεύει το ελάχιστο κόστος *minimum cost*.
- Η καμπύλη του ρίσκου Bayes πρέπει να είναι κοίλη με τα κοίλα προς τα κάτω
- Έτσι το μέσο ρίσκο δεν θα ξεπερνά την τιμή $R^*(P(\omega_1))$.
- **Λαμβάνοντας την παράγωγο (κλίση) του R ως προς το $P(\omega_1)$ και μηδενίζοντας λαμβάνουμε τον κανόνα *minimax***

Κανόνας *minimax*

$$R_{mm} = \lambda_{22} + (\lambda_{12} - \lambda_{22}) \int_{R_1} p(\mathbf{x}|\omega_2)\,\mathrm{d}\,\mathbf{x}$$

$$= \lambda_{11} + (\lambda_{21} - \lambda_{11}) \int_{R_2} p(\mathbf{x}|\omega_1)\,\mathrm{d}\,\mathbf{x}$$

Αν $\lambda_{11}=\lambda_{22}=0$ τότε: $\lambda_{12} \int_{R_1} p(\mathbf{x}|\omega_2)\,\mathrm{d}\,\mathbf{x} = \lambda_{21} \int_{R_2} p(\mathbf{x}|\omega_1)\,\mathrm{d}\,\mathbf{x}$

Και αν: $\lambda_{12}=\lambda_{21}=1$ τότε: $\int_{R_1} p(\mathbf{x}|\omega_2)\,\mathrm{d}\,\mathbf{x} = \int_{R_2} p(\mathbf{x}|\omega_1)\,\mathrm{d}\,\mathbf{x}$

$$P_{FA} = P_M$$

Και τελικά:

$$R_{mm} = P_{FA}(1 - \mathrm{P}(\omega_1)) + P_M\,\mathrm{P}(\omega_1) = P_{FA}\,\mathrm{P}(\omega_2) + P_M\,\mathrm{P}(\omega_1)$$

Δλδ: Η μέση πιθανότητα σφάλματος:

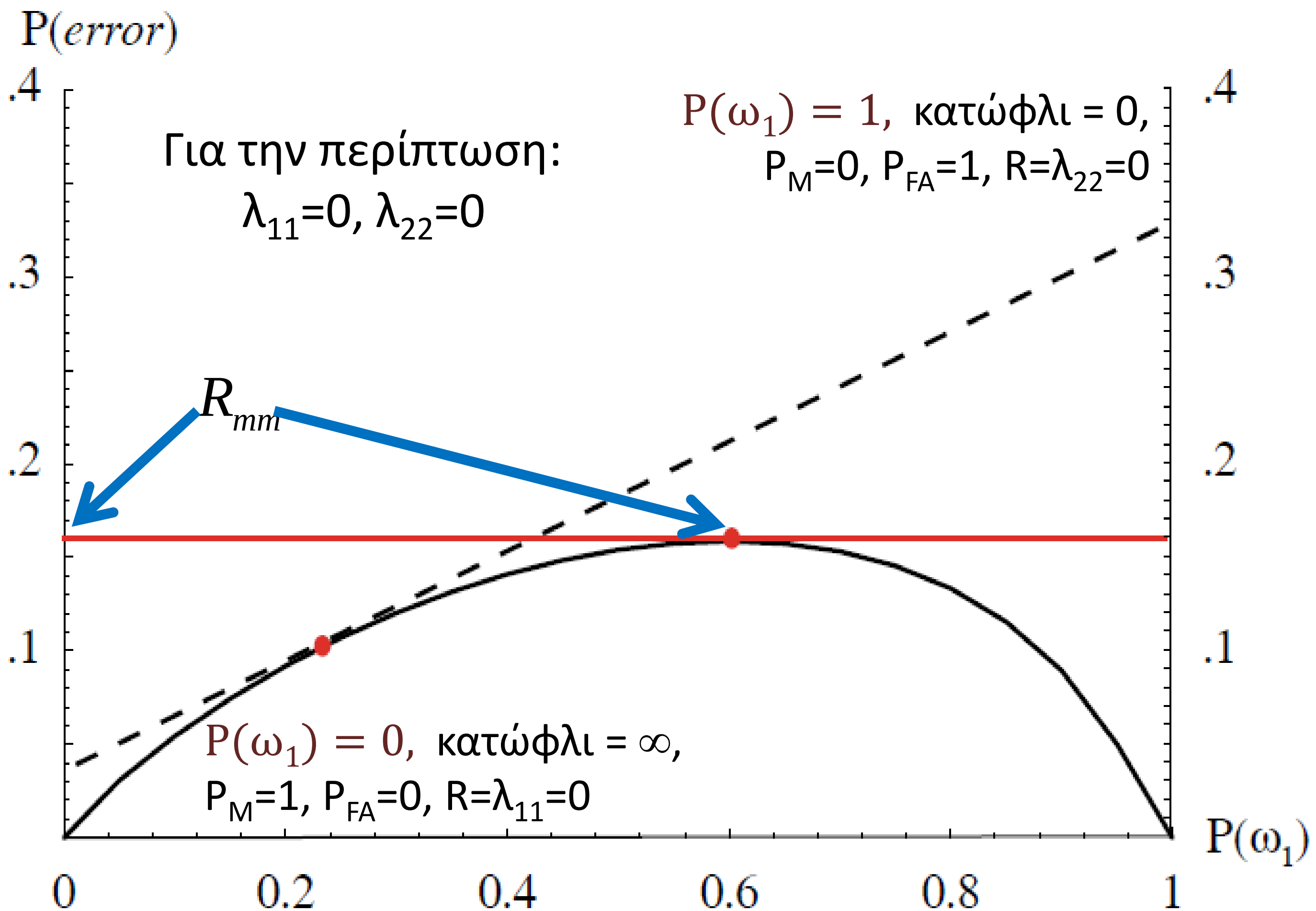


Figure 2.4: The curve at the bottom shows the minimum (Bayes) error as a function of prior probability $P(\omega_1)$ in a two-category classification problem of fixed distributions. For each value of the priors (e.g., $P(\omega_1) = 0.25$) there is a corresponding optimal decision boundary and associated Bayes error rate. For any (fixed) such boundary, if the priors are then changed, the probability of error will change as a linear function of $P(\omega_1)$ (shown by the dashed line). The maximum such error will occur at an extreme value of the prior, here at $P(\omega_1) = 1$. To minimize the maximum of such error, we should design our decision boundary for the maximum Bayes error (here $P(\omega_1) = 0.6$), and thus the error will not change as a function of prior, as shown by the solid red horizontal line.

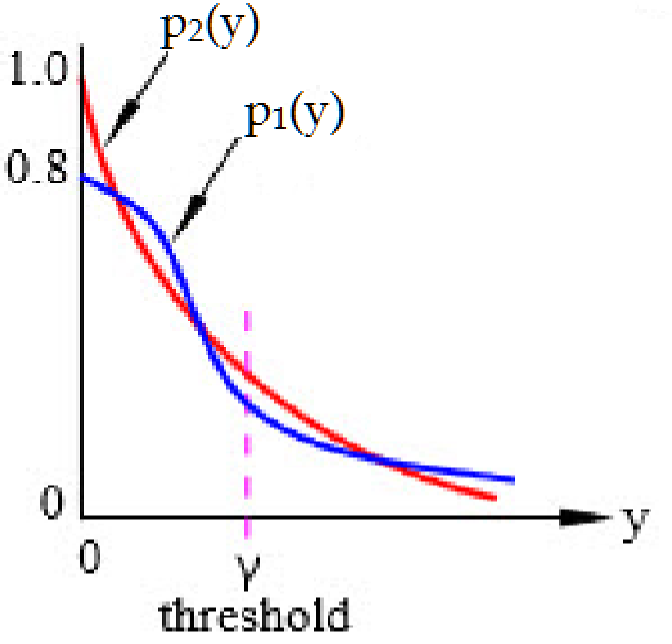
Που-χού

y είναι μια RV έτσι ώστε:

$$p(y|\omega_1) = \sqrt{\frac{2}{\pi}} e^{-\frac{y^2}{2}} u(y),$$

$$p(y|\omega_2) = e^{-y} u(y),$$

$$u(y): \text{step}$$



PDFs of the hypotheses for minimax decision rule example

Για μια τιμή κατωφλίου γ της RV y:

$$P_M = \int_0^{\gamma} \sqrt{\frac{2}{\pi}} e^{-\frac{y^2}{2}} dy \text{ και } \mathrm{P_{FA}} = \int_{\gamma}^{\infty} e^{-y} dy$$

Αν $\lambda_{11}=\lambda_{22}=0$, $\lambda_{12}=\lambda_{21}=1$ $\rightarrow$ $P_{FA} = P_M$ $\rightarrow$ $\int_0^{\gamma} \sqrt{\frac{2}{\pi}} e^{-\frac{y^2}{2}} dy = \int_{\gamma}^{\infty} e^{-y} dy$

Που-χού

Συνάρτηση σφάλματος
Error function

$$\int_0^{y}\sqrt{\frac{2}{\pi}}e^{-\frac{y^2}{2}}dy=\int_y^{\infty}e^{-y}dy \rightarrow \sqrt{\frac{2}{\pi}}[1-Q(\gamma)]=e^{-y}$$

Η τιμή του y που ικανοποιεί την παραπάνω σχέση: $y \approx 0.565$

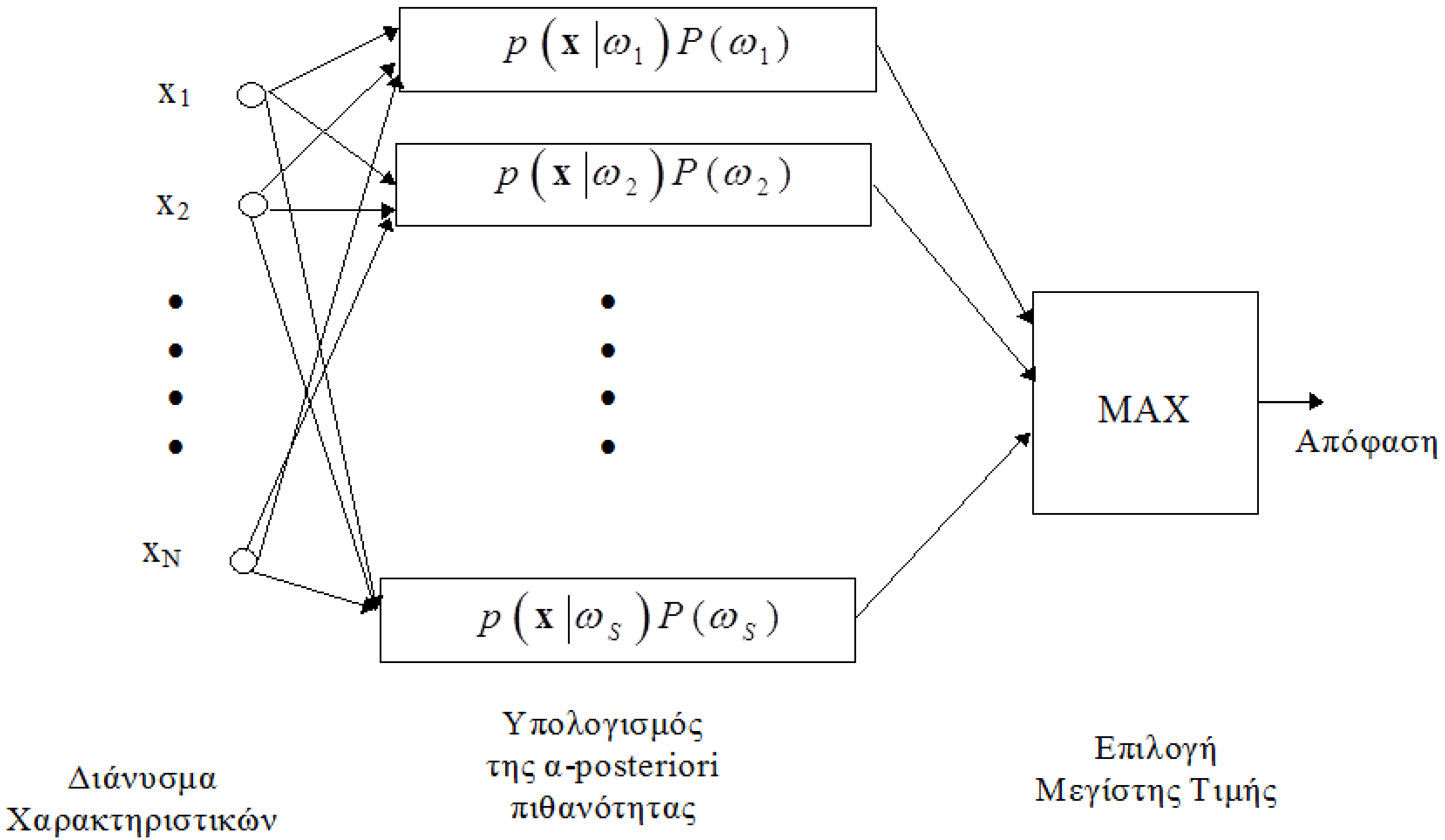
Άρα ο κανόνας minimax αποφασίζει ω_1 αν y>0.565

Ταξινομητές – Συναρτήσεις Διαχωρισμού

- Το σύστημα που δέχεται το άνυσμα των χαρακτηριστικών **x**, υπολογίζει την $P(\omega_j|\mathbf{x})$ με βάση την $p(\mathbf{x}|\omega_j)P(\omega_j)$ και αποφασίζει για την πιθανότερη κατηγορία, καλείται ταξινομητής - **classifier**.
- Ο ταξινομητής αναθέτει το προς εξέταση διάνυσμα χαρακτηριστικών **x** στην κατηγορία $\boldsymbol{\omega_i}$ εάν και μόνο εάν:

$$p\left(\mathbf{x}\middle|\omega_i\right)P(\omega_i) > p\left(\mathbf{x}\middle|\omega_j\right)P(\omega_j), \forall\ \mathrm{j} \neq \mathrm{i},\ \mathrm{j} = 1, \ldots, S$$

Ταξινομητές – Συναρτήσεις Διαχωρισμού



Ταξινομητές – Συναρτήσεις Διαχωρισμού

- Υπάρχουν πολλοί τρόποι να σχεδιάσουμε τους ταξινομητές. Ένας από αυτούς τους τρόπους περιγράφεται με την βοήθεια των συναρτήσεων διαχωρισμού (discriminant functions)

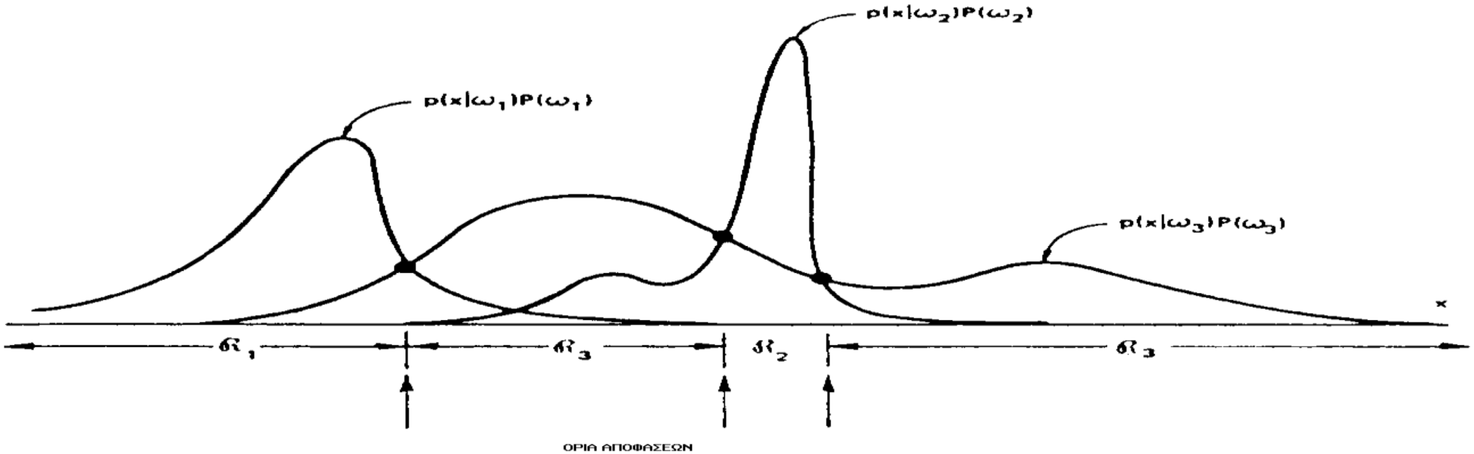
$$g_i(x) = p\left(x \middle| \omega_i\right) P(\omega_i), i = 1, \ldots, S$$

- Το πρόβλημα της επιλογής της μορφής της συνάρτησης διαχωρισμού δεν είναι μοναδικό.
- Γενικότερα αν αντικαταστήσουμε την $g_i(\mathbf{x})$ με την $f(g_i(\mathbf{x}))$ όπου f είναι μια μονότονη και αύξουσα συνάρτηση θα έχουμε τα ίδια αποτελέσματα όσο αφορά την ταξινόμηση.
- Αυτή η ανάλυση συχνά οδηγεί σε σημαντικές απλοποιήσεις.

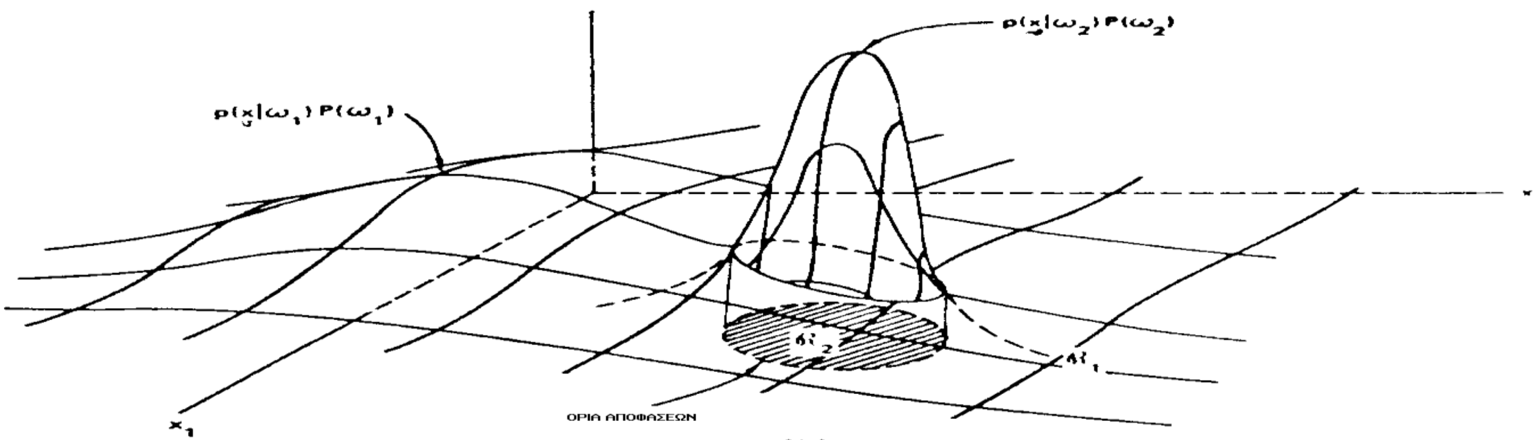
Ταξινομητές – Συναρτήσεις Διαχωρισμού

- Το σημαντικό είναι ότι ακόμα και αν οι συναρτήσεις διαχωρισμού μπορούν να γραφτούν με μια ποικιλία μορφών, η ουσία, δηλαδή η επιλογή των κατηγοριών δεν αλλάζει.
- Ο σκοπός των συναρτήσεων διαχωρισμού είναι να διαιρέσει τον χώρο χαρακτηριστικών σε s το πλήθος περιοχές αποφάσεων $R_1, \ldots, R_s$.
- Εάν $g_i(\mathbf{x}) > g_j(\mathbf{x})$ για κάθε $j \neq i$, τότε το διάνυσμα χαρακτηριστικών $\mathbf{x}$ είναι στην περιοχή R_i και ο κανόνας επιλογής αναθέτει το $\mathbf{x}$ στην κατηγορία ω_i.
- Οι περιοχές χωρίζονται από τα σύνορα αποφάσεων που είναι υπερεπιφάνειες στον χώρο των χαρακτηριστικών.

Ταξινομητές – Συναρτήσεις Διαχωρισμού



(a)



(b)

Παραδείγματα επιφανειών απόφασης

Ταξινομητές – Συναρτήσεις Διαχωρισμού

- Η εξίσωση για την εύρεση των συνόρων προκύπτει από την σχέση:

$g_i(x) = g_j(x)$

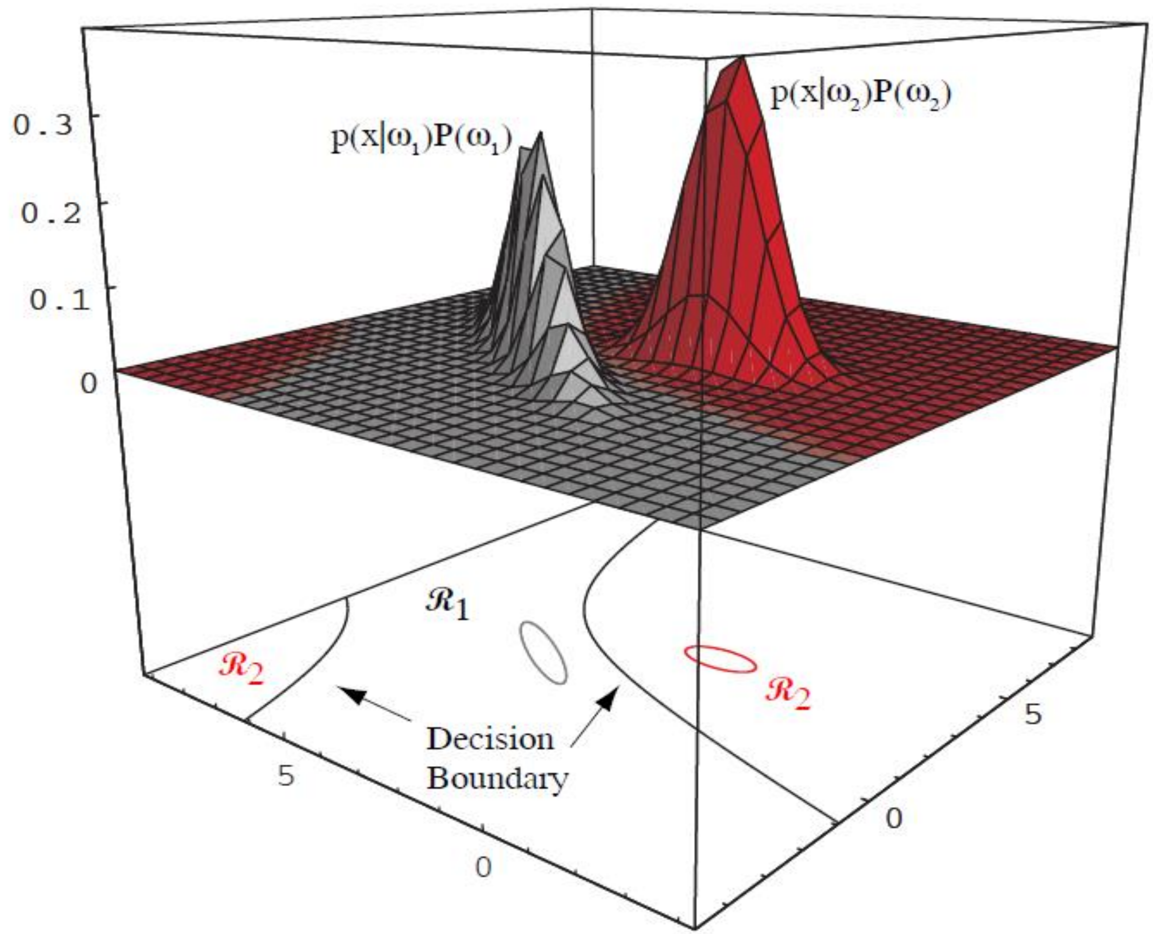


Figure 2.6: In this two-dimensional two-category classifier, the probability densities are Gaussian (with $1/e$ ellipses shown), the decision boundary consists of two hyperbolas, and thus the decision region $\mathcal{R}_2$ is not simply connected.

Ταξινομητές – Συναρτήσεις Διαχωρισμού

- Προσέγγιση κατά Bayes (γενική περίπτωση).

$$g_i(\mathbf{x}) = -R(\alpha_i/\mathbf{x})$$

  - Διαισθητικά: Η μέγιστη τιμή της συνάρτησης διαχωρισμού αντιστοιχεί στο ελάχιστο ρίσκο.

- Προσέγγιση κατά Bayes (minimum error)

$$g_i(\mathbf{x}) = \mathrm{P}(\omega_i/\mathbf{x})$$

  - Διαισθητικά: Η μέγιστη τιμή της συνάρτησης διαχωρισμού αντιστοιχεί στη μέγιστη εκ των υστέρων τιμή πιθανότητας

Ταξινομητές – Συναρτήσεις Διαχωρισμού

- Η επιλογή των ΣΔ δεν υπόκειται στο αλάθητο
- Γενικότερα αν αντικαταστήσουμε την $g_i(\mathbf{x})$ με την $f(g_i(\mathbf{x}))$ όπου f είναι μια μονότονη και αύξουσα συνάρτηση θα έχουμε τα ίδια αποτελέσματα όσο αφορά την ταξινόμηση.
- Το παραπάνω πόνημα βοηθάει μερικές φορές στην (διδακτική) απλοποίηση πολλών σημαντικών τύπων και αναλύσεων.

Ταξινομητές – Συναρτήσεις Διαχωρισμού

- Minimum Error Rate

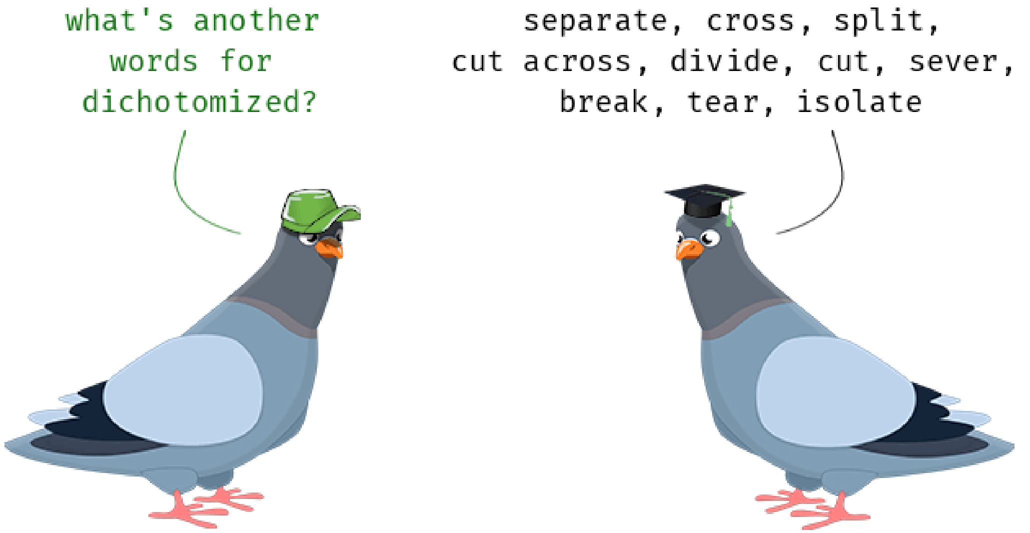
$$g_i(\mathbf{x}) = P(\omega_i|\mathbf{x}) = \frac{p\left(\mathbf{x}|\omega_i\right)P(\omega_i)}{\sum_{j=1}^{S} p\left(\mathbf{x}|\omega_j\right)P(\omega_j)}$$

$$g_i(\mathbf{x}) = p\left(\mathbf{x}|\omega_i\right)P(\omega_i),$$

$$\text{ή} \quad g_i(\mathbf{x}) = \ln p\left(\mathbf{x}|\omega_i\right) + \ln P(\omega_i)$$

Ταξινομητές – Συναρτήσεις Διαχωρισμού: Που-χού

- Ταξινόμηση σε δυο κατηγορίες
  - Π.χ Να ζει κανείς ή να μην ζει.
- Από παράδοση, είναι το πιο σεβαστό παράδειγμα.
- Όρος εις την βαρβαρική: Dichotomizer



Τ/ΣΔ: Dichotomizer

- Αντι να χρησιμοποιήσω δύο συναρτήσεις διαχωρισμού g_1, g_2 και να κάνω ανάθεση στην τάξη ω_1 αν $g_1 > g_2$ χρησιμοποιώ μία:

Την διαφορά τους g_1-g_2.

$$g(\boldsymbol{x})=g_1(\boldsymbol{x})-g_2(\boldsymbol{x})$$

Αποφάσισε ω_1
Εάν:
$R(\alpha_1|x) < R(\alpha_2|x)$ αλλιώς
αποφάσισε ω_2

Αποφάσισε αποφάσισε τελικά
τι διατάζεις να κάνω,
θέλεις να ζήσω η να πεθάνω
"LE-PA"

Τ/ΣΔ: Dichotomizer

$$g(\mathbf{x}) = P(\omega_1|\mathbf{x}) - P(\omega_2|\mathbf{x})$$

$$g(\mathbf{x}) = \ln\frac{p(\mathbf{x}|\omega_1)}{p(\mathbf{x}|\omega_2)} + \ln\frac{P(\omega_1)}{P(\omega_2)}$$

- Η δομή ενός ταξινομητή Bayes καθορίζεται από της υπό συνθήκη κατανομές p($\mathbf{x}|\omega_i$) και τις a-priori πιθανότητες P(ω_i)

Λίγη ορολογία ξανά...

- Προσδοκώμενη τιμή (Expected Value) $\mathcal{E}$ μιας βαθμωτής συνάρτησης *f(x)* της οποίας η τυχαία μεταβλητή *x* ακολουθεί την κατανομή *p(x)*:

$$\mathcal{E}[f(x)] = \int_{-\infty}^{+\infty} f(x)p(x)dx$$

- Αν έχουμε διακριτή κατανομή:

$$\mathcal{E}[f(x)] = \sum_{x \in D} f(x)P(x)$$

1D Κατανομή

- Κατανομή Γκαους μιας μεταβλητής

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

- Προσδοκώμενη τιμή (expected value):

$$\mu \equiv \mathcal{E}[x] = \int_{-\infty}^{\infty} x p(x)\ dx,$$

Προσδοκώμενο τετράγωνο της τυπικής απόκλισης (expected variance):

$$\sigma^2 \equiv \mathcal{E}[(x-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 p(x)\ dx.$$

1D Κατανομή

- Η μονοδιάστατη κατανομή Gauss καθορίζεται πλήρως από τις τιμές μ, σ^2.
- Για απλότητα την συμβολίζουμε:

$$\mathrm{p(x)} \sim N(\mu, \sigma^2)$$

- Τα δείγματα της $N(\mu, \sigma^2)$ τείνουν να κατανέμονται γύρω από το μ με μια διασπορά σχετική με την τυπική απόκλιση σ.

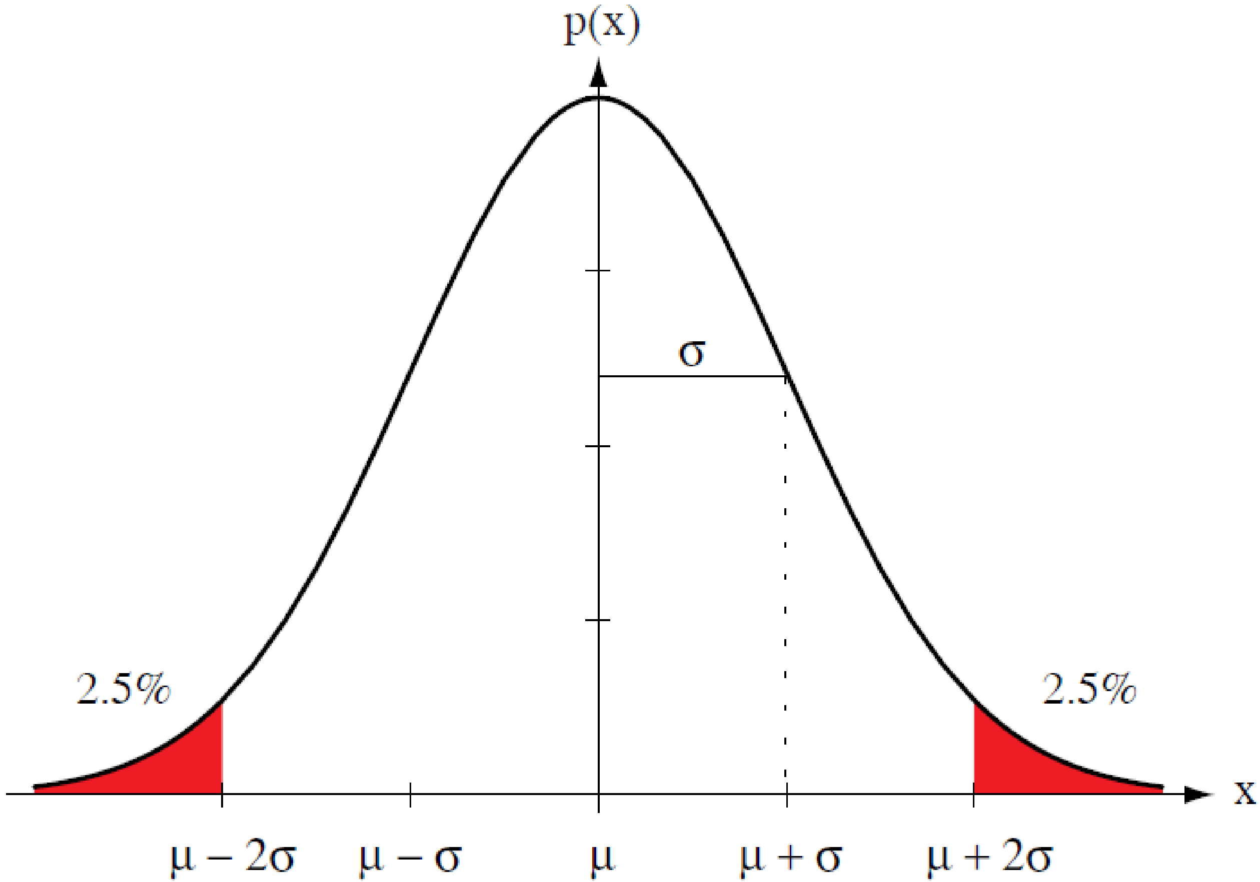


Figure 2.7: A univariate normal distribution has roughly 95% of its area in the range $|x - \mu| \leq 2\sigma$, as shown. The peak of the distribution has value $p(\mu) = 1/\sqrt{2\pi}\sigma$.

- Η εντροπία της 1D κατανομής:

$$\mathrm{H}(\mathrm{p}(\mathrm{x})) = -\int p(x) \ln p(x)\, dx$$

- Μονάδα μέτρησης: nats
- Αν γίνει χρήση του $\log_2$ τότε μετράμε σε bits
- Η κανονική κατανομή έχει την μεγαλύτερη εντροπία από οποιαδήποτε κατανομή με την ίδια μέση τιμή και διασπορά.
- Central Limit Theorem (***με εξαιρετικά απλά λόγια***)
  - Το άθροισμα πολλών ανεξαρτήτων τυχαίων μεταβλητών ακολουθεί την κανονική κατανομή.
  - Μοντελοποίηση (άλογα, βήματα, χέρια, μαχαίρια) ως σύνθεση ενός πρωτοτύπου και κάποιου θορύβου που προέρχεται από ένα μεγάλο αριθμό τυχαίων διαδικασιών.

Κανονική Κατανομή πολλών Μεταβλητών

- Η γενική μορφή:

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} e^{\left[-\frac{1}{2}(x-\mu)^t \Sigma^{-1} (x-\mu)\right]}$$

$\mathbf{x} \in R^d$, ένα διάνυσμα στήλη

$\boldsymbol{\mu} \in Rd$ η μέση τιμή

$\boldsymbol{\Sigma} \in \mathrm{R}^{d \times d}$: πίνακας συνδιακύμανσης-covariance

$\boldsymbol{\Sigma}^{-1}$ είναι ο αντίστροφος

$|\boldsymbol{\Sigma}|$ είναι η ορίζουσα.

Εσωτερικό γινόμενο (inner product)
Βαθμωτό - *Εσωτερικό* γινόμενο (dot product)

- Εσωτερικό γινόμενο: Για συναρτήσεις
- Βαθμωτό γινόμενο: Για διανύσματα
- Βαθμωτό γινόμενο για $a, b \in Rd$:

$$a^t b = \sum_{i=1}^{d} a_i b_i$$

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} e^{\left[-\frac{1}{2}(x-\mu)^t \Sigma^{-1} (x-\mu) \right]}$$

$$\mathrm{p}(\mathbf{x}) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

Κανονική Κατανομή πολλών Μεταβλητών

$$\boldsymbol{\mu} = \mathcal{E}[\boldsymbol{x}] = \int \boldsymbol{x}\, p(\boldsymbol{x})\, d\boldsymbol{x}$$

$$\Sigma = \mathcal{E}[(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^t] =$$
$$\int (\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^t p(\boldsymbol{x}) d\boldsymbol{x}$$

- **Σ**: πάντα **<u>συμμετρικός</u>** - positive semidefinite.
  - Εμείς θα ασχοληθούμε με την περίπτωση: **Σ:** positive definite έτσι ώστε η ορίζουσα να είναι θετική και μεγαλύτερη του μηδενός.

Ιδιότητα του πίνακα συνδιακύμανσης

Αν

$$\Sigma \in \mathrm{R}^{d \times d}$$

Τότε:

$$\Sigma = \mathrm{E} \times \Lambda \times \mathrm{E}^{t}$$

Όπου E = [$\mathbf{e}_1$, . . . , $\mathbf{e}_d$] είναι ένας ορθοκανονικός πίνακας με στήλες τα **ιδιοδιανύσματα** του **Σ** και Λ = diag(λ_1, . . . , λ_p) είναι ένα διαγώνιος πίνακας με στοιχεία τις αντίστοιχες ιδιοτιμές.

Άρα: $\mathbf{E^t E = I}$

Κανονική Κατανομή πολλών Μεταβλητών

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

$$\boldsymbol{x} = \{x_i\}, i = 1, \ldots d$$

$$\mu_i = \mathcal{E}[x_i]$$

$$\Sigma = \{\sigma_{ij}\}$$

$$\sigma_{ij} = \mathcal{E}[(x_i - \mu_i)(xj - \mu_j)]$$

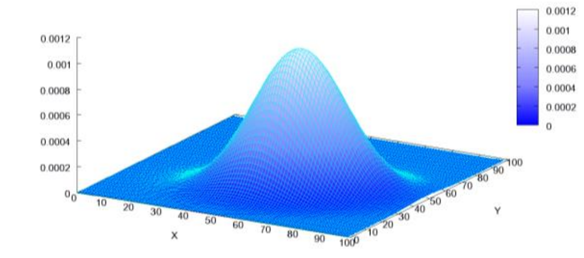
- Αν x_i, x_j στατιστικά ανεξάρτητα: σ_{ij} =0
- σ_{ii} , διαγώνια στοιχεία: απεικονίζουν την διασπορά σ_{ii}^2 του χαρακτηριστικού x_i
- σ_{ij}, μη διαγώνια στοιχεία: απεικονίζουν την συνδιακύμανση σ_{ij}^2 μεταξύ των x_i, x_j

Κανονική Κατανομή πολλών Μεταβλητών

- Αν σ_{ij} είναι 0 για κάθε $i \neq j$ τότε:

$$p(\mathbf{x}) = \prod_{i=1}^{d} \frac{1}{\sqrt{2\pi}\sigma_i} e^{\left[-\frac{1}{2}\left(\frac{x_i - \mu_i}{\sigma_i}\right)^2\right]}$$

Κανονική Κατανομή πολλών Μεταβλητών

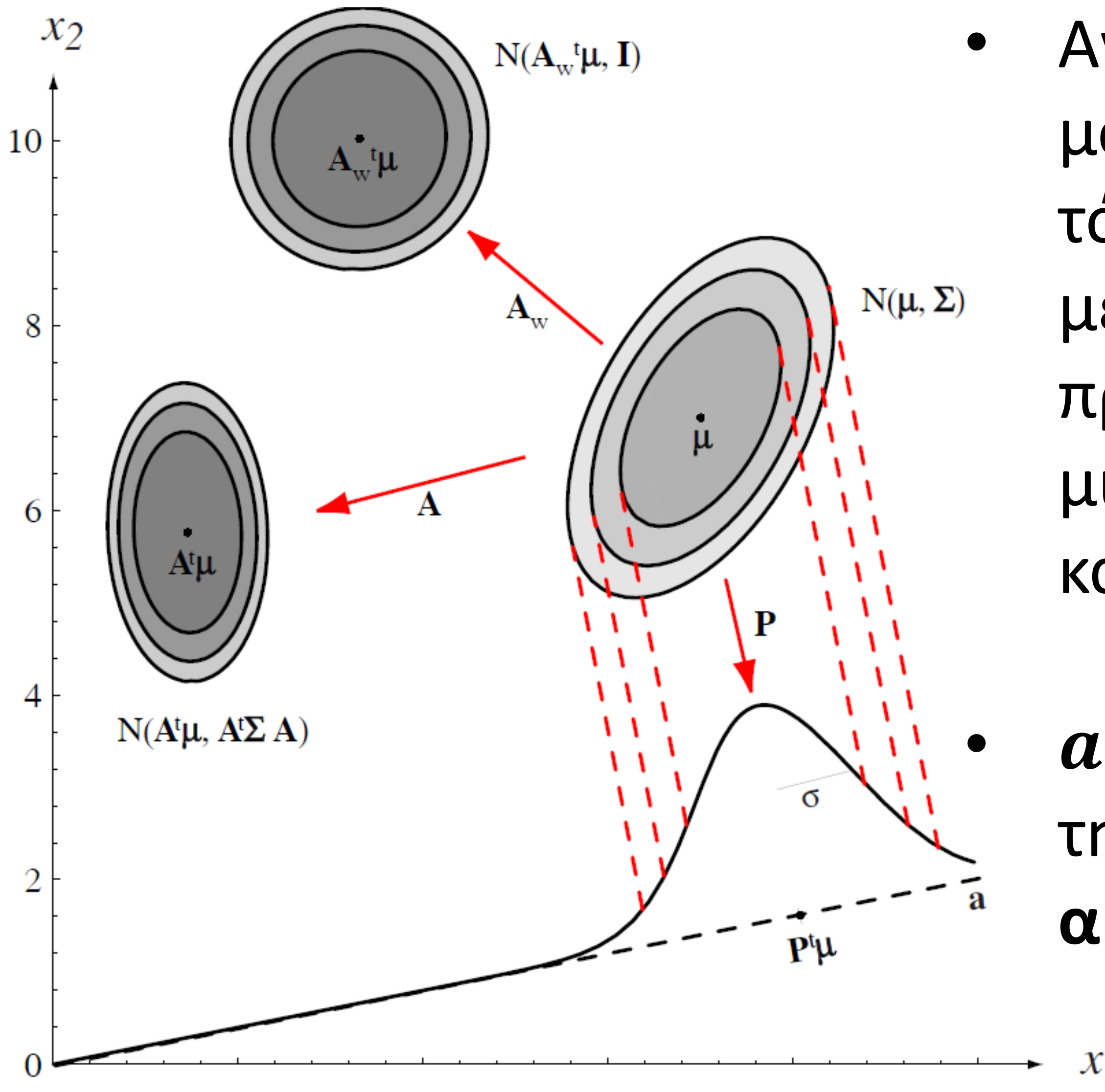


- Γραμμικός συνδυασμός τυχαίων μεταβλητών που ακολουθούν την κανονική κατανομή ανεξαρτήτων ή μη, ακολουθούν επίσης την κανονική κατανομή.
- Αν **A** είναι ένας $d \times k$ πίνακας και $\boldsymbol{y} = A^t \boldsymbol{x}$ είναι ένα νέο διάνυσμα $\in R^k$ τότε:

$$p(\mathbf{y}) \sim N(\mathbf{A}^t \boldsymbol{\mu}, \mathbf{A}^t \boldsymbol{\Sigma} \mathbf{A}),$$

$$\Sigma = \mathrm{E} \times \Lambda \times \mathrm{E}^t$$

$$\mathbf{E^t E = I}$$



- Αν k=1, και **A** είναι ένα μοναδιαίο διάνυσμα **α** τότε ο παρακάτω μετασχηματισμός προβάλει την ΚΚΠΜ σε μια γραμμή με κατεύθυνση **α** :

$$y = \boldsymbol{a}^t \boldsymbol{x}$$

- $\boldsymbol{a}^t \boldsymbol{\Sigma} \boldsymbol{a}$ είναι η διασπορά της προβολής του **x** στον **α.**

Figure 2.8: The action of a linear transformation on the feature space will convert an arbitrary normal distribution into another normal distribution. One transformation, **A**, takes the source distribution into distribution $N(\mathbf{A}^t\boldsymbol{\mu}, \mathbf{A}^t\boldsymbol{\Sigma}\mathbf{A})$. Another linear transformation — a projection **P** onto line **a** — leads to $N(\mu, \sigma^2)$ measured along **a**. While the transforms yield distributions in a different space, we show them superimposed on the original $x_1 - x_2$ space. A whitening transform leads to a circularly symmetric Gaussian, here shown displaced.

Whitening Transform

- Μερικές φορές είναι χρήσιμο το να γίνει ένας ***γραμμικός*** μετασχηματισμός οποίος μετατρέπει μια αυθαίρετη ΚΚΠΜ σε σφαιρική,
  - δηλ. ο πίνακας συνδιακύμανσης είναι ανάλογος με τον πίνακα ταυτότητας **Σ**=**I**.
- Αν: α) **Φ** είναι ο πίνακας του οποίου οι στήλες είναι τα ορθοκανονικά ιδιοδιανύσματα του **Σ** και β) **Λ** ο διαγώνιος πίνακας των ιδιοτιμών του, τότε ο μετ/μός whitening:

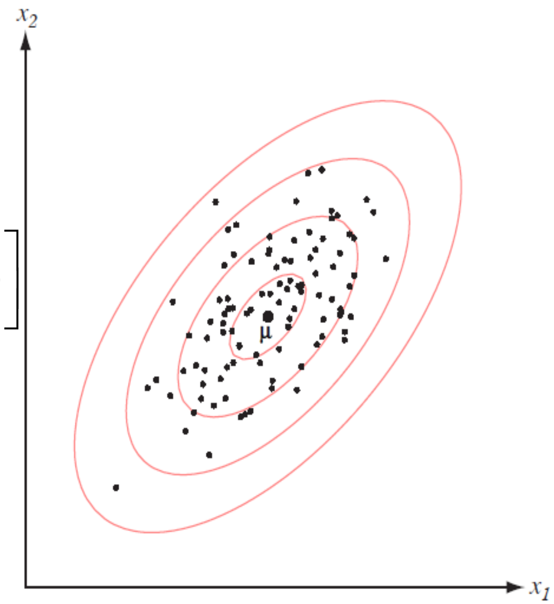
$$\mathbf{A}_w = \mathbf{\Phi}\mathbf{\Lambda}^{-1/2}$$

εγγυάται ότι ο πίνακας **Σ** της νέας RV είναι ο **I**

- **ΨΕΣ:** Ο μετ/μός whitening κάνει το φάσμα του νέου χώρου των ιδιοδιανυσμάτων ομογενές.

Χμ…

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}\,|\Sigma|^{1/2}} e^{\left[-\frac{1}{2}(x-\mu)^t \Sigma^{-1}(x-\mu)\right]}$$



$$(x-\mu)^t \Sigma^{-1}(x-\mu)$$

Figure 2.9: Samples drawn from a two-dimensional Gaussian lie in a cloud centered on the mean $\boldsymbol{\mu}$. The red ellipses show lines of equal probability density of the Gaussian.

Η κανονική κατανομή πολλών μεταβλητών καθορίζεται πλήρως από:

$$\frac{d+d(d+1)}{2} \text{ παραμέτρους}$$

- Τα δείγματα που προέρχονται από κανονική κατανομή (πολλών μεταβλητών) εμπίπτουν σε ένα ***νέφος.***
  - ***Το κέντρο του νέφους είναι η μέση τιμή μ.***
  - ***Το σχήμα του νέφους καθορίζεται από τον πίνακα Σ.***

Χμ… 

- Ο γεωμετρικός τόπος των σημείων με σταθερή πυκνότητα πιθανότητας αντιστοιχεί σε υπερ-ελλειψοειδή των οποίων η ***απόσταση που περιγράφεται από την εξίσωση***:

$$r^2 = (\boldsymbol{x} - \boldsymbol{\mu})^{\mathrm{t}} \Sigma^{-1} (\mathbf{x} - \mu)$$

είναι σταθερή.

- Οι κύριοι άξονες των υπερ-ελλειψοειδών δίνονται από τα ιδιοδιανύσματα του **Σ**: (**Φ**).
- Το μήκος των ιδιοδιανυσμάτων δίνεται από τις ιδιοτιμές του **Σ: (Λ)**

Χμ…

$$r^2 = (\mathbf{x}-\boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1} (\mathbf{x}-\boldsymbol{\mu})$$

- Η ανωτέρω ποσότητα ονομάζεται Mahalanobis απόσταση από το **x** στο **μ**:

- Η ανωτέρω ποσότητα ονομάζεται γενικευμένη απόσταση από το **x** στο **μ**:

- Ο όγκος του υπέρ-ελλειψοειδους που αντιστοιχεί σε mahalanobis απόσταση r είναι

$$V = V_d |\mathbf{\Sigma}|^{1/2} r^d,$$

- V_d είναι ο «***<u>μοναδιαίος όγκος</u>***»

$$V_d = \begin{cases} \pi^{d/2}/(d/2)! & d \text{ even} \\ \\ 2^d \pi^{(d-1)/2} (\frac{d-1}{2})!/(d)! & d \text{ odd.} \end{cases}$$

- Για δεδομένη διαστατικότητα η διασπορά των δειγμάτων είναι ανάλογη της $|\Sigma|^{(1/2)}$

- Όλες οι διαστάσεις έχουν την ίδια αξία.

- Οι διαστάσεις <u>**ΔΕΝ**</u> έχουν την ίδια αξία.

Συναρτήσεις διαχωρισμού για την κανονική κατανομή

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

Αν $p(x|\omega_i) \sim \mathrm{N}(\boldsymbol{\mu}_\mathrm{i}, \boldsymbol{\Sigma}_\mathrm{i})$ τότε:

$$g_i(\boldsymbol{x}) = -\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}_i)^t \boldsymbol{\Sigma}_i{}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_i) - \frac{d}{2}\ln 2\pi - \frac{1}{2}\ln|\boldsymbol{\Sigma}_\mathrm{i}| + \ln P(\omega_i)$$

Περίπτωση 1: $\boldsymbol{\Sigma}_i = \sigma^2 \boldsymbol{I}$

- Στατιστικώς ανεξάρτητα χαρακτηριστικά.
- Κάθε χαρακτηριστικό έχει την ίδια διασπορά σ^2.

Τότε: $\boldsymbol{\Sigma}_i = \sigma^2 I$

- Τα δείγματα βρίσκονται σε σφαιρικά νέφη ίδιου μεγέθους.
- Το κέντρο του κάθε νέφους βρίσκεται στην θέση $\boldsymbol{\mu}_{\mathrm{i}}$

$$|\boldsymbol{\Sigma}_{\mathrm{i}}| = \sigma^{2d}$$

$$\boldsymbol{\Sigma}_i^{-1} = \frac{\mathbf{1}}{\sigma^2} \boldsymbol{I}$$

- Οι όροι: Σ_{i} και $\frac{\mathrm{d}}{2}\ln 2\pi$ είναι ανεξάρτητοι της τάξης i.

$$g_i(\boldsymbol{x}) = -\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_i)^t {\Sigma_i}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_i) - \frac{d}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_{\mathrm{i}}| + \ln P(\omega_i)$$

Είναι απλές σταθερές

- Τελικά:

$$g_i(\boldsymbol{x}) = -\frac{\|\boldsymbol{x}-\boldsymbol{\mu}_i\|^2}{2\sigma^2} + \ln P(\omega_i)$$

Με την Ευκλείδεια νόρμα:

$$\|\boldsymbol{x}-\boldsymbol{\mu}_i\|^2 = (\boldsymbol{x}-\boldsymbol{\mu}_i)^t(\boldsymbol{x}-\boldsymbol{\mu}_i)$$

Περίπτωση 1: $\Sigma_i = \sigma^2 I$

- Ερμηνεία:
  - Αν οι a-priori πιθανότητες δεν είναι ίσες, τότε η απόσταση $\|\boldsymbol{x} - \boldsymbol{\mu}_i\|^2$ πρέπει να κανονικοποιηθεί με την διασπορά σ^2 και να προστεθεί ένα κατώφλι (offset) $\ln P(\omega_i)$.
  - Αν ένα άγνωστο $\boldsymbol{x}$ απέχει το ίδιο από τα κέντρα των νεφών (**ή προτύπων ή τάξεων ή κλάσεων**) τότε η απόφαση που λαμβάνεται εξαρτάται από τις οι a-priori πιθανότητες.

Περίπτωση 1: $\boldsymbol{\Sigma}_i = \sigma^2 \boldsymbol{I}$

$$\|\boldsymbol{x} - \boldsymbol{\mu}_i\|^2 = (\boldsymbol{x} - \boldsymbol{\mu}_i)^t (\boldsymbol{x} - \boldsymbol{\mu}_i)$$

$$g_i(x) = -\frac{1}{2\sigma^2}\left[\boldsymbol{x}^t\boldsymbol{x} - 2\mu_i{}^t\boldsymbol{x} + \mu_i{}^t\mu_i\right] + \ln P(\omega_i)$$

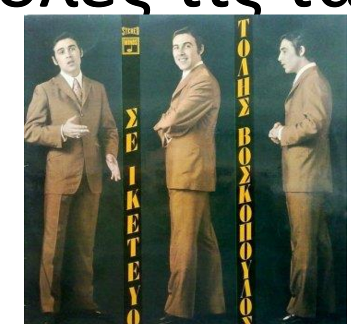
- Η παραπάνω φαίνεται να είναι μια σχέση δευτέρου βαθμού $\boldsymbol{x}^t\boldsymbol{x}$
- Όμως ο όρος $\boldsymbol{x}^t\boldsymbol{x}$ είναι ο ίδιος για όλες τις τάξεις.

Είναι το κάτι που μένει

Στίχοι: Βαρβάρα Τσιμπούλη
Μουσική: Νίκος Καρβέλας
Πρώτη εκτέλεση: Τόλης Βοσκόπουλο

Άφησες πίσω σου ένα κάτι,
κάτι από 'σένα που αγαπούσα,
κάτι που έγινε σκιά μου,

$$g_i(x) = w_i\boldsymbol{x} + w_{i0}$$



Περίπτωση 1: $\Sigma_i = \sigma^2 I$

$$\text{Όπου } w_i = \frac{1}{\sigma^2}\mu_i$$

$$\text{Και το κατώφλι: } w_{i0} = \frac{-1}{2\sigma^2}{\mu_i}^t\mu_i + \ln \mathrm{P}(\omega_\mathrm{i})$$

Περίπτωση 1 = Γραμμική Μηχανή – Linear Machine

- Η Γραμμική Μηχανή δημιουργεί επιφάνειες απόφασης με μορφή υπερεπιπέδων όπως αυτά καθορίζονται από τις σχέσεις:

$$g_i(x) = g_j(x)$$

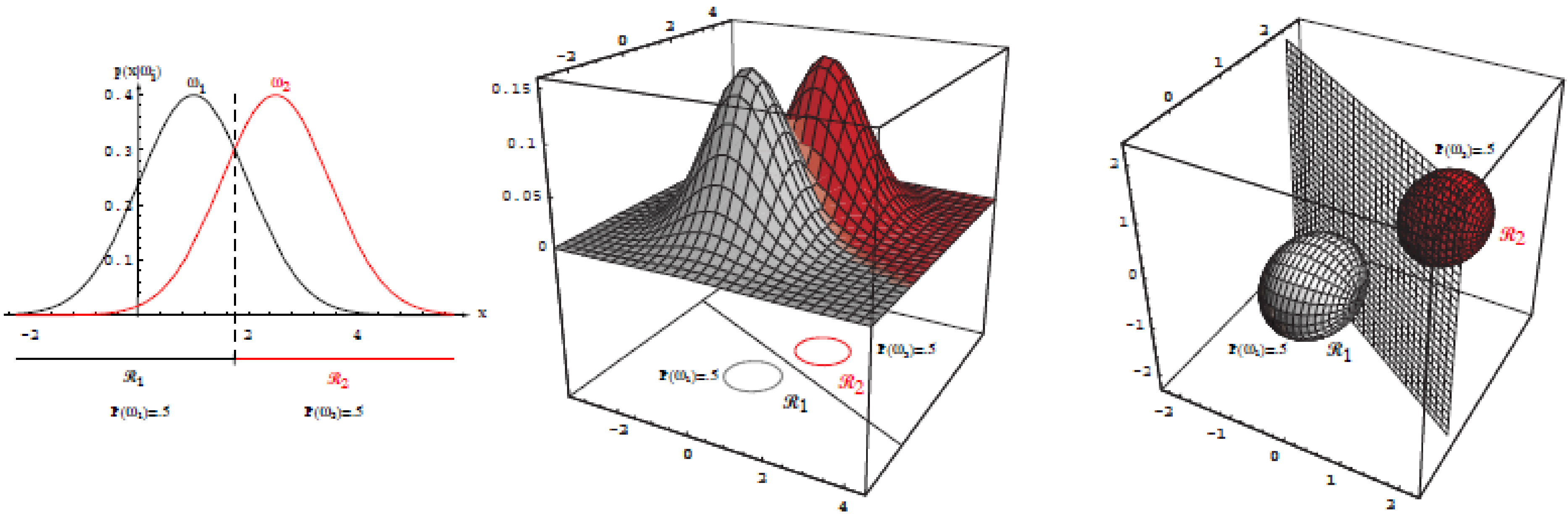


Figure 2.10: If the covariances of two distributions are equal and proportional to the identity matrix, then the distributions are spherical in d dimensions, and the boundary is a generalized hyperplane of $d-1$ dimensions, perpendicular to the line separating the means. In these 1-, 2-, and 3-dimensional examples, we indicate $p(\mathbf{x}|\omega_i)$ and the boundaries for the case $P(\omega_1) = P(\omega_2)$. In the 3-dimensional case, the grid plane separates $\mathcal{R}_1$ from $\mathcal{R}_2$.

Γραμμική Μηχανή

$$g_i(x) = g_j(x)$$

$$w^t(\boldsymbol{x} - \boldsymbol{x_0}) = 0$$

Όπου:

$$w = \mu_i - \mu_j$$

Και:

$$\mathrm{x}_0 = \frac{1}{2}(\boldsymbol{\mu}_i + \boldsymbol{\mu}_j) - \frac{\sigma^2}{\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\|^2} \ln \frac{P(\omega_i)}{P(\omega_j)} (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j).$$

Γραμμική Μηχανή

- Η εξίσωση: $$x_0 = \frac{1}{2}(\mu_i + \mu_j) - \frac{\sigma^2}{\|\mu_i - \mu_j\|^2} \ln \frac{P(\omega_i)}{P(\omega_j)} (\mu_i - \mu_j).$$

Καθορίζει ένα υπερεπίπεδο που περνάει από το σημείο $\mathbf{x}_0$ και είναι κάθετο στο διάνυσμα **w**.

- Αν $P(\omega_i)=P(\omega_j)$ τότε ο όρος γίνεται

ΛΑΚΗΣ

- Το σημείο $\mathbf{x}_0$ βρίσκεται μεταξύ των μ_i, μ_j και το υπερ-επίπεδο είναι η κάθετος διχοτόμος της γραμμής που ενώνει τα μ_i, μ_j

- Αν $P(\omega_i) \neq P(\omega_j)$ τότε το σημείο $\mathbf{x_0}$ απομακρύνεται από το κέντρο με την μεγαλύτερη a-priori πιθανότητα.

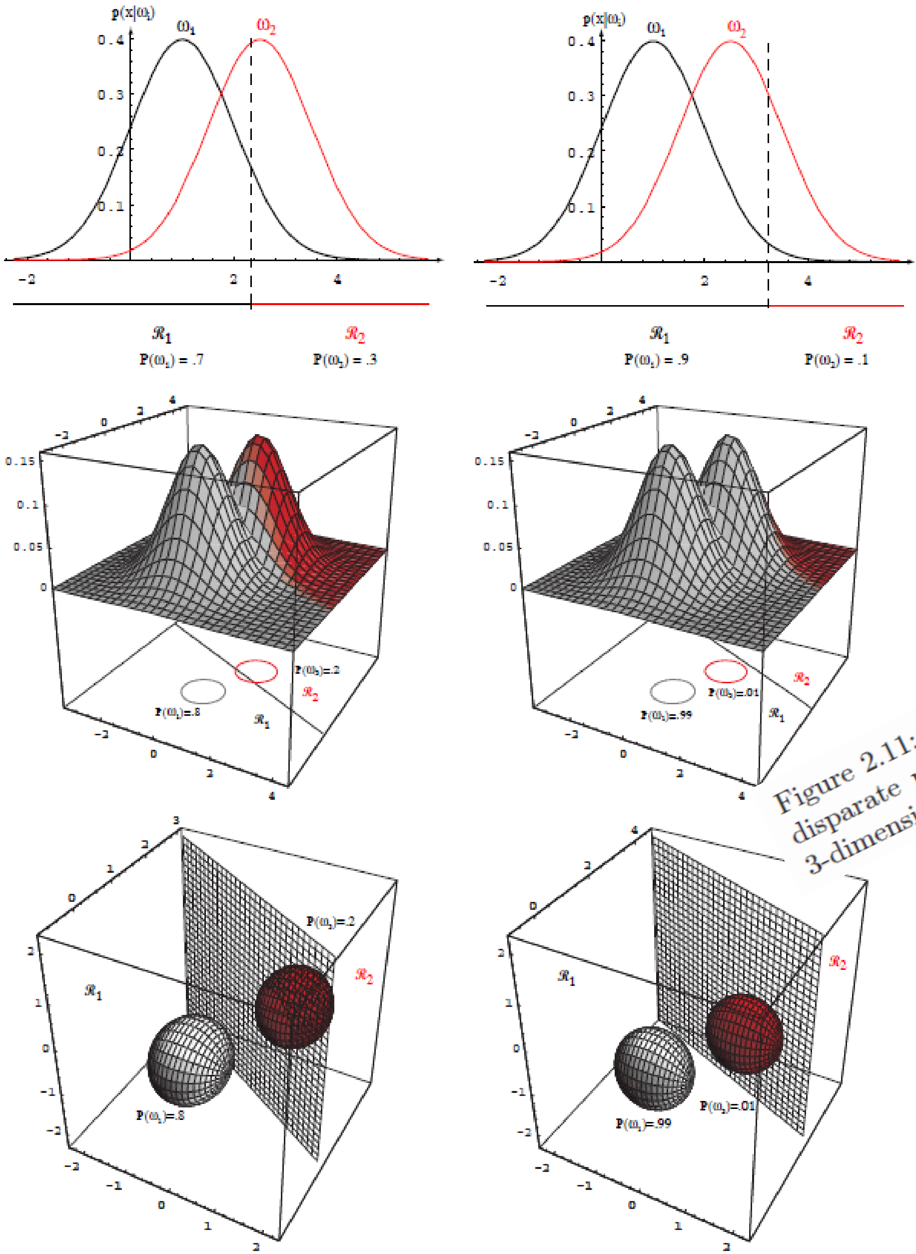


Figure 2.11: As the priors are changed, the decision boundary shifts; for sufficiently disparate priors the boundary will not lie between the means of these 1-, 2- and 3-dimensional spherical Gaussian distributions.

Περίπτωση: $\boldsymbol{\Sigma}_i=\boldsymbol{\Sigma}$

- Οι πίνακες συνδιακύμανσης για όλες τις κατηγορίες είναι ίδιες αλλά όχι διαγώνιες.
- Γεωμετρικά, αυτό αντιστοιχεί στην κατάσταση στην οποία τα δείγματα πέφτουν σε υπέρ-ελλειψοειδή συστάδες ίσου μεγέθους και σχήματος.
  - Η i-τάξη «συγκεντρώνεται» γύρω από το μέσο διάνυσμα: $\boldsymbol{\mu}_i$
- Οι όροι: Σ_i και $\frac{d}{2}\ln 2\pi$ είναι ανεξάρτητοι της τάξης i.

Περίπτωση: **Σ**$_i$=**Σ**

$$g_i(\boldsymbol{x}) = -\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_i)^t \Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_i) + \ln P(\omega_i)$$

Αν $P(\omega_i)$ ίδιες για όλες τις **c** το πλήθος κατηγορίες τότε $\ln P(\omega_i)$ δεν χρειάζεται.

Κανόνας απόφασης:

Ανάθεσε το άγνωστο διάνυσμα **x** στην τάξη με την μικρότερη απόσταση**:**

$$(\boldsymbol{x}-\boldsymbol{\mu}_i)^t \Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_i)$$

της κάθε i-κατηγορίας, i=1:...c**.**

Ανάπτυγμα της $(\boldsymbol{x}-\boldsymbol{\mu}_i)^t\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_i)$

- όρος $\boldsymbol{x}^t\Sigma^{-1}\boldsymbol{x}$ είναι ίδιος για όλες τις τάξεις.
- Τελικά:

$$g_i(\mathbf{x}) = \mathbf{w}_i^t\mathbf{x} + w_{i0},$$

Με: $$\mathbf{w}_i = \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_i$$

Και: $$w_{i0} = -\frac{1}{2}\boldsymbol{\mu}_i^t\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_i + \ln P(\omega_i).$$

Περίπτωση: $\boldsymbol{\Sigma}_i=\boldsymbol{\Sigma}$

- Δεδομένου ότι οι συναρτήσεις διαχωρισμού είναι γραμμικές, τα προκύπτοντα όρια απόφασης είναι πάλι υπέρ-επίπεδα.
- Αν R_i, R_j είναι συνεχείς το όριο μεταξύ τους έχει την εξίσωση:

$$\mathbf{w}^t(\mathbf{x}-\mathbf{x}_0)=0,$$

Με: $$\mathbf{w}=\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j)$$

και: $$\mathbf{x}_0=\frac{1}{2}(\boldsymbol{\mu}_i+\boldsymbol{\mu}_j)-\frac{\ln\left[P(\omega_i)/P(\omega_j)\right]}{(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j)^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j)}(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j).$$

Περίπτωση: $\boldsymbol{\Sigma}_i=\boldsymbol{\Sigma}$

- Καθώς το διάνυσμα $\mathbf{w} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j)$ γενικά δεν είναι στην κατεύθυνση του $(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j)$ το υπέρ-επίπεδο το οποίο διαχωρίζει τις R_i, R_j, δεν είναι κάθετο στην γραμμή των μέσων.
- Παρόλα αυτά τέμνεται με την γραμμή των μέσων στο σημείο $\mathbf{x}_0$ που είναι στο μέσο της γραμμής μεταξύ των $\boldsymbol{\mu}_i$,$\boldsymbol{\mu}_j$ αν οι a-priori ποσότητες είναι ίδιες.
- Αν οι a-priori ποσότητες δεν είναι ίδιες το βέλτιστο υπέρ-επίπεδο άπομακρύνεται από το $\boldsymbol{\mu}$ με το μεγαλύτερο prior.

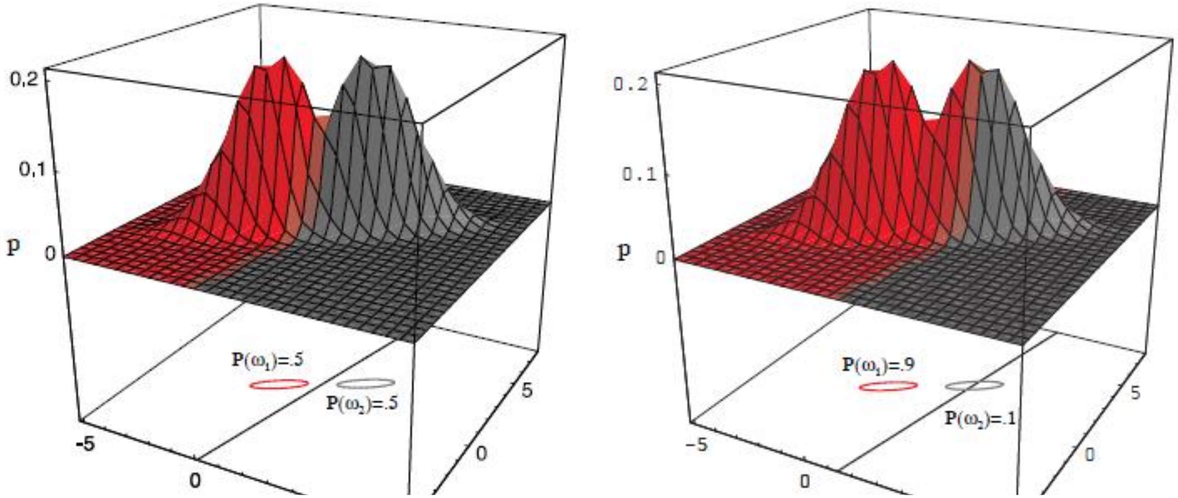
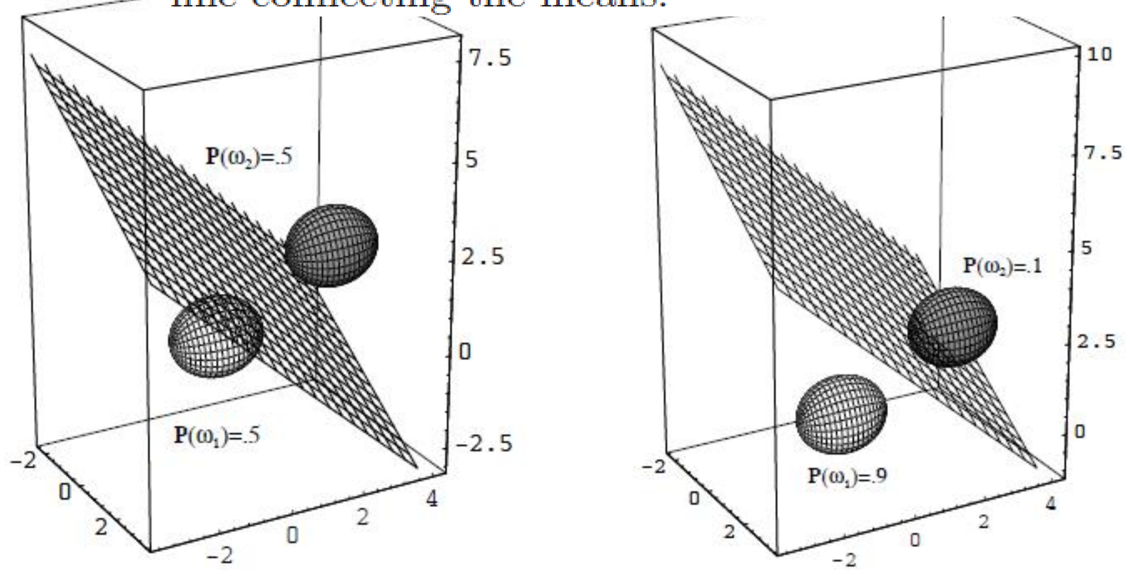


Figure 2.12: Probability densities (indicated by the surfaces in two dimensions and ellipsoidal surfaces in three dimensions) and decision regions for equal but asymmetric Gaussian distributions. The decision hyperplanes need not be perpendicular to the line connecting the means.



Περίπτωση με αυθαίρετες Σ_i

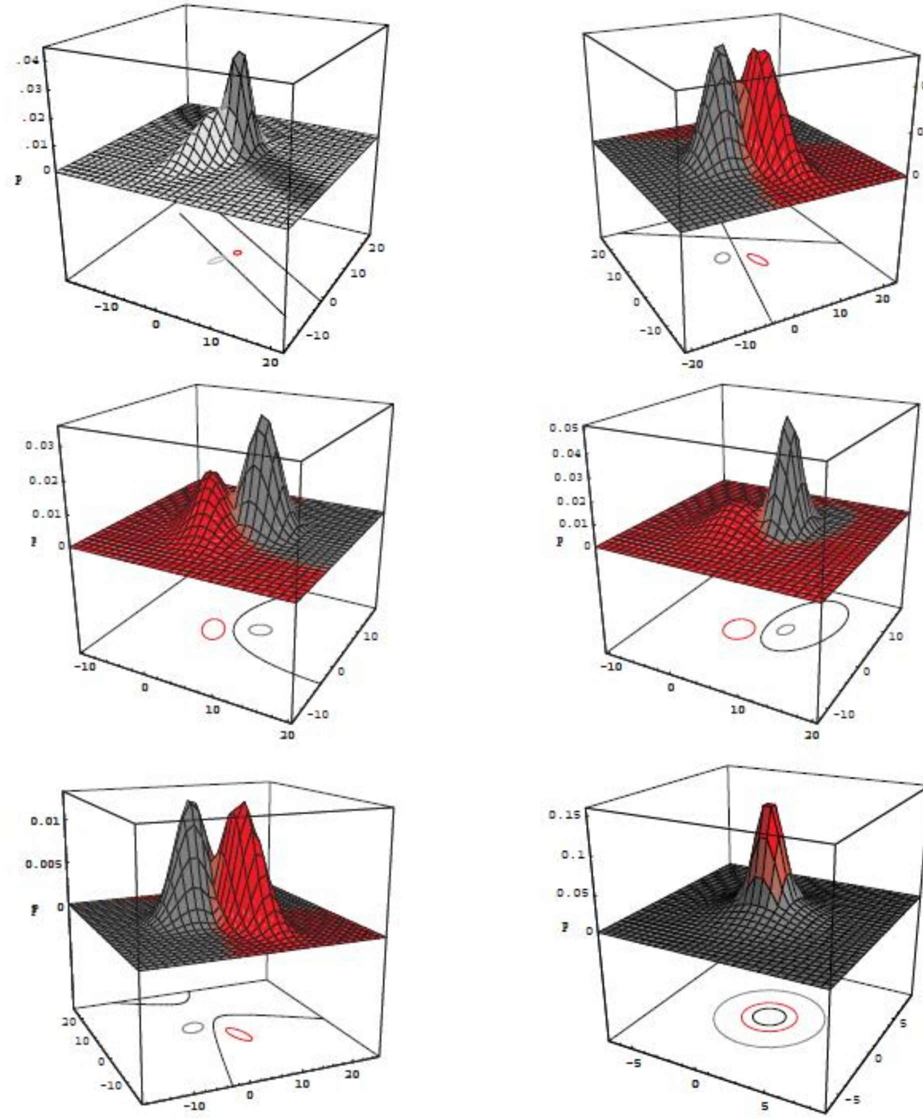
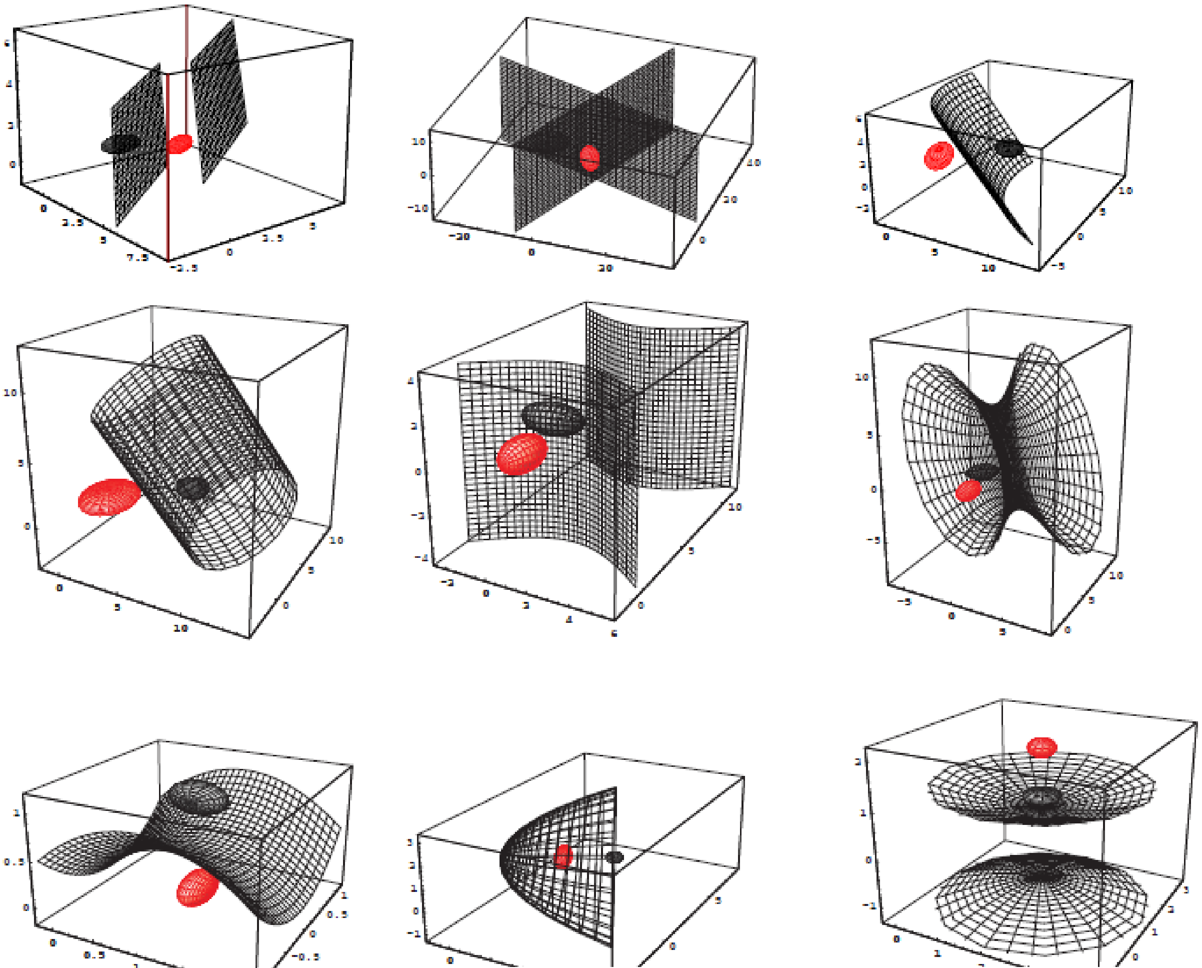


Figure 2.14: Arbitrary Gaussian distributions lead to Bayes decision boundaries that are general hyperquadrics. Conversely, given any hyperquadratic, one can find two Gaussian distributions whose Bayes decision boundary is that hyperquadric.



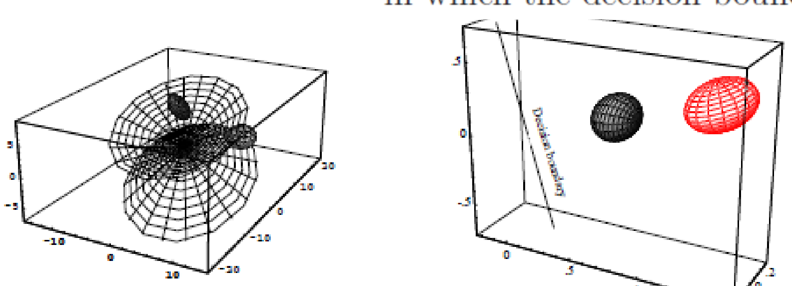


Figure 2.15: Arbitrary three-dimensional Gaussian distributions yield Bayes decision boundaries that are two-dimensional hyperquadrics. There are even degenerate cases in which the decision boundary is a line.

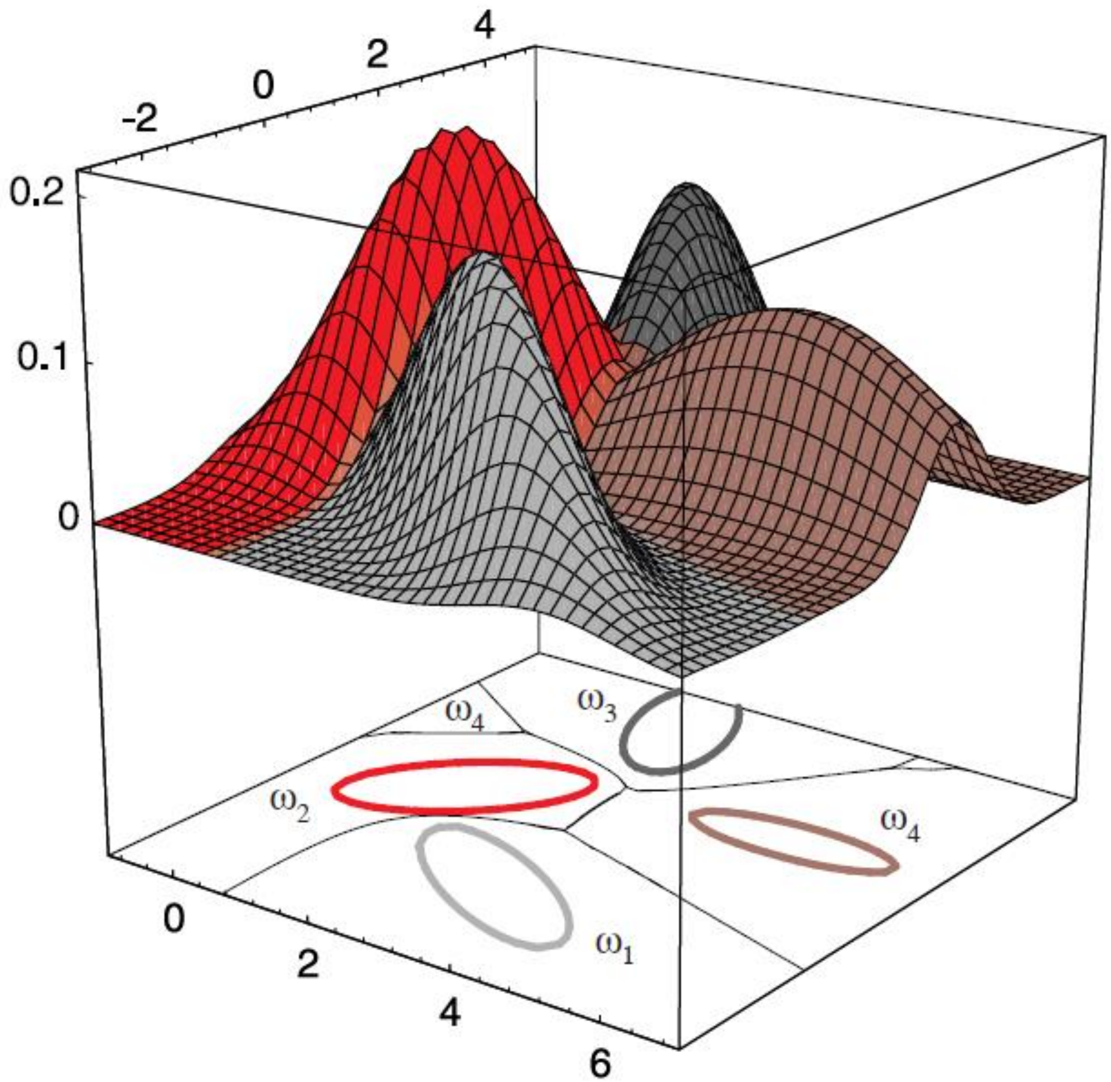
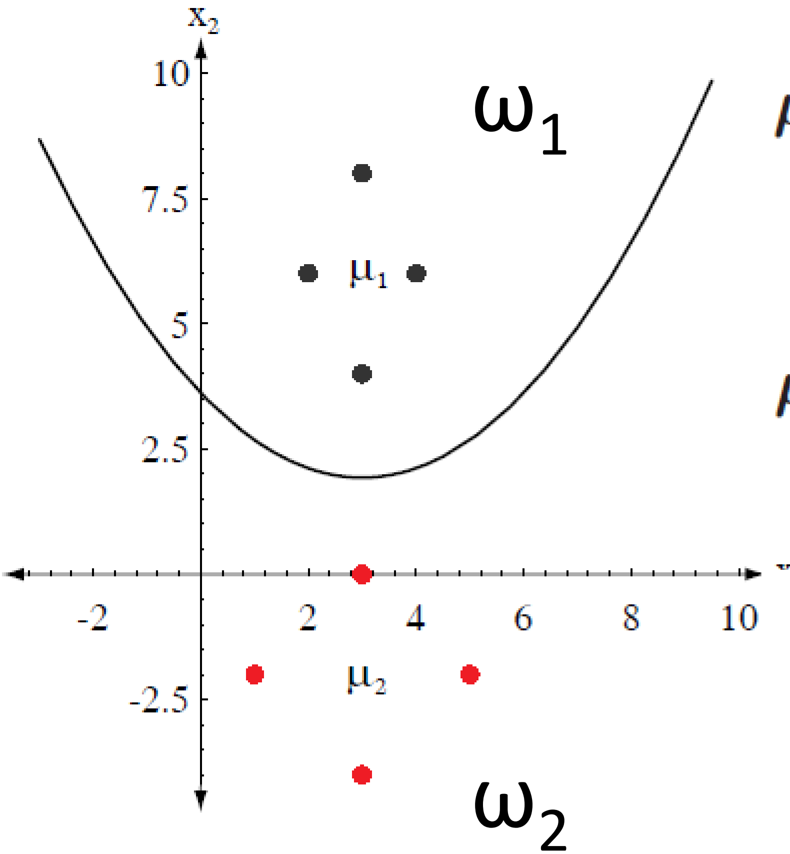


Figure 2.16: The decision regions for four normal distributions. Even with such a low number of categories, the shapes of the boundary regions can be rather complex.

Παράδειγμα



$$\boldsymbol{\mu}_1 = \begin{bmatrix} 3 \\ 6 \end{bmatrix}; \quad \boldsymbol{\Sigma}_1 = \begin{pmatrix} 1/2 & 0 \\ 0 & 2 \end{pmatrix}$$

$$\boldsymbol{\mu}_2 = \begin{bmatrix} 3 \\ -2 \end{bmatrix}; \quad \boldsymbol{\Sigma}_2 = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}.$$

$$\boldsymbol{\Sigma}_1^{-1} = \begin{pmatrix} 2 & 0 \\ 0 & 1/2 \end{pmatrix}$$

$$\boldsymbol{\Sigma}_2^{-1} = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}.$$

Παράδειγμα: έστω $P(\omega_1)= P(\omega_2)$

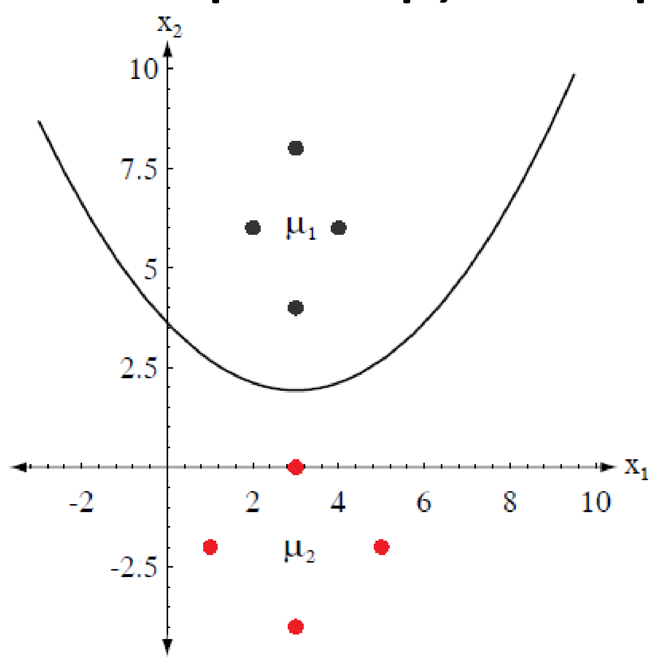
- Εξίσωση ορίου συνάρτησης διαχωρισμού:
  $\mathbf{g}_1(\mathbf{x})=\mathbf{g}_2(\mathbf{x})$

$$x_2 = 3.514 - 1.125x_1 + 0.1875x_1^2.$$

Παραβολή με ακμή στο σημείο: (3,1.83)

Σημειώστε ότι παρά το γεγονός ότι η διακύμανση των δεδομένων κατά μήκος της κατεύθυνσης $\mathbf{x_2}$ και για τις δύο κατανομές είναι ίδια, το όριο απόφασης δεν περνά μέσα από το σημείο που αντιστοιχεί στο μέσο (3,2) μεταξύ των μέσων, όπως ίσως να υποθέσουμε αφελώς.

- Επειδή οι a-priori πιθανότητες είναι οι ίδιες η κατανομή αυξάνεται κατά μήκος της κατεύθυνσης x_2 (σε σχέση με εκείνη της κατανομής ω_2).
- Έτσι, το όριο απόφασης βρίσκεται ελαφρώς χαμηλότερα από το σημείο από το σημείο που αντιστοιχεί στο μέσο (3,2) μεταξύ των μέσων, όπως φαίνεται στο όριο της απόφασης.



Άλλο Παράδειγμα:

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

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

τότε:

Ταξινομήσατε το διάνυσμα $\underline{x} = \begin{bmatrix} 1.0 \\ 2.2 \end{bmatrix}$ με Bayesian κανόνα:

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

- Mahalanobis d_m από μ_1, μ_2 :

$$d^2_{m,1} = [1.0, \quad 2.2]\Sigma^{-1}\begin{bmatrix} 1.0 \\ 2.2 \end{bmatrix} = 2.952,$$

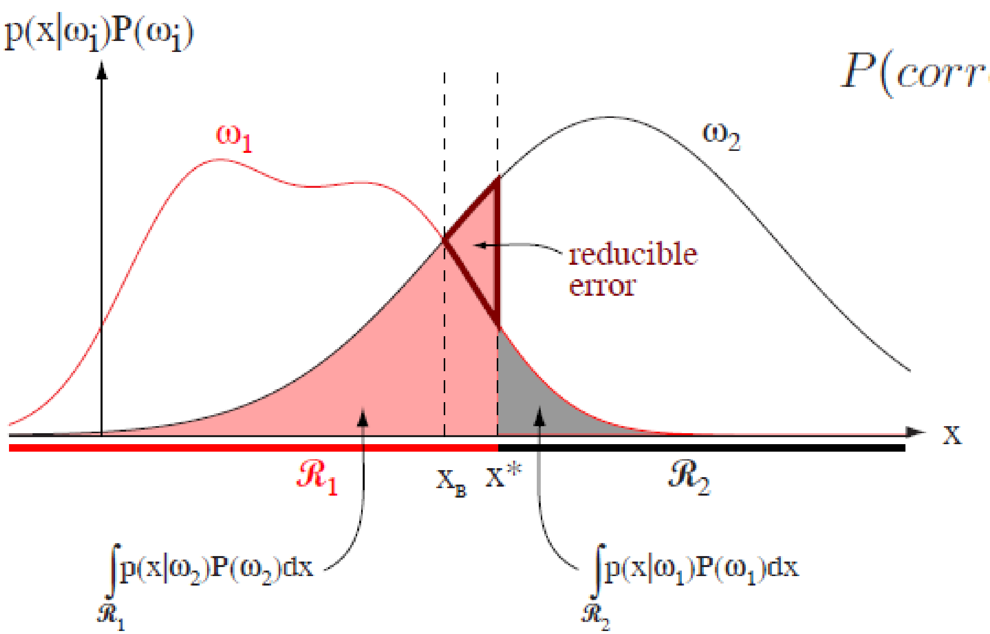
$$d^2_{m,2} = [-2.0, \quad -0.8]\,\Sigma^{-1}\begin{bmatrix} -2.0 \\ -0.8 \end{bmatrix} = 3.672$$

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

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad \text{then} \quad A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

Πιθανότητες σφάλματος

$$
\begin{aligned}
P(error) &= P(\mathbf{x} \in \mathcal{R}_2, \omega_1) + P(\mathbf{x} \in \mathcal{R}_1, \omega_2) \\
&= P(\mathbf{x} \in \mathcal{R}_2 | \omega_1) P(\omega_1) + P(\mathbf{x} \in \mathcal{R}_1 | \omega_2) P(\omega_2) \\
&= \int\limits_{\mathcal{R}_2} p(\mathbf{x}|\omega_1) P(\omega_1)\ d\mathbf{x} + \int\limits_{\mathcal{R}_1} p(\mathbf{x}|\omega_2) P(\omega_2)\ d\mathbf{x}.
\end{aligned}
$$



$$
\begin{aligned}
P(correct) &= \sum_{i-1}^{c} P(\mathbf{x} \in \mathcal{R}_i, \omega_i) \\
&= \sum_{i=1}^{c} P(\mathbf{x} \in \mathcal{R}_i | \omega_i) P(\omega_i) \\
&= \sum_{i=1}^{c} \int\limits_{\mathcal{R}_i} p(\mathbf{x}|\omega_i) P(\omega_i)\ d\mathbf{x}.
\end{aligned}
$$

Πιθανότητες σφάλματος - Bayes

- Το γενικό αποτέλεσμα της εξίσωσης δεν εξαρτάται ούτε από τον τρόπο με τον οποίο ο χώρος χαρακτηριστικών χωρίζεται σε περιοχές λήψης αποφάσεων ούτε από τη μορφή των υποκείμενων κατανομών.

$$= \sum_{i=1}^{c} \int_{\mathcal{R}_i} p(\mathbf{x}|\omega_i) P(\omega_i)\, d\mathbf{x}.$$

- Ο ταξινομητής Bayes μεγιστοποιεί αυτήν την πιθανότητα επιλέγοντας τις περιοχές έτσι ώστε το ολοκλήρωμα να είναι μέγιστο για όλα τα **x**.
- Καμία άλλη διαμέριση δεν μπορεί να αποδώσει μικρότερη πιθανότητα σφάλματος.

Θεωρία αποφάσεων κατά Bayes Διακριτά χαρακτηριστικά

- Το χαρακτηριστικό x λαμβάνει μια από τις m-διακριτές τιμές $v_1,\ldots,v_m$.

$$\int p(\mathbf{x}|\omega_j)\; d\; \alpha^* = \arg\max_i R(\alpha_i|\mathbf{x}). \quad \omega_j),$$

$$\mathbf{x}$$

Τότε: $\boxed{P(\omega_j|\mathbf{x}) = \frac{P(\mathbf{x}|\omega_j)P(\omega_j)}{P(\mathbf{x})},}$ όπου: $P(\mathbf{x}) = \sum_{j=1}^{c} P(\mathbf{x}|\omega_j)P(\omega_j).$

Ο ορισμός των ρίσκων R(α|**x**) παραμένει ίδιος και ο κανόνας του Bayes γράφεται: $\alpha^* = \arg\max_i R(\alpha_i|\mathbf{x}).$

Γραμμική Μηχανή

- Αν και οι a-priori πιθανότητες είναι ίδιες για όλες τις S - το πλήθος τάξεις τότε ο όρος $\ln P(\omega_i)$ μπορεί να επαλειφθεί.
- Τότε, για να ταξινομήσω ένα άγνωστο x μετράω την Ευκλείδια απόστασή του$\|\boldsymbol{x} - \boldsymbol{\mu}_i\|$ από όλα τα $\boldsymbol{\mu}_i$,i=1,…S κέντρα και αναθέτω το x στην κατηγορία – τάξη –κλάση κ.ο.κ που έχει την μικρότερη απόσταση.
- Minimum Distance Classifier

Γραμμική Μηχανή

- Αν θεωρήσω ότι το κάθε κέντρο είναι το ιδανικό πρωτότυπο των δειγμάτων που ανήκουν σε μια τάξη – κατηγορία – κλάση τότε η ανωτέρω τεχνική καλείται:

Template Matching