ΔΗΜΟΚΡΙΤΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΡΑΚΗΣ

Αναγνώριση Προτύπων

Baysian Θεωρία Αποφάσεων (DETECTION)

Χριστόδουλος Χαμζάς

Τα περιεχόμενα των παρουσιάσεων προέρχονται από τις παρουσιάσεις του αντίστοιχου διδακτέου μαθήματος του καθ. Παναγιώτη Τσακαλίδη, Τμ. Επιστήμης Υπολογιστών, Παν. Κρήτης. Το πρωτογενές υλικό βρίσκεται στην σελίδα http://www.csd.uoc.gr/~hy473/ και βασίζεται στο βιβλίο: "Pattern Classification", R.O. Duda, P.E. Hart, D.G. Stork, Wiley, 2nd Ed., 2001

Τμήμα Ηλεκτρολόγων Μηχανικών & Μηχανικών Υπολογιστών

Μπευζιανή θεωρία αποφάσεων

- Στατιστικά βέλτιστη ταξινόμηση.
- Βασίζεται στην περιγραφή του προβλήματος ταξινόμησης με πιθανοτικούς όρους.
- Η θεωρία υποθέτει:
  - Το πρόβλημα απόφασης μπορεί να τεθεί με όρους πιθανοτήτων
  - Είναι γνωστές όλες οι απαραίτητες τιμές και συναρτήσεις πιθανότητας (στην πράξη αυτό δεν ισχύει).

Προβλημα Εκτίμησης

- Λαμβάνουμε το σήμα (μέτρηση)

$$x(t) = \begin{cases} n(t) \\ s(t) + n(t) \end{cases}$$

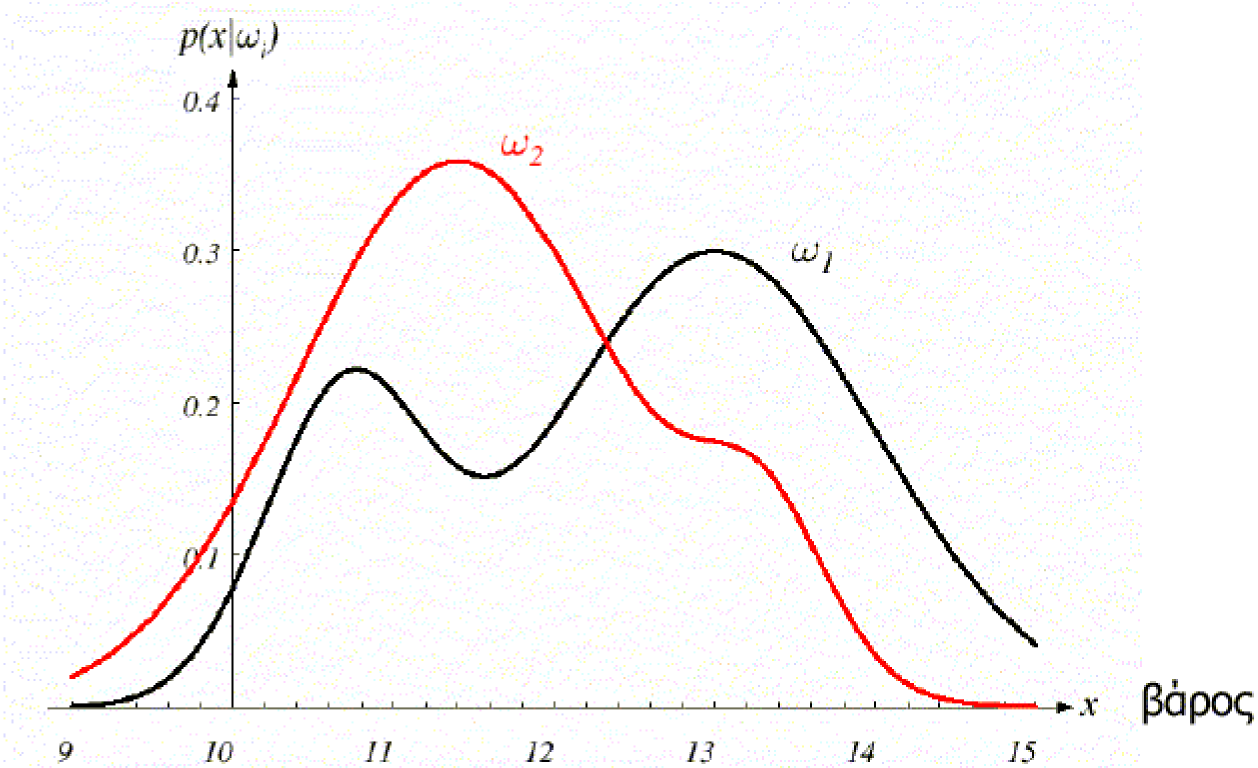
- n(t) είναι ο θόρυβος και s(t) είναι το σήμα

Δεσμευμένες πιθανότητες

ω_1: Λαβράκι
ω_2: Σολομός

•p(x | ω_2): Δεσμευμένη συνάρτηση πυκνότητας πιθανότητας (ΣΠΠ) της μεταβλητής x δεδομένης της κατάστασης της φύσης.

•Πιθανοφάνεια: Δεδομένου ότι παρατηρείται σολομός, ποιά είναι η πιθανότητα, το βάρος του να είναι μεταξύ 11 και 12 ?



Ορισμοί και κανόνας απόφασης Bayes

- Κατάσταση της φύσης (state of nature)
- Εκ των προτέρων πιθανότητα (prior)
- Εκ των υστέρων πιθανότητα (posterior)
- Πιθανοφάνεια (Likelihood)
- Evidence

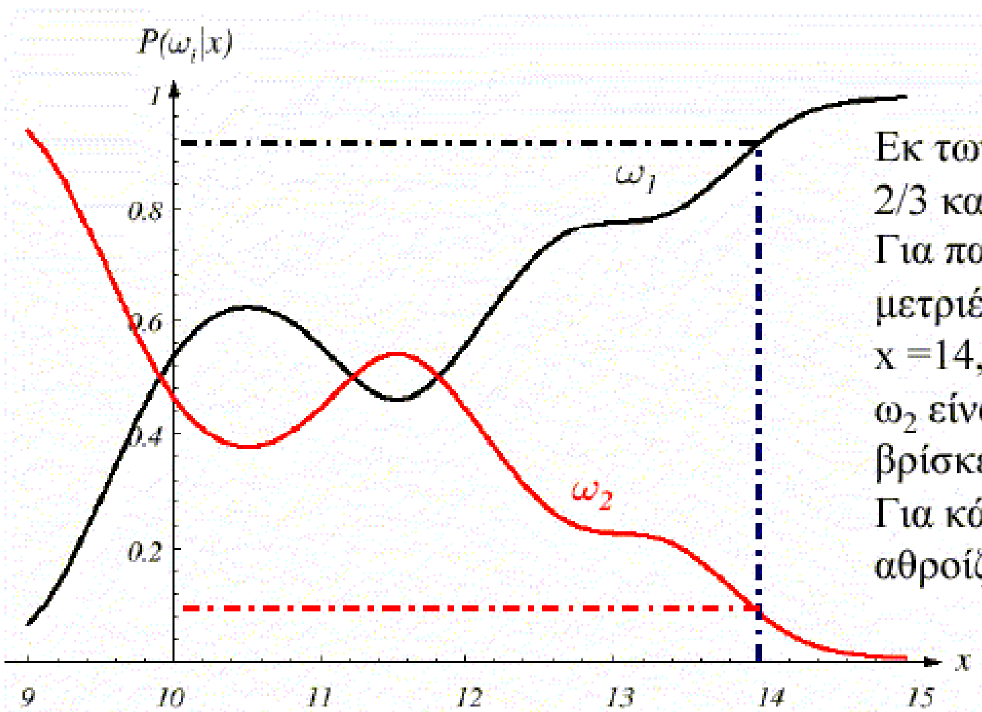
$$P(\omega_j \mid x) = \frac{p(x \mid \omega_j) \cdot P(\omega_j)}{\sum_{x \in X} p(x \mid \omega_j) \cdot P(\omega_j)} = \frac{p(x \mid \omega_j) \cdot P(\omega_j)}{p(x)}$$

Εάν έχουμε 2 κατηγορίες ω_1 και ω_2 τότε επιλέγουμε ω_1 εάν

$$\boxed{\frac{p(x/\omega_1)}{p(x/\omega_2)} \overset{\omega_1}{>} \frac{p(\omega_2)}{p(\omega_1)} = \frac{1-p(\omega_1)}{p(\omega_1)}}$$

Εκ των υστέρων πιθανότητες

Ο κανόνας Bayes επιτρέπει τον υπολογισμό της εκ των υστέρων πιθανότητας (δύσκολο να καθοριστεί) από την εκ των προτέρων πιθανότητα, την πιθανοφάνεια, και τις αποδείξεις (υπολογίζονται ευκολότερα).



Εκ των υστέρων πιθανότητες όταν $P(\omega_1) = 2/3$ και $P(\omega_2) = 1/3$.

Για παράδειγμα, δεδομένου ότι ένα πρότυπο μετριέται με τιμή χαρακτηριστικού (feature) $x = 14$, η πιθανότητα να βρίσκεται στην τάξη ω_2 είναι περίπου 0.08, ενώ η πιθανότητα να βρίσκεται στην ω_1 είναι 0.92.

Για κάθε x, οι εκ των υστέρων πιθανότητες αθροίζονται στο 1.0.

Κανόνας απόφασης Bayes

Επιλογή της κλάσης που έχει τη μεγαλύτερη εκ των υστέρων πιθανότητα !!!

Επιλογή της ω_i εάν $P(\omega_i/x) \geq P(\omega_j/x)$ για όλα τα $j=1,2,\ldots,c$

Δηλαδή $P(\omega_i/x)=\max[P(\omega_1/x), P(\omega_2/x),\ldots, P(\omega_c/x)]$ και

$$P_e = P(error) = 1 - \max[P(\omega_1/x), P(\omega_2/x),\ldots, P(\omega_c/x)]$$

Εάν υπάρχουν πολλά χαρακτηριστικά, $\mathbf{x}=\{x_1, x_1, \ldots, x_d\}$ τότε

Επιλογή της ω_i εάν $P(\omega_i/\mathbf{x}) \geq P(\omega_j/\mathbf{x})$ για όλα τα $j=1,2,\ldots,c$

$P_e = P(error) = 1 - \max[P(\omega_1/\mathbf{x}), P(\omega_2/\mathbf{x}),\ldots, P(\omega_c/\mathbf{x})]$ και

$$P_{e,total} = \int p(x)P_e(x)dx = 1 - \int \max_k(p(\omega_k, x)dx$$

και για διακριτά χαρακτηριστικά

$$P_{e,total} = \sum_i P(x_i)P_e(x_i) = 1 - \sum_i \left(\max_k(P(\omega_k, x_i))\right)$$

Συνάρτηση Κόστους

- Μαθηματική περιγραφή του κόστους κάθε επιλογής.
- Είναι κάποιες επιλογές περισσότερο «ακριβές» από άλλες?

✓ $\{\omega_1, \omega_2, \ldots, \omega_c\}$: Σύνολο φυσικών καταστάσεων (κλάσεις - classes)

✓ $\mathbf{x} = [x_1, \ldots, x_d]^T$: Διάνυσμα χαρακτηριστικών (feature vector)

✓ $\{\alpha_1, \alpha_2, \ldots \alpha_a\}$: Σύνολο δυνατών δράσεων (actions). Προσέξτε ότι το a δεν χρειάζεται να είναι το ίδιο με το c, καθώς μπορούμε να πραγματοποιούμε περισσότερες ή λιγότερες δράσεις από το πλήθος των τάξεων. Για παράδειγμα, η απόρριψη είναι επίσης μια δυνατή δράση.

✓ $\lambda(\alpha_i | \omega_j\}$: Κόστος (cost) της δράσης α_i όταν η πραγματική κατάσταση της φύσης είναι ω_j.

✓ $R(\alpha_i | \mathbf{x})$: Δεσμευμένο ρίσκο (conditional risk) – Αναμενόμενη απώλεια για την δράση α_i.

$$\boxed{R(\alpha_i \mid \mathbf{x}) = \sum_{j=1}^{c} \lambda(\alpha_i \mid \omega_j) \cdot P(\omega_j \mid \mathbf{x})}$$

Η απόφαση κατά Bayes επιλέγει τη δράση που ελαχιστοποιεί το δεσμευμένο ρίσκο!

Απόφαση κατά Bayes με βάση το δεσμευμένο ρίσκο

1. Υπολογισμός δεσμευμένου ρίσκου $R(\alpha_i | \mathbf{x})$ για κάθε δράση.
2. Επιλογή δράσης με το ελάχιστο δεσμευμένο ρίσκο. Έστω ότι είναι η δράση k.
3. Το συνολικό ρίσκο είναι:
$$R = \int_{\mathbf{x} \in X} R(\alpha_k | \mathbf{x}) \cdot p(\mathbf{x}) d\mathbf{x}$$
4. Αυτό είναι το ρίσκο του Bayes, <u>το ελάχιστο δυνατό ρίσκο</u> που μπορεί να έχει ένας οποιοσδήποτε ταξινομητής !
5. Π.χ. ταξινόμηση σε μία από δύο κλάσεις:

$$\frac{p(x/\omega_1)}{p(x/\omega_2)} \overset{\omega_1}{\geq} \frac{\lambda_{12} - \lambda_{22}}{\lambda_{21} - \lambda_{11}} \frac{P(\omega_2)}{P(\omega_1)}$$

Ταξινόμηση ελάχιστης πιθανότητας λάθους

- Εάν ως δράση α_i επιλέξουμε την ταξινόμηση στην κλάση ω_j, και εάν όλα τα κόστη λάθος ταξινόμησης είναι ίσα με μονάδα, έχουμε τη λεγόμενη **συμμετρική** ή **0-1** επιλογή:

$$\lambda(\alpha_i \mid \omega_j) = \begin{cases} 0, & \text{if } i = j \\ 1, & \text{if } i \neq j \end{cases}$$

- Αυτή η συνάρτηση κόστους προβλέπει μηδενική απώλεια για σωστή ταξινόμηση, και μοναδιαία απώλεια για λάθος ταξινόμηση. Το αντίστοιχο δεσμευμένο ρίσκο που αντιστοιχεί σε αυτή τη συνάρτηση κόστους είναι

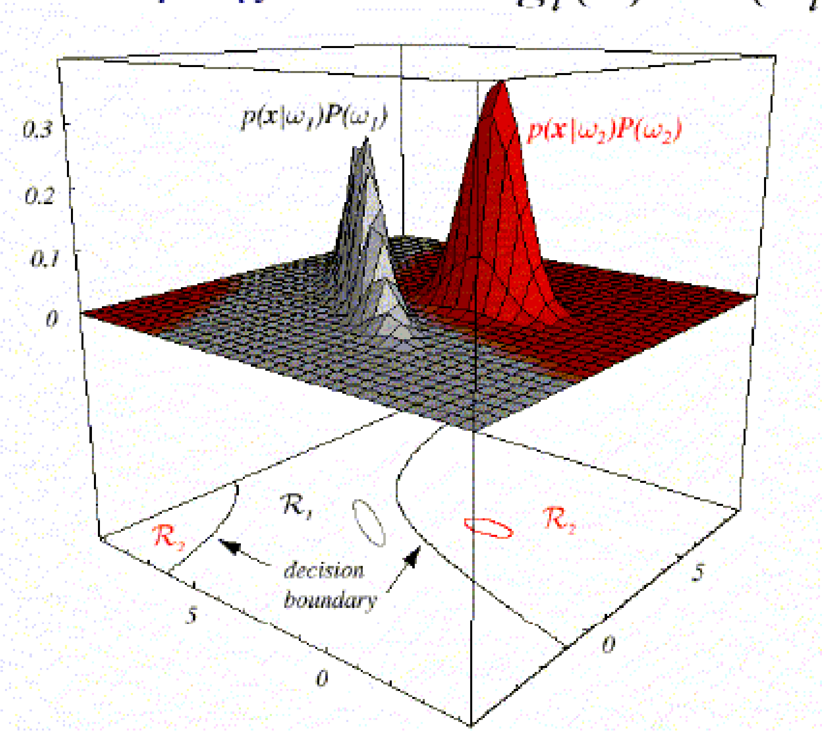
$$R(\alpha_i \mid \mathbf{x}) = \sum_{\substack{j \neq i \\ j=1,\ldots,c}} P(\omega_j \mid \mathbf{x}) = 1 - P(\omega_j \mid \mathbf{x})$$

το οποίο είναι ακριβώς η πιθανότητα λάθους. Προφανώς, για να ελαχιστοποιήσουμε το ρίσκο, θα πρέπει να επιλέξουμε την κλάση που μεγιστοποιεί την εκ των υστέρων πιθανότητα!!!

Ταξινόμηση βάσει συναρτήσεων διάκρισης

- Η συνάρτηση διάκρισης *g(x)*, διαχωρίζει τις κλάσεις μεταξύ τους. Αυτή η συνάρτηση αντιστοιχεί το διάνυσμα εισόδου σε μια κλάση σύμφωνα με τον ορισμό: Επέλεξε την τάξη *i* εάν $g_i(\mathbf{x}) > g_j(\mathbf{x}) \quad \forall i \neq j, \quad i, j = 1,2,\ldots,c$

- Ο κανόνας Bayes μπορεί να υλοποιηθεί με τη μορφή συναρτήσεων διάκρισης $g_i(\mathbf{x}) = P(\omega_i \mid \mathbf{x})$



Κάθε συνάρτηση διάκρισης δημιουργεί *c* περιοχές απόφασης, $R_1, \ldots, R_c$, οι οποίες χωρίζονται από τις *επιφάνειες απόφασης.*

Οι περιοχές απόφασης δεν απαιτείται να είναι συνεχείς.

Οι επιφάνειες απόφασης ικανοποιούν την $g_i(\mathbf{x}) = g_j(\mathbf{x})$

Γκαουσιανές συναρτήσεις ΠΠ

- Εάν οι συναρτήσεις πιθανοφάνειας ακλουθούν την πολυδιάστατη Γκαουσιανή κατανομή

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

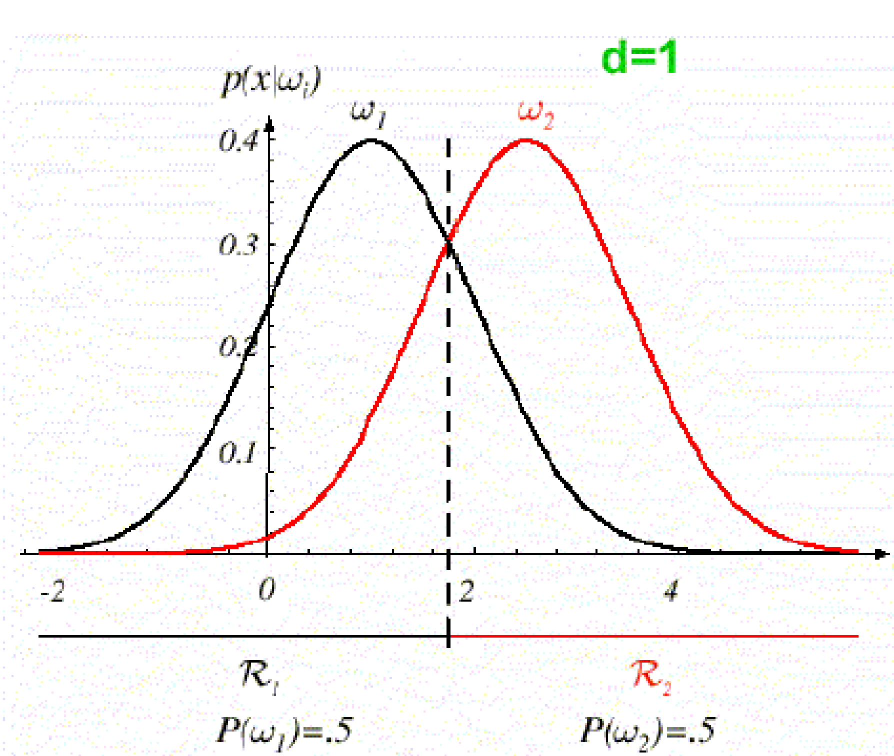
τότε η συνάρτηση διάκρισης παίρνει την μορφή

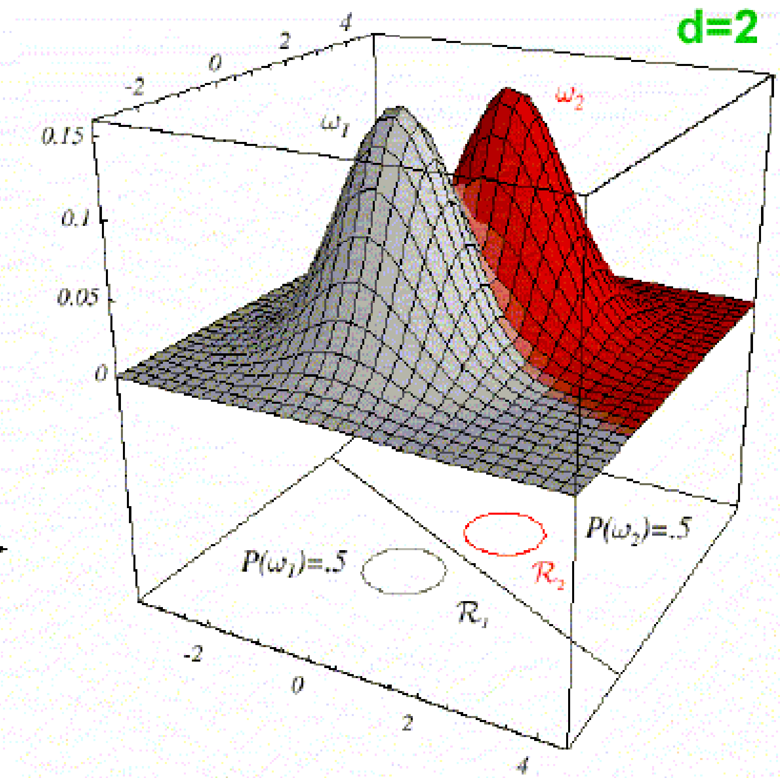
$$g_i(\mathbf{x}) = -\frac{1}{2}\left[(\mathbf{x} - \mu_i)^{\mathrm{T}} \Sigma_i^{-1} (\mathbf{x} - \mu_i) \right] - \frac{d}{2} \ln(2\pi) - \frac{1}{2} \ln|\Sigma_i| + \ln P(\omega_i)$$

- Υπάρχουν 3 περιπτώσεις ανάλογα με τον πίνακα συνδιασποράς $\mathbf{\Sigma}$

Περίπτωση 1: $\Sigma_i = \sigma^2 I$

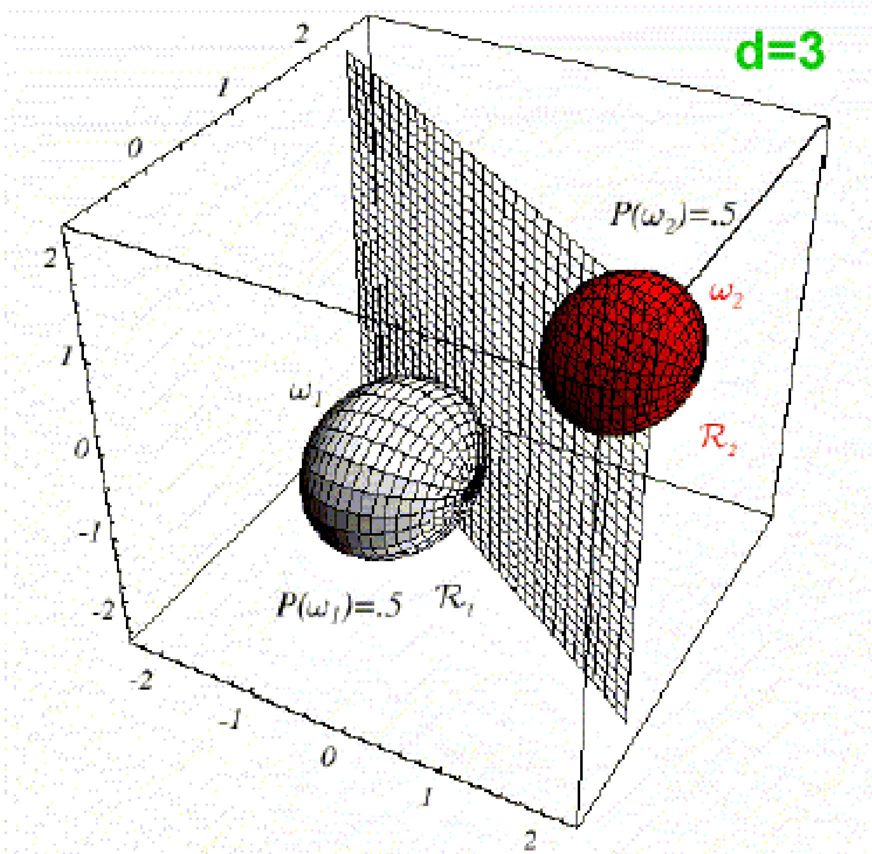
- Τα χαρακτηριστικά είναι <u>στατιστικώς ανεξάρτητα</u>, και έχουν όλα την <u>ίδια διασπορά</u>: Τα δείγματα βρίσκονται σε υπερ-σφαίρες ίσου μεγέθους, και οι επιφάνειες απόφασης είναι **υπερεπίπεδα** διάστασης *d-1*.





Περίπτωση 1: $\Sigma_i = \sigma^2 I$

- Όταν d=3, τα δείγματα βρίσκονται σε σφαίρες ίσου μεγέθους, και οι επιφάνειες απόφασης είναι **επίπεδα**.

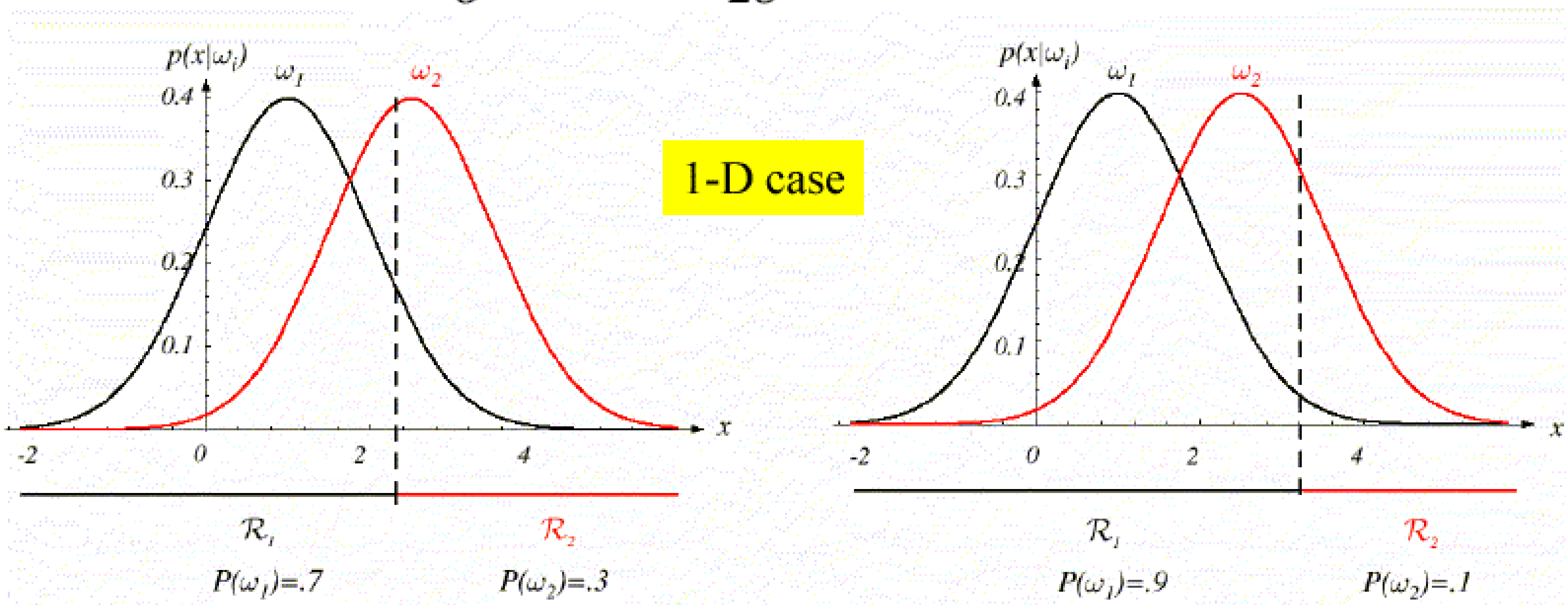


Περίπτωση 1: $\Sigma_i = \sigma^2 I$

- Αυτή η περίπτωση δημιουργεί γραμμικές συναρτήσεις διάκρισης:

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

$$\mathbf{w}_i = \frac{1}{\sigma^2}\boldsymbol{\mu}_i, \quad w_{i0} = \frac{-1}{2\sigma^2}\boldsymbol{\mu}_i^T \cdot \boldsymbol{\mu}_i + \ln P(\omega_i)$$



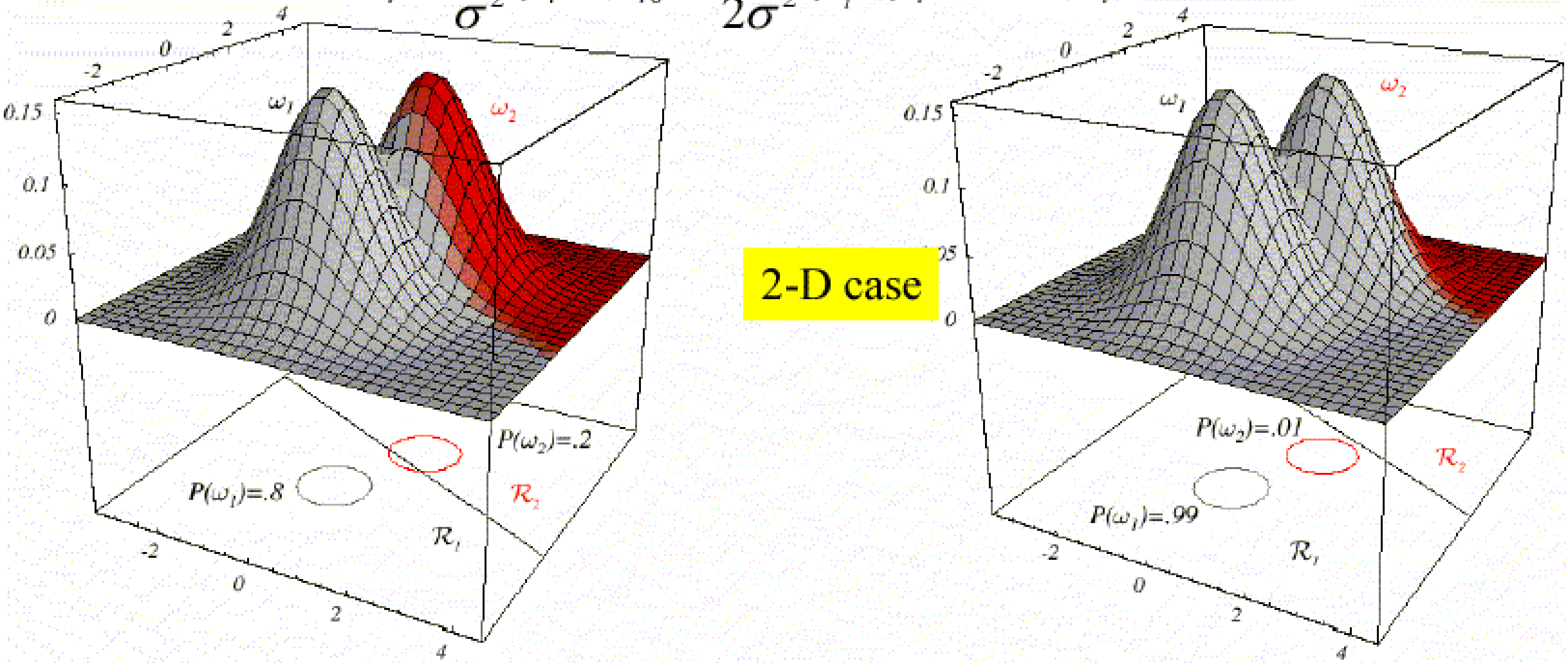
Προσέξτε πως οι εκ των προτέρων πιθανότητες απομακρύνουν το σημείο κατωφλίου από τον πιο πιθανό μέσο.

Περίπτωση 1: $\Sigma_i = \sigma^2 I$

- Αυτή η περίπτωση δημιουργεί γραμμικές συναρτήσεις διάκρισης:

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

$$\mathbf{w}_i = \frac{1}{\sigma^2}\boldsymbol{\mu}_i, \quad w_{i0} = \frac{-1}{2\sigma^2}\boldsymbol{\mu}_i^T \cdot \boldsymbol{\mu}_i + \ln P(\omega_i)$$



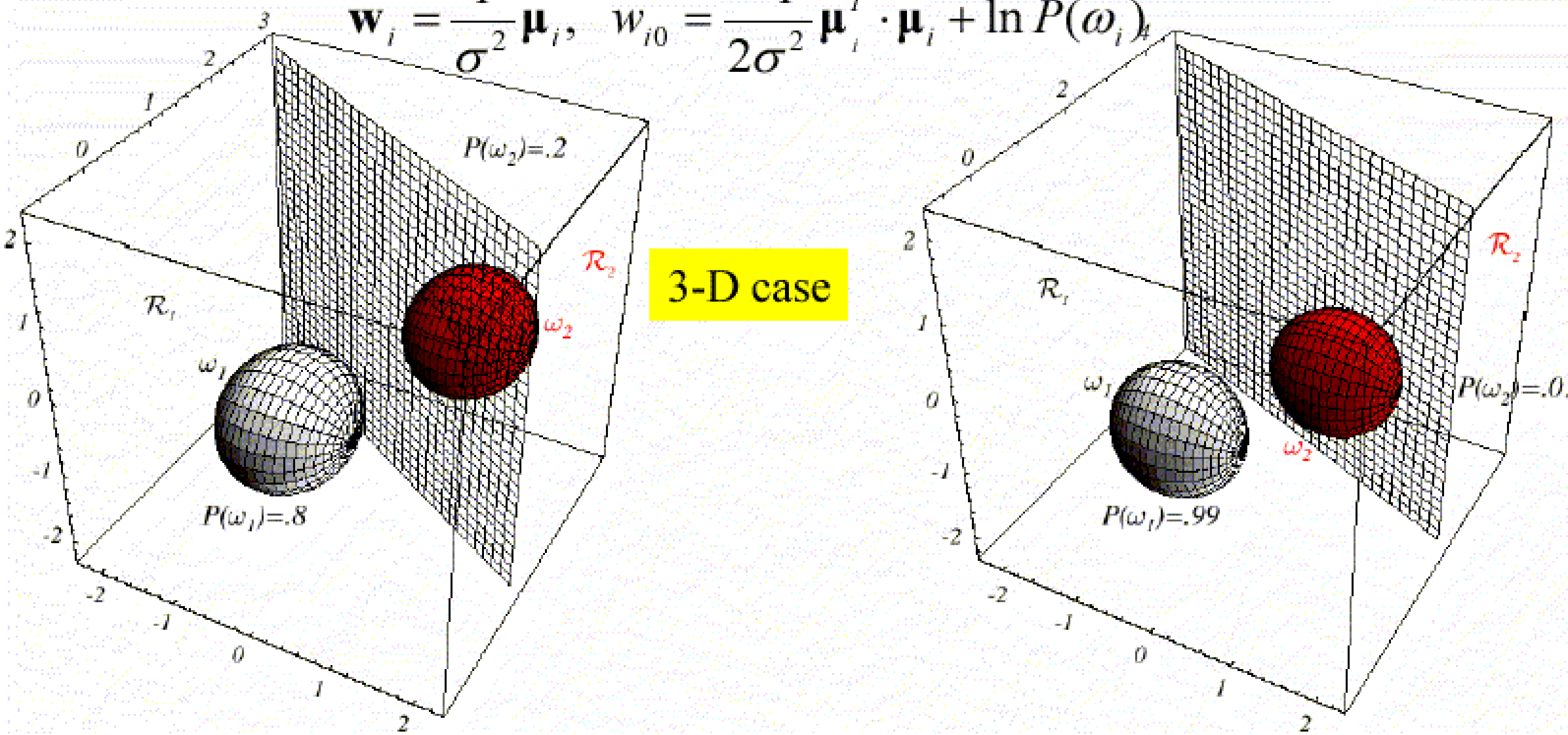
Προσέξτε πως οι εκ των προτέρων πιθανότητες απομακρύνουν την ευθεία απόφασης από τον πιο πιθανό μέσο.

Περίπτωση 1: $\Sigma_i = \sigma^2 I$

- Αυτή η περίπτωση δημιουργεί γραμμικές συναρτήσεις διάκρισης:

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

$$\mathbf{w}_i = \frac{1}{\sigma^2}\boldsymbol{\mu}_i, \quad w_{i0} = \frac{-1}{2\sigma^2}\boldsymbol{\mu}_i^T \cdot \boldsymbol{\mu}_i + \ln P(\omega_i)$$

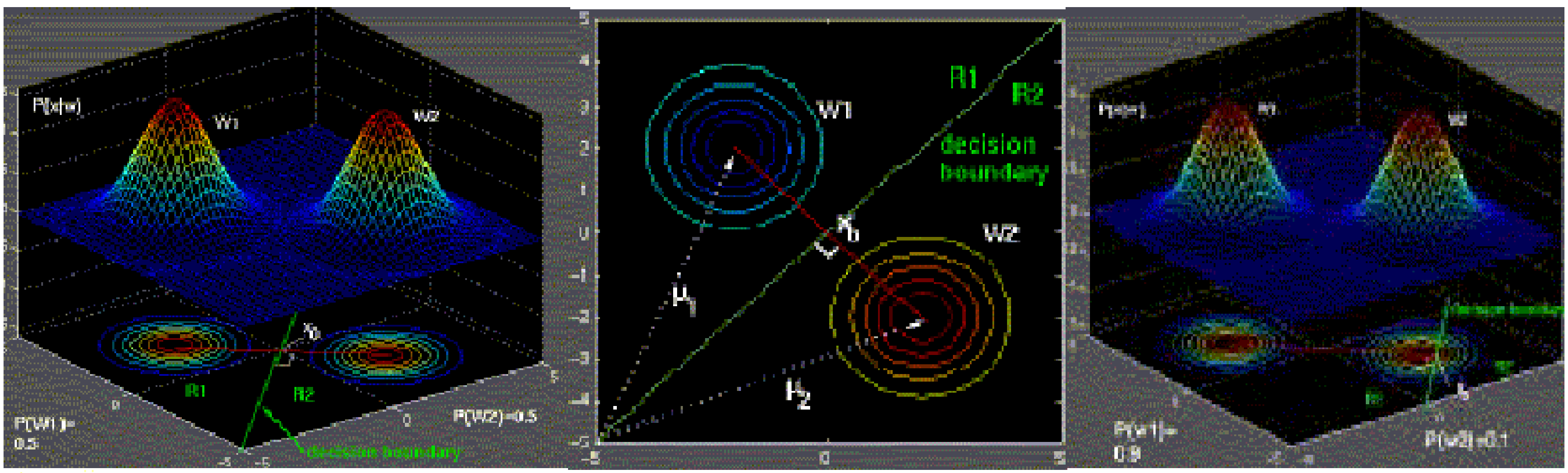


Προσέξτε πως οι εκ των προτέρων πιθανότητες απομακρύνουν το επίπεδο απόφασης από τον πιο πιθανό μέσο.

Περίπτωση 1: $\Sigma_i = \sigma^2 I$

Οι επιφάνειες απόφασης είναι υπερ-επίπεδα που ορίζονται από τις γραμμικές εξισώσεις $g_i(\boldsymbol{x}) = g_j(\boldsymbol{x})$, που γράφονται ως $\mathbf{w}^T\left(\mathbf{x}-\mathbf{x}_0\right)=0$ όπου:

$$\mathbf{w} = \boldsymbol{\mu}_i - \boldsymbol{\mu}_j, \quad \mathbf{x}_0 = \frac{1}{2}\left(\boldsymbol{\mu}_i + \boldsymbol{\mu}_j\right) - \frac{\sigma^2}{\left\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\right\|^2} \ln\left(\frac{P(\omega_i)}{P(\omega_{ji})}\right)\left(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\right)$$

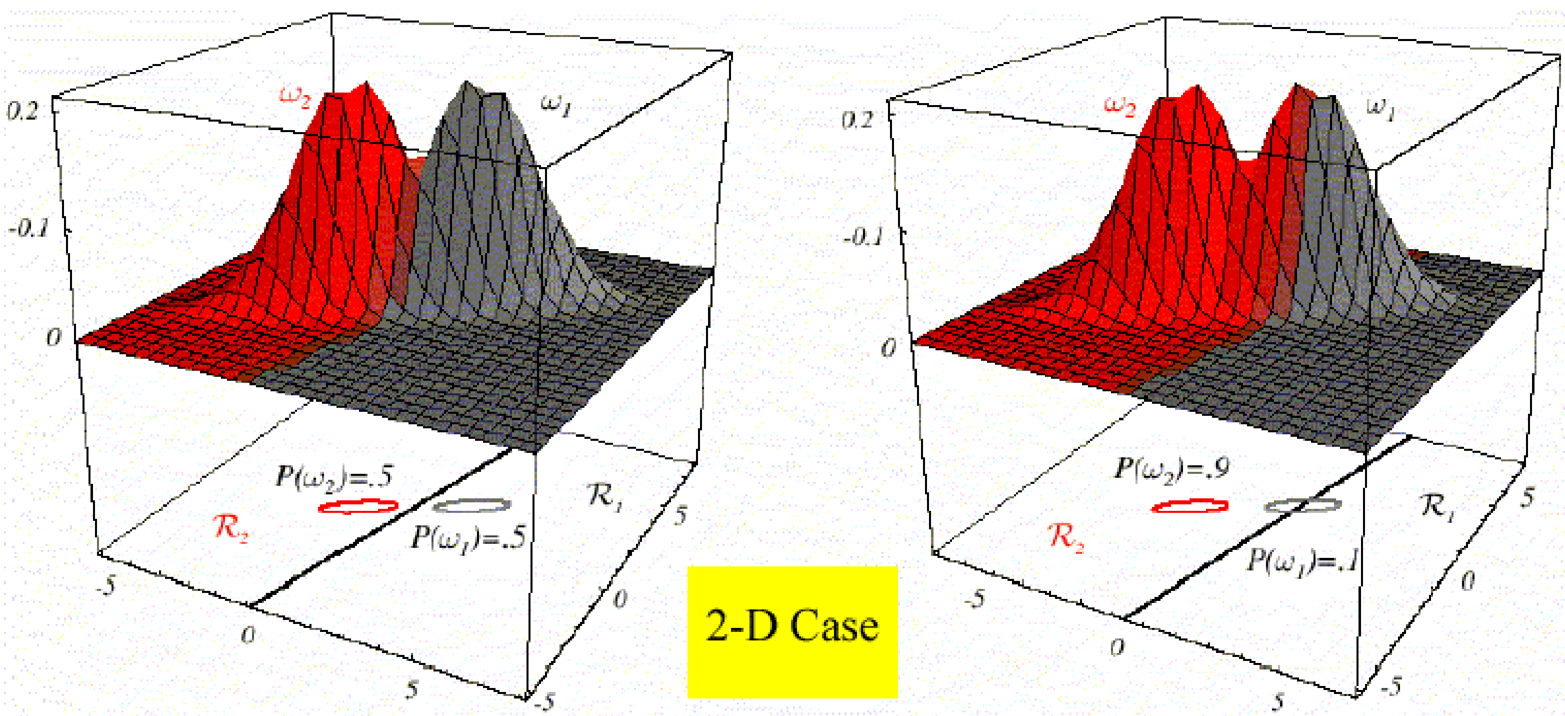


Επιφάνεια απόφασης: Υπερ-επίπεδο που περνά από το σημείο $\boldsymbol{x}_0$ και είναι κάθετο στο διάνυσμα $\boldsymbol{w}$ που ενώνει τις μέσες τιμές $\boldsymbol{\mu}_i$ και $\boldsymbol{\mu}_j$.

Περίπτωση 2: $\Sigma_i = \Sigma$

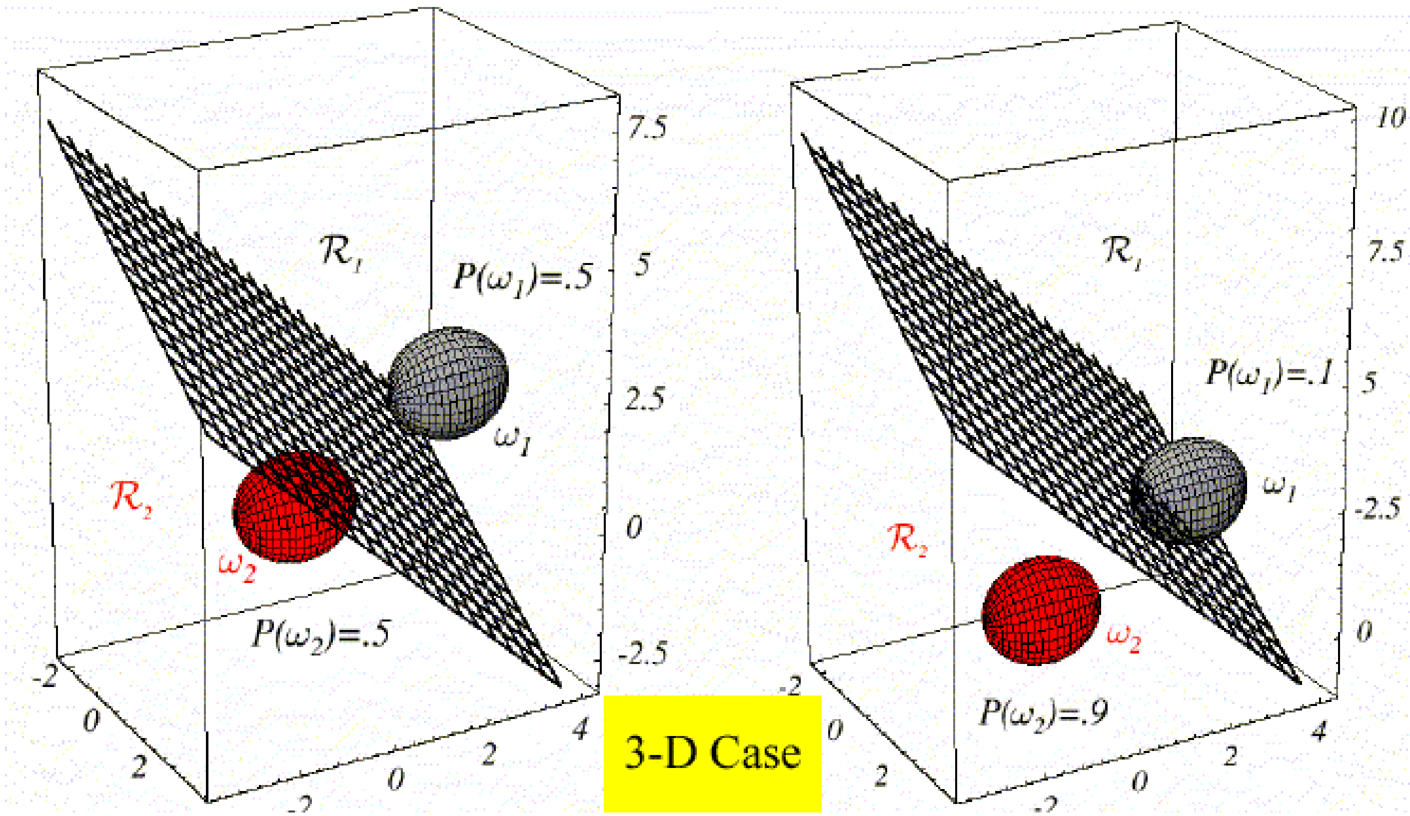
- Οι πίνακες συνδιασποράς είναι αυθαίρετοι, αλλά ίδιοι για όλες τις κλάσεις. Τα χαρακτηριστικά δημιουργούν υπερελλειψοειδείς ομάδες ίδιου μεγέθους και σχήματος με κέντρα τα $\boldsymbol{\mu}_i$.
- Γραμμικές συναρτήσεις απόφασης → Υπερεπίπεδα ως επιφάνειες απόφασης

$$g_i(\mathbf{x}) = \mathbf{w}_i^T \mathbf{x} + w_{i0} \qquad \mathbf{w}_i = \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_i, \quad w_{i0} = -\frac{1}{2}\boldsymbol{\mu}_i^T \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_i + \ln P(\omega_i)$$



Περίπτωση 2: $\Sigma_i = \Sigma$

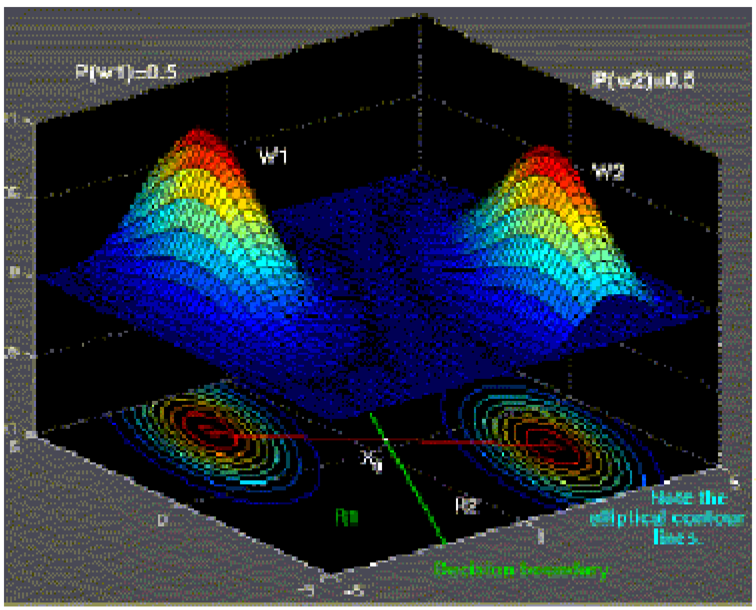
- Οι πίνακες συνδιασποράς είναι αυθαίρετοι, αλλά ίδιοι για όλες τις κλάσεις.
Τα χαρακτηριστικά δημιουργούν υπερελλειψοειδείς ομάδες ίδιου μεγέθους και σχήματος με κέντρα τα $\boldsymbol{\mu}_i$.
- Γραμμικές συναρτήσεις απόφασης → Υπερεπίπεδα ως επιφάνειες απόφασης

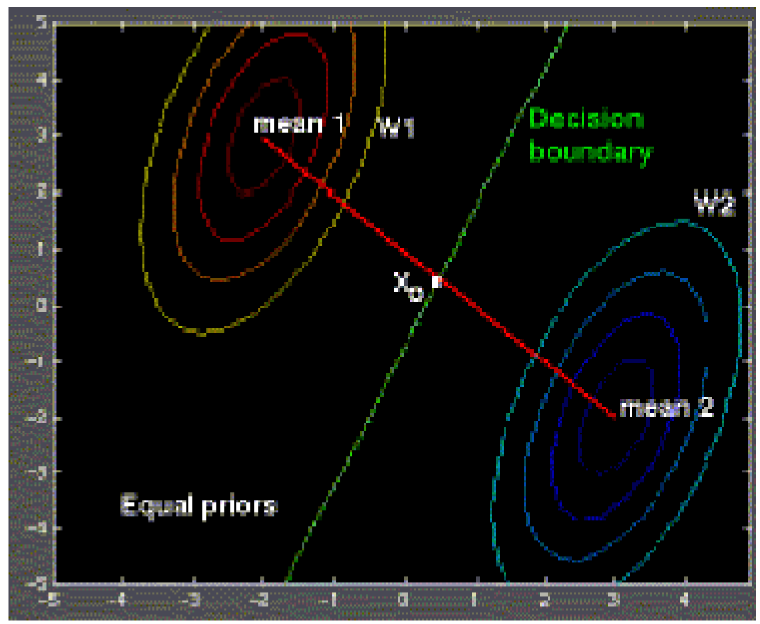


Περίπτωση 2: $\Sigma_i = \Sigma$

- Επιφάνειες απόφασης $\mathbf{w}^T(\mathbf{x}-\mathbf{x}_0)=0$ όπου:

$$\mathbf{w}=\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j),\quad \mathbf{x}_0=\frac{1}{2}(\boldsymbol{\mu}_i+\boldsymbol{\mu}_j)-\frac{\ln\left(P(\omega_i)/P(\omega_j)\right)}{(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j)}(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j)$$





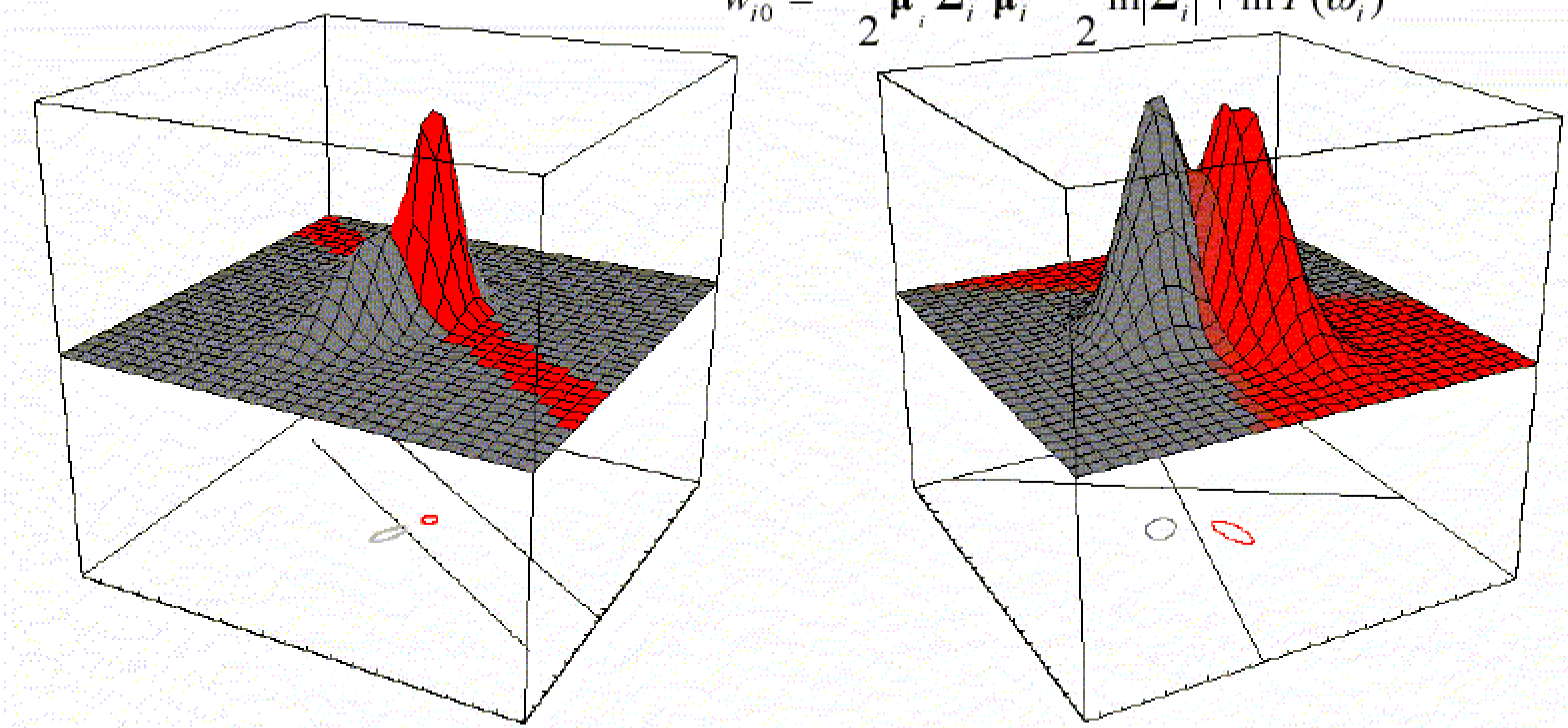
Εφόσον $\mathbf{w}=\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j)$ το υπερεπίπεδο απόφασης <u>***δεν***</u> είναι κάθετο στο διάνυσμα $\boldsymbol{w}$ που ενώνει τις μέσες τιμές $\boldsymbol{\mu}_i$ και $\boldsymbol{\mu}_j$.

Περίπτωση 3: $\Sigma_i = οποιοδήποτε$

- Μη γραμμικές αλλά τετραγωνικές συναρτήσεις απόφασης.
- Επιφάνιες απόφασης *hyperquadratics (υπερελλειψοειδή, υπερπαραβολοειδή κτλ).*

$$g_i(\mathbf{x}) = \mathbf{x}^T \mathbf{W}_i \mathbf{x} + \mathbf{w}_i^T \mathbf{x} + w_{i0} \qquad \mathbf{W}_i = -\frac{1}{2}\mathbf{\Sigma}_i^{-1}, \quad \mathbf{w}_i = \mathbf{\Sigma}_i^{-1}\boldsymbol{\mu}_i$$

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

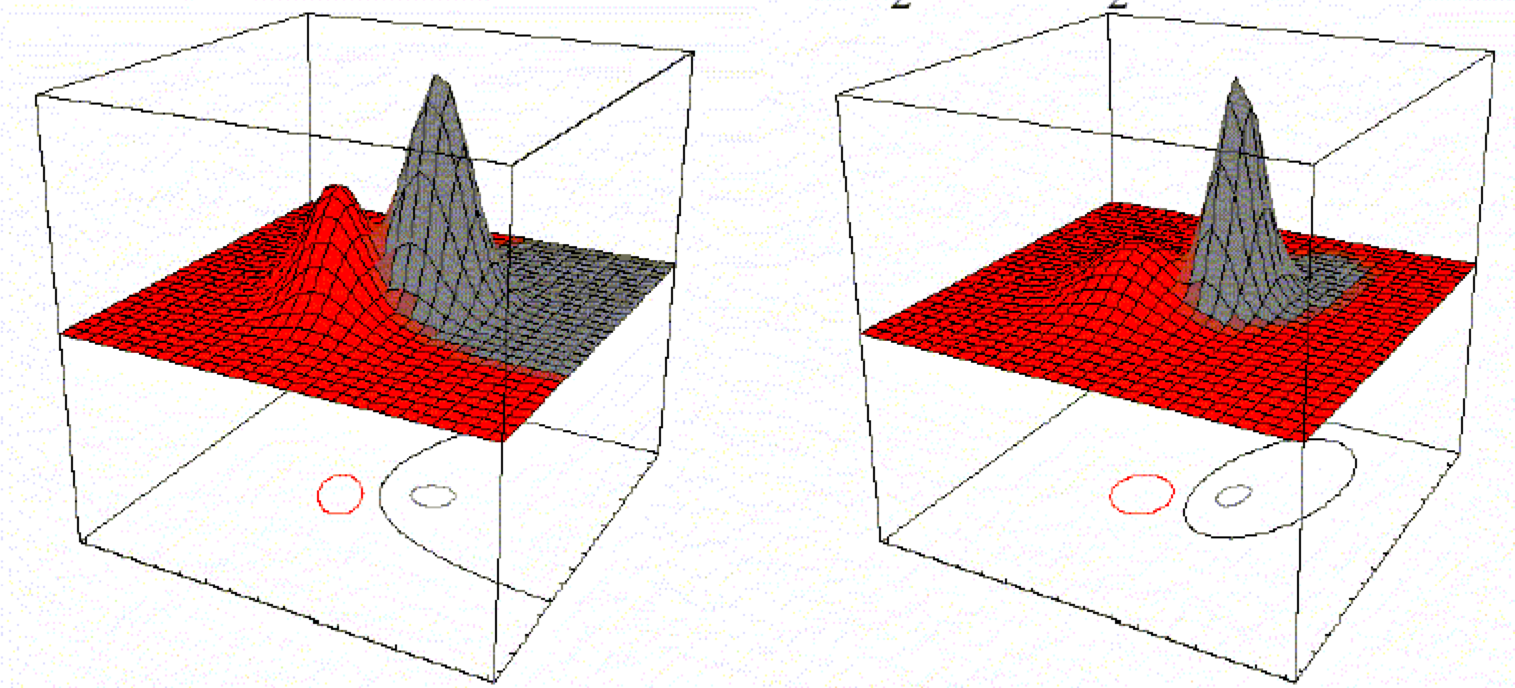


Περίπτωση 3: $\Sigma_i = \text{οποιοδήποτε}$

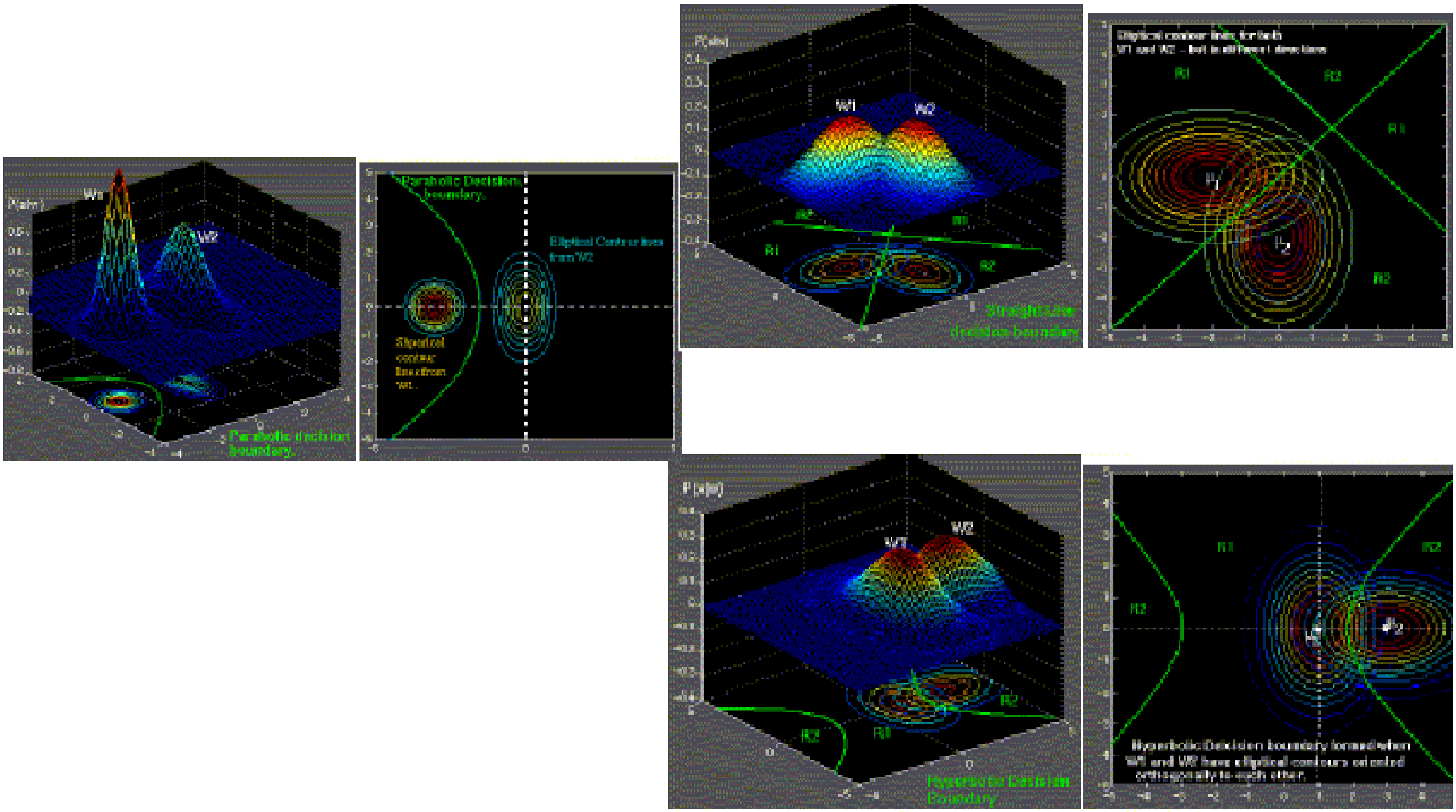
- Μη γραμμικές αλλά τετραγωνικές συναρτήσεις απόφασης.
- Επιφάνειες απόφασης *hyperquadratics (υπερελλειψοειδή, υπερπαραβολοειδή κτλ).*

$$g_i(\mathbf{x}) = \mathbf{x}^T \mathbf{W}_i \mathbf{x} + \mathbf{w}_i^T \mathbf{x} + w_{i0} \qquad \mathbf{W}_i = -\frac{1}{2}\boldsymbol{\Sigma}_i^{-1}, \quad \mathbf{w}_i = \boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i$$

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

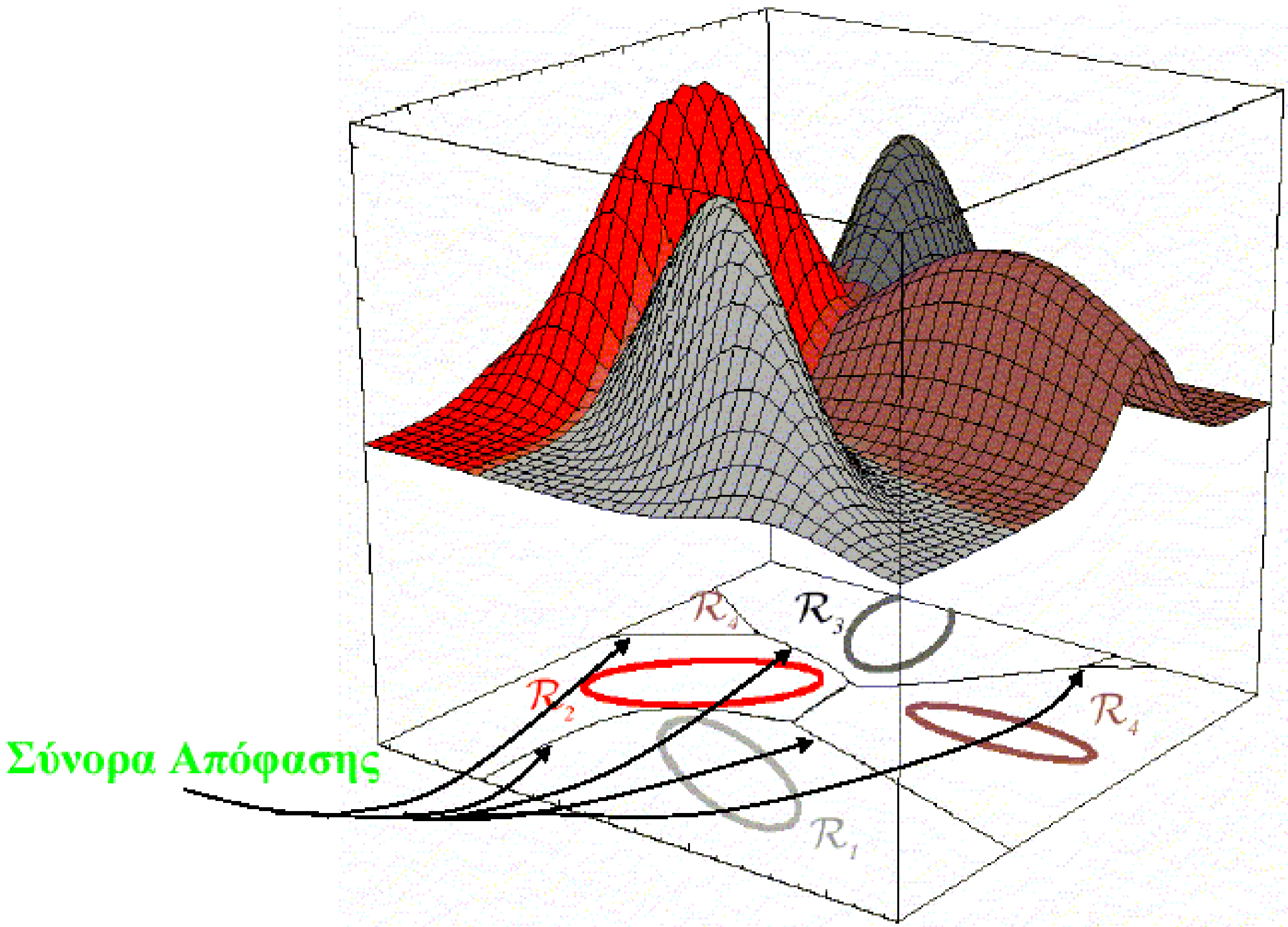


Περίπτωση 3: $\Sigma_i = \text{οποιοδήποτε}$



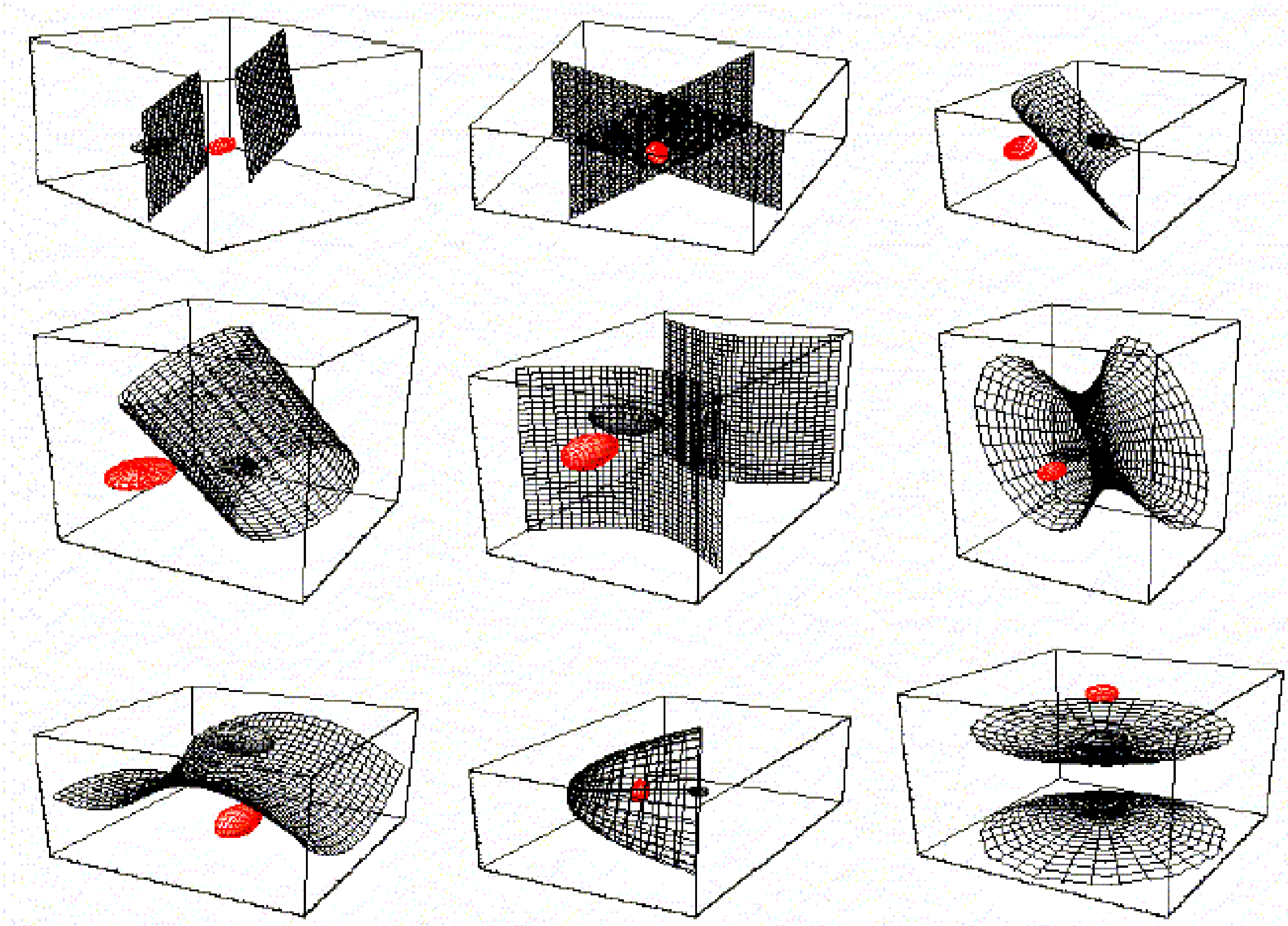
Περίπτωση 3: $\Sigma_i = \text{οποιοδήποτε}$

- Στην περίπτωση των πολλών τάξεων, τα όρια είναι ακόμα περισσότερο περίπλοκα:



Περίπτωση 3: $\Sigma_i = οποιοδήποτε$

3-D



Παράδειγμα: Πρόβλημα Εκτίμησης Σήματος σε Θόρυβο

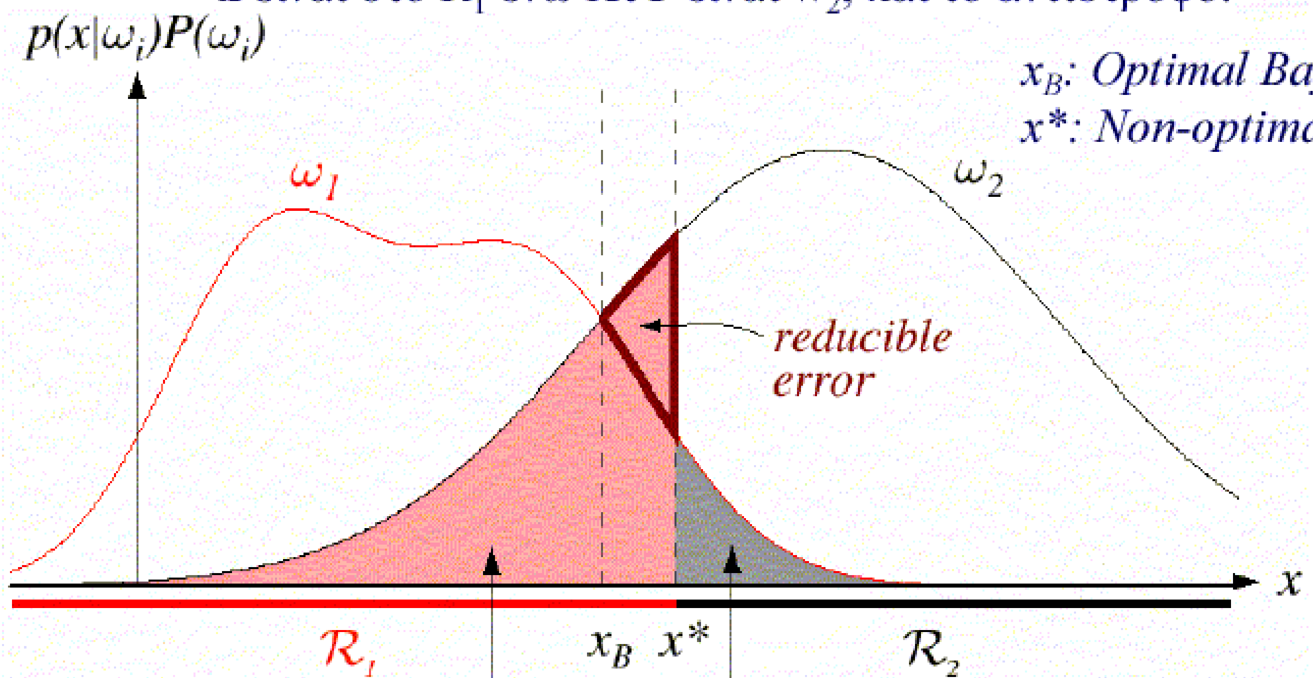
- Λαμβάνουμε το σήμα (μέτρηση)

$$x(t) = \begin{cases} n(t) \\ s(t) + n(t) \end{cases}$$

- n(t) είναι ο θόρυβος και s(t) είναι το σήμα

Πιθανότητες Σφάλματος

- Στην περίπτωση δυο κλάσεων, υπάρχουν δυο περιπτώσεις λάθους:
  x είναι στο R_1 ενώ ΚτΦ είναι w_2, και το αντίστροφο.



$$\mathbf{P(error)} = \int_{\mathcal{R}_1} p(x|\omega_2)P(\omega_2)\,dx + \int_{\mathcal{R}_2} p(x|\omega_1)P(\omega_1)\,dx$$

$$P(\mathbf{x}\in R_1,\omega_2) = P(\mathbf{x}\in R_1 \mid \omega_2)\cdot P(\omega_2)$$

$$P(\mathbf{x}\in R_2,\omega_1) = P(\mathbf{x}\in R_2 \mid \omega_1)\cdot P(\omega_1)$$

ΈΝΑ Πρόβλημα Εκτίμησης

- Λαμβάνουμε το σήμα (μέτρηση)

$$x(t) = \begin{cases} n(t) \\ s(t) + n(t) \end{cases}$$

- n(t) είναι ο θόρυβος και s(t) είναι το σήμα
- ω_0 Υπόθεση 1 H0 : Δεν υπάρχει το σήμα και x(t) είναι μόνο θόρυβος
- ω_1 Υπόθεση 2 H1 : υπάρχει το σήμα και x(t)=s(t)+n(t)

Τι ξέρουμε;

- The signal s(t) is known and we want to know if it is present or not.
- n(t) is a stochastic process with known statistics.
- The observer must base its decision on
  - Samples of x(t) (digital x(nT))
  - The entire x(t) (analog)
- We shall address the digital problem:
- We assume that
  - (a) s(t) and n(t) are independent
  - (b) The samples of n(t) are uncorrelated

Στόχος;

- Design the optimum system:

The system can be linear or nonlinear

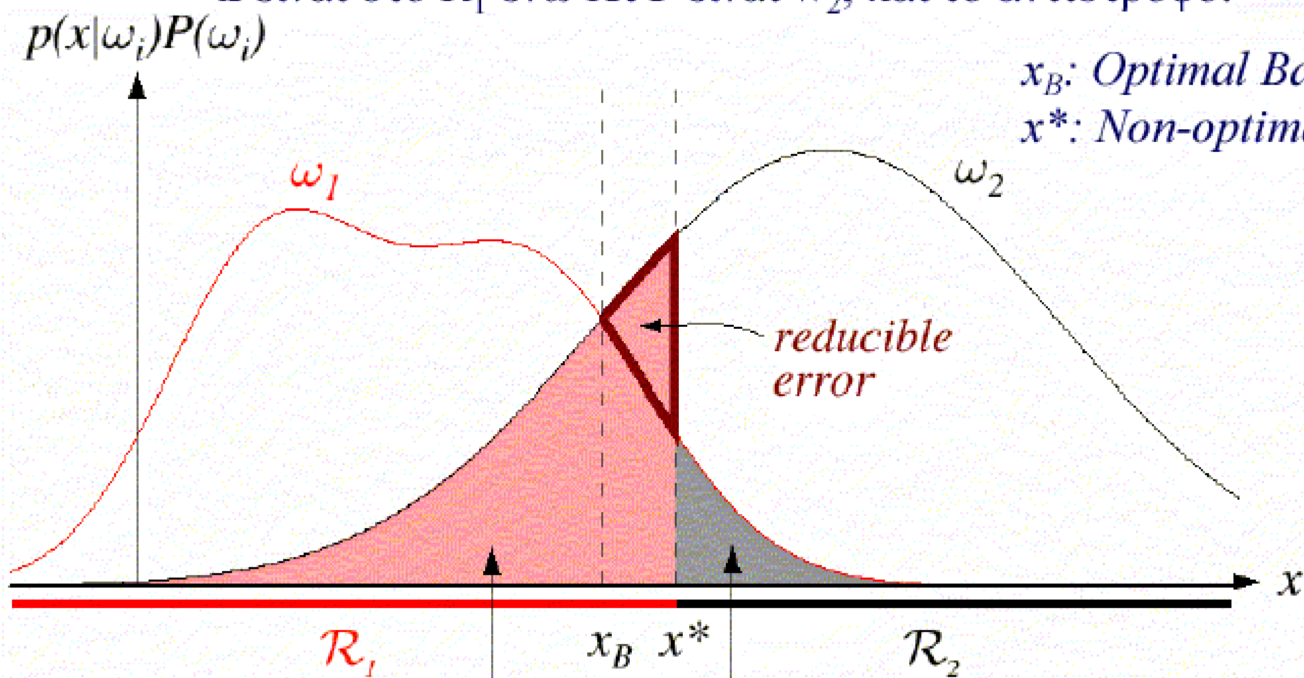
System means → Hardware device

System means → Software program

- <u>Criteria:</u>
  1. The system must be optimum with respect to a criterion which depends on the particular problem.
  2. Sometimes we are satisfied with a sub optimum system because it is much simpler than the optimum.

Πιθανότητες Σφάλματος

- Στην περίπτωση δυο κλάσεων, υπάρχουν δυο περιπτώσεις λάθους: **x** είναι στο R_1 ενώ ΚτΦ είναι w_2, και το αντίστροφο.



$$\textbf{P(error)} = \int_{\mathcal{R}_1} p(x|\omega_2)P(\omega_2)\,dx + \int_{\mathcal{R}_2} p(x|\omega_1)P(\omega_1)\,dx$$

$P(\mathbf{x} \in R_1, \omega_2) = P(\mathbf{x} \in R_1 \mid \omega_2) \cdot P(\omega_2)$

$P(\mathbf{x} \in R_2, \omega_1) = P(\mathbf{x} \in R_2 \mid \omega_1) \cdot P(\omega_1)$

ΕΙΔΗ ΛΑΘΩΝ

- Due to the presence of noise we can perform errors in our decisions
- (1). ***Error of the first kind***: P_{e0}. It is the probability to decide that the signal is present when it is not or choose H_1 when H_0 is true.
- (2). ***Error of the second kind***: P_{e1}. When the signal is present you decide that it is not or choose H_0 when H_1 is true.
- $P_e = P(\text{error}) = (1-p_1)P_{e0} + p_1 P_{e1}$

Περιπτώσεις

- s(t) είναι γνωστό
  A. s(t)=A, λαμβάνουμε μία μέτρηση
     1. *Ελάχιστη πιθανότητα λάθους*
     2. *Ελάχιστο κόστος*
     3. *Minimax (P_1 είναι άγνωστη)*
     4. *Newman-Pearson (Το συνολικό κόστος για λάθος εκτίμηση μία κατάστασης είναι δεδομένο)*
  B. Λαμβάνουμε πολλές μετρήσεις
  C. s(t) είναι γνωστή κυματομορφή
- s(t) είναι στοχαστική διεργασία (random process)
- s(t) είναι γνωστό, αλλά P_1 και θόρυβος είναι άγνωστα(matched filter)

ΠΡΟΒΛΗΜΑ 1 (Τηλεπικοινωνίες)

- Detection of a pulse (PCM or radar e.t.c)

Case A

Assumption:

1. $s(t) = \{A \text{ or } 0\}$
2. Probability of $s(t) = A = P_1$
3. Probability of $s(t) = 0 = 1-P_1$
4. Probability density function of $n(t)$ is known
5. We must base our decision on one sample

Criterion: Minimize the probability of error:

$$P_e = P_1 P_{e1} + (1-P_1) P_{e0}$$

Λύση

- Decision will be based on $x[n]=x_1$. If x_1 is in one Range R_1 then we decide that H_1 is true and our problem is to find R_1.

$$\mathrm{P_{e0}} = \{\mathrm{x_1 \text{ is in } R_1 \text{ but } s(t) \text{ is not present}}\} = \int_{R_1} f_{x_1}(x_1 / s(t) = 0)dx$$

$$P_{e1} = \int_{1-R_1} f_{x_1}(x_1 / s(t) = A)dx = 1 - \int_{R_1} f_{x_1}(x_1 / s(t) = A)dx$$

$$P_e = (1-p_1)\int_{R_1} f_{x_1}(x_1 / s(t) = 0)dx + p_1\left(1 - \int_{R_1} f_{x_1}(x_1 / s(t) = A)dx\right) =$$

$$= p_1 + \int_{R_1} [(1-p_1) f_{x_1}(x/0) - p_1 f_{x_1}(x/A)]dx$$

Λύση (συνέχεια)

- Διάλεξε το R_1 έτσι ώστε η πιθανότητα λάθους να είναι ελάχιστη

ΑΡΑ ΠΡΕΠΕΙ

$$[(1-P1)\,f_{x_1}(x/0)] - P_1\,f_{x_1}(x/A)] < 0$$

$$\boxed{l(x_1) = \frac{f_{x_1}(x/A)}{f_{x_1}(x/0)} > \frac{1-P_1}{P_1}}$$

- **Το *l(x1)* πιθανοφάνεια *(likelihood)***
- Μερικές φορές είναι καλύτερα να παίρνουμε τον λογάριθμο

$$\ln l(x_1) > \ln\frac{1-P_1}{P_1}$$

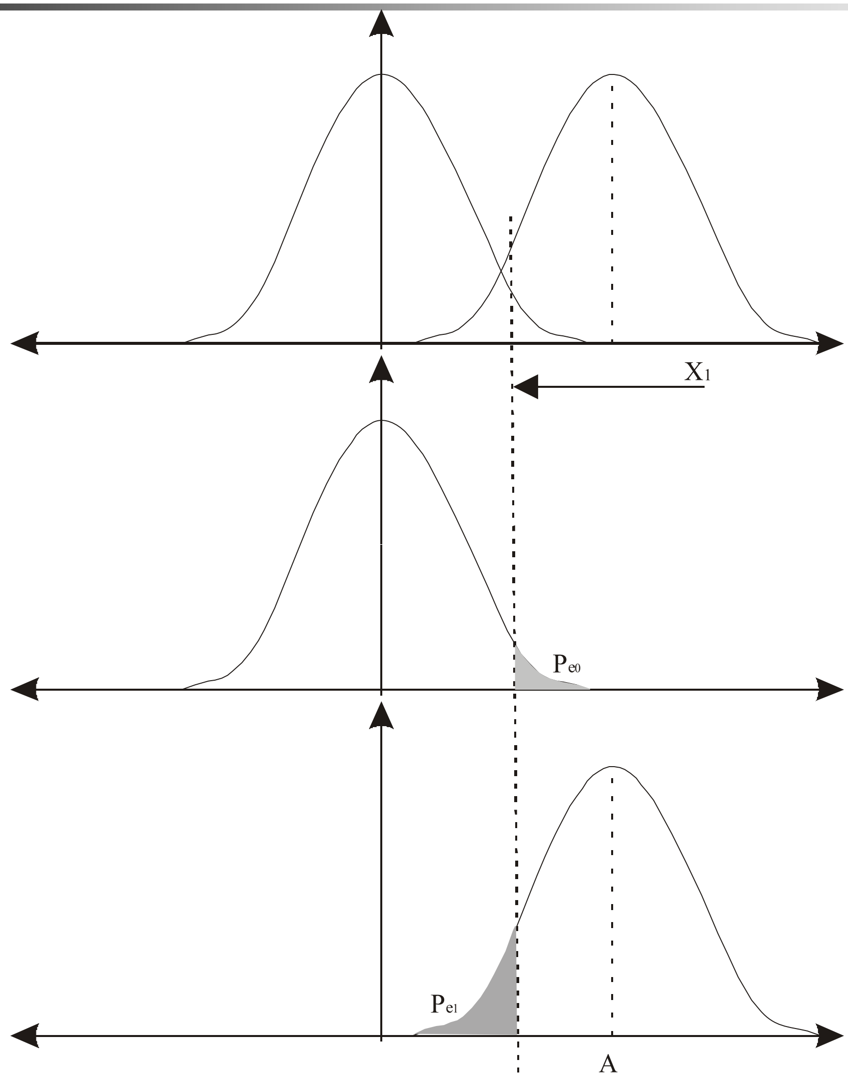
Παράδειγμα 1

- Ο θόρυβος έχει κατανομή Gaussian με μέση τιμή μηδέν και διασπορά σ².

$$\frac{\frac{1}{\sqrt{2\pi}\sigma}e^{\frac{-(x_1-A)^2}{2\sigma_x{}^2}}}{\frac{1}{\sqrt{2\pi}\sigma}e^{\frac{-x_1{}^2}{2\sigma_x{}^2}}} \geq \frac{1-P_1}{P_1} \Rightarrow \left(\frac{-(x_1-A)^2+x_1{}^2}{2\sigma_x{}^2}\right) \geq \ln\frac{1-P_1}{P_1} \Rightarrow \frac{x_1 A}{\sigma_x{}^2} - \frac{A^2}{2\sigma_x{}^2} \geq \ln\frac{1-P_1}{P_1}$$

$$\boxed{x_1 > \frac{A}{2} + \frac{\sigma_x{}^2}{A}\ln\frac{1-P_1}{P_1} = D}$$

$$P_e = \frac{1}{2} - \frac{1}{2}\left(p_1 erf\,\frac{A-D}{\sqrt{2}\sigma_x} + (1-p_1)erf\,\frac{D}{\sqrt{2}\sigma_x}\right)$$


X1
Pe0
Pe1
A

Αριθμητική εφαρμογή

- Computer transmits 0,1,… Each pulse occupies T sec and we sample every T sec with

$P_1=P_2=1/2$, $\sigma^2=2$, choose A such that $P_e=10^{-5}$

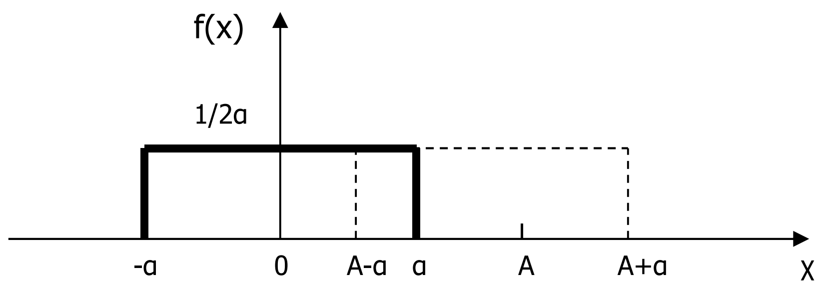
$$D=\frac{A}{2}\Rightarrow Pe=\frac{1}{2}-\frac{1}{4}\cdot\left[2erf\,\frac{A}{2\cdot 2}\right]=\frac{1}{2}-\frac{1}{2}\cdot\left[erf\,\frac{A}{4}\right]$$

$$\Rightarrow erf\,\frac{A}{4}=2\cdot\left[\frac{1}{2}-Pe\right]=0.99998\Rightarrow\frac{A}{4}=3\Rightarrow A=12$$

Let T=1μs on the average we are going to perform 10 errors every second!

Παράδειγμα 2

- Τι γίνεται εάν ο θόρυβος έχει ομοιόμορφη κατανομή στο [-α,α]



Εάν $A-\alpha \geq \alpha$ τότε $P_e=0$ και $R_1=x \geq A-\alpha$

Εάν $A-\alpha<\alpha$ $(A<2\alpha)$ τότε ?

Α.2 Ελάχιστο Κόστος

- Σε αυτή την περίπτωση κάνουμε τις ίδιες υποθέσεις με την περίπτωση Α.1, αλλά το κριτήριο θα είναι η ελαχιστοποίηση του συνολικού κόστους και όχι της συνολικής πιθανότητας λάθους.
- ***Με λ_{ij} = κόστος εάν επιλέξουμε H_i ενώ είναι το H_j σωστό έχουμε***
- ***Bayes Criterion : Ελαχιστοποίηση συνολικού κόστους C, με***

$$C = P_1\left[\lambda_{11}\int_{R_1} f_{x_1}(x_1/H_1)dx + \lambda_{01}\int_{R_0} f_{x_1}(x_1/H_1)dx\right] + (1-P_1)\left[\lambda_{00}\int_{R_0} f_{x_1}(x_1/H_0)dx + \lambda_{10}\int_{R_1} f_{x_1}(x_1/H_0)dx\right]$$

και όπως στην περίπτωση Α.1. έχουμε τελικά, ότι επιλέγουμε το H_1 εάν

$$\boxed{l(x_1) = \frac{f_{x_1}(x/A)}{f_{x_1}(x/0)} > \frac{1-P_1}{P_1}\frac{\lambda_{12}-\lambda_{22}}{\lambda_{21}-\lambda_{11}}}$$

A.3 *Minimax (P_1 είναι άγνωστη)*

- Σε αυτή την περίπτωση σχεδιάζουμε το σύστημα ώστε να ελαχιστοποιεί την χειρότερη περίπτωση κόστους

$$P_{error}(P_1)=P_1\left[\lambda_{11}\int_{R_1} f_{x_1}(x_1/H_1)dx+\lambda_{01}\int_{R_0} f_{x_1}(x_1/H_1)dx\right]+(1-P_1)\left[\lambda_{00}\int_{R_0} f_{x_1}(x_1/H_0)dx+\lambda_{10}\int_{R_1} f_{x_1}(x_1/H_0)dx\right]=$$

$$=\lambda_{00}+(\lambda_{10}-\lambda_{00})\int_{R_1} f_{x_1}(x_1/H_0)dx+P_1\left((\lambda_{01}-\lambda_{00})+(\lambda_{11}-\lambda_{01})\int_{R_1} f_{x_1}(x_1/H_1)dx-(\lambda_{10}-\lambda_{00})\int_{R_1} f_{x_1}(x_1/H_0)dx\right)=$$

$$=AP_1+B$$

Άρα θέλουμε A=0 και $P_{error,min}$ =B

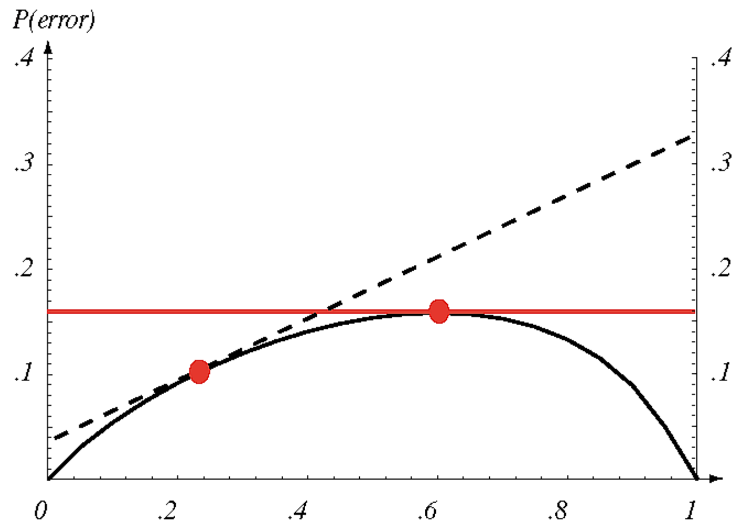


FIGURE 2.4. The curve at the bottom shows the minimum (Bayes) error as a function of prior probability $P(\omega_1)$ in a two-category classification problem of fixed distributions. For each value of the priors (e.g., $P(\omega_1) = 0.25$) there is a corresponding optimal decision boundary and associated 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. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

A.4 *Newman-Pearson Criterion*

- Σε αυτή την περίπτωση σχεδιάζουμε το σύστημα ώστε *to συνολικό κόστος για λάθος εκτίμηση μία κατάστασης είναι δεδομένο*

  Με false alarm $P_{e0} = k$ (δεδομένο), ελαχιστοποίησε $P_{e1} = min$

  Το παραπάνω πρόβλημα ελαχιστοποίησης μπορεί να λυθεί μέσω του δυαδικού του προβλήματος Lagrange. Σε πρώτο στάδιο γράφεται η Lagrangian συνάρτηση που του αντιστοιχεί

$$L(R_1,\lambda) = P_{e1} + \lambda P_{e0} = 1 - \int_{R_1} \big(f(x\,|\,H_1) - \lambda f(x\,|\,H_0) \big) dx$$

  όπου λ ο πολλαπλασιαστής Lagrange που αντιστοιχεί στον περιορισμό ισότητας $P_{e0} = k$.

  Συνεπώς $R_1 = \left\{ x \,\middle|\, f(x|H_1) - \lambda f(x\,|\,H_0) > 0 \right\}$ και το λ το υπολογίζουμε από την σχέση

$$\int_{R_1(\lambda)} f(x\,|\,H_0) dx = k$$

Αριθμητικό Παράδειγμα

Ο θόρυβος έχει κατανομή Gaussian με μέση τιμή μηδέν και διασπορά σ^2_n και το σήμα παίρνει τις τιμές 0 και A

Άρα $x > \frac{A}{2} + \frac{\sigma_n^2}{A}\ln(\lambda) = D$ και $R_1 = \{x \mid x > D\}$

Όμως $P_{e0} = \int_D^{\infty} f(x \mid H_0)dx = k \Rightarrow \frac{1}{2}\left[1 - erf\,\frac{D}{\sigma_n\sqrt{2}}\right] = k \Rightarrow erf\,\frac{D}{\sigma_n\sqrt{2}} = 1 - 2k$

Εάν k=10^{-5} τότε προκύπτει ότι $D \cong 3\sigma_n\sqrt{2}$

Β. Λαμβάνουμε πολλές μετρήσεις και s(t) είναι γνωστή κυματομορφή

Εάν s(t) = {s(t) or 0} και λαμβάνουμε πάνω από Ν δείγματα για κάθε σύμβολο, τότε

$\mathbf{x}=\{x_1, x_2,\ldots, x_N\}$ και

$\boldsymbol{\mu}_\mathbf{1}=\{s_1=s(t_1), s_2=s(t_2),\ldots, s_N=s(t_N)\}$ και $\boldsymbol{\mu}_\mathbf{0}=0$

μπορούμε να χρησιμοποιήσουμε τις συναρτήσεις διάκρισης για Γκαουσιανές κατανομές ΠΠ με $\boldsymbol{\Sigma=\sigma_n I}$ περίπτωση 1, δηλαδή

$$g_i(\mathbf{x}) = -\frac{1}{2\sigma_n^2}\left\|\mathbf{x}-\boldsymbol{\mu}_\mathbf{i}\right\|^2 + \ln(P_i)$$

και επιλέγουμε το ω_1 εάν $g_1(\mathbf{x})\geq g_0(\mathbf{x})$, μετά από πράξεις

$$\boxed{\sum_{i=1}^{N} x_i s_i \geq \frac{\sum_{i=1}^{N} s_i^2}{2} + \sigma_n^2 \ln\left(\frac{1-P_1}{P}\right) = D}$$

Τι κάνουμε?

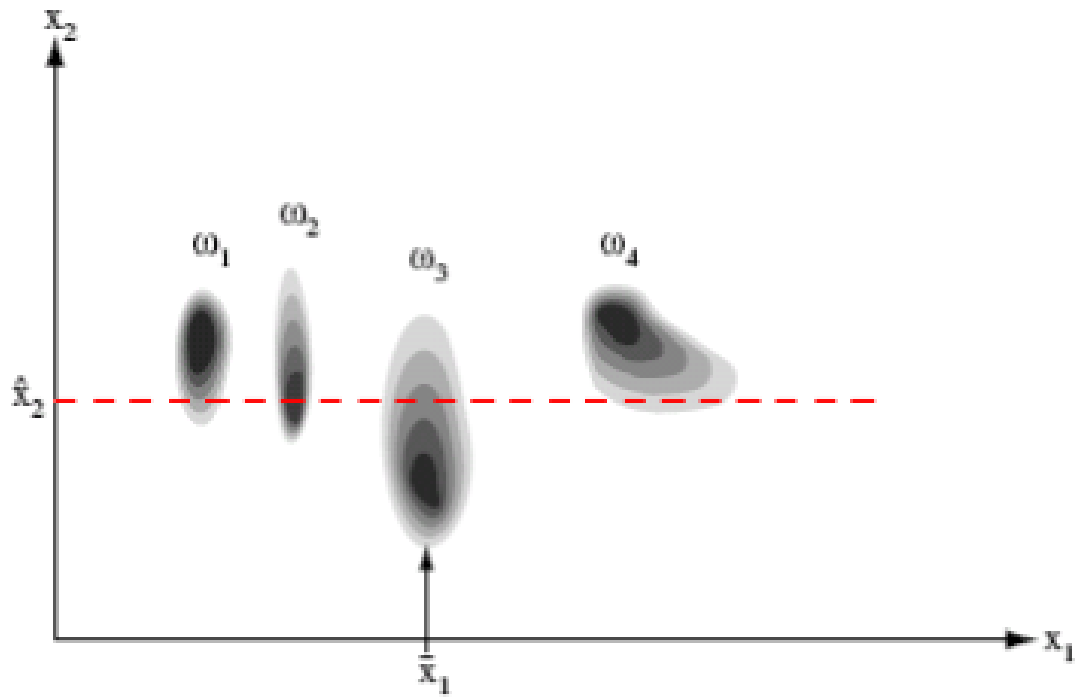
- Λείπουν χαρακτηριστικά.
- Μερικά χαρακτηριστικά έχουν πολύ θόρυβο.

Σε μία μέτρηση των χαρακτηριστικών των ψαριών με χαρακτηριστικά (x_1,x_2) (μήκος, φωτεινότητα) δεν μπορούμε να δούμε το x_2, τότε αντιμετωπίζουμε το πρόβλημα σαν πρόβλημα ενός χαρακτηριστικού και υπολογίζουμε τα pdf χρησιμοποιώντας την σχέση (marginal pdf)

$$f_{\mathbf{x}}(x) = \int f_{\mathbf{xy}}(x, y)dy$$

$$p_x(x_i) = \sum_j p_{xy}(x_i, y_j)$$

Εάν αντί για κάποιες μεταβλητές x_2 έχουν και θόρυβο και εμείς παρατηρούμε τις x_b, τότε γνωρίζοντας το $p(x_b|x_2)$ υπολογίζουμε το $P(\omega_i \mid x_1, x_b)$



Εάν έχουμε μόνο το x_2 και σαν x_1 πάρουμε την μέση τιμή του x_1 τότε οδηγούμεθα στο ω_3, ενώ εάν λάβουμε υπ' όψιν μόνο το x_2 τότε επιλέγουμε το ω_2 (σωστό) διότι το $p(x_2|\omega_2)$ είναι μεγαλύτερο

Bayesian Belief Δίκτυα

Υπάρχουν πολλές περιπτώσεις όπου γνωρίζουμε ή μπορούμε να προβλέψουμε με βεβαιότητα ποιες μεταβλητές είναι ή δεν είναι ανεξάρτητα, ακόμη και χωρίς δείγμα δεδομένων.

Ας υποθέσουμε, για παράδειγμα, ότι περιγράφουμε την κατάσταση ενός αυτοκινήτου - θερμοκρασία του κινητήρα, οι πιέσεις στα υγρά και τα ελαστικά, οι τάσεις στα καλώδια, και ούτω καθεξής. Είναι γνωστό ότι η πίεση λαδιού στον κινητήρα και η πίεση του αέρα σε ένα τροχό δεν σχετίζονται λειτουργικά, και ως εκ τούτου μπορούμε με ασφάλεια να υποθέσουμε ότι είναι στατιστικά ανεξάρτητα.
Ωστόσο, η θερμοκρασία του λαδιού και η θερμοκρασία του κινητήρα δεν είναι ανεξάρτητα (αλλά θα μπορούσε να είναι ανεξάρτητα υπό όρους). Επίσης μπορούμε να γνωρίζουμε πολλές μεταβλητές που ενδέχεται να επηρεάζουν η μία την άλλη: η θερμοκρασία του ψυκτικού υγρού επηρεάζεται από τη θερμοκρασία του κινητήρα, την ταχύτητα του ανεμιστήρα του ψυγείου και ούτω καθεξής.
Εμείς θα εκπροσωπούμε αυτές τις εξαρτήσεις γραφικά, μέσω των Bayesian Belief Δίκτυα, ονομάζονται επίσης αιτιατά δίκτυα, ή απλά δίκτυα

Bayesian Belief Δίκτυα

Παράδειγμα με το σολομό και το λαυράκι:

Εποχή A={a_1,a_2,a_3,a_4}={Χειμώνας, Άνοιξη, Καλοκαίρι, Φθινόπωρο},

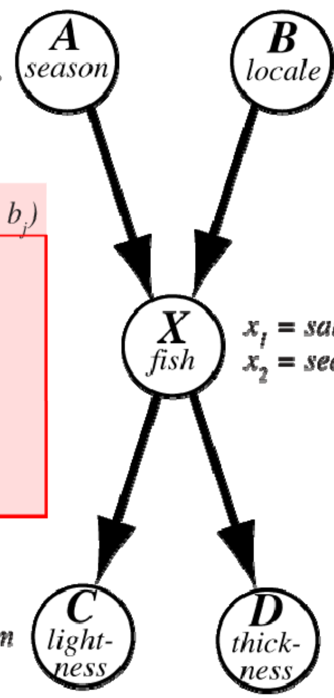
Περιοχή B={β_1,β_2}={Βόρειος, Νότιος Ατλαντικός}, Ψάρι X={χ_1,χ_2}= {σολομός, λαυράκι}

Φωτεινότητα C={c_1,c_2,c_3}={φωτεινό, μέτριο, σκοτεινό}, Πλάτος D={d_1,d_2}={πλατύ, στενό}

$P(a)$

$P(a_1)$	$P(a_2)$	$P(a_3)$	$P(a_4)$
0.25	0.25	0.25	0.25

a_1 = winter
a_2 = spring
a_3 = summer
a_4 = autumn



b_1 = north Atlantic
b_2 = south Atlantic

$P(b)$

$P(b_1)$	$P(b_2)$
0.6	0.4

$P(x|a,b)$

	$P(x_1 \mid a_i, b_j)$	$P(x_2 \mid a_i, b_j)$
a_1, b_1	0.5	0.5
a_1, b_2	0.7	0.3
a_2, b_1	0.6	0.4
a_2, b_2	0.8	0.2
a_3, b_1	0.4	0.6
a_3, b_2	0.1	0.9
a_4, b_1	0.2	0.8
a_4, b_2	0.3	0.7

x_1 = salmon
x_2 = sea bass

$$P(a_3,b_1,x_2,c_3,d_2) =$$
$$= P(a_3)P(b_1)P(x_2 \mid a_3,b_1)P(c_3 \mid x_2)P(d_2 \mid x_2) =$$
$$= 0.25 \times 0.6 \times 0.6 \times 0.5 \times 0.4 = 0.018$$

$P(c|x)$

	$P(c_1 \mid x_k)$	$P(c_2 \mid x_k)$	$P(c_3 \mid x_k)$
x_1	0.6	0.2	0.2
x_2	0.2	0.3	0.5

c_1 = light
c_2 = medium
c_3 = dark

d_1 = wide
d_2 = thin

$P(d|x)$

	$P(d_1 \mid x_k)$	$P(d_2 \mid x_k)$
x_1	0.3	0.7
x_2	0.6	0.4

Now we turn to the problem of using such a belief net to infer the identity of a fish. We have no direct information about the identity of the fish, and thus $P(x_1) = P(x_2) = 0.5$. This might be a reasonable starting point, expressing our lack of knowledge of the identity of the fish. Our goal now is to estimate the probabilities $P(x_1|\mathbf{e})$ and $P(x_2|\mathbf{e})$. Note that without any evidence we have

$$\begin{aligned}
P(x_1) &= \sum_{i,j,k,l} P(x_1, a_i, b_j, c_k, d_l) \\
&= \sum_{i,j,k,l} P(a_i)P(b_j)P(x_1|a_i, b_j)P(c_k|x_1)P(d_l|x_1) \\
&= \sum_{i,j} P(a_i)P(b_j)P(x_1|a_i, b_j) \\
&= (0.25)(0.5)\sum_{i,j} P(x_1|a_i, b_j) \\
&= (0.25)(0.5)(0.9 + 0.3 + 0.4 + 0.7 + 0.8 + 0.2 + 0.1 + 0.6) \\
&= 0.5,
\end{aligned}$$

and thus $P(x_1) = P(x_2)$, as we would expect.

Now we collect evidence for each node, $\{e_{\mathbf{A}}, e_{\mathrm{B}}, e_{\mathrm{C}}, e_{\mathrm{D}}\}$, assuming they are independent of each other. Suppose we know that it is winter, i.e., $P(a_1|e_{\mathbf{A}}) = 1$ and $P(a_i|e_{\mathbf{A}}) = 0$ for $i = 2, 3, 4$. Suppose we do not know which fishing area the boat came from but found that the particular fishing crew prefers to fish in the south Atlantic; we assume, then, that $P(b_1|e_{\mathrm{B}}) = 0.2$ and $P(b_2|e_{\mathrm{B}}) = 0.8$. We measure the fish and find that it is fairly light, and set by hand to be $P(e_{\mathrm{C}}|c_1) = 1$, $P(e_{\mathrm{C}}|c_2) = 0.5$, and $P(e_{\mathrm{C}}|c_3) = 0$. Suppose that due to occlusion, we cannot measure the width of the fish; we thus set $P(e_{\mathrm{D}}|d_1) = P(e_{\mathrm{D}}|d_2)$.

By Eq. 82, we have the estimated probability of each fish due to the parents $\mathcal{P}$ is, in full expanded form

$$
\begin{aligned}
P_{\mathcal{P}}(x_1) \quad \propto \quad & P(x_1|a_1, b_1)P(a_1)P(b_1) \\
& + P(x_1|a_1, b_2)P(a_1)P(b_2) \\
& + P(x_1|a_2, b_1)P(a_2)P(b_1) \\
& + P(x_1|a_2, b_2)P(a_2)P(b_2) \\
& + P(x_1|a_3, b_1)P(a_3)P(b_1) \\
& + P(x_1|a_3, b_2)P(a_3)P(b_2) \\
& + P(x_1|a_4, b_1)P(a_4)P(b_1) \\
& + P(x_1|a_4, b_2)P(a_4)P(b_2) \\
= \quad & 0.82.
\end{aligned}
$$

A similar calculation gives $P_{\mathcal{P}}(x_2) = 0.18$.

We now turn to the children nodes and find by Eq. 84

$$\begin{aligned}
P_{\mathcal{C}}(x_1) &\propto P(e_{\mathrm{C}}|x_1)P(e_{\mathrm{D}}|x_1) \\
&= [P(e_{\mathrm{C}}|c_1)P(c_1|x_1) + P(e_{\mathcal{C}}|c_2)P(c_2|x_1) + P(e_{\mathrm{C}}|c_3)P(c_3|x_1)] \\
&\quad \times [P(e_{\mathrm{D}}|d_1)P(d_1|x_1) + P(e_{\mathrm{D}}|d_2)P(d_2|x_1)] \\
&= [(1.0)(0.33) + (0.5)(0.33) + (0)(0.34)] \times [(1.0)(0.4) + (1.0)(0.6)] \\
&= 0.495.
\end{aligned}$$

A similar calculation gives $P_{\mathcal{C}}(x_2) \propto 0.85$. We put these estimates together by Eq. 79 as products $P(x_i) \propto P_{\mathcal{C}}(x_i)P_{\mathcal{P}}(x_i)$ and renormalize (i.e., divide by their sum). Thus our final estimates for node $\mathbf{X}$ are

$$P(x_1|\mathrm{e}) = \frac{(0.82)(0.495)}{(0.82)(0.495) + (0.18)(0.85)} = 0.726$$

$$P(x_2|\mathrm{e}) = \frac{(0.18)(0.85)}{(0.82)(0.495) + (0.18)(0.85)} = 0.274.$$

Thus given all the evidence throughout the belief net, the most probable outcome is $x_1 = salmon$.