# 2: p.d.f. BASED MODELS

CLASIFICACIÓN DE PATRONES (CLP)

P07

Profesores: M. Cabrera, J. Vidal

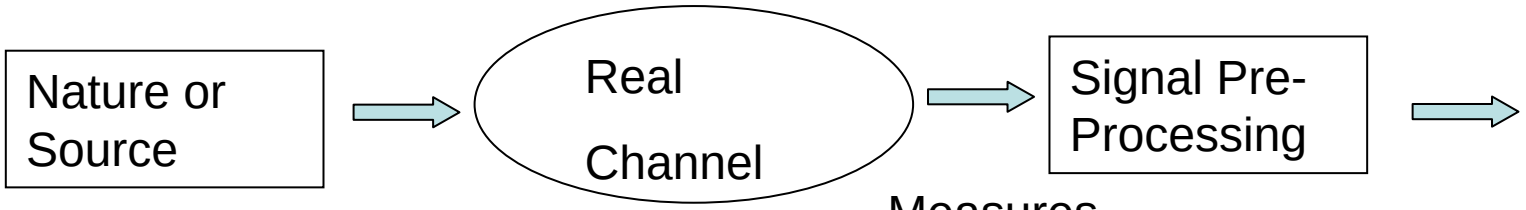
ETSETB-UPC

Optativa de 2º ciclo

(Some figures of this document are obtained from the reference book: “Pattern Classification” 2nd Edition by Duda, Hart and Stork, Ed. Wiley)

# Tema 1
# Introducción al tema 2

## CLASSIFICATION:

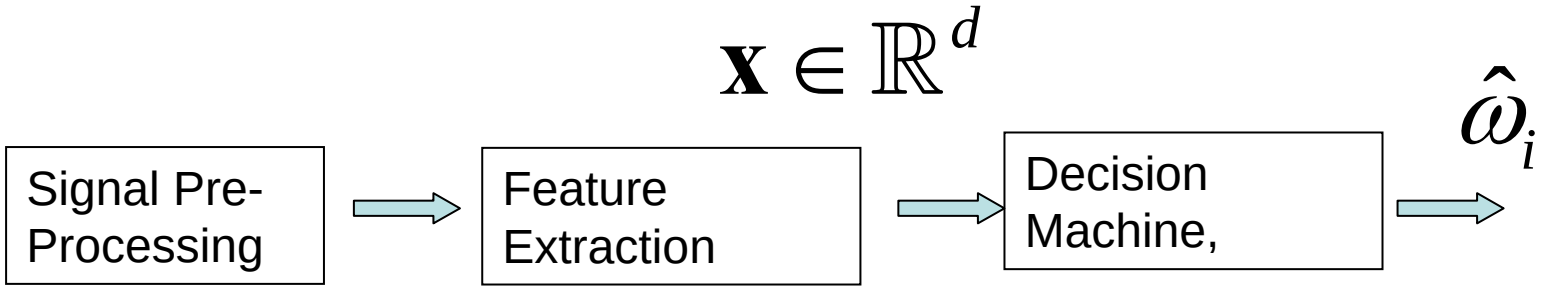



$\omega_1, \omega_2, ..., \omega_C$

C states or classes

# Tema 1
# Introducción al tema 2

CLASSIFICATION:

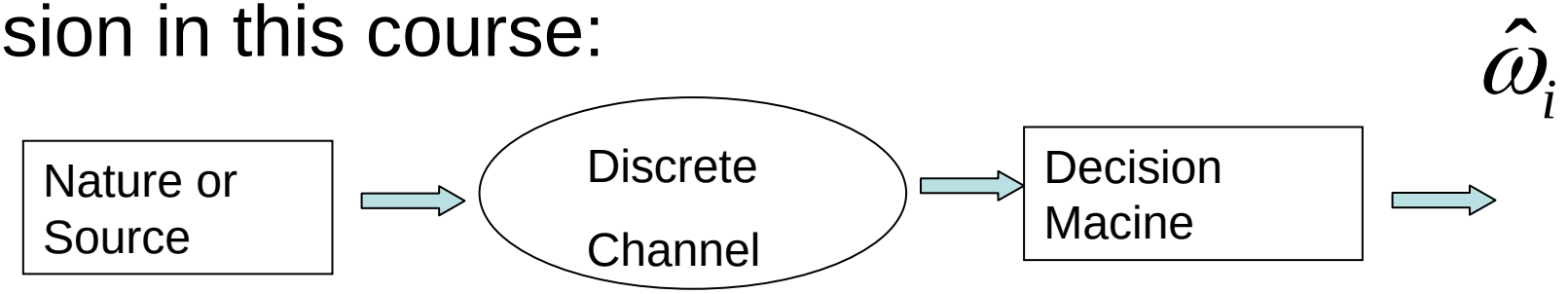



State or Class ??

# Tema 1
# Introducción al tema 2

CLASSIFICATION:

Vision in this course:



# Tema 1
# Introducción al tema 2

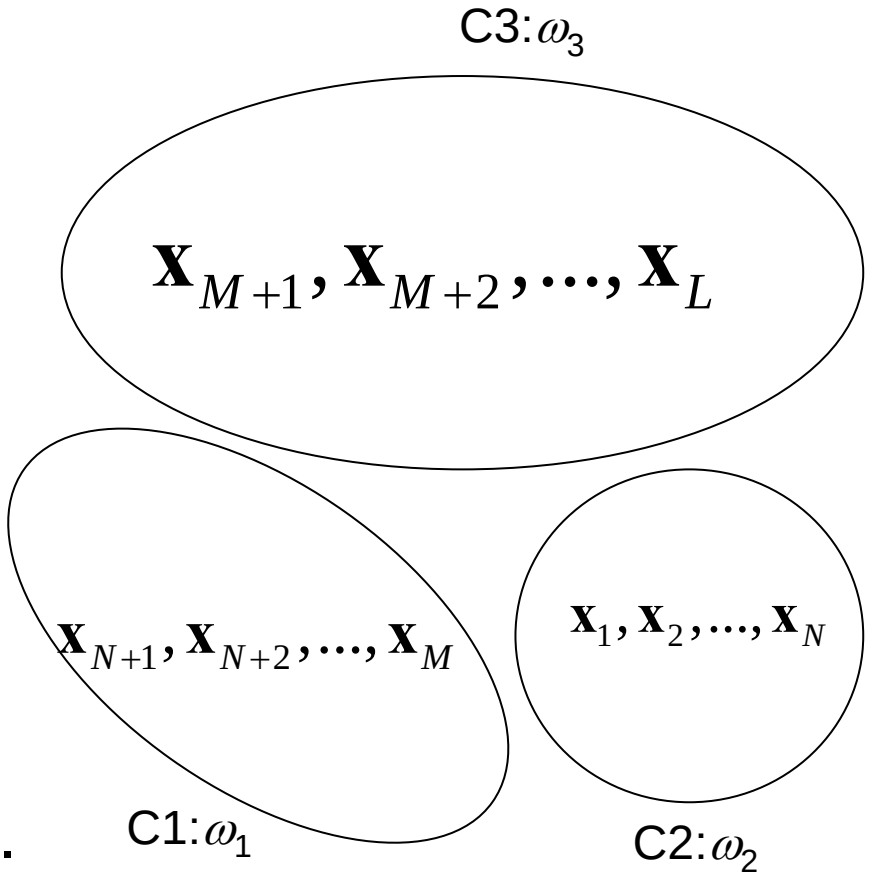
Designing the Decision Machine?

Data Base:

Vector set.

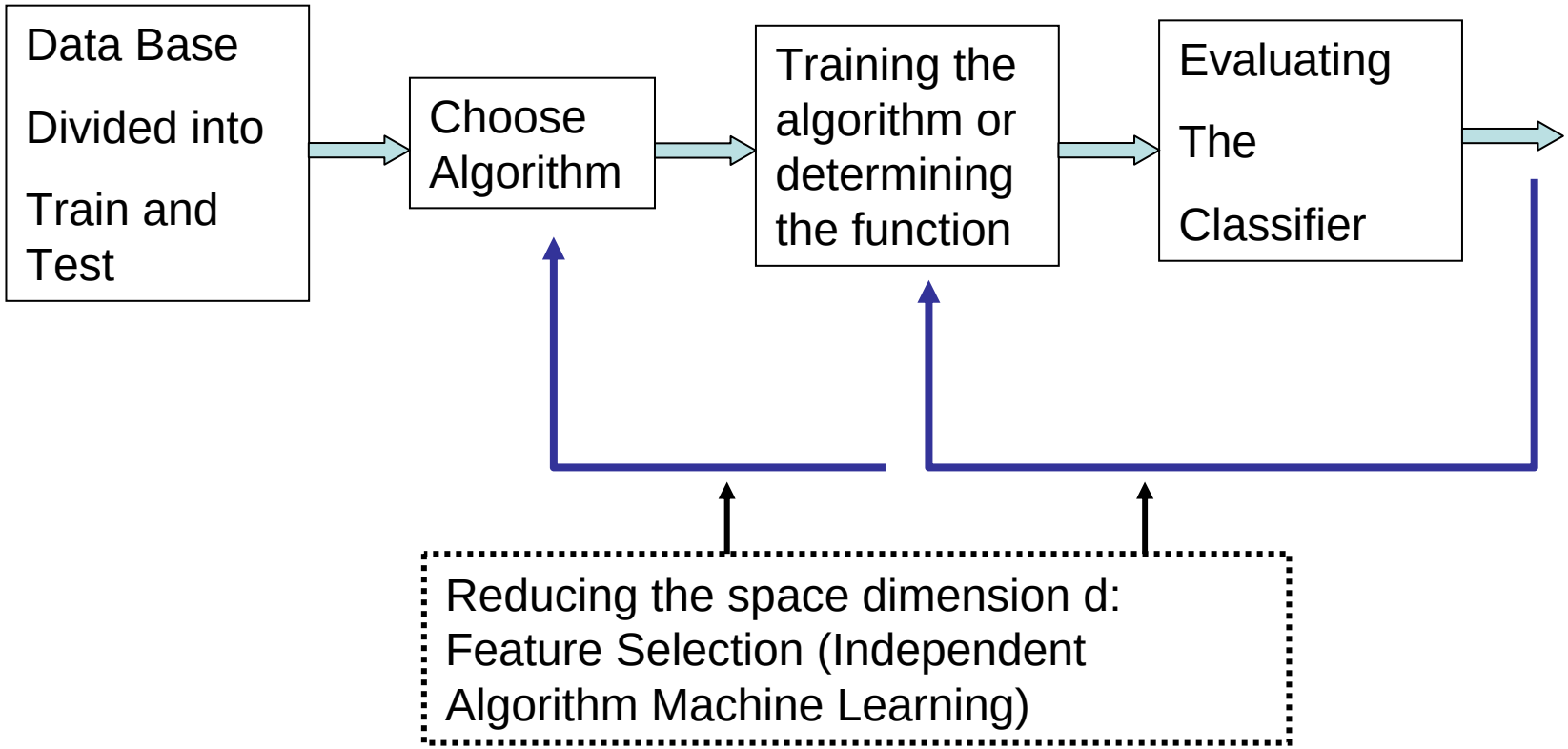
Coordinates or features

If The Data Base is organized in classes, categories, symbols or types. (Supervised Learning), we are lucky.



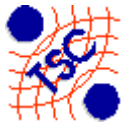
# Tema 1
# Introducción al tema 2

## Steps to design the classifier:



# Tipos de Clasificadores

- **MODELOS BASADOS en *f.d.p*.**
- **(ANÁLISIS DE COMPONENTES)**
- **TECNICAS NO basadas en *f.d.p*.**
  **APRENDIZAJE SUPERVISADO**
- **APRENDIZAJE NO SUPERVISADO**
- **(APRENDIZAJE INDEPENDIENTE DEL ALGORITMO)**



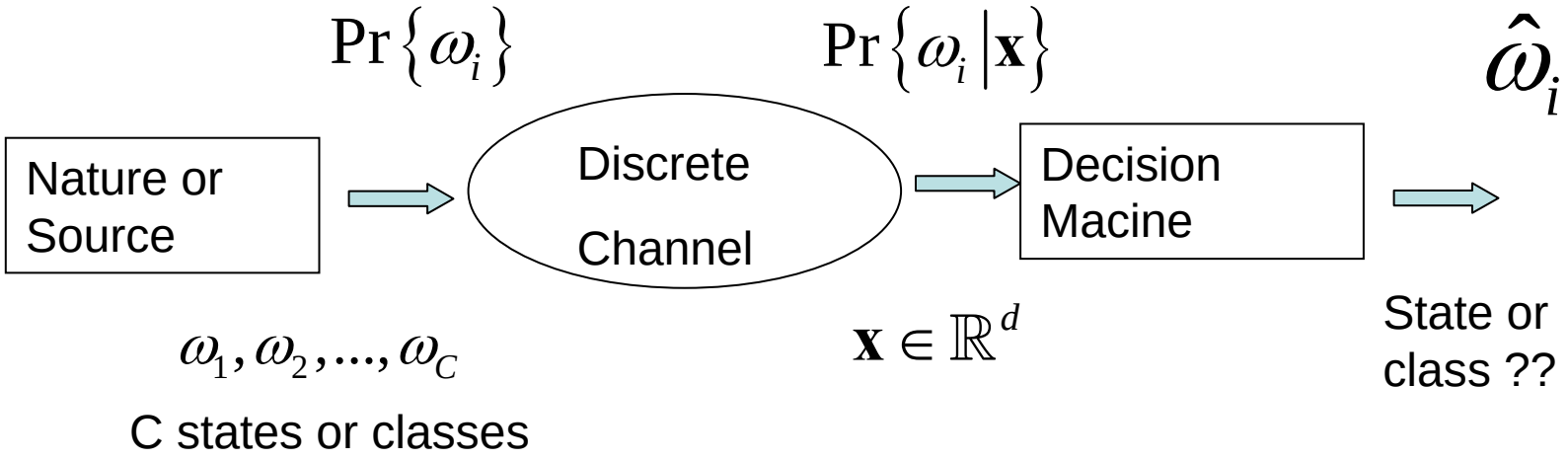

# Tema 2:
# Models with known Probability Density Funtion:

## 2.1 Bayesian Decision Theory: MAP
## 2.2 Maximum Likelihood ML and Bayesian Parameter Estimation

# INDICE:

## Bayesian Decision Theory: MAP

# Tema 1
# Introducción al tema 2



# 1 INTRODUCCIÓN

- Measures Dimension Vector *d*
- Nature State (Random Vector, *C* classes): Salmon or Sea Bass
- A priori probabilities.
- Class conditional p.d.f.
- Posterior Probabilities
- Evidence

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \text{Lightness} \\ \text{Length} \end{pmatrix} \in \mathbb{R}^d$$

$$C = 2 \Rightarrow \omega_1; \omega_2$$

$$\Pr\{\omega_1\}; \Pr\{\omega_2\} \quad \sum_{i=1}^{C} \Pr\{\omega_i\} = 1$$

$$f_{\mathbf{x}}(\mathbf{x}|\omega_1);\ f_{\mathbf{x}}(\mathbf{x}|\omega_2)$$

$$\Pr(\omega_j|\mathbf{x}) = \frac{f_{\mathbf{x}}(\mathbf{x}|\omega_j)\Pr(\omega_j)}{f_{\mathbf{x}}(\mathbf{x})}$$

$$f_{\mathbf{x}}(\mathbf{x}) = \sum_{i=1}^{C} f_{\mathbf{x}}(\mathbf{x}|\omega_j)\Pr\{\omega_1\}$$

# 1 INTRODUCCIÓN

$$\Pr\left(\omega_j \middle| \mathbf{x}\right) = \frac{f_{\mathbf{x}}\left(\mathbf{x} \middle| \omega_j\right)\Pr\left(\omega_j\right)}{f_{\mathbf{x}}\left(\mathbf{x}\right)}$$

$$POSTERIOR = \frac{LIKELIHOOD.PRIOR}{EVIDENCE}$$

$$\sum_{i=1}^{C} \Pr\left\{\omega_i \middle| \mathbf{x}\right\} = 1$$

# 1 INTRODUCCIÓN

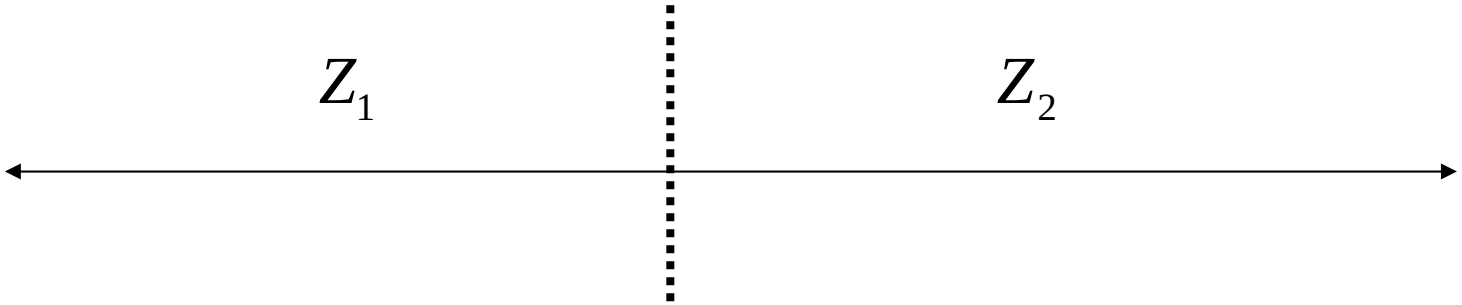
- **PRIOR**: Conocimiento a priori del estado, en ocasiones denominado Prejuicio.
- **POSTERIOR**: Probabilidad de que el estado de la naturaleza sea uno determinado cuando ya se han recibido los datos.
- **LIKELIHOOD**: Agrupa características que son comunes a todos los datos de una categoría determinada. Representa el modelo que el diseñador tiene sobre el comportamiento de la naturaleza.
- **EVIDENCE:** Factor de escala, no influye en las decisiones.

# 2 REGLA DE DECISIÓN DE BAYES (MAP)

Two Category Case

- Decision Rule: Vector Data **x**.

$$\Pr(\omega_1|\mathbf{x}) \underset{<}{\overset{>}{}} \Pr(\omega_2|\mathbf{x}) \quad \begin{matrix}\omega_1 \\ \\ \omega_2\end{matrix}$$



# 2 REGLA DE DECISIÓN DE BAYES (MAP)

- Probabilidad de error condicionada a **x**.

$$P_e = \Pr(error|\mathbf{x}) = \begin{cases} \mathbf{x} \in Z_1 & \hat{\omega} = \omega_1 & \Pr(\omega_2|\mathbf{x}) \\ \mathbf{x} \in Z_2 & \hat{\omega} = \omega_2 & \Pr(\omega_1|\mathbf{x}) \end{cases}$$

- Probabilidad de error promedio.

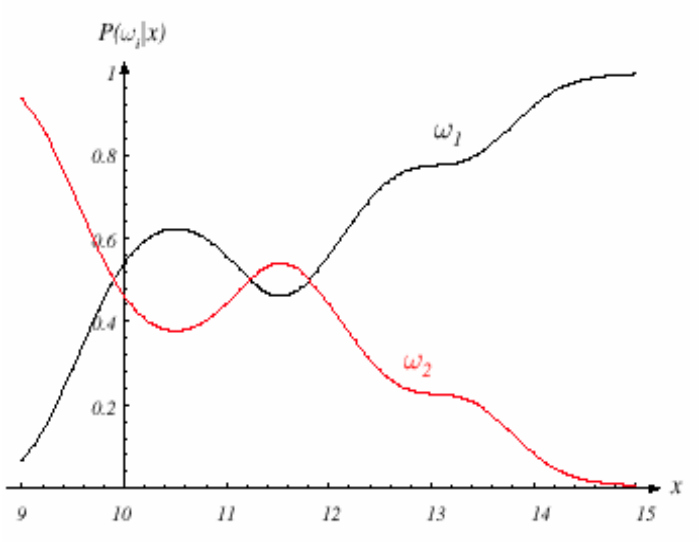
$$\Pr(error) = P_e =$$

$$E\left[\Pr(error|\mathbf{x})\right] = \int \Pr(error|\mathbf{x}) f_{\mathbf{x}}(\mathbf{x}) d\mathbf{x} =$$

$$\int_{Z_1} \Pr(\omega_2|\mathbf{x}) f_{\mathbf{x}}(\mathbf{x}) d\mathbf{x} + \int_{Z_2} \Pr(\omega_1|\mathbf{x}) f_{\mathbf{x}}(\mathbf{x}) d\mathbf{x}$$

# 2 REGLA DE DECISIÓN DE BAYES (MAP)

MAP




$$\Pr(\omega_1|\mathbf{x}) \underset{\omega_2}{\overset{\omega_1}{\gtrless}} \Pr(\omega_2|\mathbf{x})$$

Mínima Probabilidad de error.

$$\Pr(\omega_1)\, f_{\mathbf{x}}(\mathbf{x}|\omega_1) \underset{\omega_2}{\overset{\omega_1}{\gtrless}} \Pr(\omega_2)\, f_{\mathbf{x}}(\mathbf{x}|\omega_2)$$

# 2 REGLA DE DECISIÓN DE BAYES (MAP)

- If for some **x**, … particular observation give us no information about the state of nature

$$\Pr\left(\mathbf{x}_0 \middle| \omega_1\right) = \Pr\left(\mathbf{x}_0 \middle| \omega_2\right) \Rightarrow \Pr\left(\omega_i \middle| \mathbf{x}_0\right) = \Pr\left(\omega_i\right)$$

- If ……. the decision is based entirely on the likelihoods.

$$\Pr(\omega_1) = \Pr(\omega_2); \Rightarrow \Pr\left(\mathbf{x} \middle| \omega_1\right) \begin{matrix} > \omega_1 \\ < \omega_2 \end{matrix} \Pr\left(\mathbf{x} \middle| \omega_2\right)$$

# 2 REGLA DE DECISIÓN DE BAYES (MAP)

- General Case (Continuous Features):
- C Classes

$$\hat{\omega}_i = \max_{\omega_i} \left\{ \Pr\left(\omega_i \left| \mathbf{x} \right.\right) \right\} = \max_{\omega_i} \left\{ \Pr\left(\omega_i\right) f_{\mathbf{x}}\left(\mathbf{x} \left| \omega_i \right.\right) \right\}$$

# 2.2 REGLA DE DECISIÓN DE BAYES (MAP)

- Caso General:
- El criterio MAP es equivalente a minimizar la Probabilidad de error en el clasificador:
- *Demostración:*

$$\Pr\left\{e\middle|\mathbf{x}\in Z_i\right\}=\sum_{\substack{j=1\\ j\neq i}}^{C}\Pr\left\{\omega_j\middle|\mathbf{x}\right\}=1-\Pr\left\{\omega_i\middle|\mathbf{x}\right\}$$

# 2 REGLA DE DECISIÓN DE BAYES (MAP)

$$\Pr\{e\} = \int_{R^d} \Pr\{e|\mathbf{x}\} f_{\mathbf{x}}(\mathbf{x}) d\mathbf{x} = \sum_{i=1}^{C} \int_{Z_i} \Pr\{e|\mathbf{x} \in Z_i\} f_{\mathbf{x}}(\mathbf{x}) d\mathbf{x} =$$

$$\sum_{i=1}^{C} \int_{Z_i} \left(1 - \Pr\{\omega_i|\mathbf{x}\}\right) f_{\mathbf{x}}(\mathbf{x}) d\mathbf{x} =$$

$$\sum_{i=1}^{C} \int_{Z_i} f_{\mathbf{x}}(\mathbf{x}) d\mathbf{x} - \sum_{i=1}^{C} \int_{Z_i} \Pr\{\omega_i|\mathbf{x}\} f_{\mathbf{x}}(\mathbf{x}) d\mathbf{x} =$$

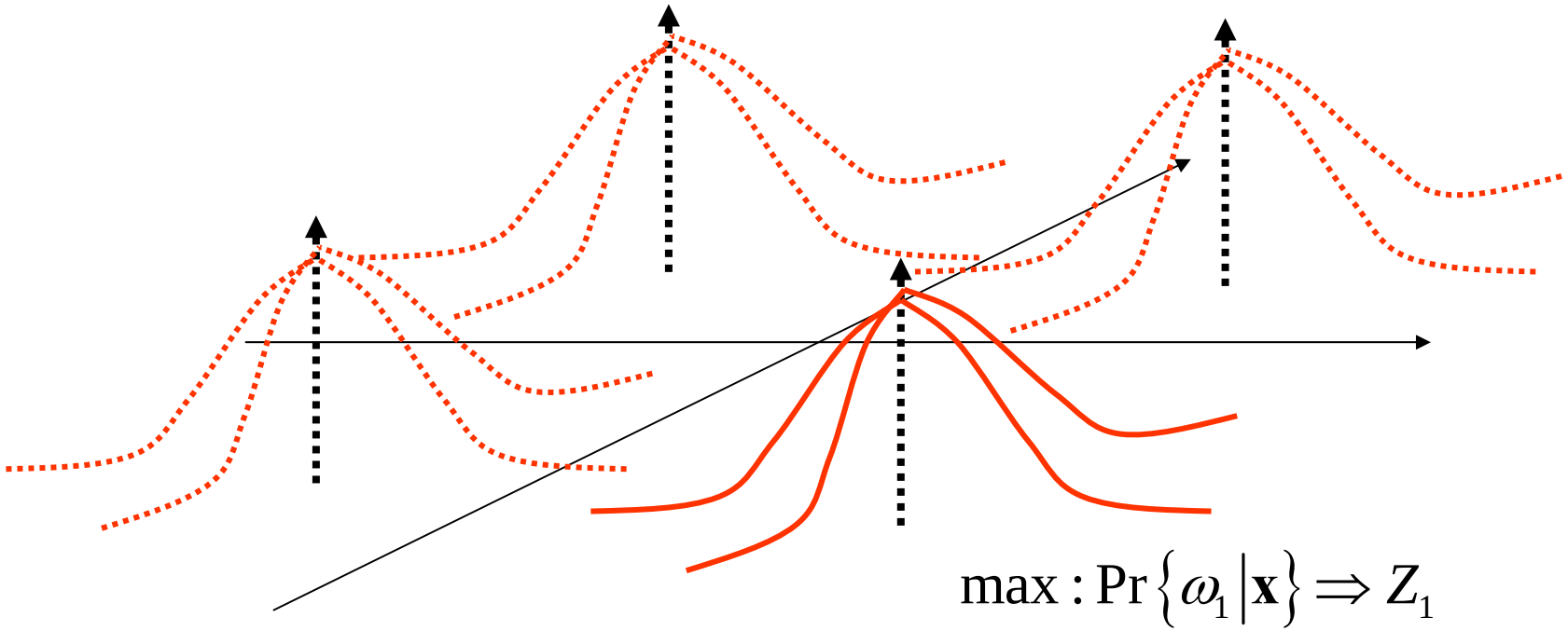
$$\int_{R^d} f_{\mathbf{x}}(\mathbf{x}) d\mathbf{x} - \sum_{i=1}^{C} \int_{Z_i} \Pr\{\omega_i|\mathbf{x}\} f_{\mathbf{x}}(\mathbf{x}) d\mathbf{x} =$$

$$1 - \sum_{i=1}^{C} \int_{Z_i} \Pr\{\omega_i|\mathbf{x}\} f_{\mathbf{x}}(\mathbf{x}) d\mathbf{x}$$

# 2 REGLA DE DECISIÓN DE BAYES (MAP)

$$R^d = Z_1 \cup Z_2 \ldots \cup Z_C = \bigcup_{i=1}^{C} Z_i$$

$$Z_i \cap Z_j = \varnothing$$



# 2 REGLA DE DECISIÓN DE BAYES (MAP)

- Permitir o Realizar acciones distintas a la toma de decisiones

$$\alpha_1, \ldots, \alpha_a$$

- Se define una función de coste en función de estas acciones.

# 2 REGLA DE DECISIÓN DE BAYES (MAP)

Ejemplos

- Bases de datos biomédicas ¿Penalizo por igual los errores 1) sano/enfermo 2) enfermo/sano?
- SPAM
- OCR ¿Penalizo por igual error en consonante que error en vocal?
- RADAR

# 3 CLASIFICADORES DE MÍNIMO RIESGO

- Pérdida que genera la decisión i cuando el estado verdadero es j

$$\lambda\left(\alpha_i \left| \omega_j \right.\right)$$

- Pérdida asociada a la acción i, Riesgo condicional

$$R\left(\alpha_i \left| \mathbf{x} \right.\right) = \sum_{j=1}^{C} \lambda\left(\alpha_i \left| \omega_j \right.\right) \Pr\left(\omega_j \left| \mathbf{x} \right.\right)$$

- Riesgo Total

$$R = \sum_{i=1}^{a} \int R\left(\alpha_i \left| \mathbf{x} \right.\right) f_{\mathbf{X}}\left(\mathbf{x}\right) d\mathbf{x}$$

- Mínimo Riesgo, equivale a elegir

$$\alpha_i \Rightarrow \min\left(R\left(\alpha_i \left| \mathbf{x} \right.\right)\right)$$

# 3 CLASIFICADORES DE MÍNIMO RIESGO

- C = 2 categorías:

$$\left.\begin{array}{l}\alpha_1 : \text{decidir } \omega_1 \\ \alpha_2 : \text{decidir } \omega_2\end{array}\right\} \Rightarrow \lambda_{ij} = \lambda\left(\alpha_i \left| \omega_j \right.\right)$$

- Riesgo Condicional

$$R\left(\alpha_1 \left| \mathbf{x} \right.\right) = \lambda_{11} \Pr\left(\omega_1 \left| \mathbf{x} \right.\right) + \lambda_{12} \Pr\left(\omega_2 \left| \mathbf{x} \right.\right)$$
$$R\left(\alpha_2 \left| \mathbf{x} \right.\right) = \lambda_{21} \Pr\left(\omega_1 \left| \mathbf{x} \right.\right) + \lambda_{22} \Pr\left(\omega_2 \left| \mathbf{x} \right.\right)$$

- Regla de Decisión:

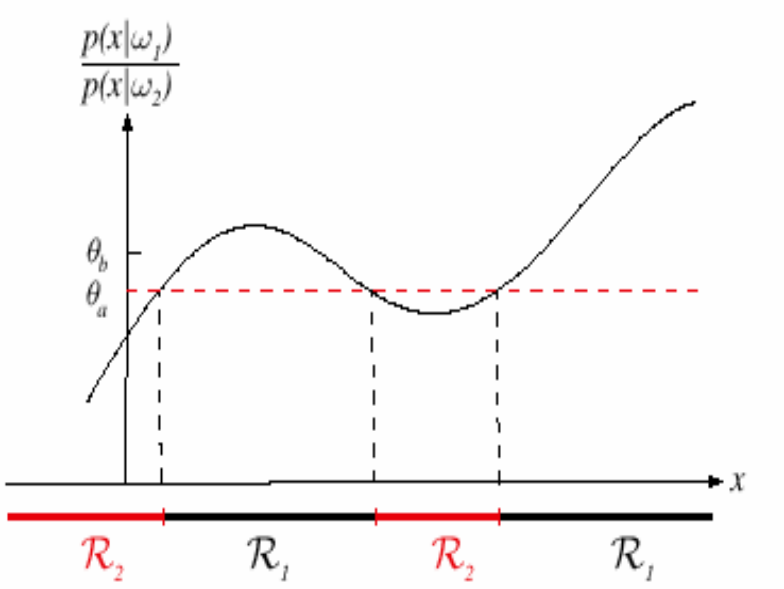
$$R\left(\alpha_1 \left| \mathbf{x} \right.\right) \underset{\alpha_1}{\overset{\alpha_2}{\gtrless}} R\left(\alpha_2 \left| \mathbf{x} \right.\right)$$

# 3 CLASIFICADORES DE MÍNIMO RIESGO

- LIKELIHOOD Ratio

$$\frac{f_{\mathbf{x}}\left(\mathbf{x}\middle|\omega_1\right)}{f_{\mathbf{x}}\left(\mathbf{x}\middle|\omega_2\right)} \underset{\alpha_2}{\overset{\alpha_1}{\gtrless}} \left(\frac{\lambda_{12}-\lambda_{22}}{\lambda_{21}-\lambda_{11}}\right)\frac{\Pr\left(\omega_2\right)}{\Pr\left(\omega_1\right)} = \gamma$$

Umbral o Threshold independiente de x



# 2.3 CLASIFICADORES DE MÍNIMO RIESGO

- LIKELIHOOD Ratio
- Mínima Pr(error): MAP

$$\lambda_{ij} = \begin{cases} 0 & i = j \\ 1 & i \neq j \end{cases}$$

$$\frac{f_{\mathbf{x}}\left(\mathbf{x}\middle|\omega_1\right)}{f_{\mathbf{x}}\left(\mathbf{x}\middle|\omega_2\right)} \underset{\alpha_2}{\overset{\alpha_1}{\gtrless}} \frac{\Pr\left(\omega_2\right)}{\Pr\left(\omega_1\right)} = \gamma$$

# 2.3 CLASIFICADORES DE MÍNIMO RIESGO

- Mínimo Riesgo = Mínima Probabilidad de error

$$\lambda_{ij} = \begin{cases} 0 & i = j \\ 1 & i \neq j \end{cases}$$

$$R(\alpha_i \mid \mathbf{x}) = \sum_{j=1}^{C} \lambda(\alpha_i \mid \omega_j) \Pr(\omega_j \mid \mathbf{x})$$

$$= \sum_{j=1,\neq i}^{C} \Pr(\omega_j \mid \mathbf{x}) = 1 - \Pr(\omega_i \mid \mathbf{x})$$

# 3 CLASIFICADORES DE MÍNIMO RIESGO

Otros Criterios:

- MINIMAX:
  - Tiene sentido cuando no se conocen las probabilidades a priori.
  - Minimiza el Máximo Riesgo, eligiendo las regiones de decisión para que la función de riesgo no dependa de las probabilidades a priori.
  - Ejemplo para C=2 Categorías

# 3 CLASIFICADORES DE MÍNIMO RIESGO

- MINIMAX: Ejemplo para C=2 Categorías

$$R = \int_{Z_1} \left( \lambda_{11} \Pr(\omega_1)\, f_{\mathbf{x}}(\mathbf{x}|\omega_1) + \lambda_{12} \Pr(\omega_2)\, f_{\mathbf{x}}(\mathbf{x}|\omega_2) \right) d\mathbf{x} +$$

$$\int_{Z_2} \left( \lambda_{21} \Pr(\omega_1)\, f_{\mathbf{x}}(\mathbf{x}|\omega_1) + \lambda_{22} \Pr(\omega_2)\, f_{\mathbf{x}}(\mathbf{x}|\omega_2) \right) d\mathbf{x} =$$

$$\left\{ \begin{array}{l} \Pr(\omega_1) + \Pr(\omega_2) = 1 \\ \int_{Z_1} f_{\mathbf{x}}(\mathbf{x}|\omega_1) d\mathbf{x} + \int_{Z_2} f_{\mathbf{x}}(\mathbf{x}|\omega_1) d\mathbf{x} = 1 \end{array} \right\}$$

$$\lambda_{22} + (\lambda_{12} - \lambda_{22}) \int_{Z_1} f_{\mathbf{x}}(\mathbf{x}|\omega_2) d\mathbf{x} +$$

$$\Pr(\omega_1) \left( \lambda_{11} - \lambda_{22} + (\lambda_{21} - \lambda_{11}) \int_{Z_2} f_{\mathbf{x}}(\mathbf{x}|\omega_1) d\mathbf{x} + (\lambda_{22} - \lambda_{12}) \int_{Z_1} f_{\mathbf{x}}(\mathbf{x}|\omega_2) d\mathbf{x} \right)$$

$$R = K_1(Z_1, Z_2) + \Pr(\omega_1) K_2(Z_1, Z_2)$$

# 3 CLASIFICADORES DE MÍNIMO RIESGO

- MINIMAX: Ejemplo para C=2 Categorías

$$R = K_1(Z_1, Z_2) + \Pr(\omega_1) K_2(Z_1, Z_2) = K_1(Z_1, Z_2)$$

$$K_2(Z_1, Z_2) = 0$$

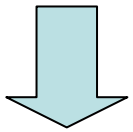
$$R_{\text{minimax}} = \lambda_{22} + (\lambda_{12} - \lambda_{22}) \int_{Z_1} f_{\mathbf{x}}(\mathbf{x} | \omega_2) d\mathbf{x}$$

# 3 CLASIFICADORES DE MÍNIMO RIESGO

- MINIMAX: Ejemplo para C=2 Categorías y mínima probabilidad de error

$$\lambda_{11}-\lambda_{22}+(\lambda_{21}-\lambda_{11})\int_{Z_2} f_{\mathbf{x}}(\mathbf{x}|\omega_1)d\mathbf{x}+(\lambda_{22}-\lambda_{12})\int_{Z_1} f_{\mathbf{x}}(\mathbf{x}|\omega_2)d\mathbf{x}=0$$

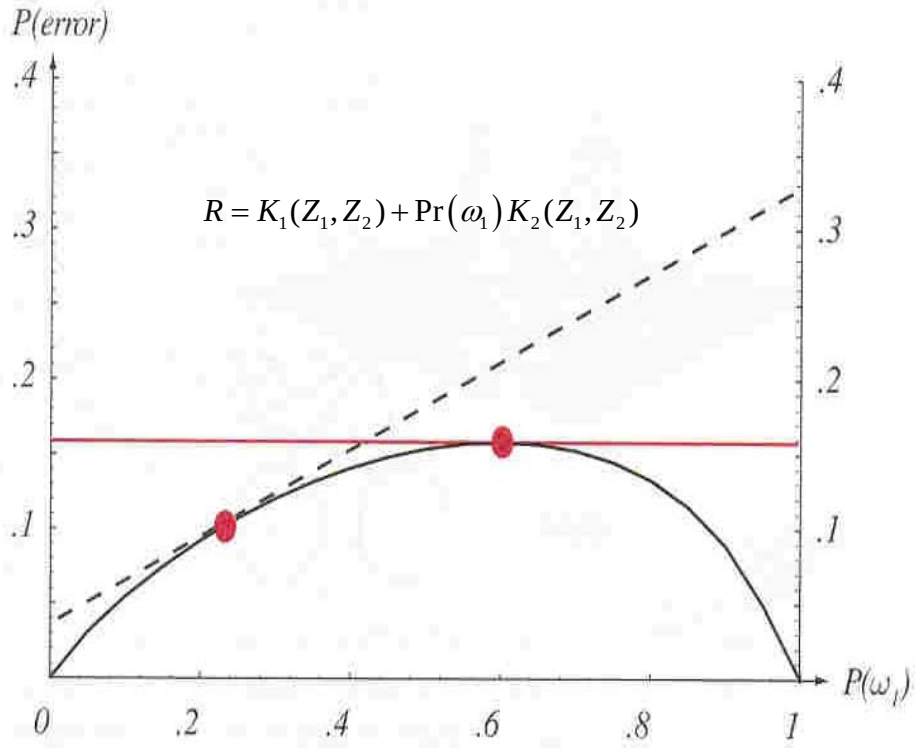
$$\lambda_{ij}=\begin{cases}0 & i=j\\ 1 & i\neq j\end{cases}$$



$$\int_{Z_2} f_{\mathbf{x}}(\mathbf{x}|\omega_1)d\mathbf{x}-\int_{Z_1} f_{\mathbf{x}}(\mathbf{x}|\omega_2)d\mathbf{x}=0$$

# 3 CLASIFICADORES DE MÍNIMO RIESGO

- MINIMAX: Ejemplo para C=2 Categorías



# 3 CLASIFICADORES DE MÍNIMO RIESGO

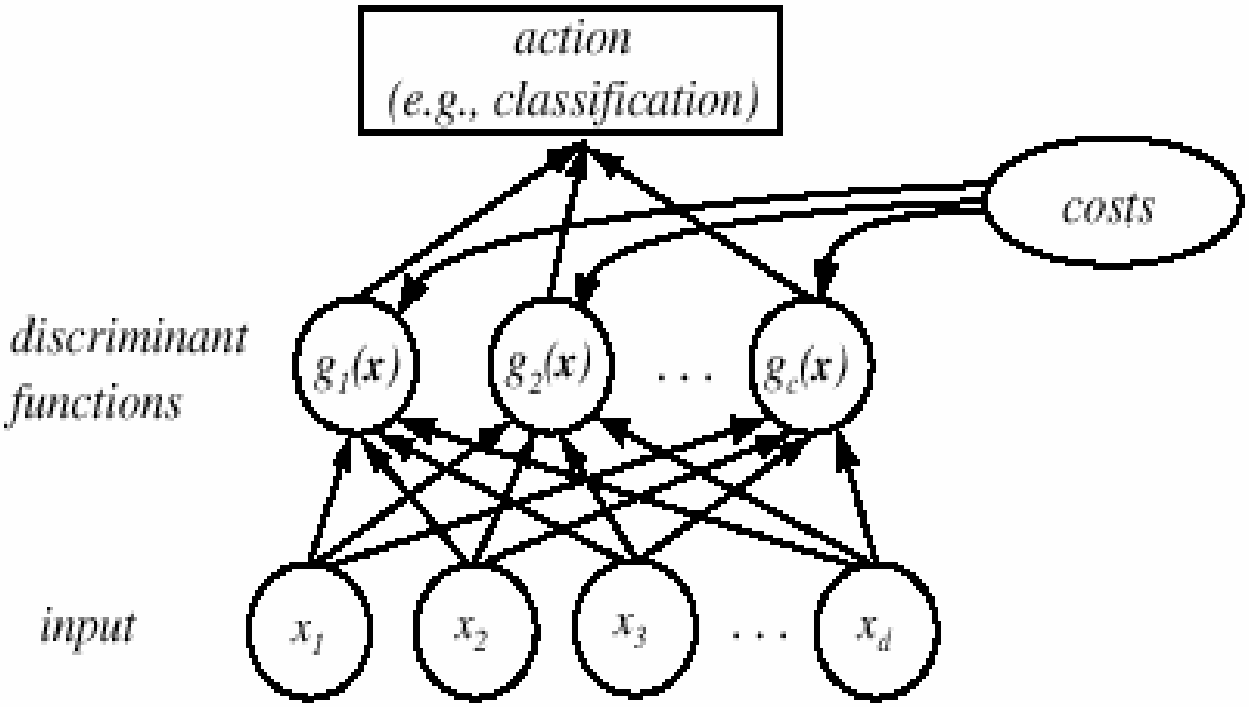
Otros Criterios:

- NEYMAN PEARSON:
  - Se minimiza el riesgo total sujeto a alguna restricción.

$$\int R\left(\alpha_i \middle| \mathbf{x}\right) d\mathbf{x} < \text{cte}$$

# 2.4 FUNCIONES DISCRIMINANTES Y REGIONES DE DECISIÓN

- Caso de múltiples categorías C:



# 4 FUNCIONES DISCRIMINANTES Y REGIONES DE DECISIÓN

- Definición de Función Discriminante ($g_i$):
  - El clasificador asigna una clase $\omega_i$ a un vector de características **x**.
  - Criterio de clasificación.

$$g_i(\mathbf{x}) > g_j(\mathbf{x}) \qquad \forall j \neq i$$

# 4 FUNCIONES DISCRIMINANTES Y REGIONES DE DECISIÓN

- Casos Particulares:
  - MAP (Equivalente a mínima probabilidad de error)

$$g_i(\mathbf{x}) = \Pr\left(\omega_i \left| \mathbf{x} \right.\right)$$

  - Mínimo Riesgo.

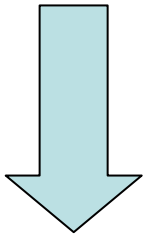
$$g_i(\mathbf{x}) = -R\left(\alpha_i \left| \mathbf{x} \right.\right)$$

# 4 FUNCIONES DISCRIMINANTES Y REGIONES DE DECISIÓN

- Casos Particulares: MAP
  - Un mismo criterio puede realizarse mediante diferentes funciones discriminantes:

$$g_i(\mathbf{x}) = \Pr\left(\omega_i \left| \mathbf{x}\right.\right)$$

Ln (log neperiano)es una función convexa



$$h(\mathbf{x}) = \ln\left(g(\mathbf{x})\right)$$

$$h_i(\mathbf{x}) = \ln\left(f_{\mathbf{x}}(\mathbf{x}\left|\omega_i\right.)\right) + \ln\left(\Pr\left(\omega_i\right)\right)$$

# 4 FUNCIONES DISCRIMINANTES Y REGIONES DE DECISIÓN

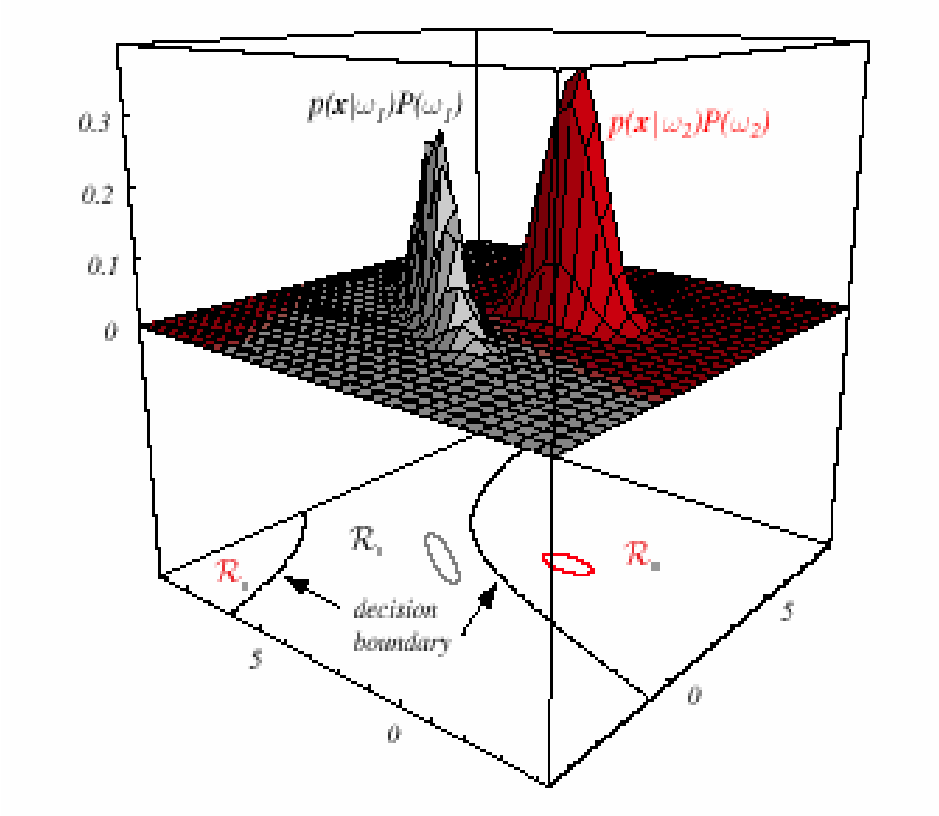
- Casos C=2 Categorías: DICOTOMIZADOR

$$g(\mathbf{x}) \equiv g_1(\mathbf{x}) - g_2(\mathbf{x}) \begin{matrix} > \alpha_1 \\ \\ < \alpha_2 \end{matrix} 0$$

- Ejemplos
  - Comunicaciones Binarias BPSK, 2FSK
  - Detección de Enfermedades SI/NO

# 4 FUNCIONES DISCRIMINANTES Y REGIONES DE DECISIÓN

- Casos C=2 Categorías: DICOTOMIZADOR

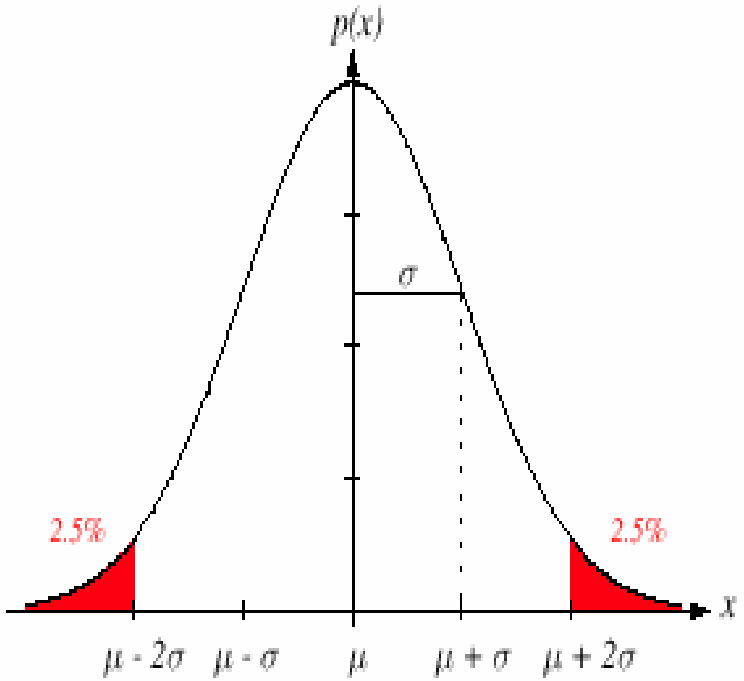


# 5 f.d.p. NORMAL O GAUSSIANA

- UNIVARIABLE

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

$$\mu = \mathbf{E}[x]; \quad \sigma^2 = \mathbf{E}\left[(x-\mu)^2\right]$$



# 5 f.d.p. NORMAL O GAUSSIANA

- MULTIVARIABLE $\mathbf{x} : N(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x)$
  - Momentos estadísticos

$$\mathbf{x} \in \mathbb{R}^d; \quad \boldsymbol{\mu}_x = \mathbf{E}[\mathbf{x}] \in \mathbb{R}^d; \quad \boldsymbol{\Sigma}_x = \mathbf{E}\left[(\mathbf{x} - \mu_x)(\mathbf{x} - \mu_x)^{\mathbf{T}}\right] \in \mathbb{R}^{dxd}$$

  - La matriz de covarianza es definida positiva (Autovalores reales y positivos)
  - f.d.p. del vector **x**:

$$f_{\mathbf{x}}(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}_x|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \mu_x)^{\mathbf{T}} \boldsymbol{\Sigma}_x^{-1}(\mathbf{x} - \mu_x)\right\}$$

# 5 f.d.p. NORMAL O GAUSIANA

- MULTIVARIABLE
  - Las transformaciones lineales de v.a.gausianas presentan distribución normal

$$\mathbf{A} \in \mathbb{R}^{dxk}; \quad \mathbf{y} = \mathbf{A}^{\mathbf{T}}\mathbf{x} \in \mathbb{R}^{k}$$

$$\boldsymbol{\mu}_{\mathbf{y}} = \mathbf{E}[\mathbf{y}] = \mathbf{E}\left[\mathbf{A}^{\mathbf{T}}\mathbf{x}\right] = \mathbf{A}^{\mathbf{T}}\mathbf{E}[\mathbf{x}] = \mathbf{A}^{\mathbf{T}}\boldsymbol{\mu}_{\mathbf{x}}$$

$$\boldsymbol{\Sigma}_{\mathbf{y}} = \mathbf{E}\left[\left(\mathbf{y}-\boldsymbol{\mu}_{\mathbf{y}}\right)\left(\mathbf{y}-\boldsymbol{\mu}_{\mathbf{y}}\right)^{\mathbf{T}}\right] =$$

$$\mathbf{E}\left[\left(\mathbf{A}^{\mathbf{T}}\mathbf{x}-\mathbf{A}^{\mathbf{T}}\boldsymbol{\mu}_{\mathbf{x}}\right)\left(\mathbf{A}^{\mathbf{T}}\mathbf{x}-\mathbf{A}^{\mathbf{T}}\boldsymbol{\mu}_{\mathbf{x}}\right)^{\mathbf{T}}\right] = \mathbf{E}\left[\mathbf{A}^{\mathbf{T}}\left(\mathbf{x}-\boldsymbol{\mu}_{\mathbf{x}}\right)\left(\mathbf{x}-\boldsymbol{\mu}_{\mathbf{x}}\right)^{\mathbf{T}}\mathbf{A}\right] = \mathbf{A}^{\mathbf{T}}\boldsymbol{\Sigma}_{x}\mathbf{A}$$

# 5 f.d.p. NORMAL O GAUSIANA

- MULTIVARIABLE
  - Blanqueo a partir de la diagonalización de la matriz de covarianza:

$$\boldsymbol{\Sigma}\mathbf{v}_i = \lambda_i \mathbf{v}_i$$

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_x = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{\mathbf{T}}$$

  - Matriz de Autovectores, ortonormales entre sí

$$\mathbf{U} = \left(\mathbf{v}_1, \mathbf{v}_2, .., \mathbf{v}_d\right); \quad \mathbf{U}\mathbf{U}^{\mathbf{T}} = \mathbf{I}$$

  - Valores Propios

$$\boldsymbol{\Lambda} = diag(\lambda_1, \lambda_2, ..., \lambda_d)$$

  - Transformación:

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Lambda}^{-1/2}; \qquad \boldsymbol{\Lambda}^{-1/2} = diag(1/\sqrt{\lambda}_1, 1/\sqrt{\lambda}_2, ..., 1/\sqrt{\lambda}_d)$$

# 5 f.d.p. NORMAL O GAUSIANA

- MULTIVARIABLE
  - Media:

$$\boldsymbol{\mu}_y = \mathbf{A}^T\boldsymbol{\mu}_x = \boldsymbol{\Lambda}^{-1/2}\mathbf{U}^T\boldsymbol{\mu}_x$$

$$\boldsymbol{\mu}_y = \mathbf{A}^T\boldsymbol{\mu}_x = \mathbf{U}^T\boldsymbol{\mu}_x$$

  - Matriz de Covarianza:

$$\boldsymbol{\Sigma}_y = \mathbf{A}^T\boldsymbol{\Sigma}_x\mathbf{A} = \boldsymbol{\Lambda}^{-1/2}\mathbf{U}^T\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T\mathbf{U}\boldsymbol{\Lambda}^{-1/2}$$

$$= \boldsymbol{\Lambda}^{-1/2}\boldsymbol{\Lambda}\boldsymbol{\Lambda}^{-1/2} = \mathbf{I}$$

$$\boldsymbol{\Sigma}_y = \mathbf{A}^T\boldsymbol{\Sigma}_x\mathbf{A} = \mathbf{U}^T\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T\mathbf{U} = \boldsymbol{\Lambda}$$

# 5 f.d.p. NORMAL O GAUSIANA

- MULTIVARIABLE
  - f.d.p

$$f_{\mathbf{y}}(\mathbf{y}) = \frac{1}{(2\pi)^{d/2}\left|\boldsymbol{\Sigma}_{\mathbf{y}}\right|^{1/2}} \exp\left\{-\frac{1}{2}\left(\mathbf{y}-\mu_{\mathbf{y}}\right)^{\mathbf{T}} \boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\left(\mathbf{y}-\mu_{\mathbf{y}}\right)\right\} =$$

$$\frac{1}{(2\pi)^{d/2}} \exp\left\{-\frac{1}{2}\left(\mathbf{y}-\mu_{\mathbf{y}}\right)^{\mathbf{T}}\left(\mathbf{y}-\mu_{\mathbf{y}}\right)\right\} = \prod_{i=1}^{d} f_{y_i}(y_i)$$

$$y_i : N(\mu_i, 1)$$

$$y_i : N(\mu_i, \lambda_i)$$

# 5 f.d.p. NORMAL O GAUSIANA

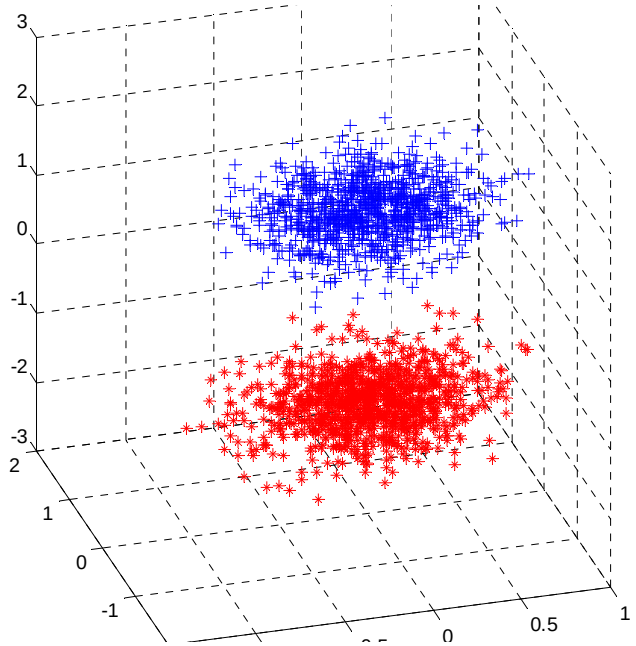
- MULTIVARIABLE
  - Las muestras de una población normal se agrupan en clusters alrededor de la media $\mu$
  - Los ejes principales de los hiper-elipsoides son los autovectores de la matriz de covarianza.
  - La distancia cuadrática de Mahalanobis constituye el término del exponente de la f.d.p., ayuda a evaluar i/o interpretar los clusters

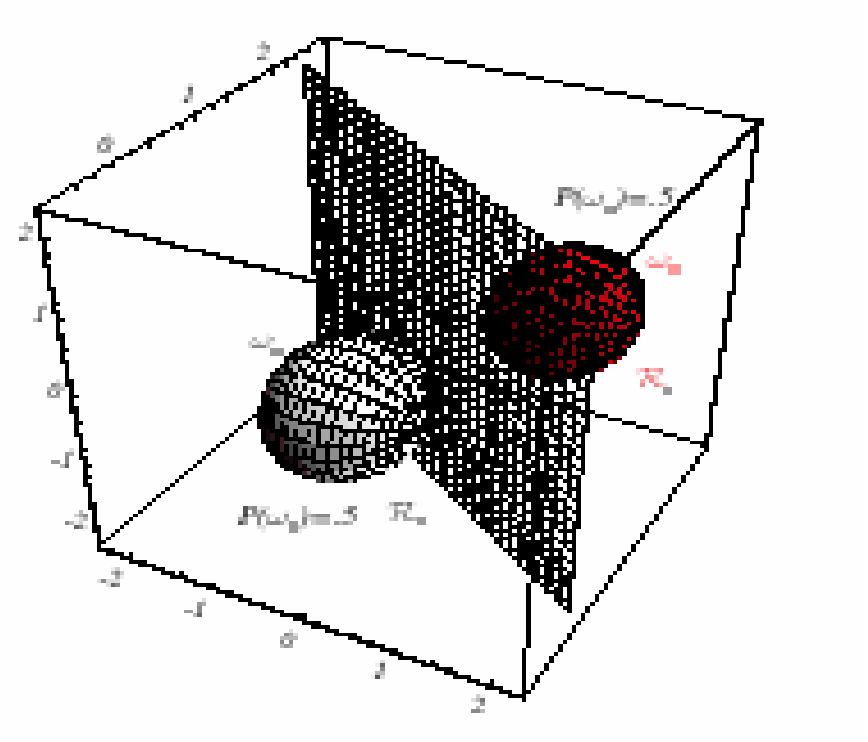
$$d_M^2 = (\mathbf{x} - \mu)^{\mathbf{T}} \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \mu)$$

  - Blanqueo convierte hiper-elipsoides en hiper-esferas
  - Si **A=U** los clusters mantienen la forma de elipsoides con semi-ejes paralelos a los ejes de coordenadas.

# 5 f.d.p. NORMAL O GAUSIANA

- CLUSTERS:



# 6 FUNC. DISCRIMINANTES: f.d.p. NORMAL

- f.d.p. Condicionada: $f_{\mathbf{x}}(\mathbf{x}|\omega_i) : N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$
- Probabilidad a Priori $\Pr(\omega_i)$
- Función discriminante MAP

$$g_i(\mathbf{x}) = \ln\left(f_{\mathbf{x}}(\mathbf{x}|\omega_i)\right) + \ln\left(\Pr(\omega_i)\right) =$$
$$-\tfrac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^{\mathbf{T}}\boldsymbol{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i) - \tfrac{d}{2}\ln(2\pi) - \tfrac{1}{2}\ln\left(|\boldsymbol{\Sigma}_i|\right) + \ln\left(\Pr(\omega_i)\right)$$

# 6 FUNC. DISCRIMINANTES: f.d.p. NORMAL

- 3 Casos respecto a la Matriz de covarianza

  - Caso 1 $\boldsymbol{\Sigma}_i = \sigma^2 \mathbf{I}$
  - Caso 2 $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}$
  - Caso 3 $\boldsymbol{\Sigma}_i$ Arbitrario

# 6 F. DISCRIMINANTES: f.d.p. NORMAL

Caso 1 $\quad \boldsymbol{\Sigma}_i = \sigma^2 \mathbf{I}$

- La función discriminante:
  - depende de la distancia euclídea

$$g_i(\mathbf{x}) = -\frac{1}{2\sigma^2}\left(\mathbf{x}-\boldsymbol{\mu}_i\right)^{\mathbf{T}}\left(\mathbf{x}-\boldsymbol{\mu}_i\right)+\ln\left(\Pr\left(\omega_i\right)\right)$$

  - Es LINEAL con el vector de datos recibido:

$$h_i(\mathbf{x}) = +\frac{1}{\sigma^2}\boldsymbol{\mu}_i{}^{\mathbf{T}}\mathbf{x}-\frac{1}{2\sigma^2}\boldsymbol{\mu}_i{}^{\mathbf{T}}\boldsymbol{\mu}_i+\ln\left(\Pr\left(\omega_i\right)\right)=\mathbf{w}_i{}^{\mathbf{T}}\mathbf{x}+w_{i0}$$

  - Las Fronteras de decisión son HIPERplanos:

$$h_i(\mathbf{x}) = h_j(\mathbf{x}) \Rightarrow \mathbf{w}^{\mathbf{T}}\left(\mathbf{x}-\mathbf{x}_0\right)=0$$

# 6 F. DISCRIMINANTES: f.d.p. NORMAL

Caso 1 $\quad \boldsymbol{\Sigma}_i = \sigma^2 \mathbf{I}$

- Las Fronteras de decisión son HIPERplanos:

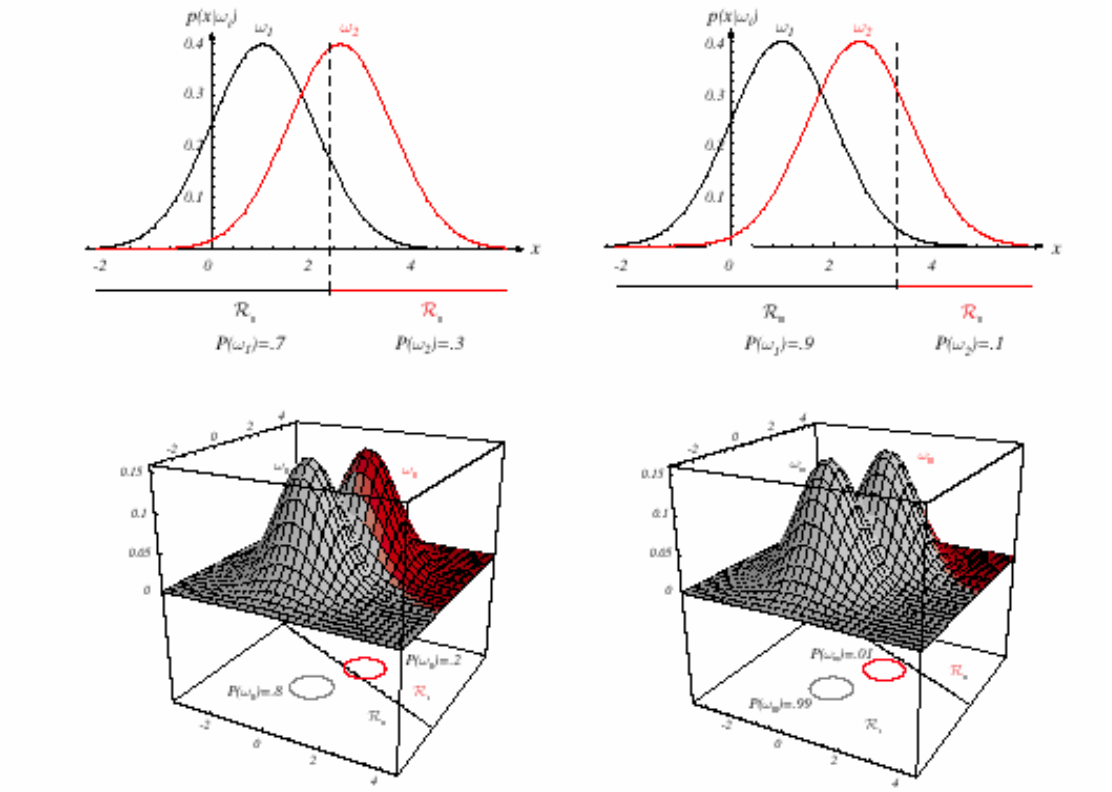
$$h_i(\mathbf{x}) = h_j(\mathbf{x}) \Rightarrow \mathbf{w}^{\mathbf{T}}\left(\mathbf{x} - \mathbf{x}_0\right) = 0$$

$$\mathbf{w} = \boldsymbol{\mu}_i - \boldsymbol{\mu}_j;$$

$$\mathbf{X}_0 = \tfrac{1}{2}\left(\boldsymbol{\mu}_i + \boldsymbol{\mu}_j\right) - \frac{\sigma^2}{\left\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\right\|^2} \ln\left(\frac{\Pr(\omega_i)}{\Pr(\omega_j)}\right)\left(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\right)$$

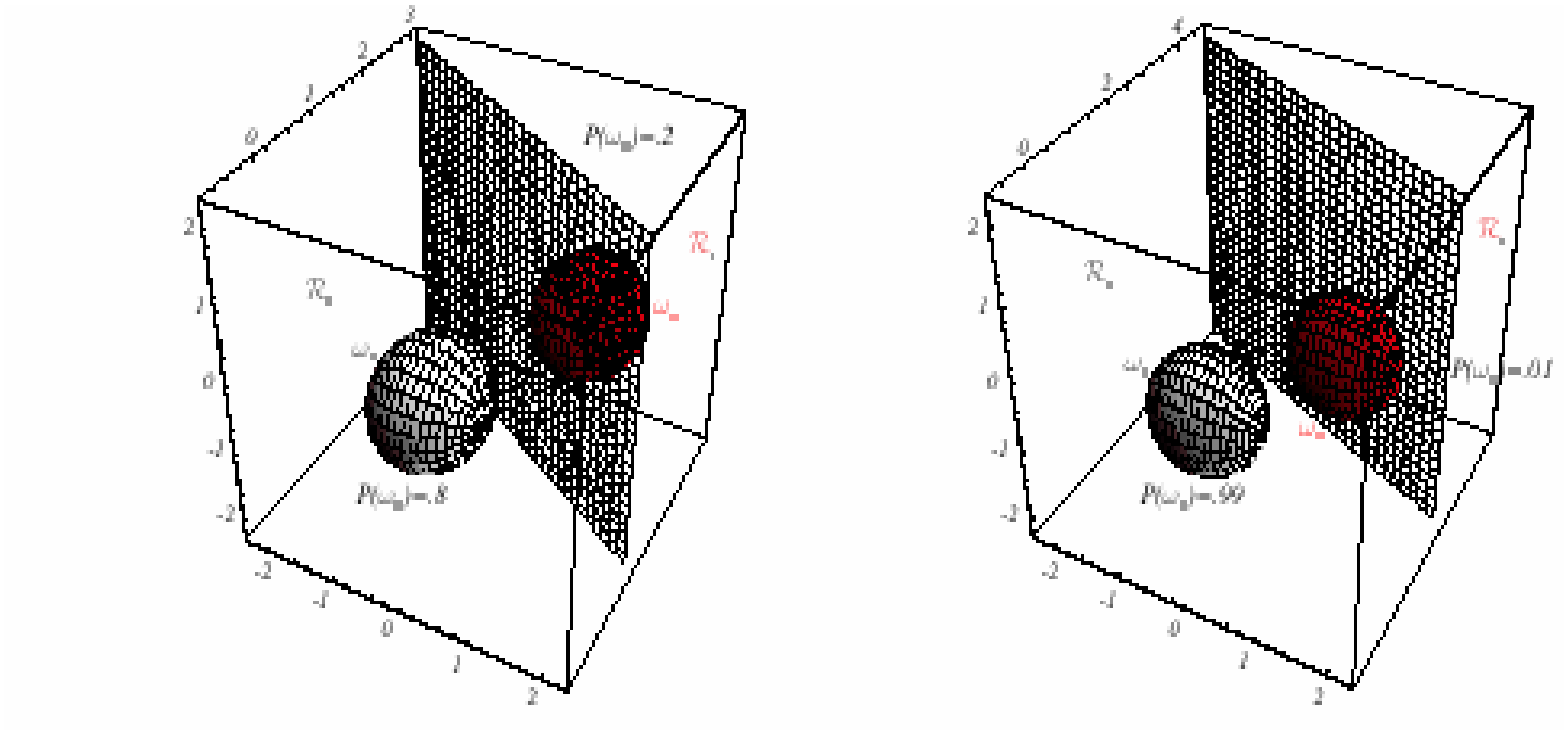
# 6 FUNC. DISCRIMINANTES: f.d.p. NORMAL

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



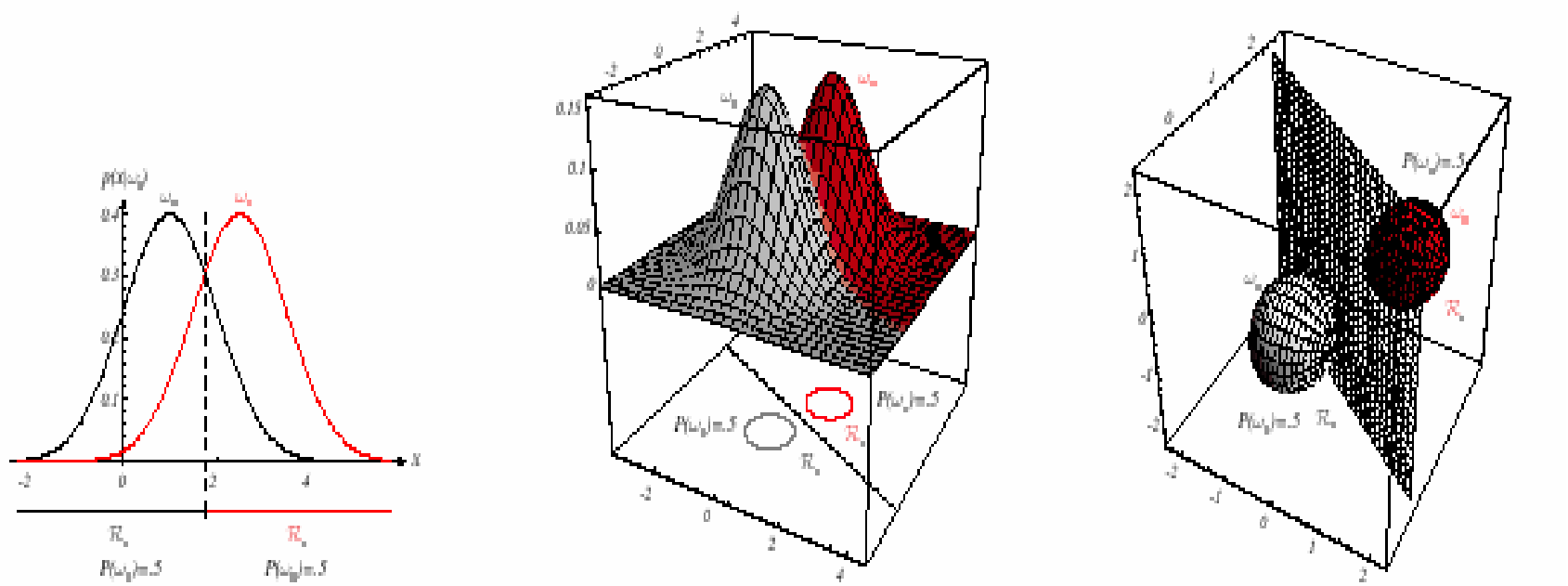
# 6 FUNC. DISCRIMINANTES: f.d.p. NORMAL

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



# 6 FUNC. DISCRIMINANTES: f.d.p. NORMAL

- Caso 1 $\boldsymbol{\Sigma}_i = \sigma^2 \mathbf{I}$
  - Categorías equiprobables: $\Pr(\omega_i) = \frac{1}{C}$
  - Clasificador de Mínima Distancia Euclídea



# 6 F. DISCRIMINANTES: f.d.p. NORMAL

Caso 2 $\quad \boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}$

- La función discriminante:

$$g_i(\mathbf{x}) = -\tfrac{1}{2}\left(\mathbf{x}-\boldsymbol{\mu}_i\right)^{\mathbf{T}} \boldsymbol{\Sigma}^{-1}\left(\mathbf{x}-\boldsymbol{\mu}_i\right) + \ln\left(\Pr\left(\omega_i\right)\right)$$

  - Es LINEAL con el vector de datos recibido:

$$h_i(\mathbf{x}) = \left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_i\right)^{\mathbf{T}} \mathbf{x} - \tfrac{1}{2}\boldsymbol{\mu}_i{}^{\mathbf{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_i + \ln\left(\Pr\left(\omega_i\right)\right) = \mathbf{w}_i{}^{\mathbf{T}}\mathbf{x} + w_{i0}$$

  - Las Fronteras de decisión son HIPERplanos:

$$h_i(\mathbf{x}) = h_j(\mathbf{x}) \Rightarrow \mathbf{w}^{\mathbf{T}}\left(\mathbf{x}-\mathbf{x}_0\right) = 0$$

# 6 F. DISCRIMINANTES: f.d.p. NORMAL

Caso 2 $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}$

- Las Fronteras de decisión son HIPERplanos:

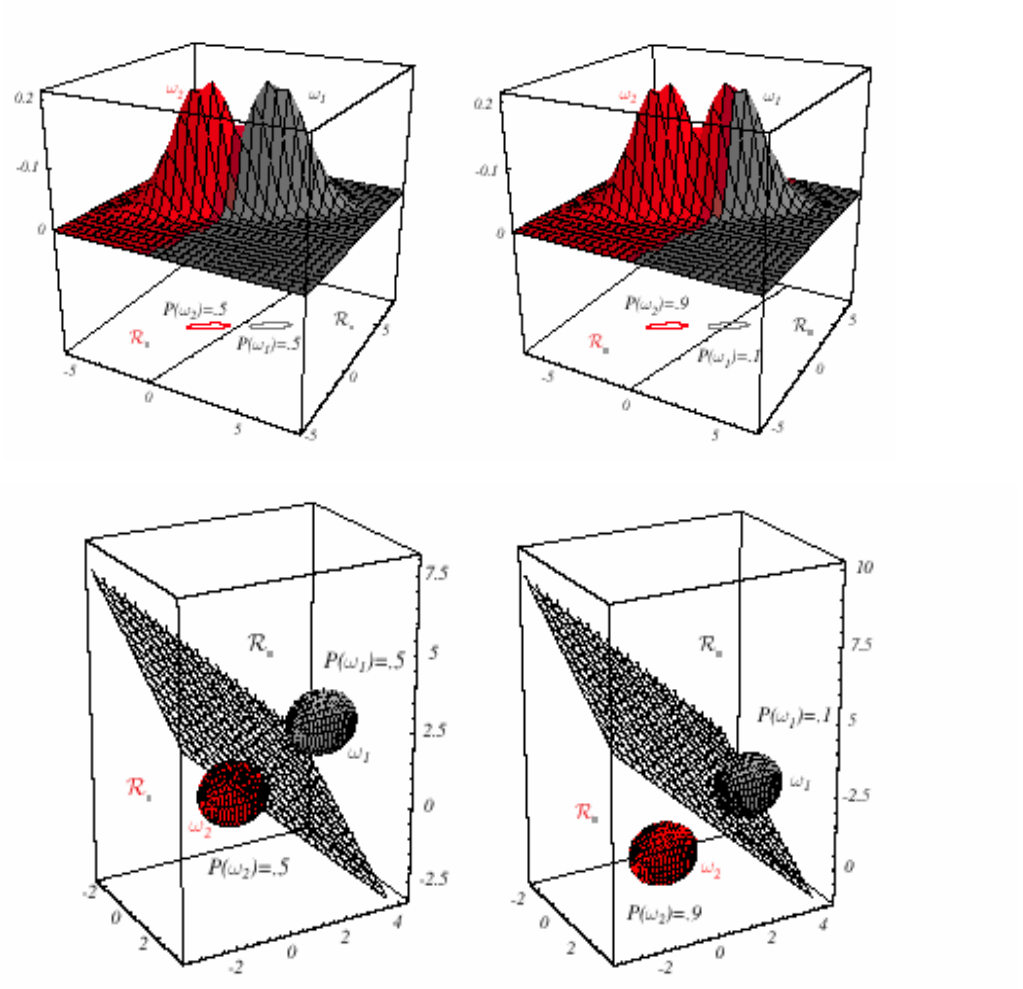
$$h_i(\mathbf{x}) = h_j(\mathbf{x}) \Rightarrow \mathbf{w}^{\mathbf{T}}\left(\mathbf{x} - \mathbf{x}_0\right) = 0$$

$$\mathbf{w} = \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\right);$$

$$\mathbf{X}_0 = \tfrac{1}{2}\left(\boldsymbol{\mu}_i + \boldsymbol{\mu}_j\right) - \frac{\ln\left(\Pr(\omega_i)\right) - \ln\left(\Pr(\omega_j)\right)}{\left(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\right)^{\mathbf{T}} \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\right)}\left(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\right)$$

# 6 F. DISCRIMINANTES: f.d.p. NORMAL

Caso 2



# 6 F. DISCRIMINANTES: f.d.p. NORMAL

Caso 3 $\boldsymbol{\Sigma}_i$ arbitrario

$$g_i(\mathbf{x}) =$$

$$-\tfrac{1}{2}\mathbf{x}^{\mathbf{T}}\boldsymbol{\Sigma}_i^{-1}\mathbf{x} + \boldsymbol{\mu}_i^{\mathbf{T}}\boldsymbol{\Sigma}_i^{-1}\mathbf{x} - \tfrac{1}{2}\boldsymbol{\mu}_i^{\mathbf{T}}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i - \tfrac{1}{2}\ln\left(\left|\boldsymbol{\Sigma}_i\right|\right) + \ln\left(\Pr\left(\omega_i\right)\right)$$

- Las superficies que separan 2 zonas son hiperquadráticas:
  - Hiperplanos
  - Hiperesferas
  - Hiperelipsoides
  - Hiperparaboloides
  - hiperhiperboloides

# 6 F. DISCRIMINANTES: f.d.p. NORMAL

## Caso 3 $\boldsymbol{\Sigma}_i$ arbitrario

- Cálculo de las superficies que separan 2 zonas

$$g_i(\mathbf{x}) = g_j(\mathbf{x}) \Rightarrow$$

$$-\tfrac{1}{2}\mathbf{x}^{\mathbf{T}}\boldsymbol{\Sigma}_i^{-1}\mathbf{x} + \boldsymbol{\mu}_i^{\mathbf{T}}\boldsymbol{\Sigma}_i^{-1}\mathbf{x} - \tfrac{1}{2}\boldsymbol{\mu}_i^{\mathbf{T}}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i - \tfrac{1}{2}\ln\left(\left|\boldsymbol{\Sigma}_i\right|\right) + \ln\left(\Pr\left(\omega_i\right)\right)$$

$$+\tfrac{1}{2}\mathbf{x}^{\mathbf{T}}\boldsymbol{\Sigma}_j^{-1}\mathbf{x} - \boldsymbol{\mu}_j^{\mathbf{T}}\boldsymbol{\Sigma}_j^{-1}\mathbf{x} + \tfrac{1}{2}\boldsymbol{\mu}_j^{\mathbf{T}}\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\mu}_j + \tfrac{1}{2}\ln\left(\left|\boldsymbol{\Sigma}_j\right|\right) - \ln\left(\Pr\left(\omega_j\right)\right) = 0$$

$$\Rightarrow$$

$$\mathbf{x}^{\mathbf{T}}\left(\tfrac{1}{2}\boldsymbol{\Sigma}_j^{-1} - \tfrac{1}{2}\boldsymbol{\Sigma}_i^{-1}\right)\mathbf{x} + \left(\boldsymbol{\mu}_i^{\mathbf{T}}\boldsymbol{\Sigma}_i^{-1} - \boldsymbol{\mu}_j^{\mathbf{T}}\boldsymbol{\Sigma}_j^{-1}\right)\mathbf{x}$$

$$-\tfrac{1}{2}\boldsymbol{\mu}_i^{\mathbf{T}}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i + \tfrac{1}{2}\boldsymbol{\mu}_j^{\mathbf{T}}\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\mu}_j - \tfrac{1}{2}\ln\left(\frac{\left|\boldsymbol{\Sigma}_i\right|}{\left|\boldsymbol{\Sigma}_j\right|}\right) + \ln\left(\frac{\Pr\left(\omega_i\right)}{\Pr\left(\omega_j\right)}\right) = 0$$

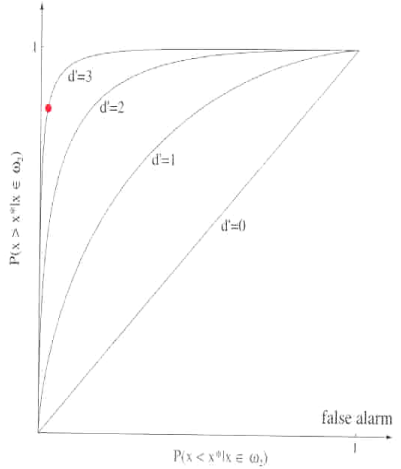
$$\Rightarrow h(\mathbf{x}) = g_i(\mathbf{x}) - g_j(\mathbf{x}) = \mathbf{x}^{\mathbf{T}}\mathbf{A}\mathbf{x} + \mathbf{v}^{\mathbf{T}}\mathbf{x} + e = 0$$

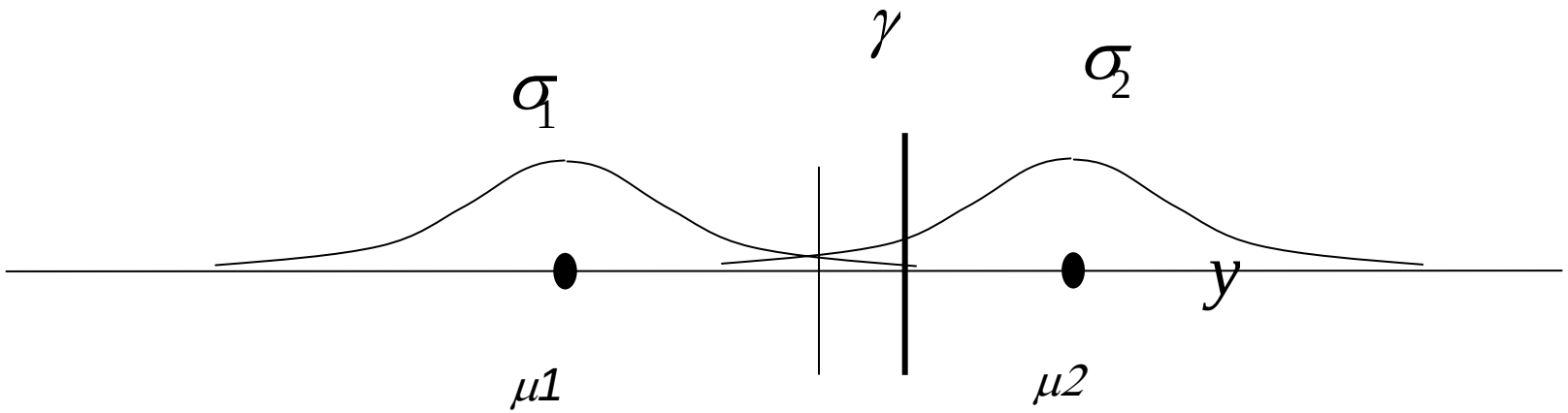
# 7 ROC: Característica de Operación del Receptor

- Caso Binario y escalar (d=1,C=2) $x|\omega_1 : N\left(\mu_1, \sigma^2\right)$
- El clasificador utiliza un umbral $\gamma$ $x|\omega_2 : N\left(\mu_2, \sigma^2\right)$
- Los experimentadores no conocen el umbral $\gamma$, ni los parámetros de la distribución, pero tienen acceso a medir las 4 probabilidades.
  - Hit $\Pr\left(x > \gamma | x \in \omega_2\right)$
  - Falsa Alarma $\Pr\left(x > \gamma | x \in \omega_1\right)$
  - Pérdida $\Pr\left(x < \gamma | x \in \omega_2\right)$
  - Rechazo Correcto $\Pr\left(x < \gamma | x \in \omega_1\right)$
- Medida de Discriminabilidad $d' = \frac{|\mu_2 - \mu_1|}{\sigma}$

# 7 ROC: Característica de Operación del Receptor

- La ROC es la representación de la probabilidad de acierto (Hit) respecto a la probabilidad de falsa alarma y en general depende de la discriminabilidad.



# 7 ROC: Característica de Operación del Receptor

- Caso Gaussiano:

$$\Pr(Hit) = P_2 \int_{\gamma}^{+\infty} \frac{1}{\sqrt{2\pi\sigma_2^{\,2}}} \exp\left(-\frac{(y-\mu_2)^2}{2\sigma_2^{\,2}}\right) dy; \quad \Pr(FA) = P_1 \int_{\gamma}^{+\infty} \frac{1}{\sqrt{2\pi\sigma_1^{\,2}}} \exp\left(-\frac{(y-\mu_1)^2}{2\sigma_1^{\,2}}\right) dy$$

$\gamma < \mu_1 < \mu_2$

$$\Pr(Hit) = P_2\left(1 - Q\left(\frac{\mu_2-\gamma}{\sigma_2}\right)\right); \quad \Pr(FA) = P_1\left(1 - Q\left(\frac{\mu_1-\gamma}{\sigma_1}\right)\right)$$

$\mu_1 < 0$

$\mu_1 < \gamma < \mu_2$

$$\Pr(Hit) = P_2\left(1 - Q\left(\frac{\mu_2-\gamma}{\sigma_2}\right)\right); \quad \Pr(FA) = P_1 Q\left(\frac{\gamma-\mu_1}{\sigma_1}\right)$$

$0 < \mu_2$

$\mu_1 < \mu_2 < \gamma$

$$\Pr(Hit) = P_2 Q\left(\frac{\gamma-\mu_2}{\sigma_2}\right); \quad \Pr(FA) = P_1 Q\left(\frac{\gamma-\mu_1}{\sigma_1}\right)$$

# 7 ROC: Característica de Operación del Receptor

- Para el caso multidimiensional para un valor dado de Probabilidad de Hit existen diferentes posibles valores de la Probabilidad de Falsa Alarma.
- Propuesta sencilla de medida de discriminabilidad.

$$D\left(c_i, c_j\right)_d = \left| \frac{(\boldsymbol{\mu}_i)_d}{(\sigma_i)_d} - \frac{(\boldsymbol{\mu}_j)_d}{(\sigma_j)_d} \right|$$

  - Distancia de Mahalanobis entre $d_M\left(\boldsymbol{\mu}_i, \boldsymbol{\mu}_j\right)$

# 8 VECTOR DE CARACTERÍSTICAS DE VALORES DISCRETOS

- Las componentes del vector x, son de valores binarios o enteros
- Caso binario de C=dos categorías y dimensión *d*
- Componentes estadísticamente independientes entre sí.

$$\mathbf{x} = \begin{pmatrix} x_1 \\ : \\ x_d \end{pmatrix}$$

$$p_i = \Pr\left(x_i = 1 \middle| \omega_1\right) = 1 - \Pr\left(x_i = 0 \middle| \omega_1\right)$$

$$q_i = \Pr\left(x_i = 1 \middle| \omega_2\right) = 1 - \Pr\left(x_i = 0 \middle| \omega_2\right)$$

$$\Pr\left(\mathbf{x} \middle| \omega_1\right) = \prod_{i=1}^{d} \left(p_i\right)^{x_i} \left(1 - p_i\right)^{1-x_i}$$

$$\Pr\left(\mathbf{x} \middle| \omega_2\right) = \prod_{i=1}^{d} \left(q_i\right)^{x_i} \left(1 - q_i\right)^{1-x_i}$$

# 8 VECTOR DE CARACTERÍSTICAS DE VALORES DISCRETOS

- Likelihood ratio

$$\frac{\Pr\left(\mathbf{x}\middle|\omega_1\right)}{\Pr\left(\mathbf{x}\middle|\omega_2\right)} = \prod_{i=1}^{d}\left(\frac{p_i}{q_i}\right)^{x_i}\left(\frac{1-p_i}{1-q_i}\right)^{1-x_i}$$

- Función discriminante LINEAL con $x_i$

$$g(x) \equiv \ln(\Pr(\omega_1|x)) - \ln(\Pr(\omega_2|x)) \begin{matrix} > \omega_1 \\ < \omega_2 \end{matrix} 0$$

$$g(\mathbf{x}) = \sum_{i=1}^{d}\left(x_i \ln\left(\frac{p_i}{q_i}\right) + (1-x_i)\ln\left(\frac{1-p_i}{1-q_i}\right)\right) + \ln\frac{\Pr\left(\omega_1\right)}{\Pr\left(\omega_2\right)}$$

# 8 VECTOR DE CARACTERÍSTICAS DE VALORES DISCRETOS

- Dado que la función discriminante para el caso de C=2 categorías y d dimensiones estadísticamente independientes resulta lineal, determine el valor del vector y del escalar que determinan dicha función:

$$g(\mathbf{x}) = \mathbf{w}^{\mathrm{T}}\mathbf{x} + w$$

# 9 CONCLUSIONES

- Interesan funciones de discriminación lineales
- Podemos encontrarnos con vectores de características híbridas en cuanto a valores continuos/valores discretos

# 9 CONCLUSIONES

- Se maximiza una función Discriminante:

$$\hat{\omega}_i = \max_i \{g_i(\mathbf{x})\}; \qquad i = 1..C$$

  - MAP (Equivalente a mínima probabilidad de error)

$$g_i(\mathbf{x}) = \Pr(\omega_i | \mathbf{x})$$

  - Mínimo Riesgo. $$g_i(\mathbf{x}) = -R(\alpha_i | \mathbf{x})$$

# 9 CONCLUSIONES

- Interesan funciones lineales con los datos:
  - C Categorías:

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

  - Regiones de decisión son hiperplanos (dimensión: *d*-1).

$$h_i(\mathbf{x}) = h_j(\mathbf{x}) \Rightarrow \mathbf{w}^{\mathrm{T}}\left(\mathbf{x} - \mathbf{x}_0\right) = 0$$