

## Sistemas de Visión Artificial

### Tema 5. Técnicas de reconocimiento (2ª parte)

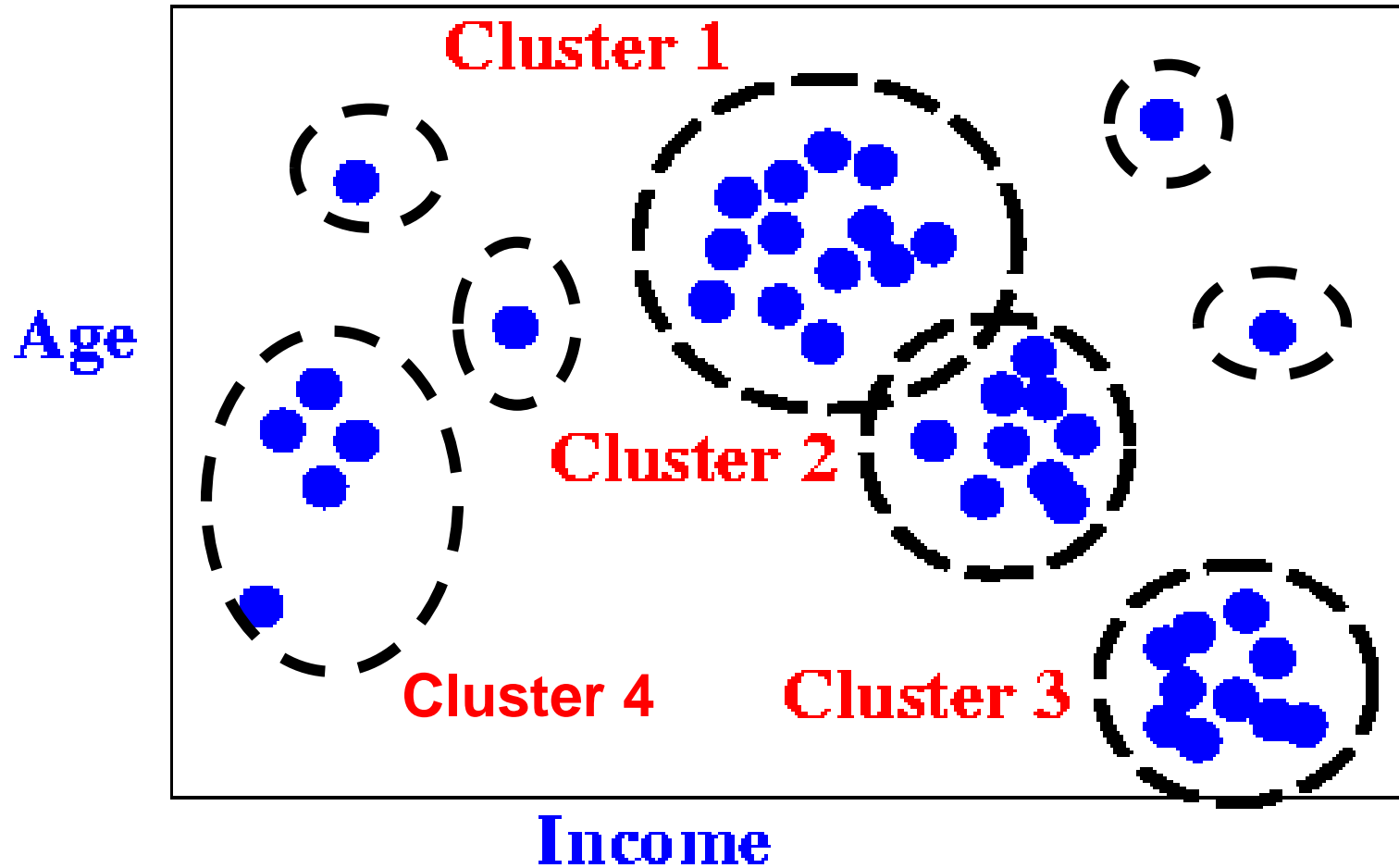
Autores: Sira Palazuelos, Luis M. Bergasa , Manuel Mazo,  
M. Ángel García, Marisol Escudero, J. Manuel Miguel  
Departamento de Electrónica. Universidad de Alcalá.



## 3. Clasificadores basados en distancias



### 3. Clasificadores basados en distancias



[academic.uprm.edu/eacuna/dm15.ppt](http://academic.uprm.edu/eacuna/dm15.ppt)

□ Hay muchas definiciones **distancia**. P.ej.:

□ **Distancia Minkowski o norma  $L_p$ .**

$$D_p(\mathbf{x}, \mathbf{y}) = \left( \sum_{i=1}^M (x_i - y_i)^p \right)^{1/p}$$

□ Casos particulares

□ **Distancia euclídea:  $p=2$**   $d_E(X, Z_i) = \sqrt{\sum_{j=1}^N (x_j - z_{ij})^2}$

□ **Distancia de Manhattan o “city block”:  $p=1$**   $d_M(x_i, x_j) = \sum_{k=1}^d |x_{i,k} - x_{j,k}|$

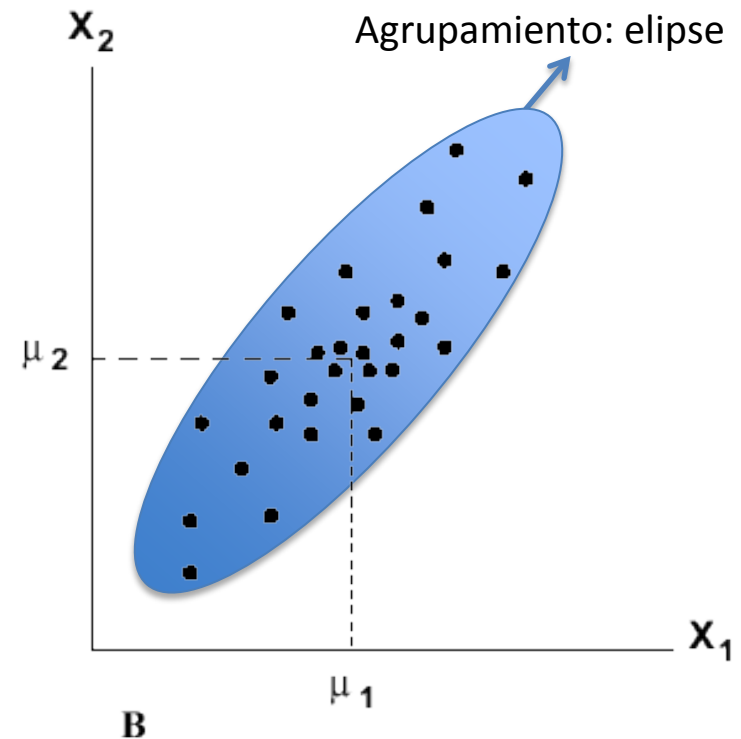
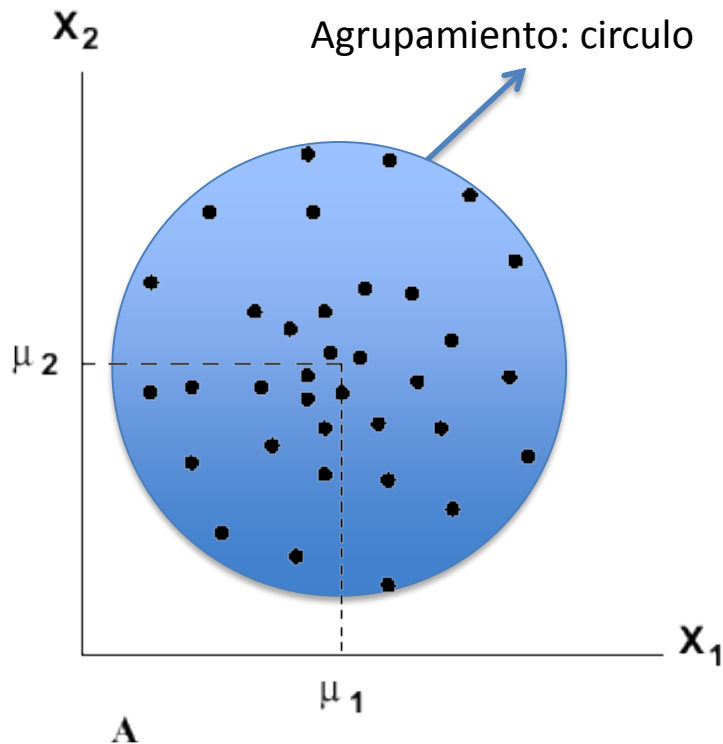
□ **Distancia Chebychev:  $p=\infty$** ,  $D_\infty = \max_{1 \leq i \leq M} |x_i - y_i|$

□ **Distancia ponderada Minkowski**  $D_p(\mathbf{x}, \mathbf{y}) = \left( \sum_{i=1}^M w_i (x_i - y_i)^p \right)^{1/p}$

Se utiliza en casos en que las distintas componentes varían en márgenes muy diferentes: tienen distinta desviación típica. P. ej., Si  $p=2$  y ponderamos dividiendo cada componente por su varianza, igualamos la capacidad de discriminación de cada una.

## □ Problema:

- A veces las variables no son independientes:



A) Variables independientes    B) Variables correladas

Adaptado de <http://www.inf.utfsm.cl/~hallende/download/Pattern/Cap-2.1-2002-1.ppt>

## □ Más distancias:

- **Distancia de Mahalanobis:** tiene en cuenta la **correlación** entre las variables aleatorias. Se utiliza cuando las variables no son independientes.

donde: 
$$d_E(X_1, X_2) = \sqrt{(X_1 - X_2)^T \Sigma^{-1} (X_1 - X_2)}$$

Media:  $\mu_i = E(X_i)$

Matriz de covarianzas:  $\Sigma_{ij} = E[(X_i - \mu_i)(X_j - \mu_j)]$

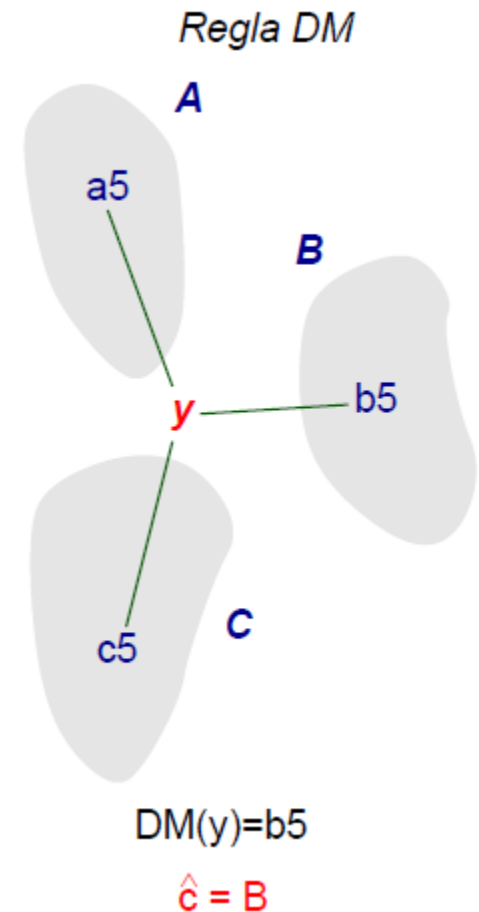
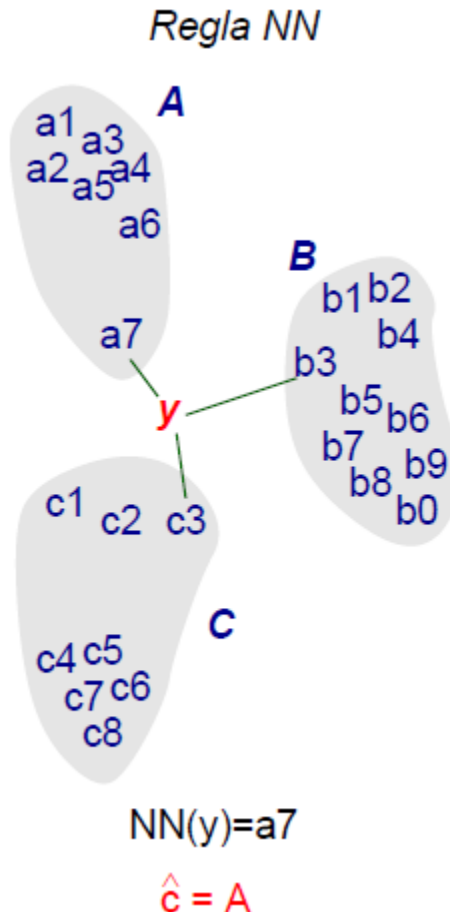
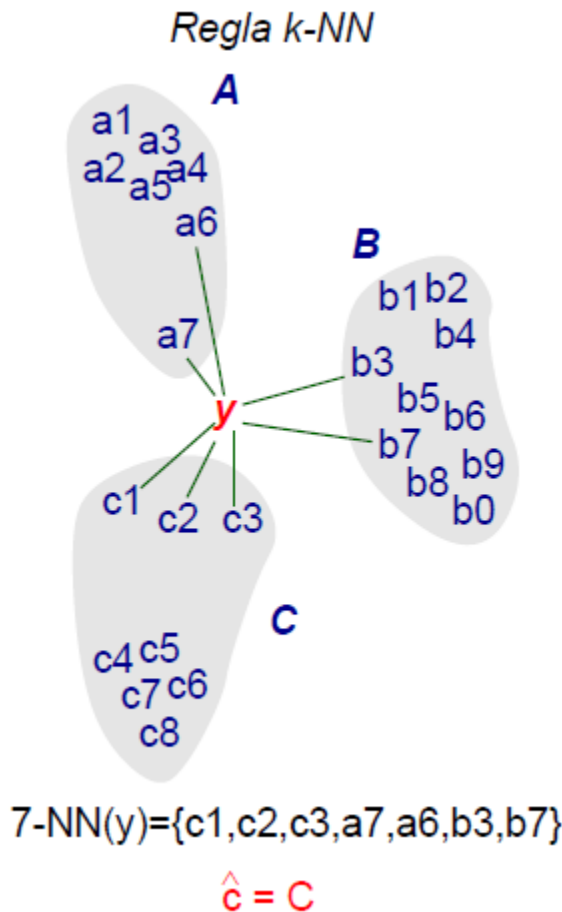
$$\Sigma = \begin{bmatrix} E[(X_1 - \mu_1)(X_1 - \mu_1)] & E[(X_1 - \mu_1)(X_2 - \mu_2)] & \cdots & E[(X_1 - \mu_1)(X_n - \mu_n)] \\ E[(X_2 - \mu_2)(X_1 - \mu_1)] & E[(X_2 - \mu_2)(X_2 - \mu_2)] & \cdots & E[(X_2 - \mu_2)(X_n - \mu_n)] \\ \vdots & \vdots & \ddots & \vdots \\ E[(X_n - \mu_n)(X_1 - \mu_1)] & E[(X_n - \mu_n)(X_2 - \mu_2)] & \cdots & E[(X_n - \mu_n)(X_n - \mu_n)] \end{bmatrix}.$$

Si las variables son independientes, la matriz será diagonal  $\rightarrow$  será una distancia con cada componente ponderada por su varianza<sup>-1</sup>. Si, además, todas las componentes tienen la misma varianza  $\rightarrow$  distancia euclídea (con un factor de escala).

- Hay **varios tipos de clasificadores** basados en distancias:
  - **Clasificador de distancia mínima (DM).**
  - **Clasificador por los k vecinos más cercanos (k-NN).**
  - **Clasificador por el vecino más cercano (NN).**
- Seleccionaremos el **más adecuado** para nuestro problema dependiendo de la forma de las clases.



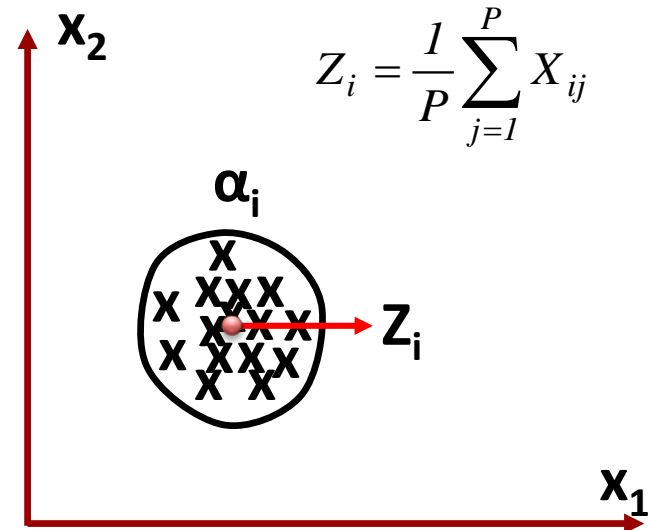
## □ Ejemplos de los tipos de clasificadores



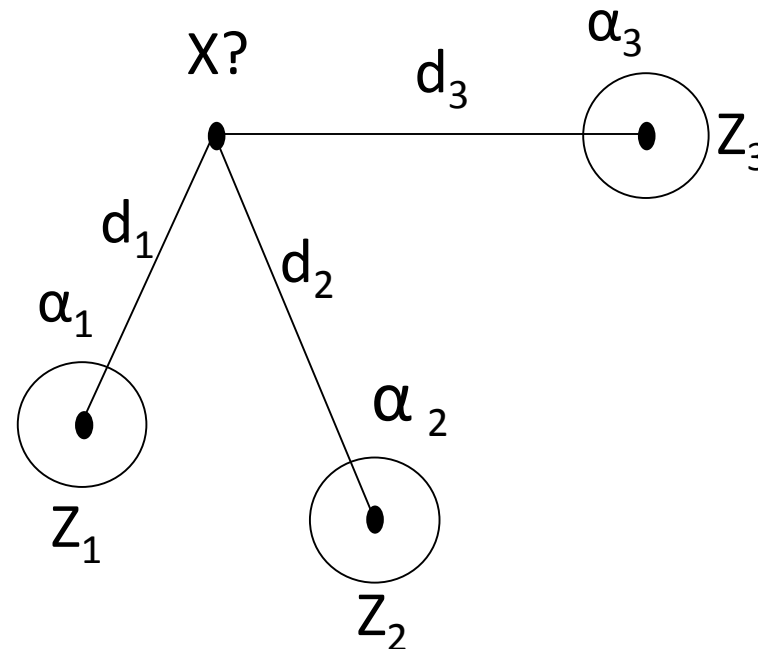
<http://web.iti.upv.es/~evidal/students/app/tema4/>

## □ Clasificador por distancia mínima

- Sea  $X$  un vector a clasificar: el **reconocedor por distancia mínima** asociará el vector  $X$  a la clase cuyo **prototipo** esté más cerca según la distancia utilizada.
- Una **clase**  $\alpha_i = (X_{i1}, X_{i2}, \dots, X_{iP})$  formada por  $P$  elementos estará representada por un único vector: **prototipo** o **centroide** ( $Z_i$ ) que será la **media** de los elementos.
- Se utiliza cuando las **desviaciones típicas** de los elementos de una clase difieren **menos del 10 %** de su media.



- Ejemplo: Supongamos que existen **M** clases ( $\alpha_1, \alpha_2, \dots, \alpha_M$ ) con sus respectivos **prototipos** ( $Z_1, Z_2, \dots, Z_M$ ), y que utilizamos la distancia euclídea (**clasificador por distancia mínima euclídea**).



- Asignaremos  $X?$  a  $\alpha_1$ , porque su centroide es el más cercano a  $X?$  según la distancia euclídea.

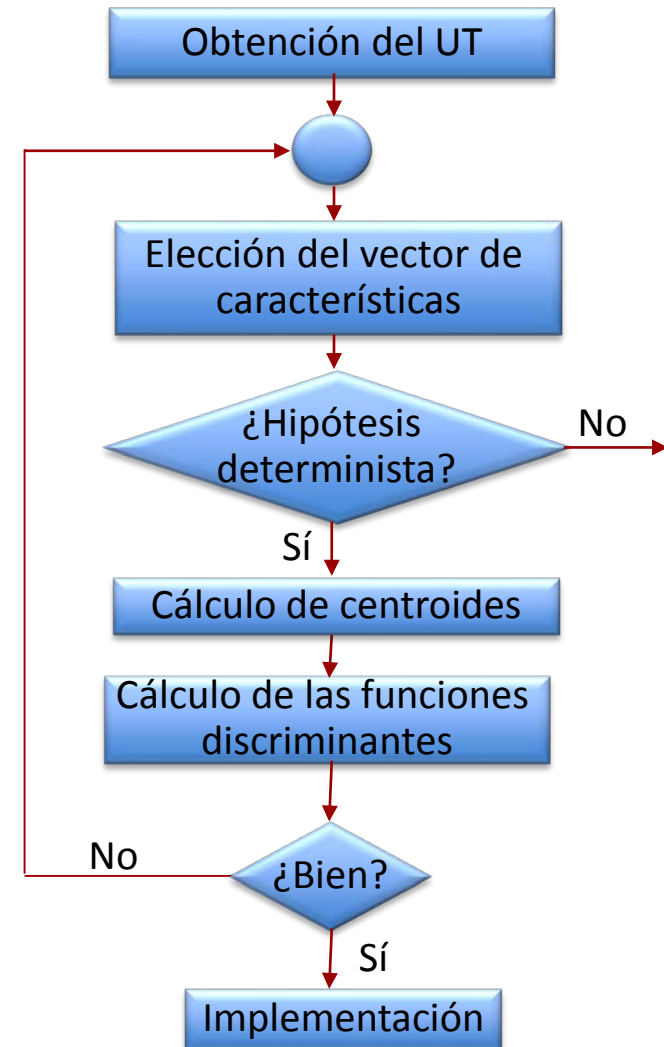
□ **Entrenamiento de un clasificador por distancia mínima:** selección de las clases → los centroides.

1. Se hace **manualmente** según las características del clasificador a diseñar.
2. **No existen reglas formales**, se trata más de un arte que de una ciencia. Existen paquetes software de ayuda (MATLAB, Tooldiag, etc).
3. Hay que estudiar las **matrices de covarianza** de cada clase.

$$C_i = E\left[\left[X - E[X]\right]^T \left[X - E[X]\right]\right] = \begin{bmatrix} C_{11} & C_{12} & \dots & C_{1N} \\ C_{21} & C_{22} & \dots & C_{2N} \\ \dots & \dots & \dots & \dots \\ C_{N1} & C_{N2} & \dots & C_{NN} \end{bmatrix}$$

$$C_{ij} = (x_i - \bar{x}_i)(x_j - \bar{x}_j) \quad E(x_i) = \bar{x}_i \quad \longrightarrow \quad \frac{\sqrt{C_{jj}}}{Z_j} \leq 0.1$$

El centroide de cada clase será:  $Z_i = \frac{1}{P} \sum_{j=1}^P X_{ij}$



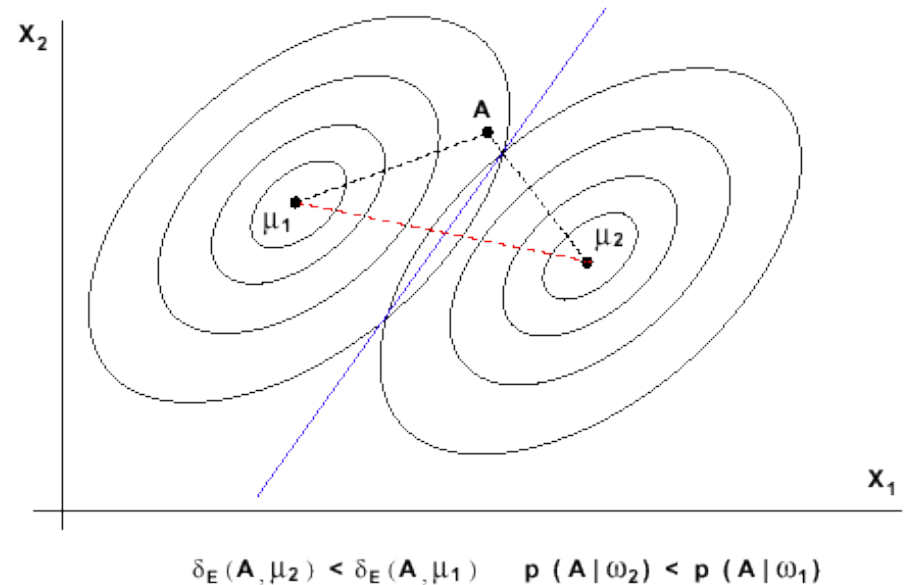
- **Entrenamiento de un clasificador por distancia mínima:**  
selección de la distancia a utilizar.
- [Suponemos que a priori todas las clases son equiprobables].
- Como tenemos las matrices de covarianza, podemos saber si las variables son estadísticamente dependientes o independientes:

## 1. Distancia euclídea:

- Variables estadísticamente independientes.
- Variables igualmente escaladas en todas las direcciones.

## 2. Distancia de Mahalanobis:

- Variables correladas.
- Variables posiblemente escaladas de forma diferente.



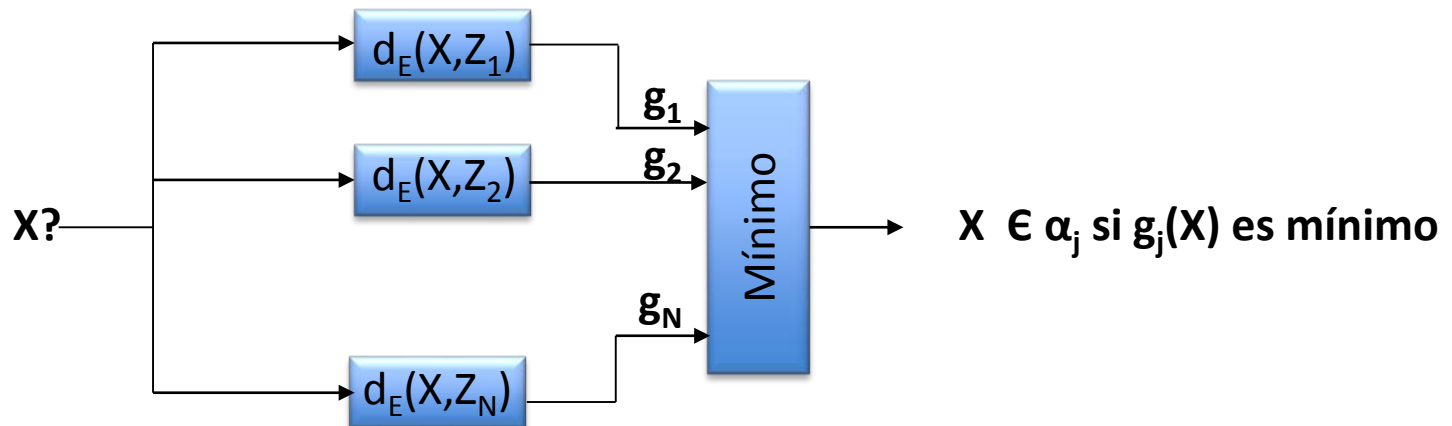
Adaptado de <http://www.inf.utfsm.cl/~hallende/download/Pattern/Cap-2.1-2002-1.ppt>

## □ Clasificador de distancia mínima euclídea. Etapa de clasificación:

1. Calcular la fd para cada **prototipo**:

$$g_i(X) = d_E(X, Z_i)$$

2. Asignar X a la clase que obtenga una **mínima distancia**:



□ El número de fds es siempre **fijo e igual al número de clases**.

## □ Posibles **optimizaciones** para acelerar los cálculos:

- La decisión final es la misma con la distancia que con la **distancia al cuadrado** → utilizando  $D^2$  eliminamos la necesidad de hacer la raíz.

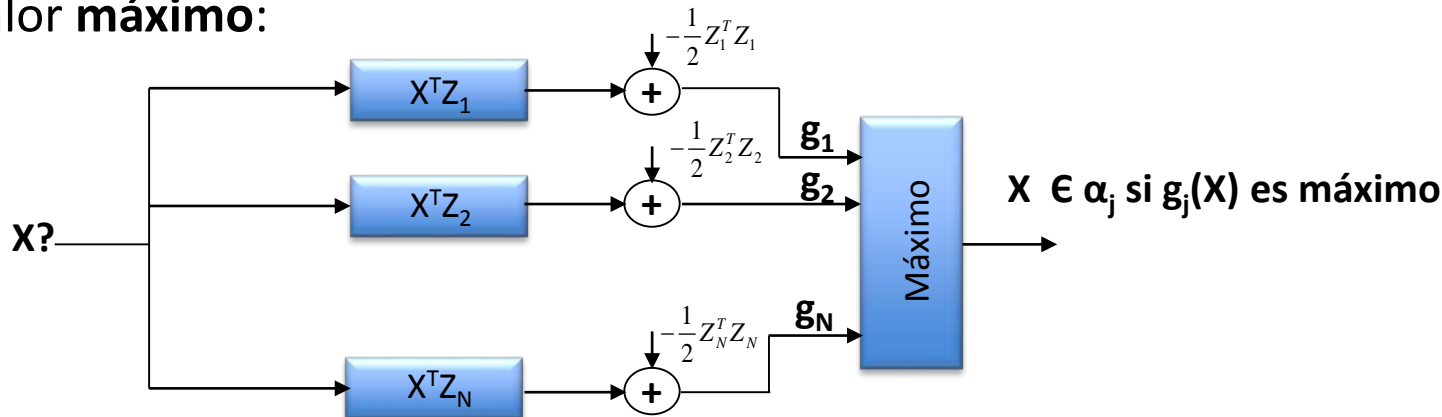
$$d_E^2(X, Z_i) = (X - Z_i)^T (X - Z_i) = X^T X - 2X^T Z_i + Z_i^T Z_i$$

- Las fds **no son lineales**. Dado que  $X^T X$  es un término no discriminante se puede **eliminar, linealizando** la función. También cambiaremos el signo y dividiremos por dos para simplificar el primer producto:

$$g_i(X) = X^T Z_i - \frac{1}{2} Z_i^T Z_i$$

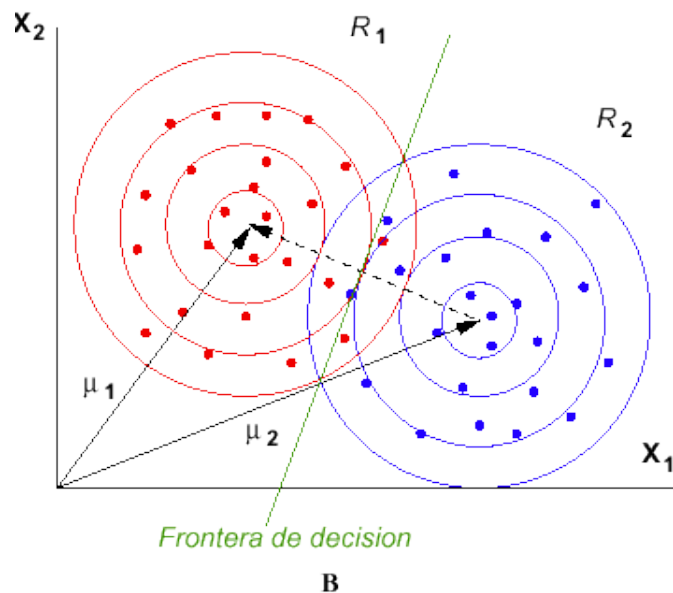
Estos términos se precaculan para acelerar el proceso

- Por el cambio de signo, ahora hay que buscar la fd que proporcione el valor **máximo**:

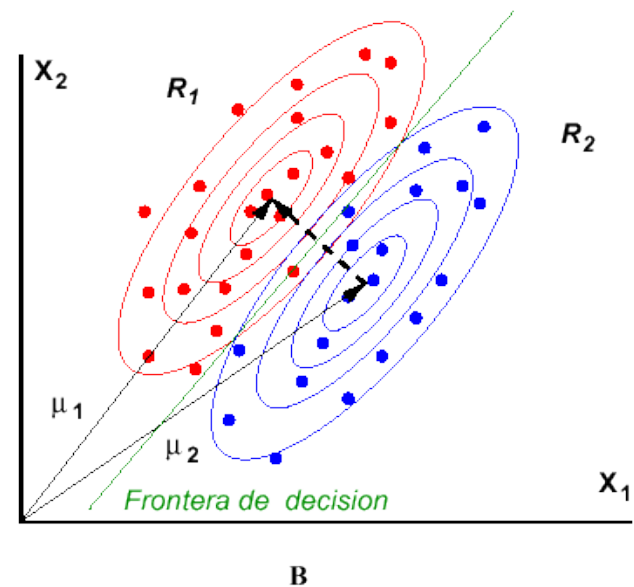


## □ Frontera de decisión

- Frontera de decisión asociada al clasificador de distancia mínima: lineal, hiperplano.



Distancia euclídea, cuando las variables son estadísticamente independientes y tienen la misma varianza.

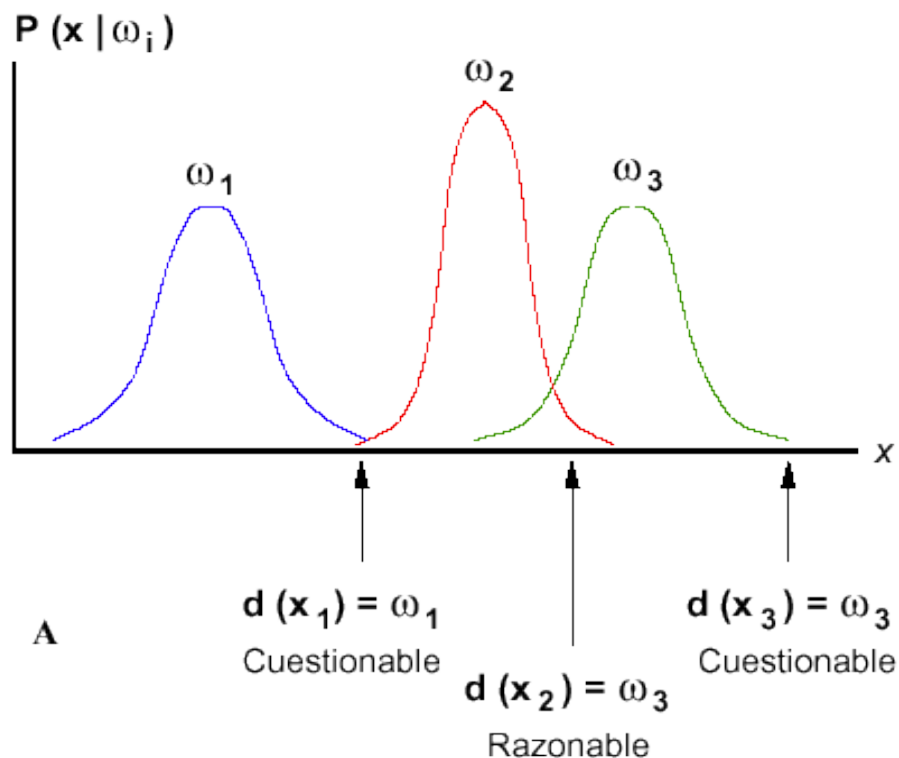


Distancia de Mahalanobis, cuando las variables no son estadísticamente independientes y las varianzas individuales son diferentes.



## □ Detección de puntos dudosos

- En el modelo actual, un punto se asigna a la clase cuyo centroide esté más cerca, pero a veces no es razonable → hay patrones que deben descartarse.
- Creamos la clase  $w_0$  y se los asignamos a ella.



Gráfica adaptada de <http://www.inf.utfsm.cl/~hallende/download/Pattern/Cap-2.1-2002-1.ppt>

## □ Detección de puntos dudosos

- Definimos un umbral tal que si la probabilidad de  $X?$  de pertenecer a la clase más cercana está por debajo de dicho umbral, el dato se rechaza.

1. Suponemos que  $X?$  ha sido asignado a la clase  $w_c$  (porque su centroide es el más cercano).
2. Conocemos la media y varianza de la gaussiana que modela  $w_c$ : calculamos  $P(x|w_c)$ .
3. Para el umbral de rechazo definido (p.ej. que asegure una prob. del 95%):

$$clase(X?) \begin{cases} w_c & \text{si } P(x|w_c) > T \\ w_0 & \text{si } P(x|w_c) \leq T \end{cases}$$

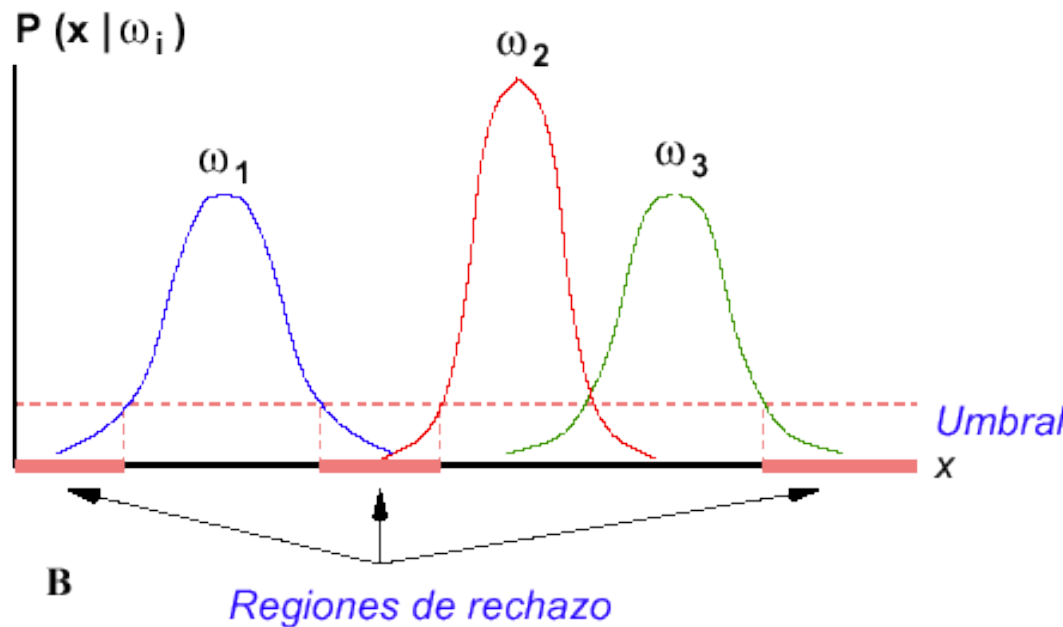


Gráfico adaptado de <http://www.inf.utfsm.cl/~hallende/download/Pattern/Cap-2.1-2002-1.ppt>

## □ Clasificador por los K vecinos más cercanos

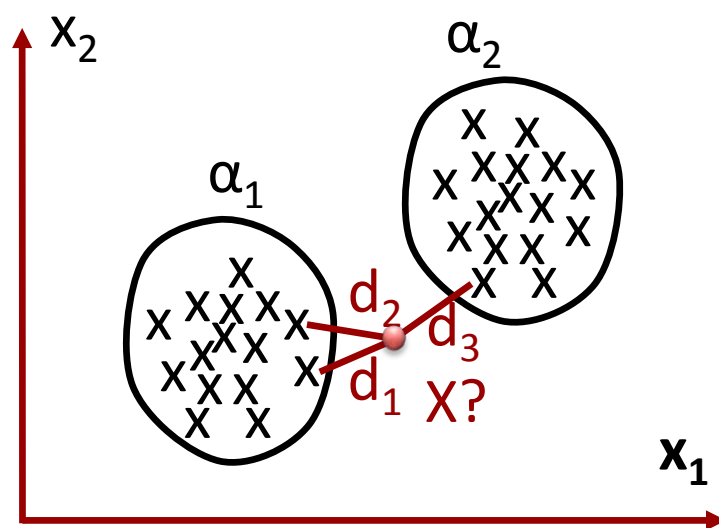
- Se calcula la distancia de  $X?$  a los  $K$  vectores más cercanos y se **clasifica  $X?$  como perteneciente a la clase más representada entre los  $K$  vecinos.**
- Cuanto mayor sea  $K$ , menor será la tasa de error, aunque el tiempo de cálculo será mayor.

$$d_i(X, X_i) = \|X - X_i\| \quad 0 < i < K$$

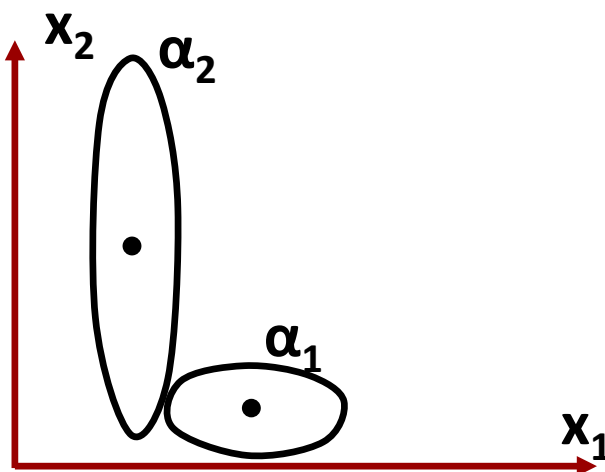
$K = n^\circ$  de puntos que se evalúan

$$d_{ordenada} = (do_1, do_2, \dots, do_K)$$

$$X? \in \alpha_i \text{ si } \max\_veces(\alpha(do_1, do_2, \dots, do_K)) = i$$



- El uso del clasificador por K vecinos más cercanos está justificado cuando los **centroides** de las clases **no son representativos** de todos los elementos de las clases. Por ejemplo cuando la desviación típica de una de las características es mucho mayor que la de las otras.
- En caso de **empate** X? se asignará, entre las empatadas, a la clase que tenga mayor número de prototipos.



## □ Regla k-NN con rechazo (I)

- Objetivo: aumentar la **efectividad** de la regla k-NN, **descartando la clasificación de las muestras “dudosas”**, es decir, la de aquellas que se encuentran próximas a las fronteras de decisión, para las que no se tiene garantía de clasificación correcta.
- Se añade una **clase artificial**  $\Theta_0$  a la que se asignarán los vectores rechazados.
- El problema pasa a tener **(C + 1) clases**.
- En general se fija un **umbral**: si al hacer la comparación con los K vecinos más cercanos no pertenecen a la misma clase al menos un número de ellos superior a umbral, la muestra es asignada a  $\Theta_0$ .

## □ Regla k-NN con rechazo (II)

- **Regla k-NN con rechazo I:** se utiliza un umbral común para todas las clases:  $L$  entero positivo, tal que  $\lceil K/2 \rceil < L \leq K$ . Si de los  $K$  vecinos más próximos a  $X$ ,  **$L$  o más pertenecen a la misma clase**, clasificaremos  $X$  como perteneciente a esa clase. Si no, se rechazará y se asignará a  $\Theta_0$ .
- **Regla k-NN con rechazo II:** se establece un **umbral  $L_i$  diferente para cada clase**.
- **Regla k-NN con rechazo III:** se establece un umbral absoluto  $MP_1$ . Si ninguna de las clases votadas **supera en  $M$  votos a las demás**, la muestra será rechazada. Requiere un  $K$  grande pues, de lo contrario, pocas muestras superarían dicho umbral.

## □ Clasificador por el vecino más próximo:

- Clasificador por los K vecinos más cercanos particularizado para  $L=1$ .
- $X?$  se asigna a la clase a la que pertenezca su vecino más próximo.
- La principal ventaja respecto a la regla por distancia mínima radica en que aquí se clasifica la muestra  $X?$  a partir de **todos** los prototipos del conjunto  $T$ .
- La **efectividad** de esta regla se verá fuertemente condicionada por la disponibilidad de un número **suficientemente grande de prototipos** y que éstos hayan sido **correctamente etiquetados**.

- **Fronteras de decisión asociadas a los clasificadores NN y k-NN:**  
Funciones discriminantes lineales a trozos (LT):
  - Existen fronteras de decisión LT que no pueden obtenerse mediante ningún clasificador NN.
  - Toda frontera de decisión LT puede obtenerse mediante FDs k-NN.

