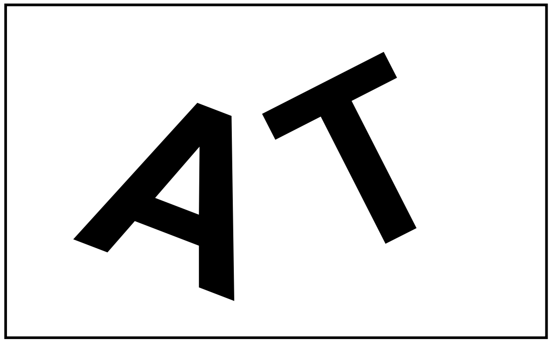
Lección 4: Reconocimiento basado en descriptores

1. Introducción
2. Clasificación Bayesiana
3. Clasificación por distancias
4. Árboles de decisión
5. Otros métodos
6. Conclusiones

1. Introducción

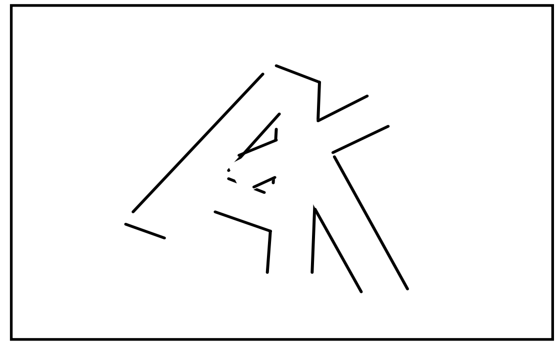
- **Reconocimiento 2D:** etiquetado de regiones correspondientes a objetos 2D en la imagen.

- **Reconocimiento basado en descriptores**: propiedades globales invariantes, objetos completamente visibles



  » Área
  » Perímetro
  » No. de Euler
  » Elongación
  » ...
- Más sencillo, menos robusto

- **Reconocimiento Geométrico:** propiedades locales

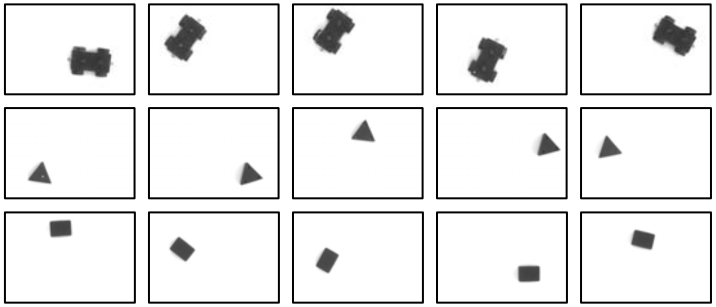


  » Vértices
  » Aristas
  » Círculos
  » Esquinas
  » ...
- Más complejo, más robusto

Esquema general

Aprendizaje:

- Tomar varias imágenes de cada objeto, y calcular sus descriptores



Explotación:

- Calcular de los descriptores de cada objeto en la imagen
- Comparar con los valores estimados de los descriptores de los objetos conocidos

Diferentes métodos usan diferentes técnicas para efectuar esta comparación

Notación:

- *n* objetos conocidos (clases):

$$\omega = \{\omega_1, \cdots, \omega_i, \cdots, \omega_n\}$$

- *m* descriptores utilizados:

$$\mathbf{x} = \left(x_1, \cdots, x_j, \cdots, x_m\right)^T$$

- *N* muestras del descriptor *j* de la clase *i*:

$$\mathbf{x}_{ij} = \left(x_{ij1}, \cdots, x_{ijk}, \cdots, x_{ijN}\right)^T$$

Dada una muestra $\mathbf{x}$, ¿a qué clase ω_i corresponde?

¿se trata de un objeto desconocido?

2. Clasificación Bayesiana

- Los descriptores se consideran variables aleatorias.
- Modelo probabilístico de cada clase:

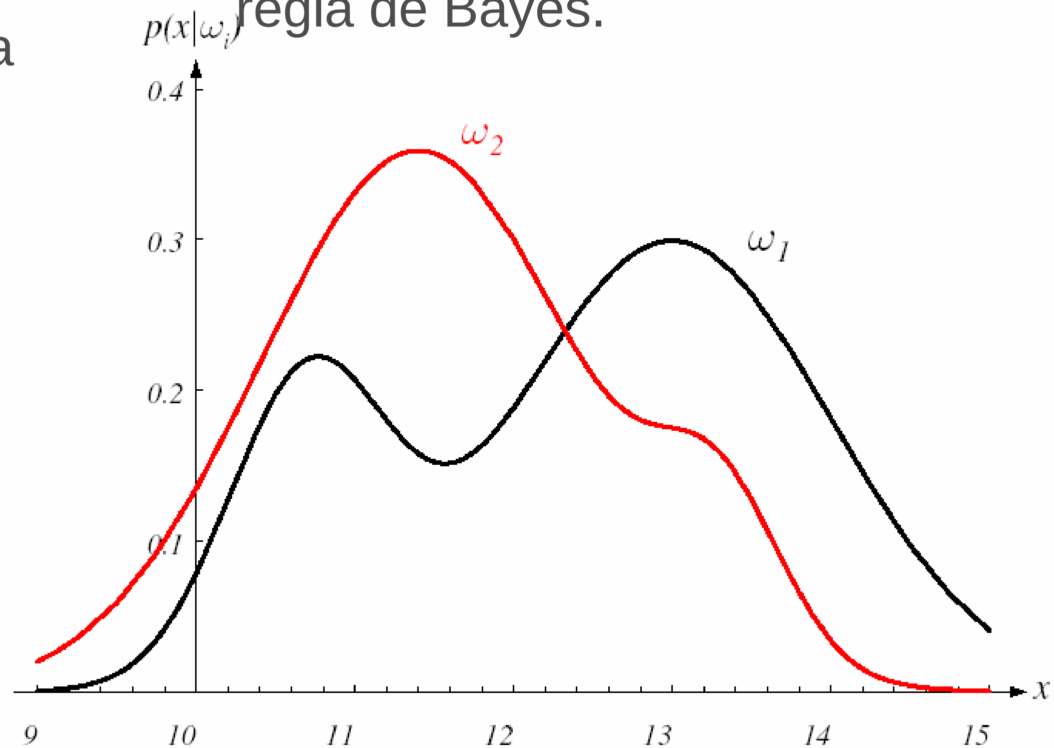
$$p(\mathbf{x}/\omega_i)$$

- Probabilidad a priori de cada clase:

$$P(\omega_i)$$

$$\sum_{i=1}^{n} P(\omega_i) = 1$$

- Se calcula la probabilidad de que la muestra provenga de la observación de cada clase utilizando la regla de Bayes.



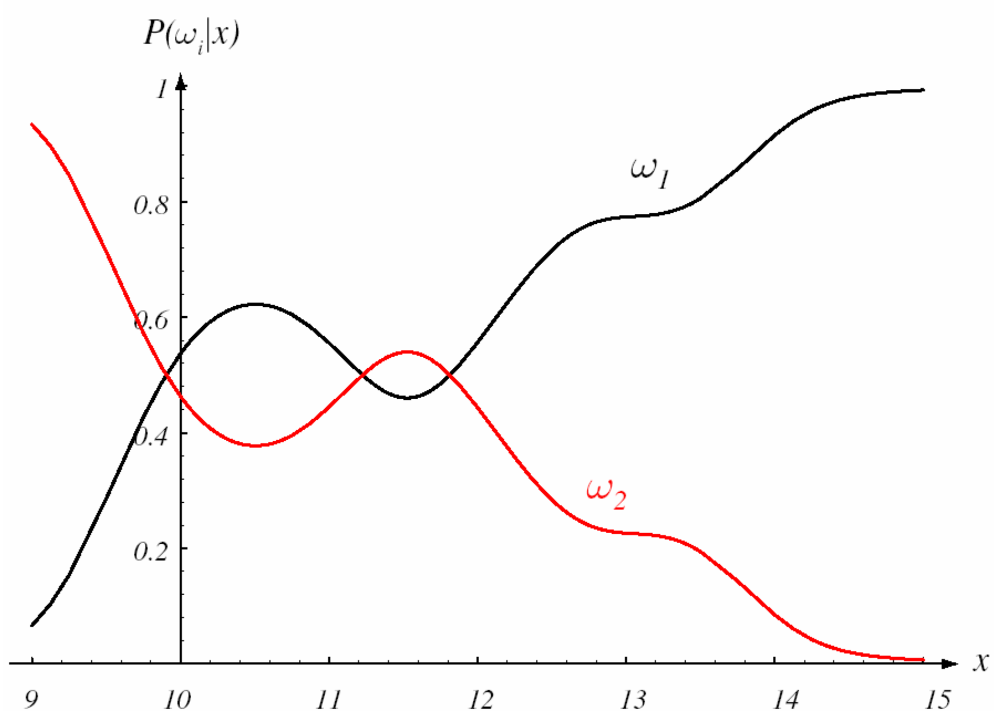
$$p(\mathbf{x}/\omega_i) = p(\mathbf{x}/\omega_j)?$$
$$P(\omega_i) = P(\omega_j)?$$

Clasificación Bayesiana

- **Regla de Bayes:**

$$P(\omega_i/\mathbf{x}) = \frac{p(\mathbf{x}/\omega_i)P(\omega_i)}{p(\mathbf{x})}$$

$$p(\mathbf{x}) = \sum_{j=1}^{n} p(\mathbf{x}/\omega_j)P(\omega_j)$$



- Distribuye la probabilidad entre todas las clases:

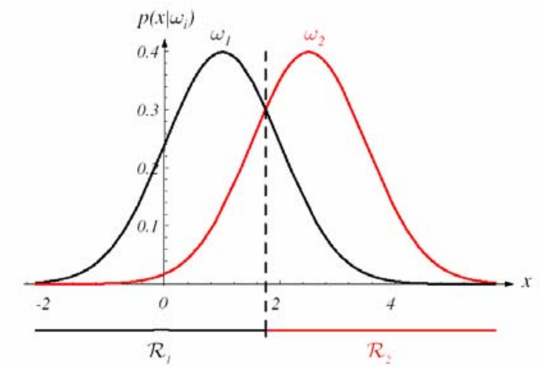
$$\sum_{i=1}^{n} P(\omega_i/\mathbf{x}) = 1$$

- Decidimos ω_i si:

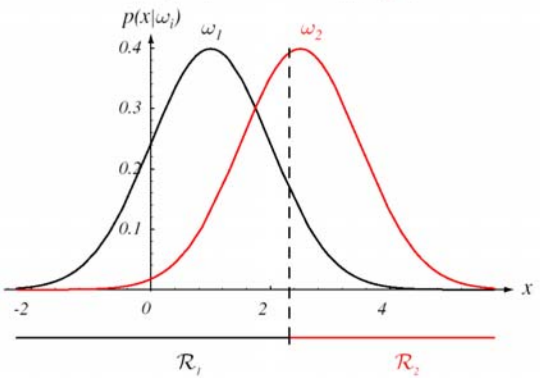
$$P(\omega_i/\mathbf{x}) = \max_{j=1}^{n} P(\omega_j/\mathbf{x})$$

Clasificación Bayesiana

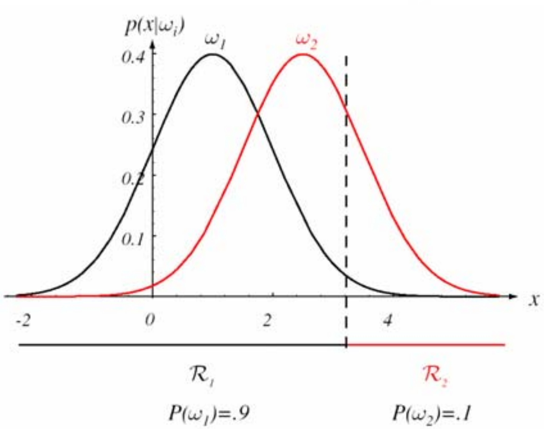
- Influencia de $P(\omega_i)$:



$$P(\omega_1) = 0.5$$
$$P(\omega_2) = 0.5$$

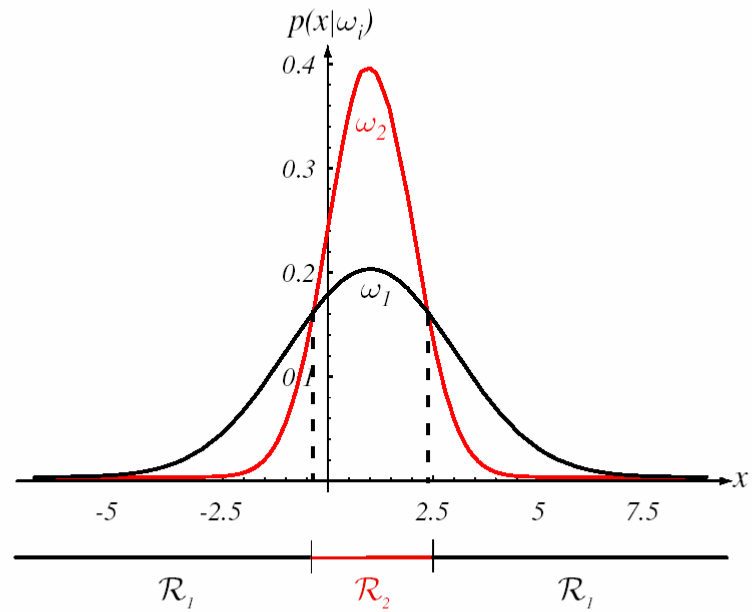


$$P(\omega_1) = 0.7$$
$$P(\omega_2) = 0.3$$



$$P(\omega_1) = 0.9$$
$$P(\omega_2) = 0.1$$

- Influencia de $p(x/\omega_i)$:

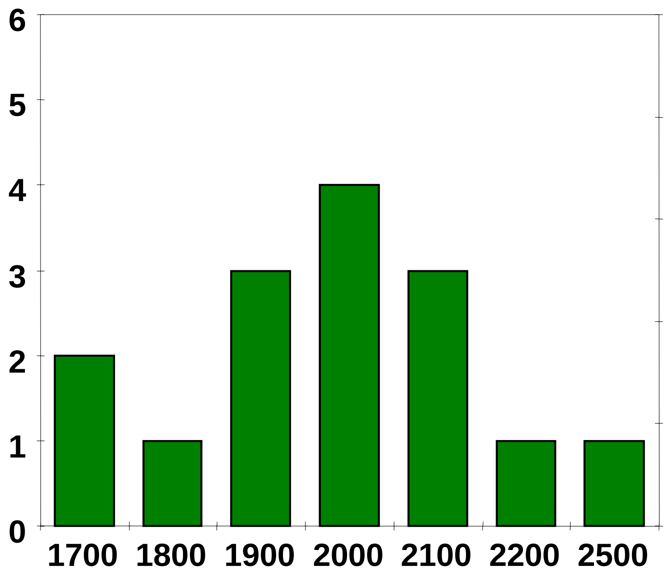


Descriptores

- Ejemplo: el área

N=15

1727
1683
1766
1924
1886
1914
2038
1965
1967
1999
2139
2089
2146
2158
2460



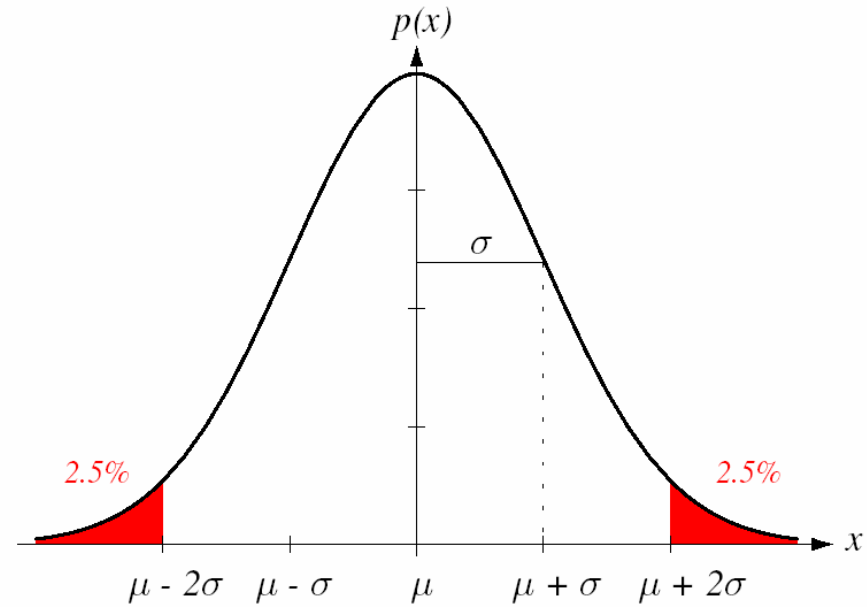
$$\hat{\mu} = 1990.7$$

$$\hat{\sigma} = 197.04$$

- En ausencia de más información, caracterizamos un descriptor por medio de una *distribución Gaussiana:*

$$x \sim N(\mu, \sigma^2)$$

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



$$Pr\{\mid x - \mu \mid \leq 2\sigma\} \simeq 0.95$$

Descriptores

- **Media muestral:** valor representativo:

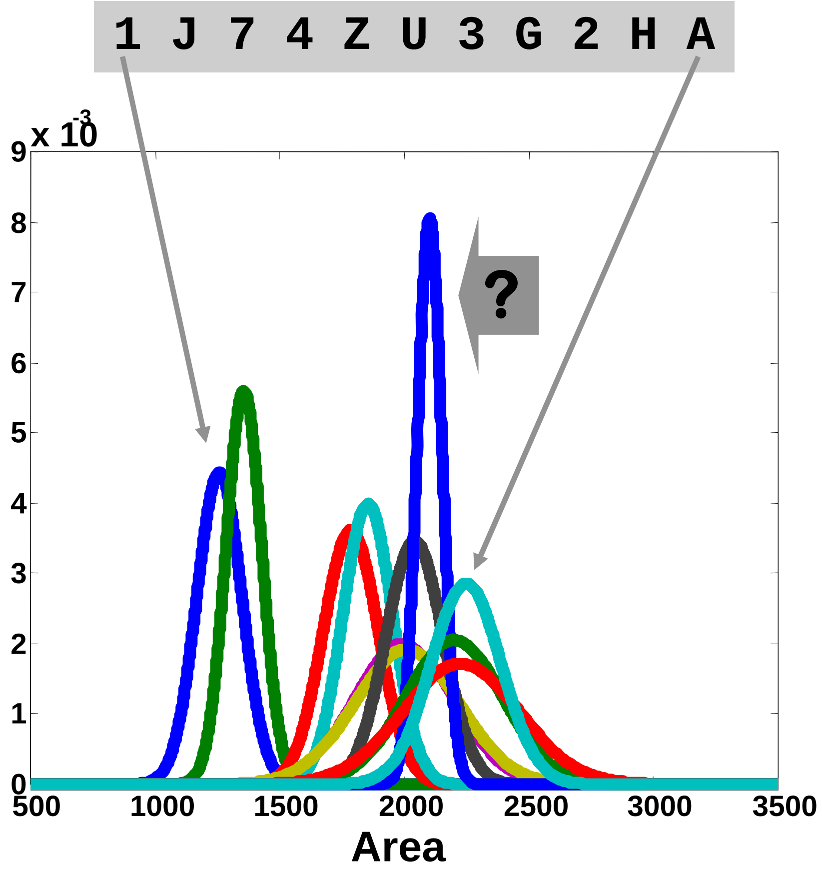
$$\hat{\mu}_{ij} = \frac{\sum_{k=1}^{N} x_{ijk}}{N}$$

$$x_{ij} \sim N(\hat{\mu}_{ij}, \hat{\sigma}_{ij}^2)$$

- **Varianza muestral:** dispersión o variabilidad

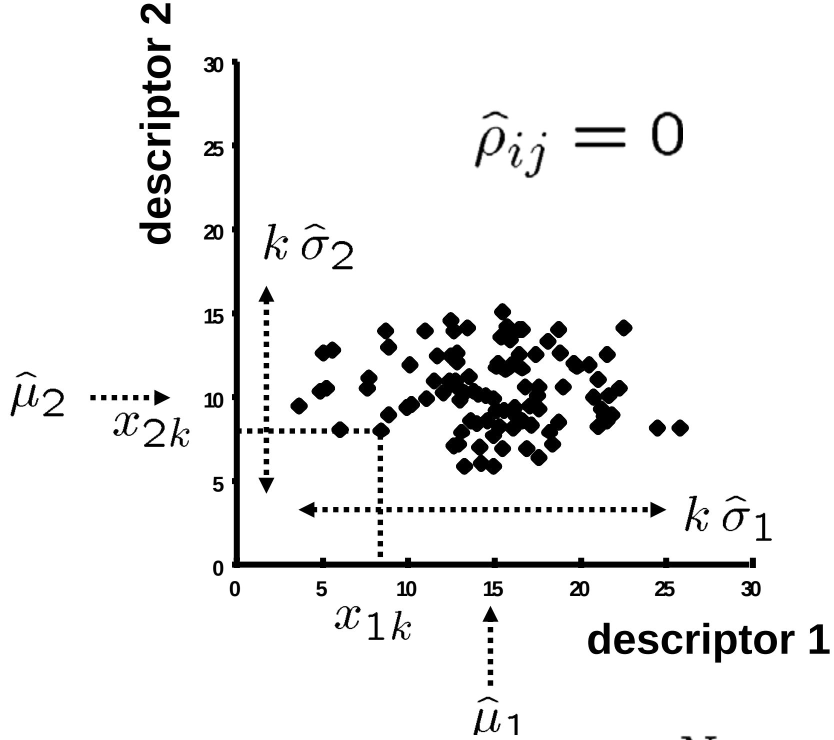
$$\hat{\sigma}_{ij}^2 = \frac{\sum_{k=1}^{N} (x_{ijk} - \hat{\mu}_{ij})^2}{N-1} = \frac{\sum_{k=1}^{N} x_{ijk}^2 - N\hat{\mu}_{ij}^2}{N-1}$$

?

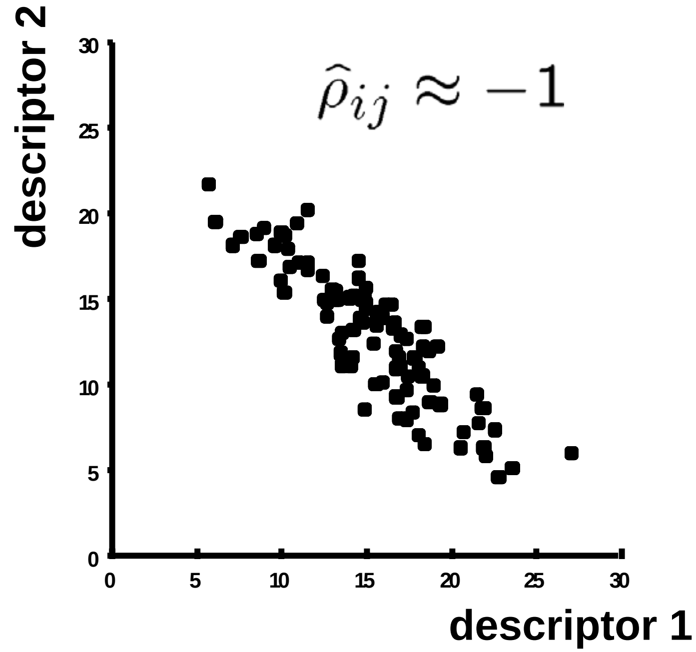


Independencia de los descriptores

- Independientes:



- Linealmente correlados



$$\hat{\rho}_{ij} = \frac{\sum_{k=1}^{N}(x_{ik}-\hat{\mu}_i)(x_{jk}-\hat{\mu}_j)}{\sqrt{\sum_{k=1}^{N}(x_{ik}-\hat{\mu}_i)^2}\sqrt{\sum_{k=1}^{N}(x_{jk}-\hat{\mu}_j)^2}}$$

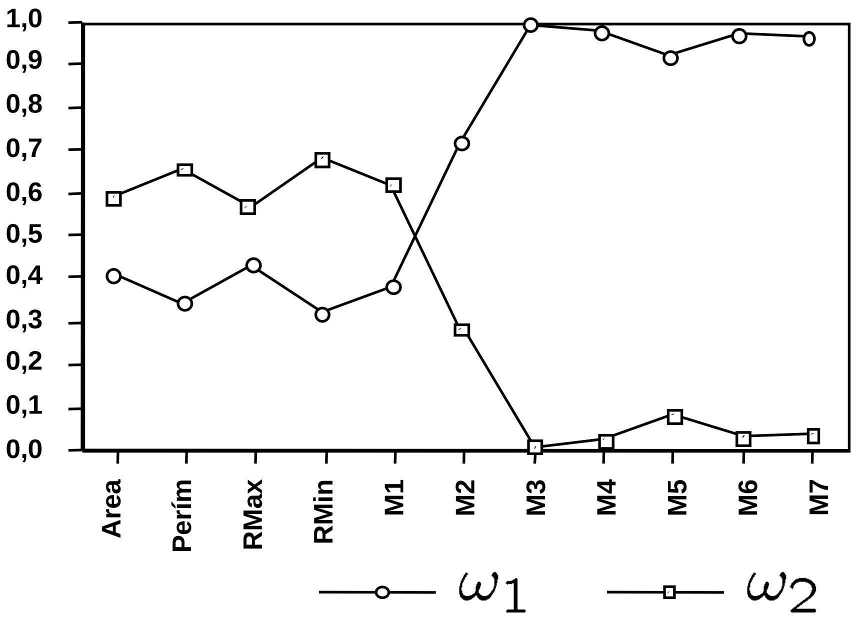
Los descriptores independientes
permiten discriminar mejor entre clases.

Clasificación Bayesiana

- Suponiendo Gaussianidad e independencia estadística:

$$p(\mathbf{x}/\omega_i) = p(x_1/\omega_i)p(x_2/\omega_i)\cdots p(x_m/\omega_i)$$

$$p(x_j/\omega_i) = \frac{1}{\sqrt{2\pi}\sigma_{ij}} e^{-\frac{1}{2}\left(\frac{x_j-\mu_{ij}}{\sigma_{ij}}\right)^2}$$

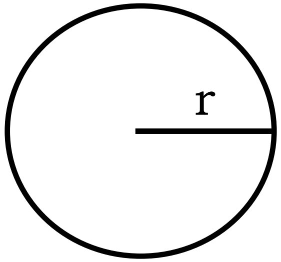


- Método iterativo:

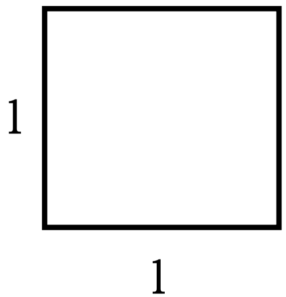
$$\begin{aligned} P^0(\omega_i) &= P(\omega_i) \\ P^d(\omega_i) &= P(\omega_i/x_1,\cdots,x_d) \\ &= \frac{p(x_d/\omega_i)P^{d-1}(\omega_i)}{\sum_{j=1}^m p(x_d/\omega_j)P^{d-1}(\omega_j)} \end{aligned}$$

- Se itera hasta que el objeto más probable supere un umbral (puede no ser necesario calcular todos los descriptores).

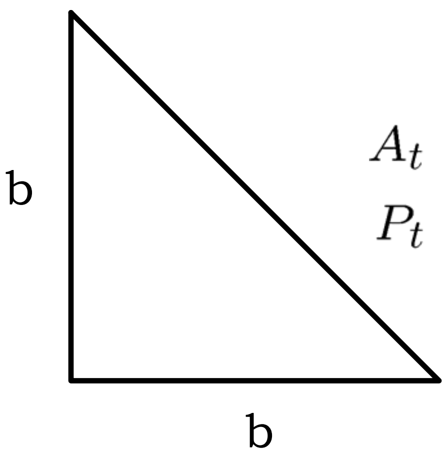
Medición secuencial de descriptores



$$\begin{aligned} A_c &= \pi r^2 \\ P_c &= 2\pi r \end{aligned}$$



$$\begin{aligned} A_d &= l^2 \\ P_d &= 4l \end{aligned}$$



$$\begin{aligned} A_t &= b^2/2 \\ P_t &= b(2+\sqrt{2}) \end{aligned}$$

- Suponiendo que:

$$l = 1, r = \frac{1}{\sqrt{\pi}}, b = \sqrt{2}$$

- El área no es un parámetro discriminante en este caso:

$$A_c = A_d = A_t$$

- El perímetro si lo es:

$$\begin{aligned} P_c &= 2\sqrt{\pi} = 1.772453851 \\ P_d &= 4 \\ P_t &= 4.828427123 \end{aligned}$$

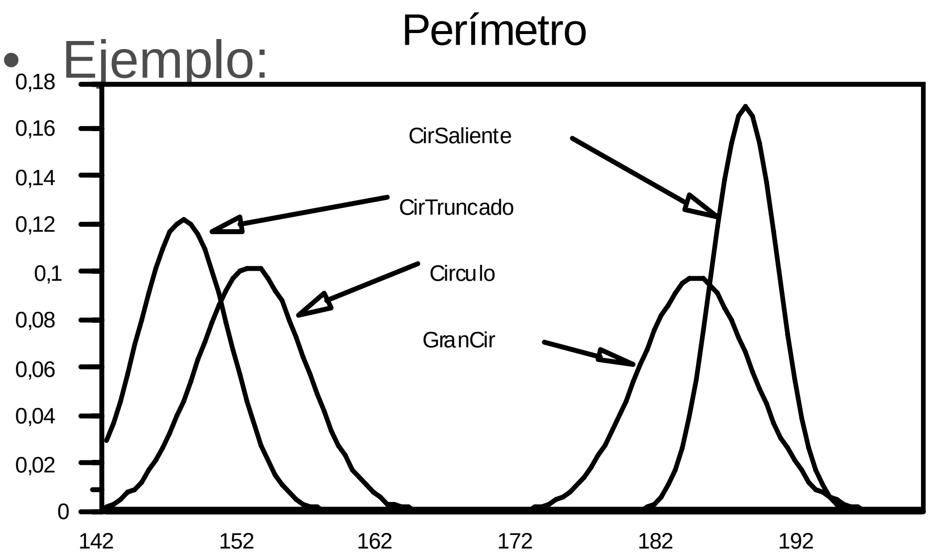
¿En qué orden calcular los descriptores?

Medición secuencial de descriptores

- Poder discriminante del parámetro *k* entre las clases ω_i y ω_j:

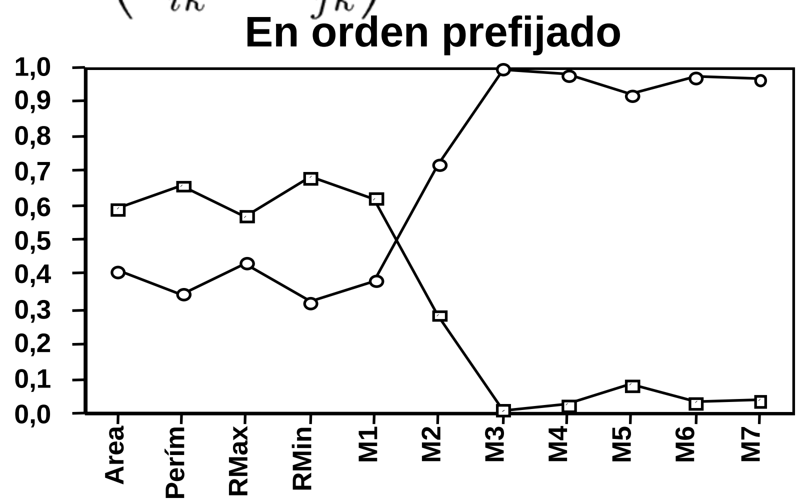
$$J_{ij}^{k} = \frac{1}{2}\left(\frac{1}{\sigma_{ik}^2} + \frac{1}{\sigma_{jk}^2}\right)\left(\mu_{ik} - \mu_{jk}\right)^2 + \frac{1}{2}\left(\frac{1}{\sigma_{ik}^2} - \frac{1}{\sigma_{jk}^2}\right)\left(\sigma_{ik}^2 - \sigma_{jk}^2\right)$$

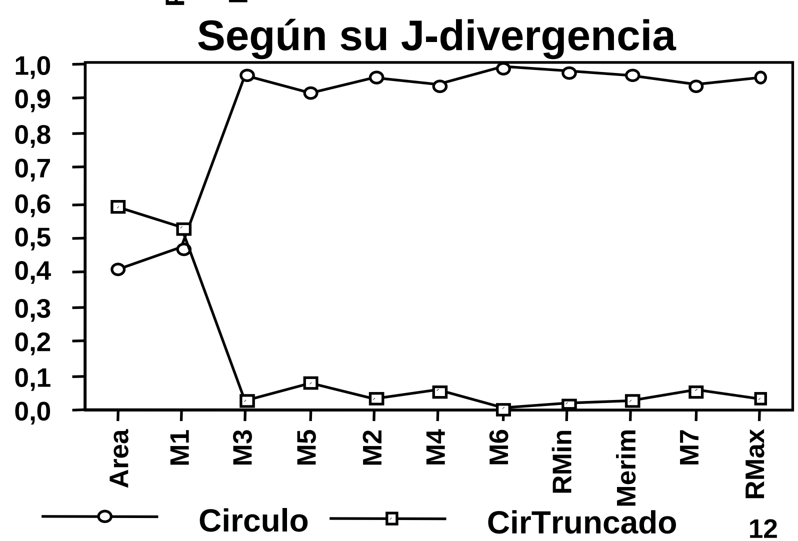
- Ejemplo:



J-Divergencia

	Círculo	GranCir	CirTruncad
GranCir	63.47		
CirTruncad	1.94	102.64	
CirSaliente	153.58	2.20	220.13





Tratamiento de objetos desconocidos

- Clase ω_0 que representa al objeto desconocido:

$$p(x_j/\omega_0) = U(x_j^{min}, x_j^{max})$$

- Distribución uniforme entre los valores mínimo y máximo del descriptor:

$$p(x_j/\omega_0) = \begin{cases} 0, & x_j < x_j^{min} \\ \frac{1}{x_j^{max} - x_j^{min}}, & x_j^{min} \leq x_j \leq x_j^{max} \\ 0, & x_j > x_j^{max} \end{cases}$$

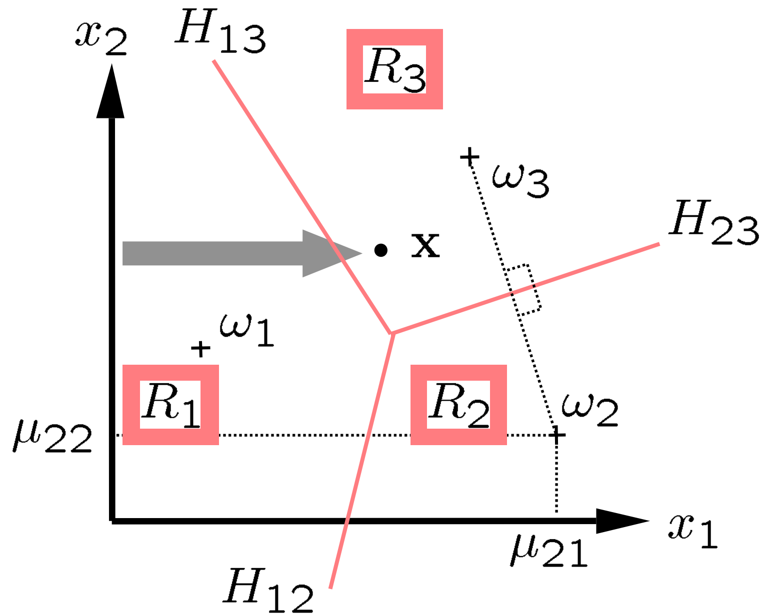
3. Distancias mínimas

- Se calcula la ***distancia*** del vector de descriptores a cada uno de los objetos conocidos.
- **Distancia Euclidiana:**

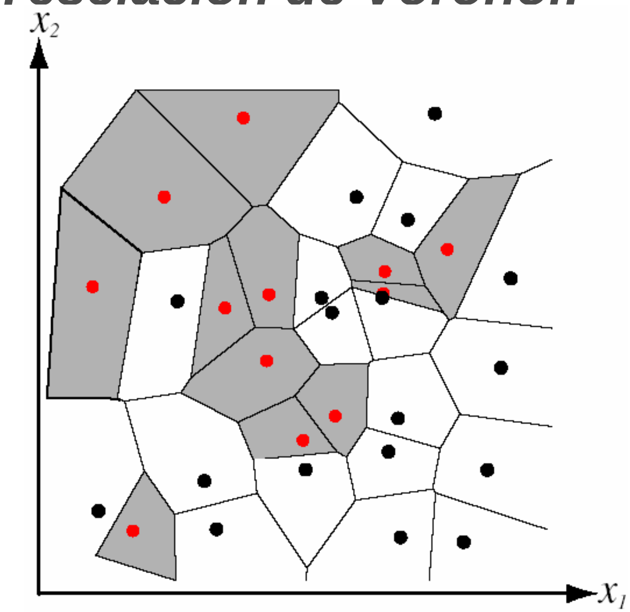
$$
\begin{aligned}
D^2(\mathbf{x}, \omega_i) &= \|\mathbf{x} - \omega_i\|^2 \\
&= \sum_{j=1}^{m} (x_j - \mu_{ij})^2
\end{aligned}
$$

- Se escoge la clase con distancia mínima (vecino más próximo).
- *L*a frontera de decisión entre dos clases estará dada por:

$$H_{ij} : D^2(\mathbf{x}, \omega_i) = D^2(\mathbf{x}, \omega_j)$$



- ***Teselación de Voronoi:***



Ejercicio

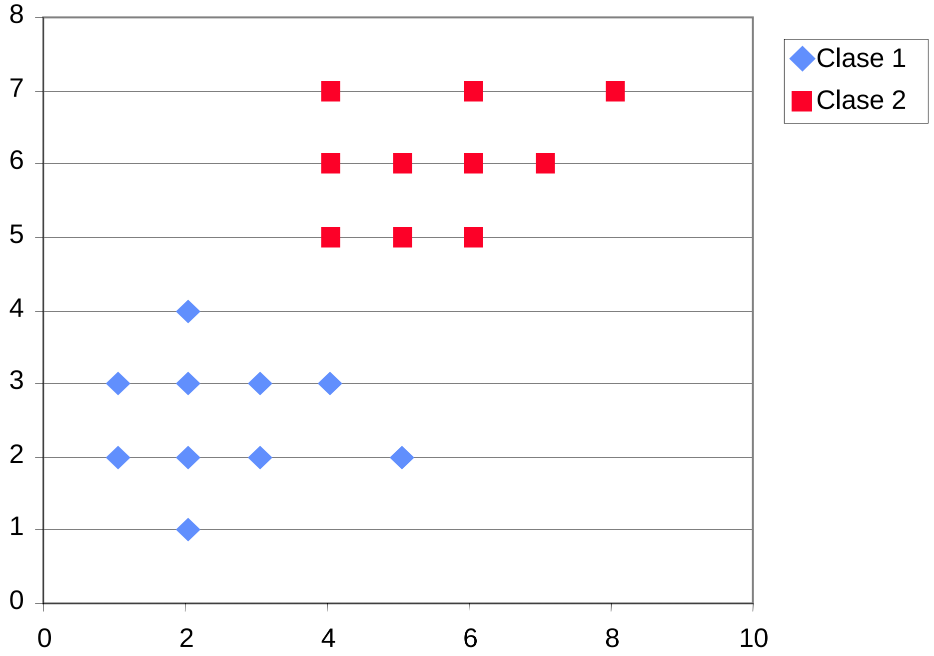
- Dadas las siguientes muestras de los descriptores $\mathbf{x_1}$ y $\mathbf{x_2}$ de dos clases:

Clase 1

x1	*x2*
1	3
2	1
2	2
2	3
2	4
3	2
3	3
4	3
5	2
1	2

Clase 2

x1	*x2*
4	5
5	5
5	6
4	7
6	5
6	6
6	7
7	6
4	6
8	7



- Determinar la ecuación de la curva que representa la frontera de decisión entre las dos clases de acuerdo con la distancia Euclidiana, y graficarla.

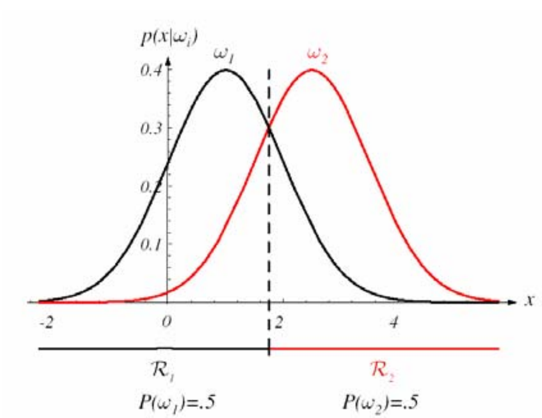
Distancia Euclidiana

- Equivalente a la clasificación Bayesiana cuando:
  - Todas las clases son equiprobables *a priori*

$$P(\omega_i) = P(\omega_j) = 1/n$$

  - Todos los descriptores tienen la misma varianza

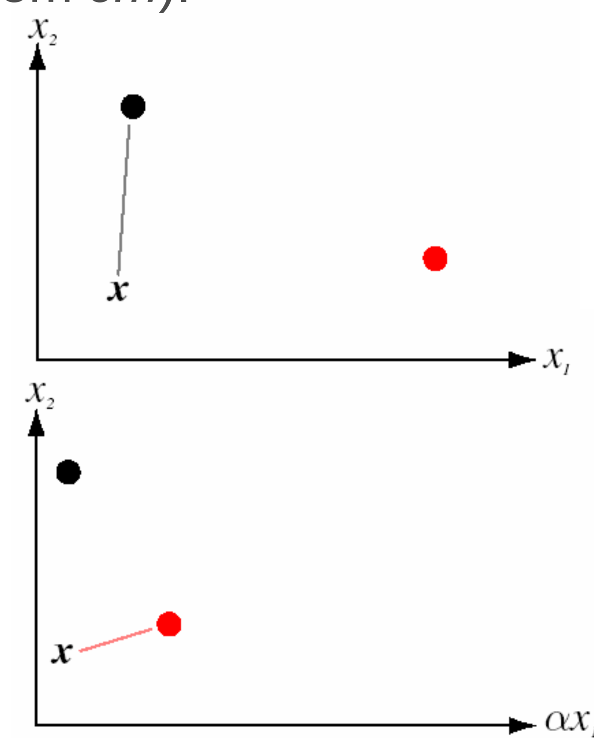
$$\sigma_{ik} = \sigma_{kl} = \sigma$$



- Siempre se escoge una clase, o es necesario definir una distancia máxima arbitraria.

- Interpretación clara sólo si x_1 y x_2 tienen las mismas unidades
  - x_1 = perímetro
  - x_2 = área

- Sensible a cambios de escala (p.e. perímetro medido en *mm* o em *cm*).

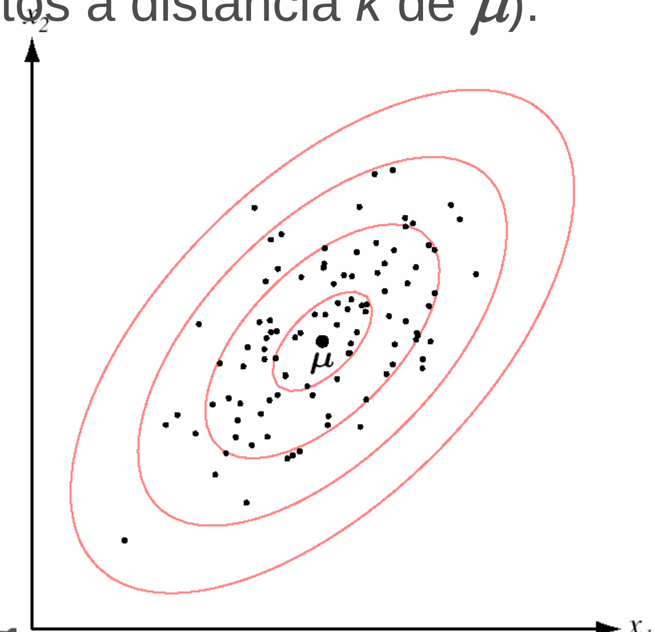


Distancias mínimas

- **Distancia de Mahalanobis:**

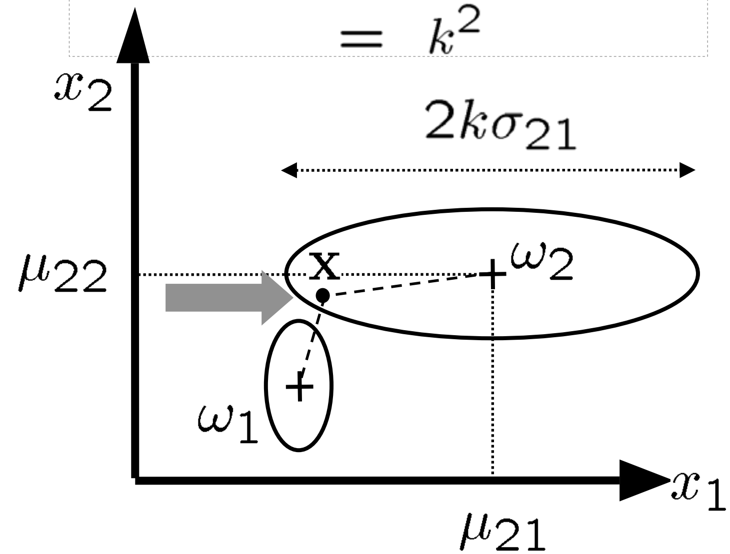
$$D^2(\mathbf{x},\omega_i) = \sum_{j=1}^{m} \frac{(x_j-\mu_{ij})^2}{\sigma_{ij}^2}$$

- Distancia adimensional; considera la imprecisión del descriptor (la varianza)
- Caso general, define una familia de elipses (ecuación de los puntos a distancia *k* de μ):



- $m = 2$:

$$\begin{aligned} D^2(\mathbf{x},\omega_i) &= \frac{(x_1-\mu_{i1})^2}{\sigma_{i1}^2} \\ &+ \frac{(x_2-\mu_{i2})^2}{\sigma_{i2}^2} \\ &= k^2 \end{aligned}$$



- ω_2: X_1 menos preciso que X_2 (influye menos en D^2).

$$\sigma_{21} > \sigma_{22}$$

- Puede contradecir a la distancia Euclidiana.

Distancia de Mahalanobis

- **Test de hipótesis:**
  - μ_{ij}: valor esperado
  - σ_{ij}: desviación esperada
  - x_j: valor observado

$$
\begin{aligned}
x_j &\sim N(\mu_{ij}, \sigma_{ij}^2) \\
\frac{x_j - \mu_{ij}}{\sigma_{ij}} &\sim N(0,1) \\
\frac{(x_j - \mu_{ij})^2}{\sigma_{ij}^2} &\sim \chi_1^2 \\
D^2(\mathbf{x}, \omega_i) &\sim \chi_m^2
\end{aligned}
$$

- H_0: {$\mathbf{x}$ proviene de ω_i}
- α: nivel de confianza

$$Pr\left\{D^2(\mathbf{x}, \omega_i) < \chi^2_{\alpha(m)}\right\} = 1 - \alpha$$

Tablas de Chi-cuadrado

m\ α	0.05	0.025	0.01
1	3.84	5.02	6.64
2	5.99	7.38	9.22
3	7.82	9.36	11.32

- $m = 2$:
  $$Pr\left\{D^2(\mathbf{x}, \omega_i) < 5.99\right\} = .95$$

- Se escoge la clase de menor distancia que pase este test de hipótesis.

Una hipótesis correcta será rechazada con probabilidad α (% de falsos negativos).

- La detección de falsos positivos es más compleja (involucra al resto de las clases).

Ejercicio

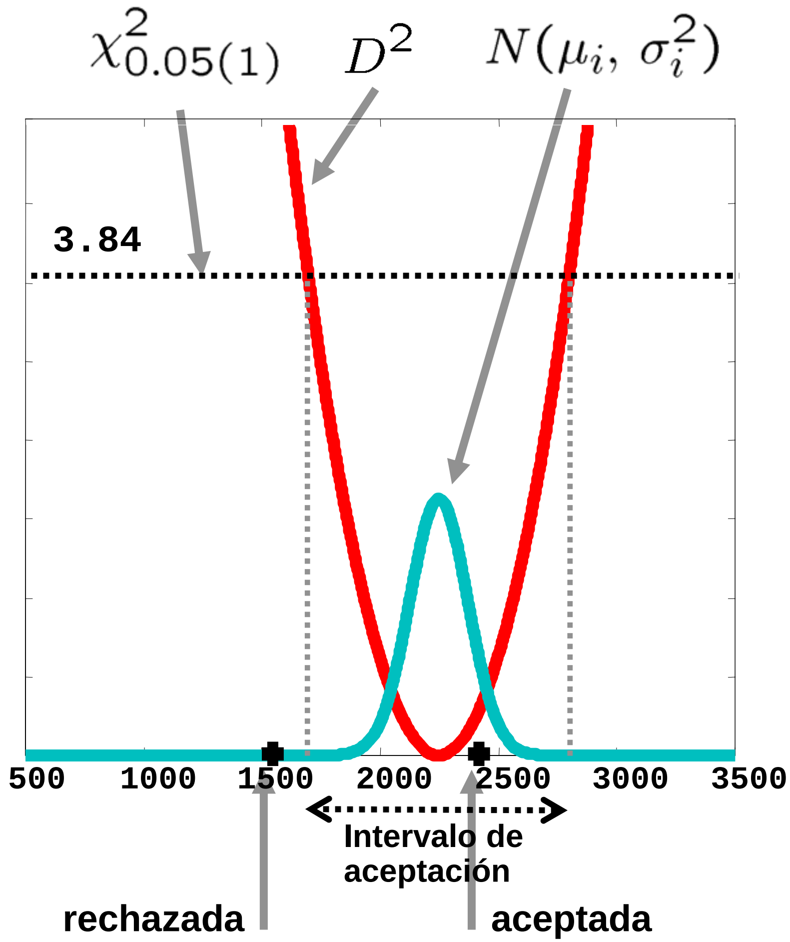
- Dado un objeto caracterizado por *m* descriptores independientes, y dada una muestra de cada descriptor:

- ¿Puede cada muestra ser estadísticamente compatible con el correspondiente descriptor, y sin embargo ser las muestras conjuntamente incompatibles con los descriptores del objeto?

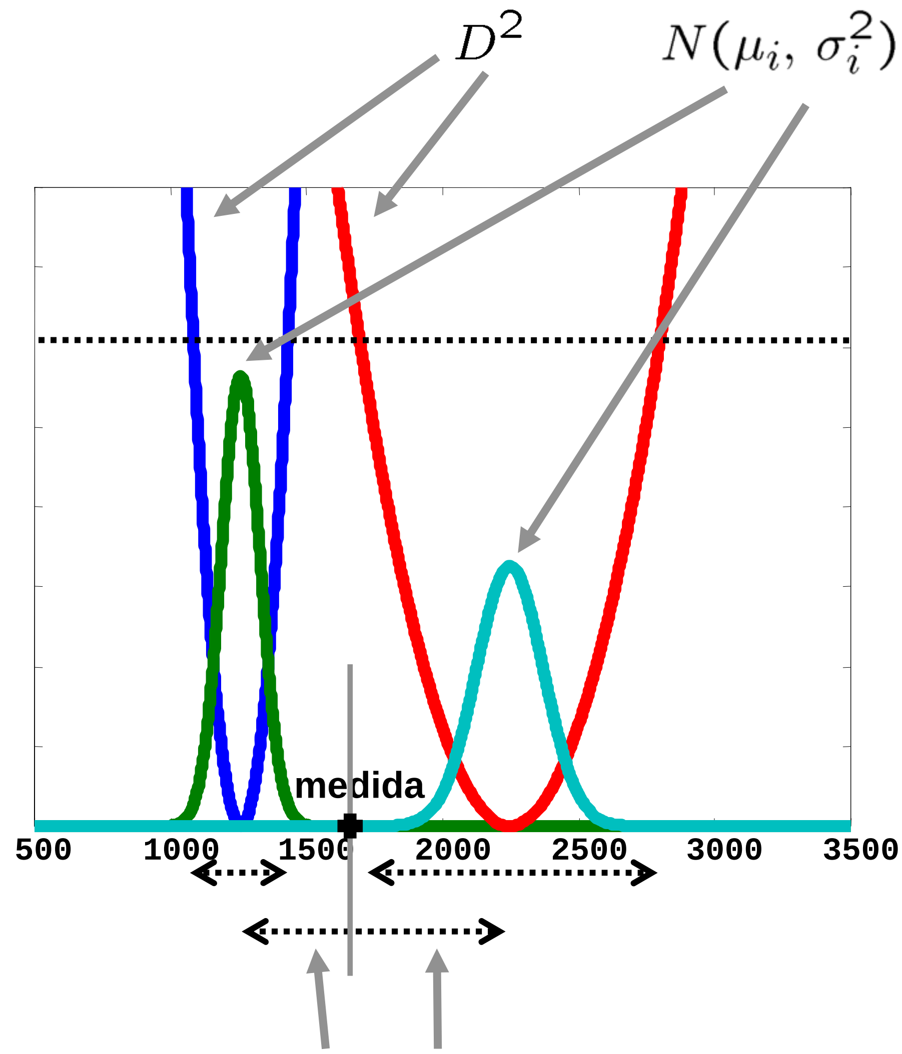
- ¿Puede la muestra de algunos descriptores ser estadísticamente incompatible con el correspondiente descriptor del objeto, y sin embargo ser las muestras conjuntamente compatibles con los descriptores del objeto?

Euclidiana .vs. Mahalanobis

- *m = 1*: $D^2 = \frac{(x-\mu_i)^2}{\sigma_i^2}$



- Euclidiana .vs. Mahalanobis

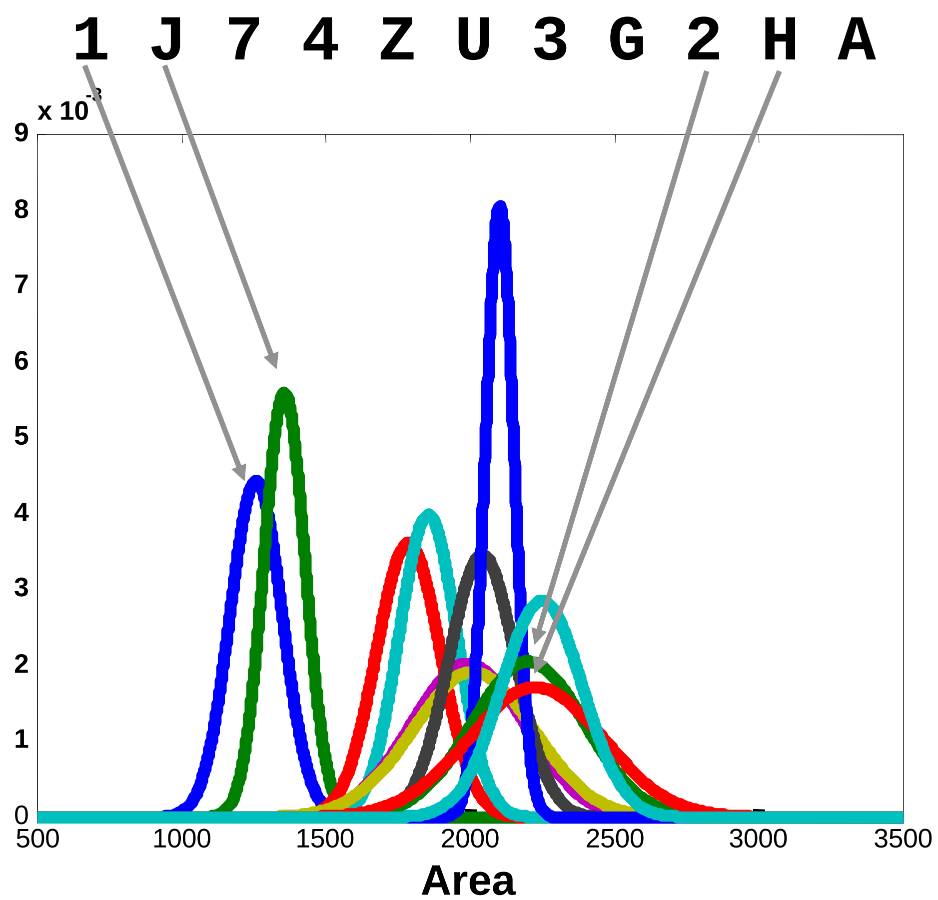


Distinguibilidad entre objetos

- Distancia entre dos clases:

$$D^2(\omega_i, \omega_k) = \sum_{j=1}^{m} \frac{(\mu_{ij} - \mu_{kj})^2}{\sigma_{ij}^2 + \sigma_{kj}^2}$$

- Permite evaluar el potencial de un conjunto de descrip- tores



$$Pr\left\{D^2(\omega_i, \omega_k) < \chi^2_{\alpha(m)}\right\} = 1 - \alpha$$

Si la hipótesis es aceptable, las clases ω_i y ω_k son indistinguibles

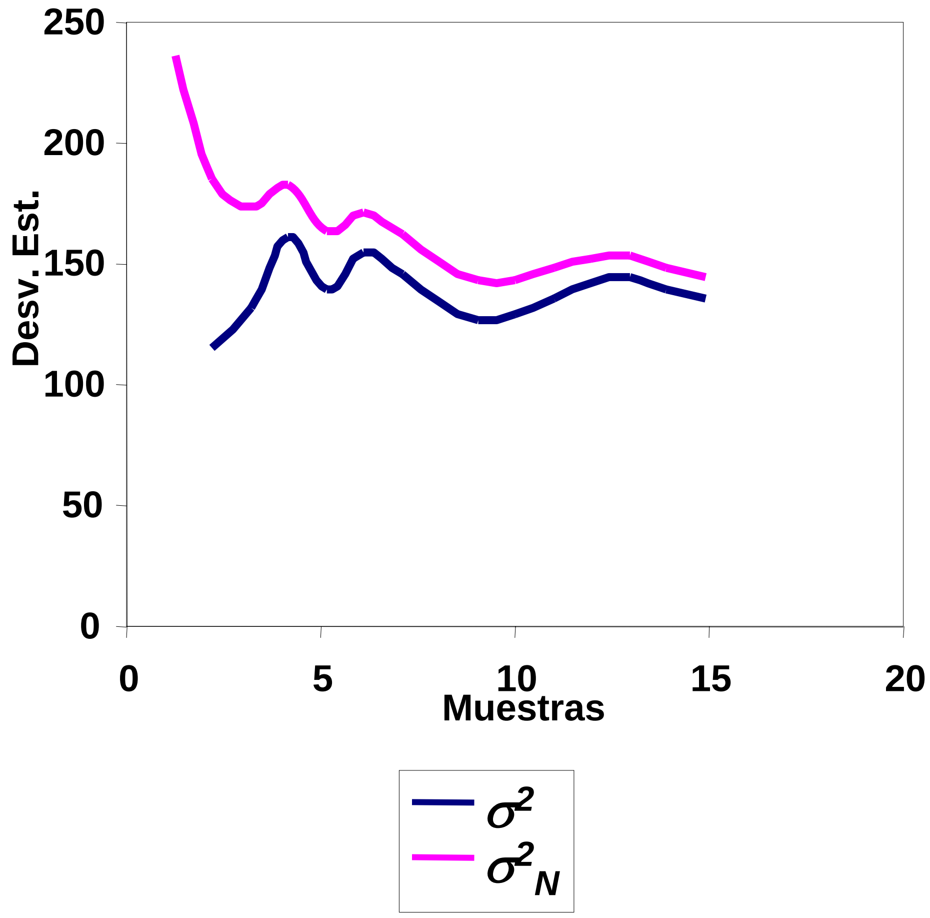
Número de muestras

- La varianza σ^2 no está definida para N=1
- Si N es pequeño, tiende a ser optimista (subestima la varianza)

$$
\begin{aligned}
\sigma^2_N \\
&= \frac{\sigma^2_{0N} + \sum_{k=1}^{N}(x_k - \hat{\mu})^2}{N} \\
&= \frac{\sigma^2_{0N}}{N} + \frac{N-1}{N}\hat{\sigma}^2
\end{aligned}
$$

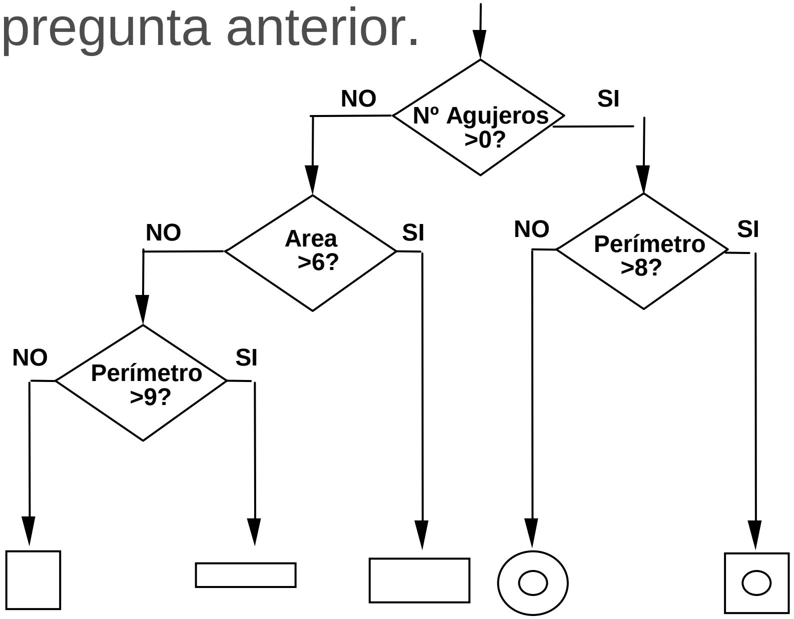
- σ^2_0: estimación a priori de la varianza (p. e. el cuadrado del 1% del valor del descriptor)

- Para N grande, tiende a la varianza clásica



4. Árboles de Decisión

- Secuencia de preguntas condi-cionadas a la respuesta de la pregunta anterior.



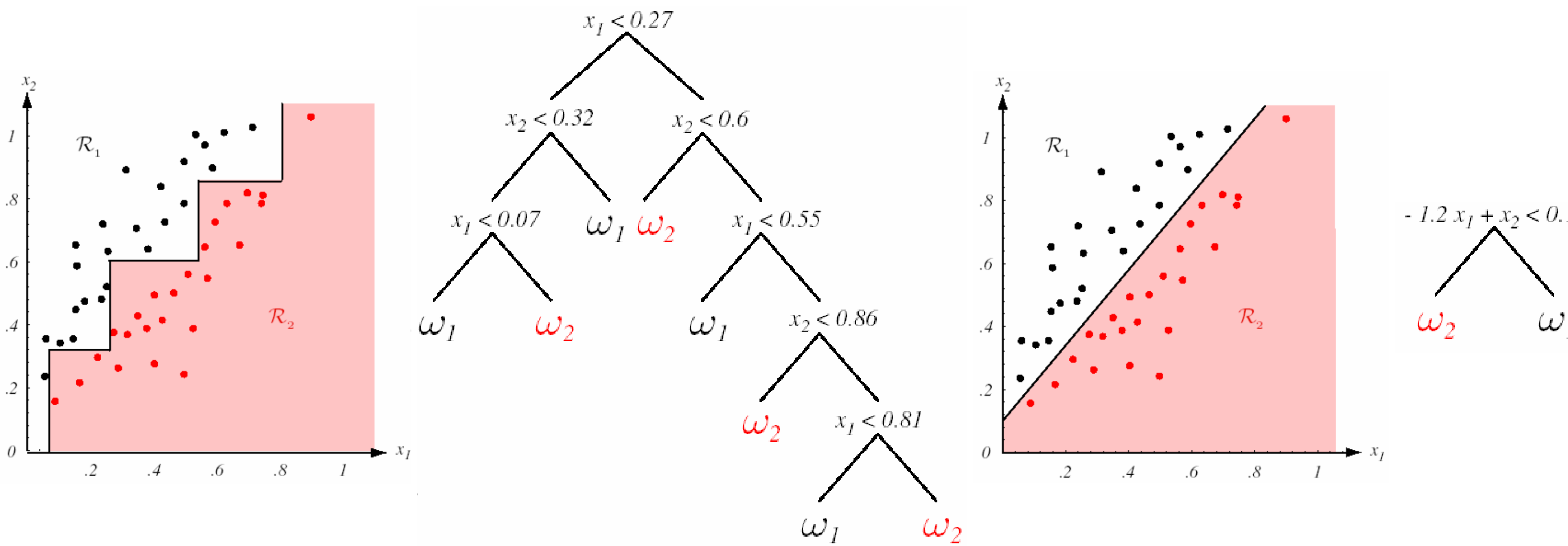
- **Método no métrico** (no se basa en mediciones de distancias).
- **Ventajas:**
  - Interpretabilidad
  - Puede incluir descriptores no numéricos
  - Permite incorporar conocimiento de expertos

- **CART** (*Classification and Regression Trees*): dados un conjunto de muestras de *m* descriptores de *n* clases:
  - División recursiva de las muestras de acuerdo con un descriptor.

1. ¿qué factor de división usar?
2. ¿qué descriptor consultar en un nodo?
3. ¿cuándo detener la división?
4. ¿cómo optimizar un árbol (hacerlo más pequeño, más simple)?
5. ¿qué hacer con nodos *impuros*?
6. ¿cómo manejar información no disponible?

Árboles de decisión

- Importancia de la selección de los descriptores



- Desventajas:
  - Un error en un descriptor puede impedir reconocer el objeto correctamente.
  - Siempre escoge una clase (o hay que introducir el objeto desconocido en muchos sitios).

5. Otros Métodos

- **Métodos paramétricos** (los anteriores): estimación de los parámetros de la función $p(x/\omega_i)$ que caracteriza a cada clase y que se supone conocida.
- **Métodos NO paramétricos:**
  - Estimación de la función de probabilidad:

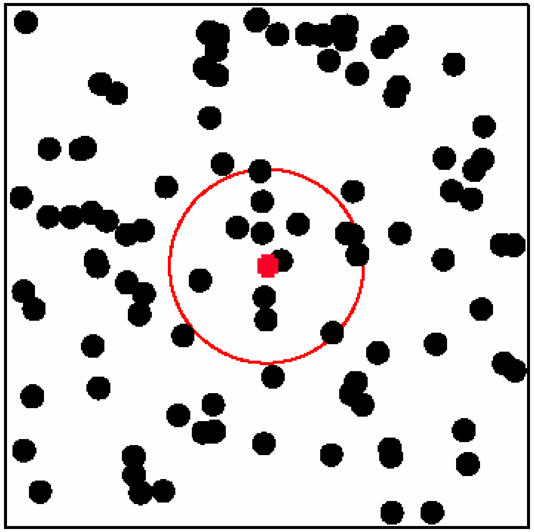
$$p(\mathbf{x}/\omega_i)?$$

  - Estimación de la probabilidad *a posteriori*:

$$P(\omega_i/\mathbf{x})?$$

- **Probabilidad *a posteriori***: Dada una ventana *V* alrededor de ***x***, que incluye *k* muestras, k_i de las cuales pertenecen a ω_i:

$$P(\omega_i/\mathbf{x}) \simeq \frac{k_i}{k}$$



El vecino más próximo

- Dadas *N* muestras de *m* descriptores de *n* clases:

$$\mathbf{x}_{i \cdot k} = (x_{i \cdot 1}, \cdots, x_{i \cdot N})^T$$

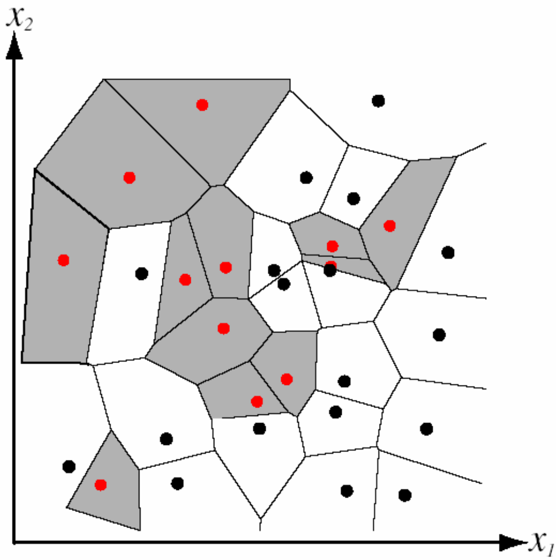
- Dados *m* descriptores del objeto a identificar:

$$\mathbf{x} = (x_1, \cdots, x_m)^T$$

- Escoger la clase ω_i tal que:

$$\mathbf{x}_{i \cdot k} = \min_k D^2(\mathbf{x}, \mathbf{x}_{i \cdot k})$$

- Usar teselación de Voronoi

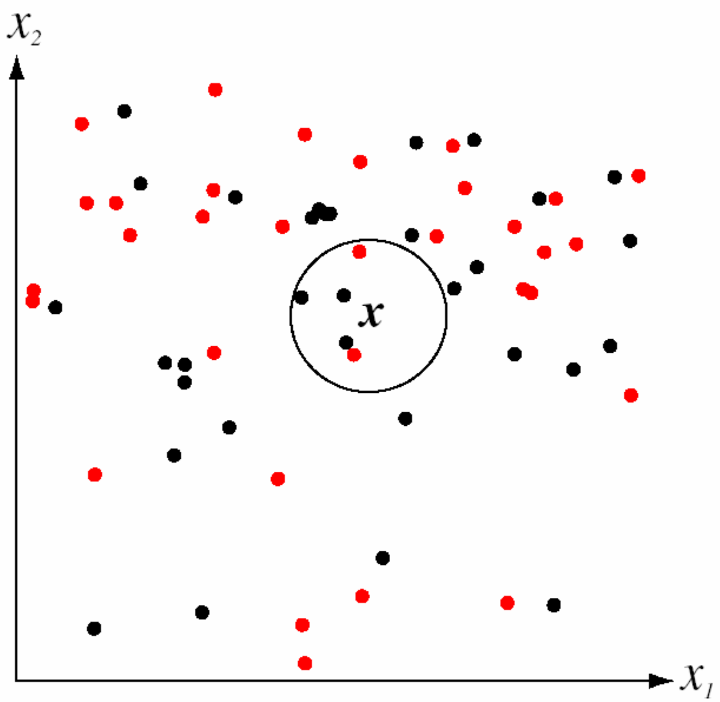


- **Método subóptimo**: tasa de errores mayor que la Clasificación Bayesiana.

¡La tasa de errores no es más del doble!

Los *k* vecinos más próximos

- Escoger la clase ω_i más fre-cuente en los k vecinos más próximos a $\mathbf{x}$.



- Los empates se pueden evitar aumentando *k* (*n*=2, *k* impar).

- Para *n*, *m* y *N* grande, es computacionalmente costo-so.

$$O(n\,m\,N)$$

- Distancias parciales:

$$D_r^2(\mathbf{a},\mathbf{b}) \;=\; \sum_{i=1}^{r}(a_i - b_i)^2$$

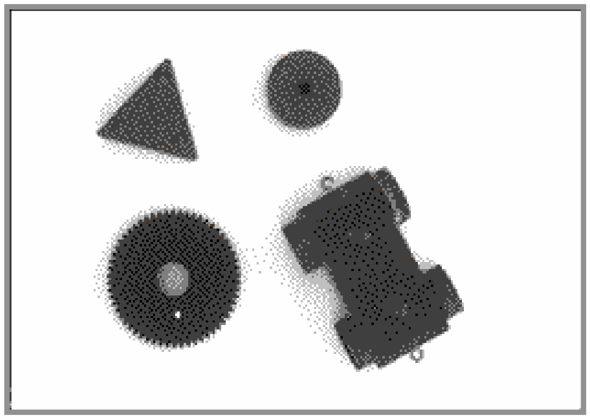
$$r \;<\; m$$

- Se abandona el cálculo de la distancia a una muestra cuan-do la distancia parcial es mayor que la distancia total al *k*-ésimo vecino.

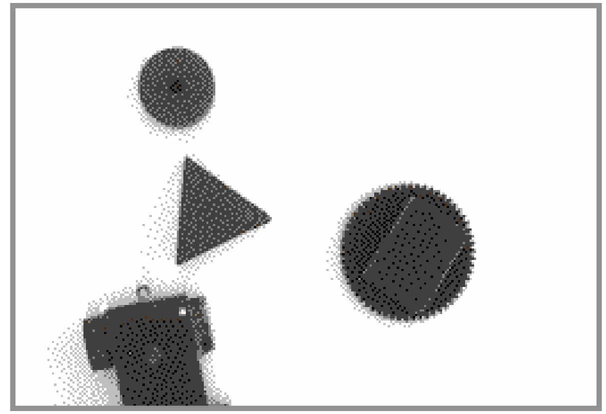
6. Conclusiones

- Métodos de **complejidad lineal**
- Descriptores invariantes en 2D

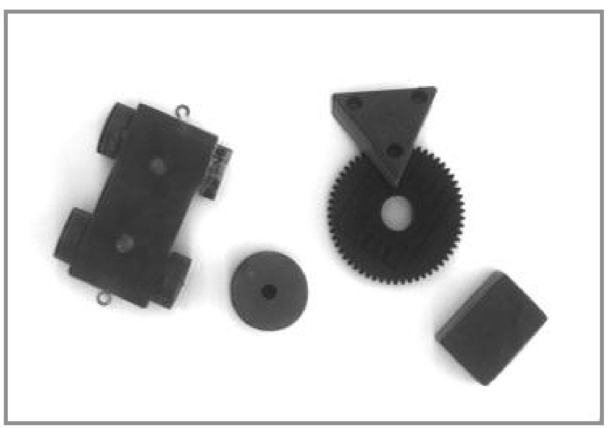
OK



!Visibilidad parcial!



!Solapamiento!



Reconocimiento geométrico