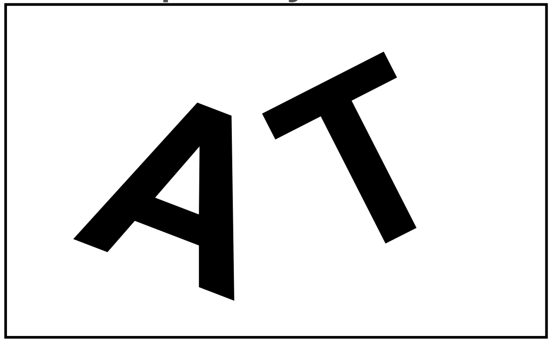
Lesson 4: Parametric Recognition

1. Introduction
2. Bayesian Classification
3. Classification based on distances
4. Decision trees
5. Other methods
6. Conclusions

1. Introduction

- **2D recognition:** Labeling of regions corresponding to 2D objects in the image.

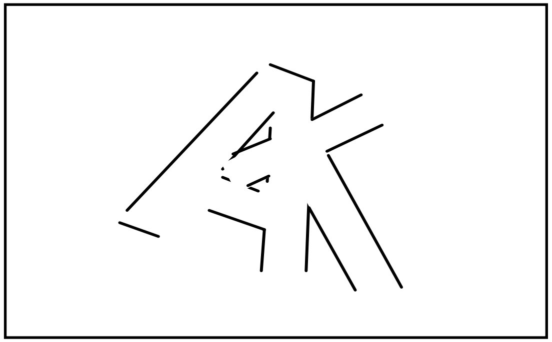
- **Parametric recognition**: global invariant properties, objects completely visible



  » Area
  » Perimeter
  » Euler number
  » Elongation
  » ...

- Simpler, less robust

- **Geometric recognition:** local properties



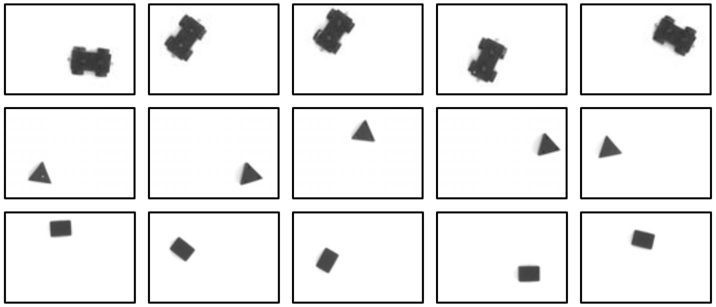
  » Vertices
  » Edges
  » Circles
  » Corners
  » ...

- More complex, more robust

General System Scheme

Learning phase:

- Acquire several images of each object in isolation, and compute its descriptors.



Exploitation phase:

- Compute the descriptors of each blob in a given image
- Compare the estimated values of the descriptors with those of known objects

Different methods use different techniques to carry out this comparison

Notation:

- *n* known objects (classes):

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

- *m* descriptors used:

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

- *N* samples of descriptor *j* of class *i*:

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

Given a sample $\mathbf{x}$, which class ω_i does it correspond to?

Is it an unknown object?

2. Bayesian Classification

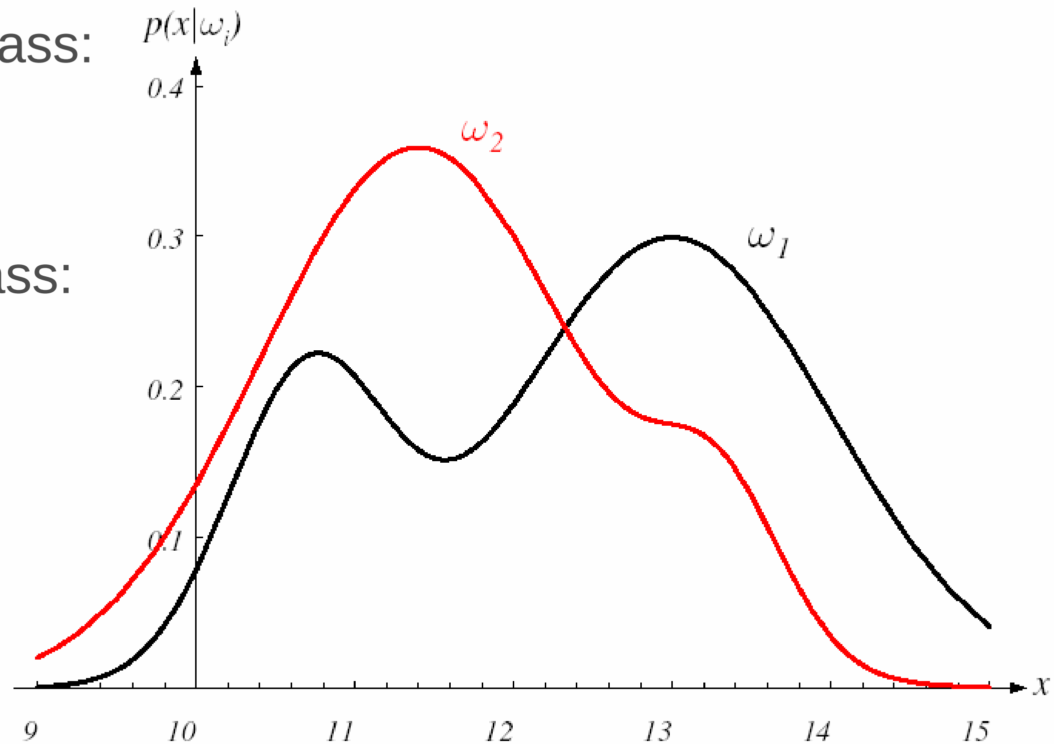
- Descriptors are considered random variables.
- Compute the probability that the sample comes from each class using Bayes' Rule.
- Probabilistic model of each class:

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

- A priori probability of each class:

$$P(\omega_i)$$

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



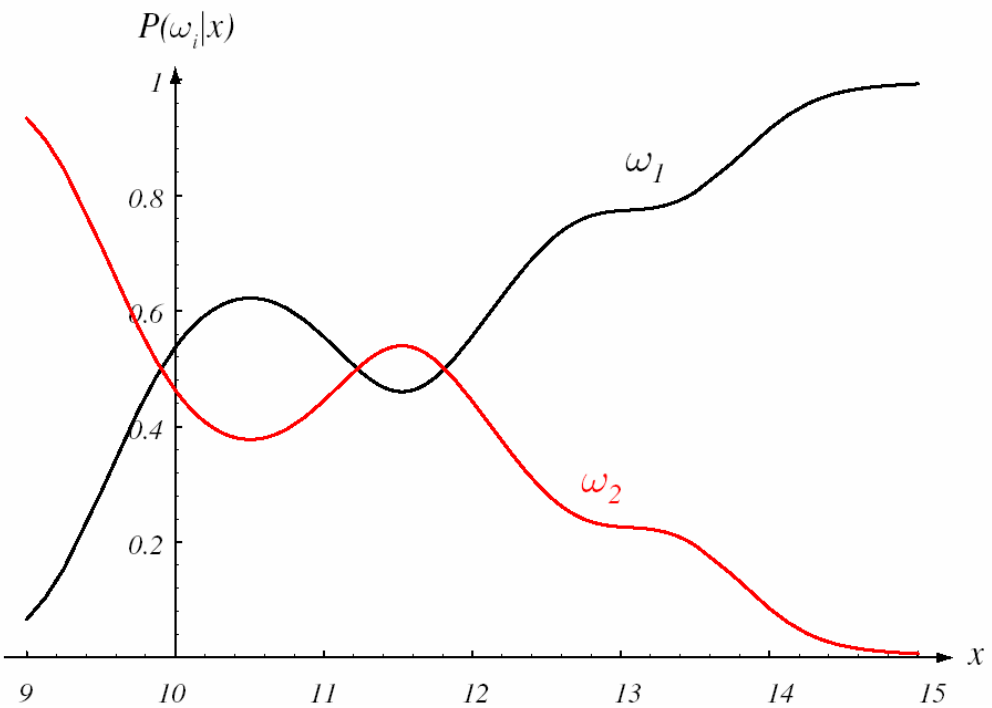
$$p(\mathbf{x}/\omega_i) = p(\mathbf{x}/\omega_j)?$$
$$P(\omega_i) = P(\omega_j)?$$

Bayesian Classification

- **Bayes' Rule:**

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

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



- It distributes the probability among all classes:

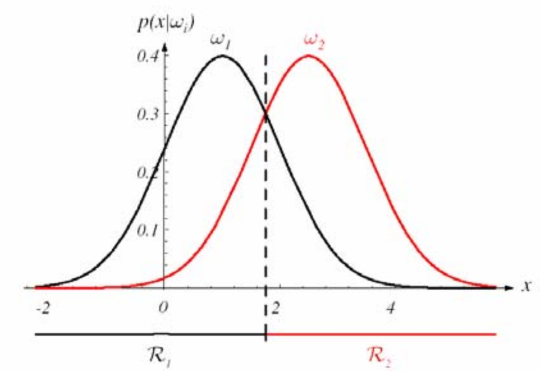
$$\sum_{i=1}^{n} P(\omega_i/\mathbf{x}) = 1$$

- We decide ω_i if:

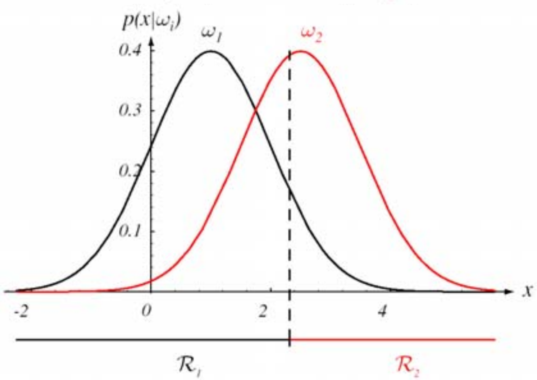
$$P(\omega_i/\mathbf{x}) = \max_{j=1}^{n} P(\omega_j/\mathbf{x})$$

Bayesian Classification

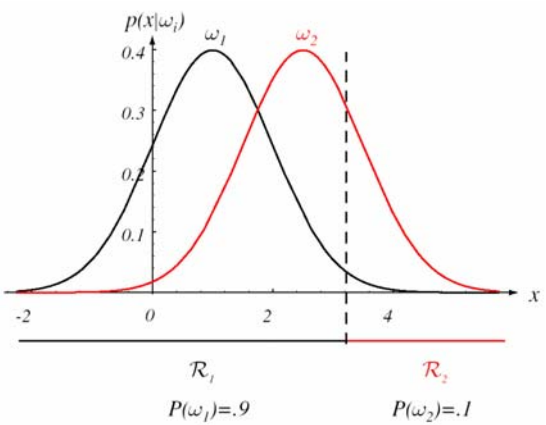
- Influence of $P(\omega_i)$:



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

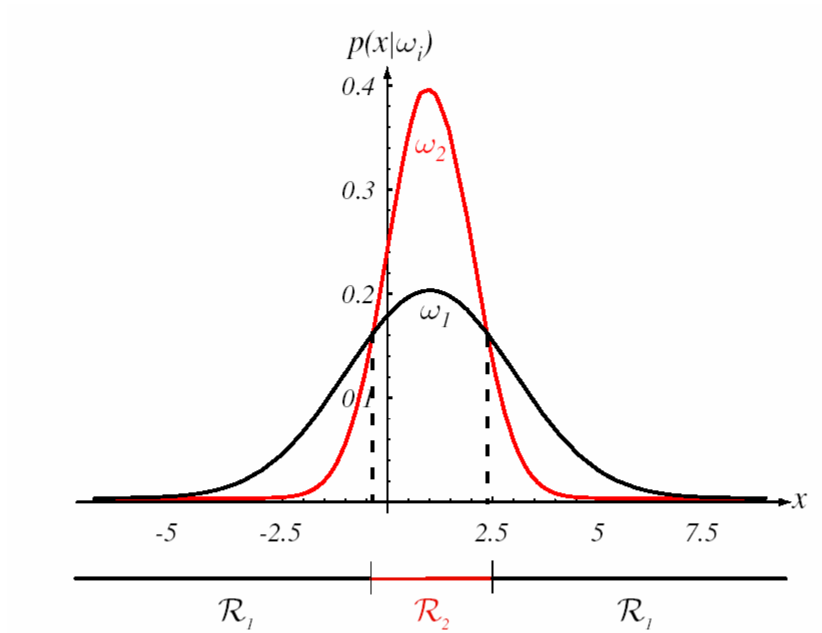


$$P(\omega_1) = 0.7$$
$$P(\omega_2) = 0.3$$



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

- Influence of $p(x/\omega_i)$:

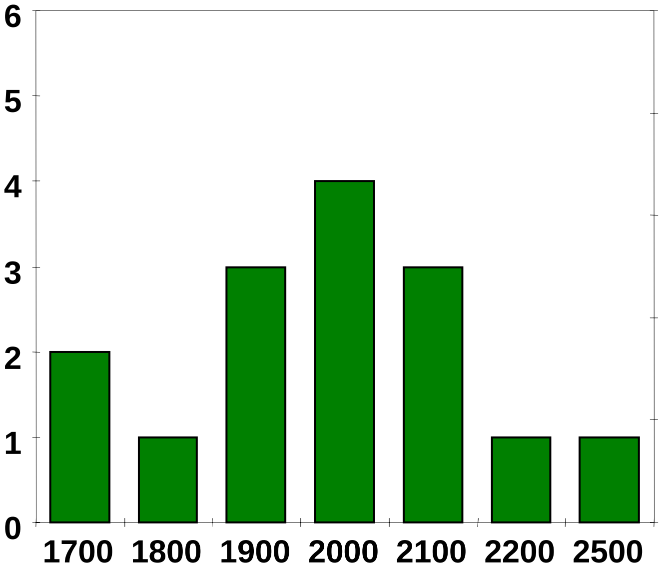


Descriptors

- Example: area

N=15

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



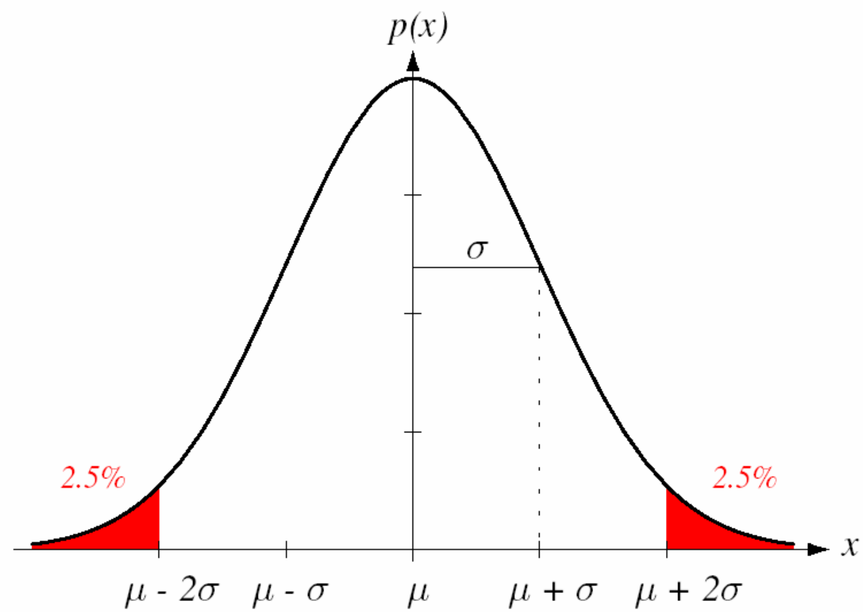
$$\hat{\mu} = 1990.7$$

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

- Given no other information, we characterize a descriptor using a *Gaussian distribution*:

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

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



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

Descriptors

- **Sample mean:** representative value:

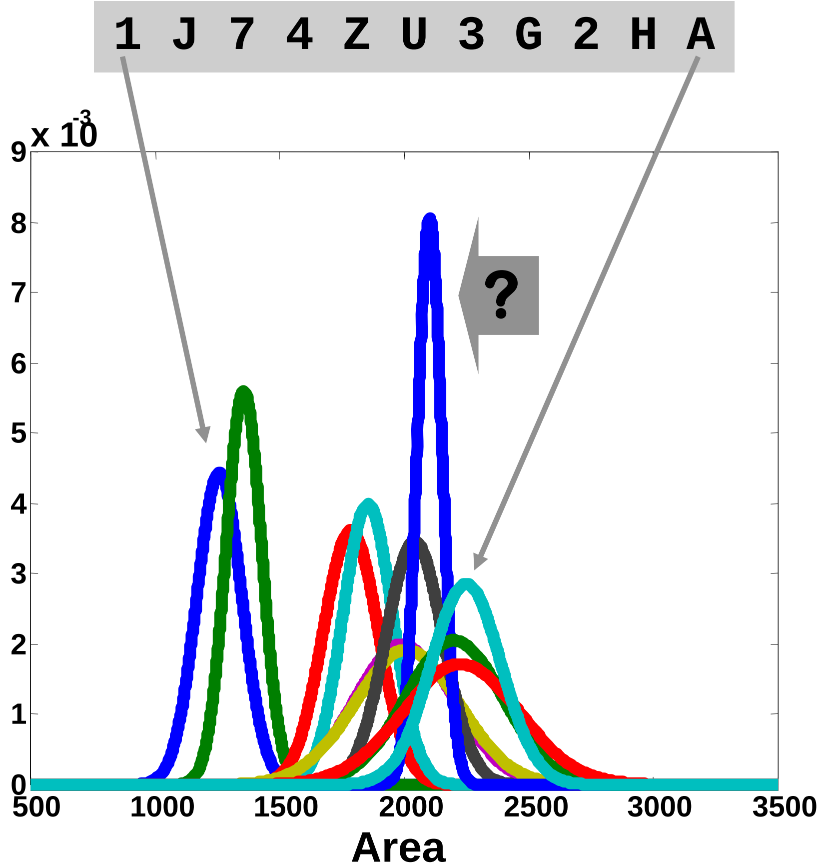
$$\hat{\mu}_{ij} = \frac{\sum_{k=1}^{N} x_{ijk}}{N}$$

- **Sample variance:** dispersion or variability:

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

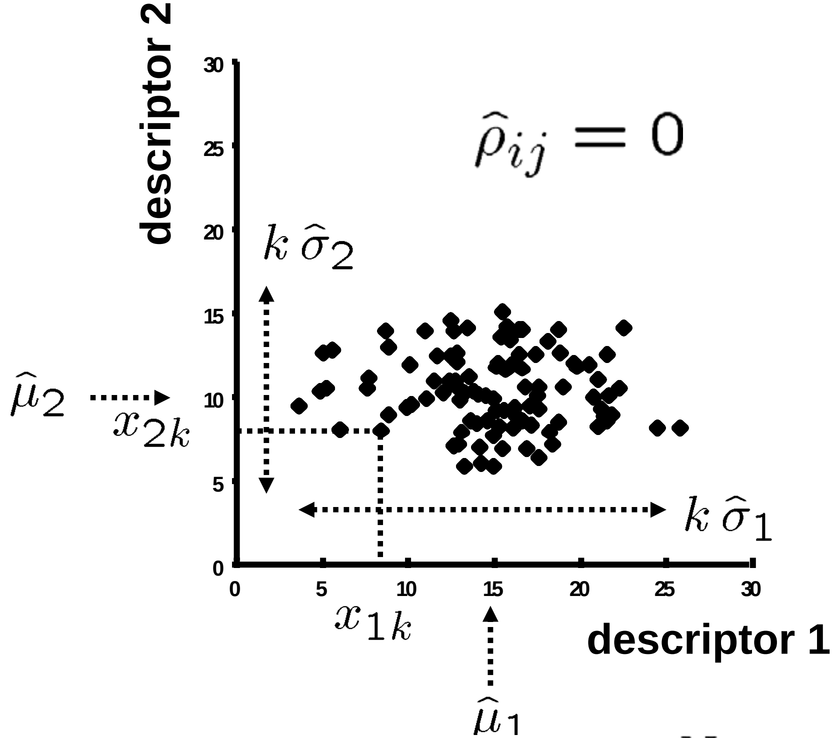
?

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

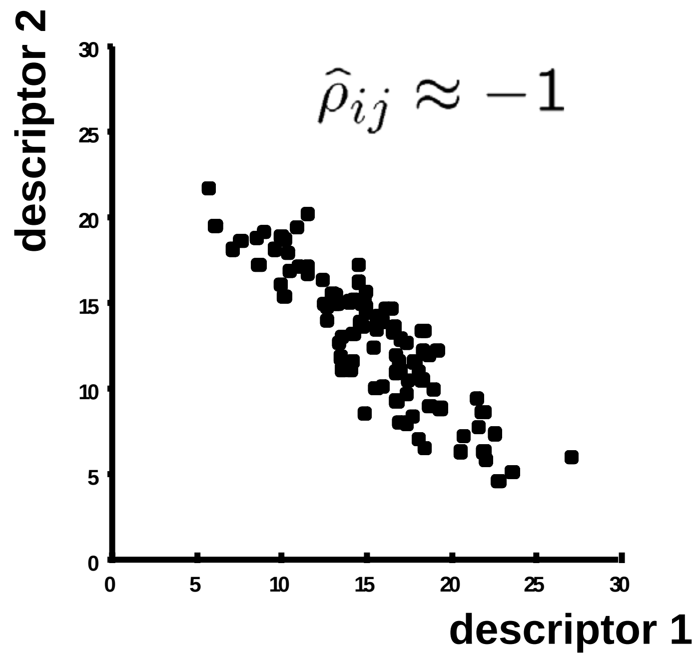


Independence of descriptors

- Independent:



- Linearly correlated:



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

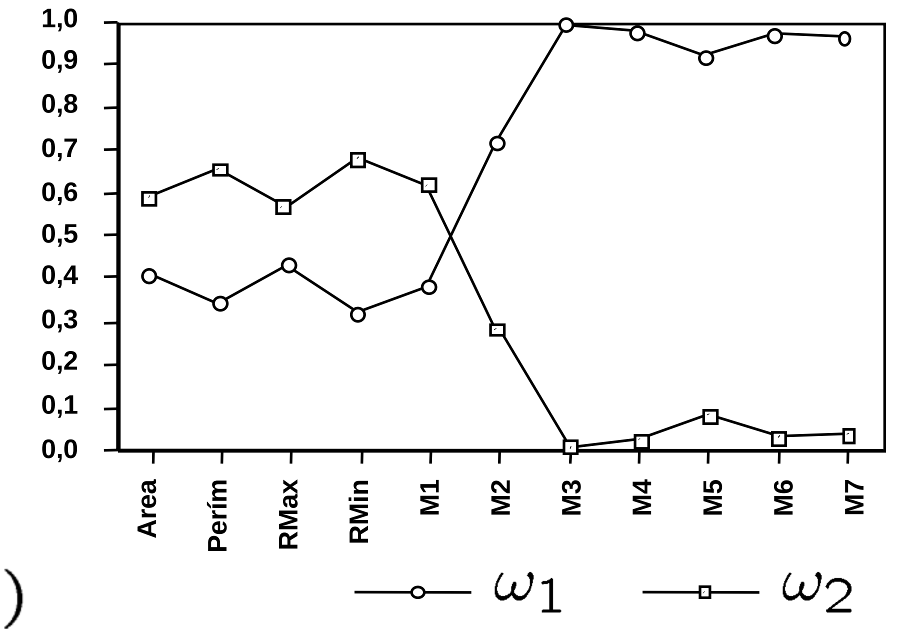
Independent descriptors allow to discriminate more clearly between classes.

Bayesian Classification

- Assuming Gaussianity and statistical independence:

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

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

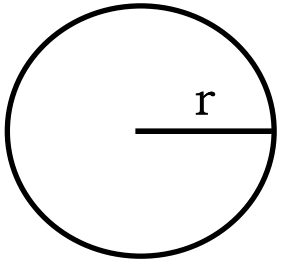


- Iterative method:

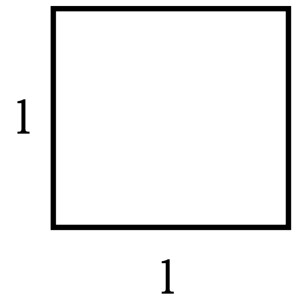
$$\begin{aligned} P^0(\omega_i) &= P(\omega_i) \\ P^d(\omega_i) &= P(\omega_i/x_1,\cdots,x_d) \\ &= \frac{p(x_d/\omega_i)P^{d-1}(\omega_i)}{\sum_{j=1}^{m} p(x_d/\omega_j)P^{d-1}(\omega_j)} \end{aligned}$$

- Iterate until the most probable object surpasses the threshold (computing all descriptors might not be necessary).

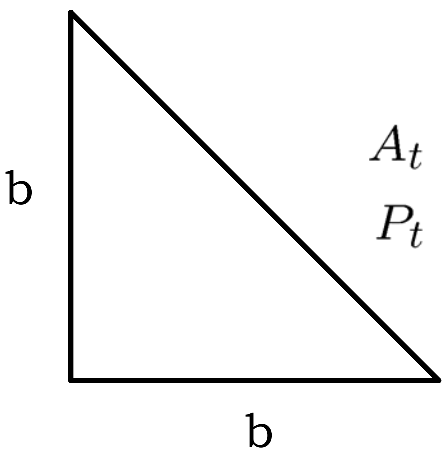
Sequential measurement of descriptors



$$A_c = \pi r^2$$
$$P_c = 2\pi r$$



$$A_d = l^2$$
$$P_d = 4l$$



$$A_t = b^2/2$$
$$P_t = b(2+\sqrt{2})$$

- Assuming that:

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

- Area is not a discerning parameter in this case:

$$A_c = A_d = A_t$$

- But perimeter is:

$$P_c = 2\sqrt{\pi} = 1.772453851$$
$$P_d = 4$$
$$P_t = 4.828427123$$

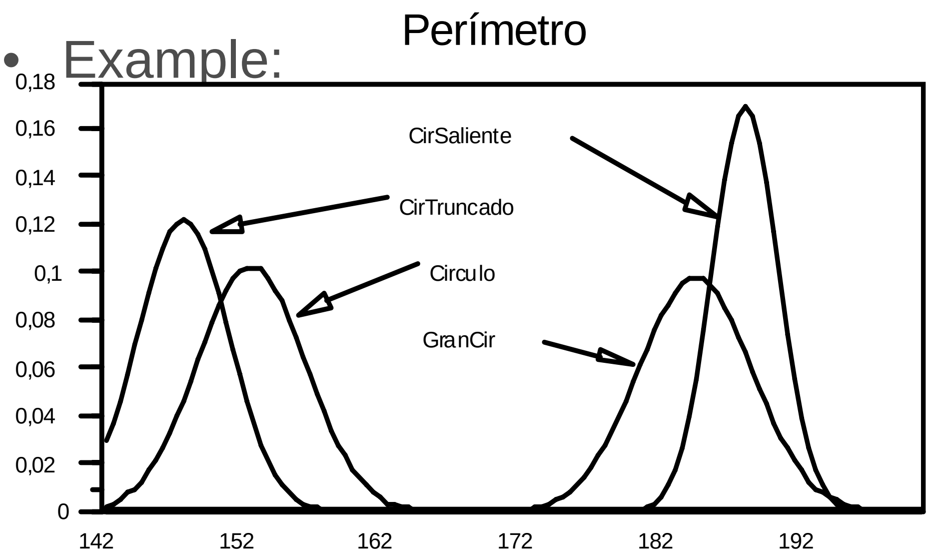
In which order should we compute descriptors?

Sequential measurement of descriptors

- Discerning power of parameter *k* between clases ω_i y ω_j:

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

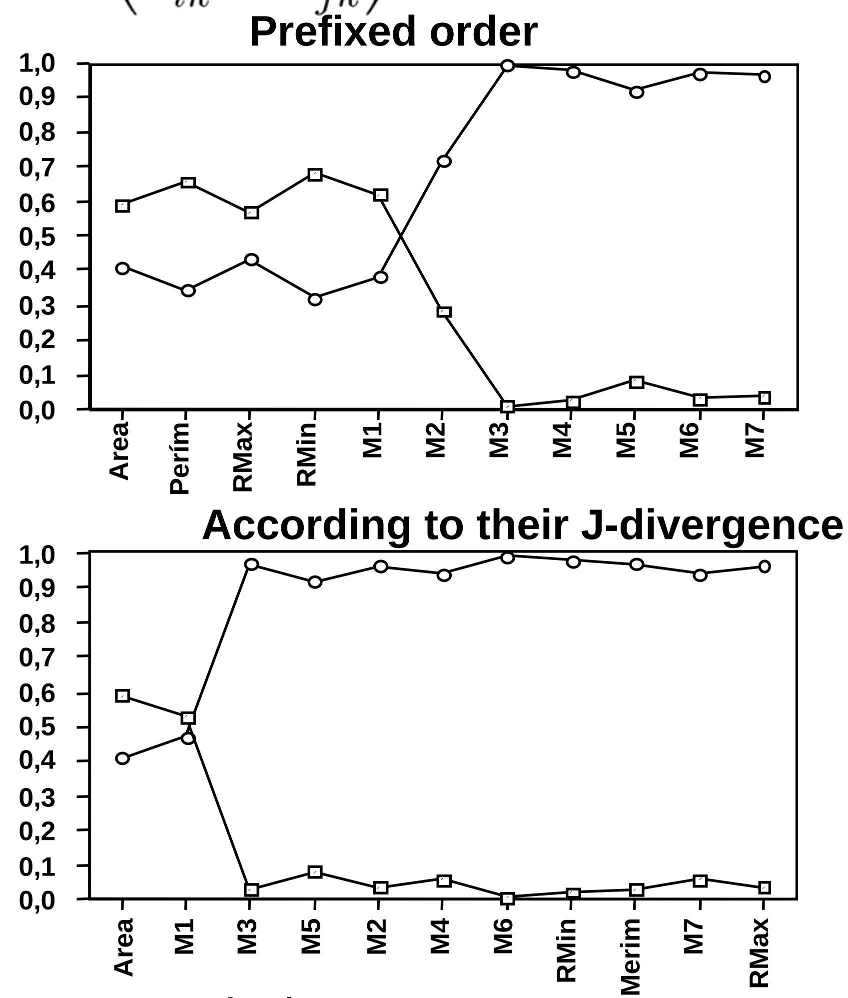
- Example:



J-Divergencia

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

- Can be computed *a priori*



Unkown object management

- Class ω_0 representing the unknown object:

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

- Uniform distribution between smallest and largest possible value for the descriptor:

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

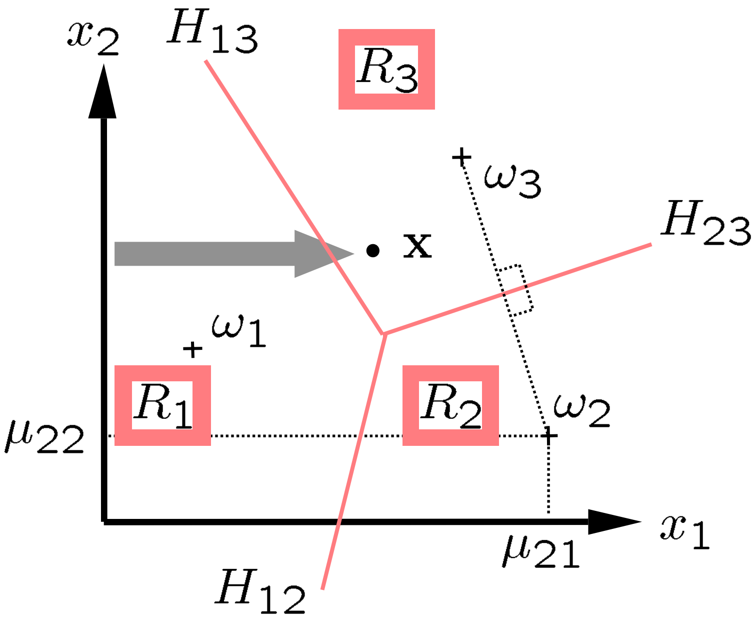
3. Minimal distances

- Compute the ***distance*** of the descriptor vector to each of the known objects.
- **Euclidean distance:**

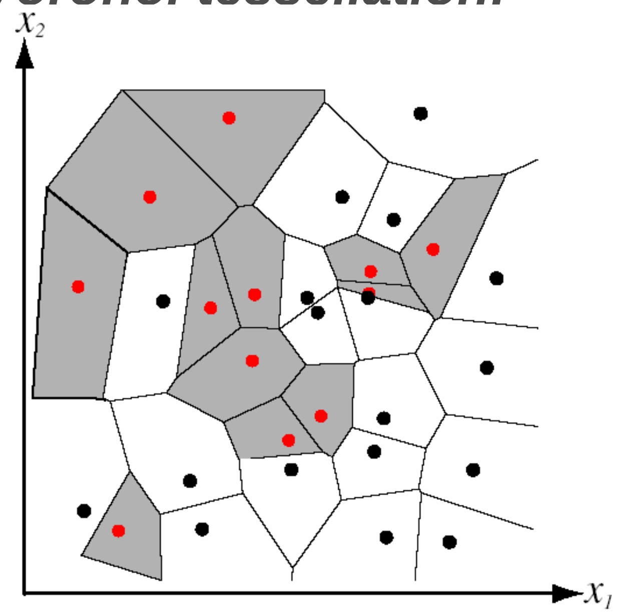
$$D^2(\mathbf{x}, \omega_i) = ||\mathbf{x} - \omega_i||^2 = \sum_{j=1}^{m} (x_j - \mu_{ij})^2$$

- Choose the class of minimal distance (nearest neighbor).
- The decision frontier between two classes will be given :

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



- ***Voronoi tessellation:***



Exercise

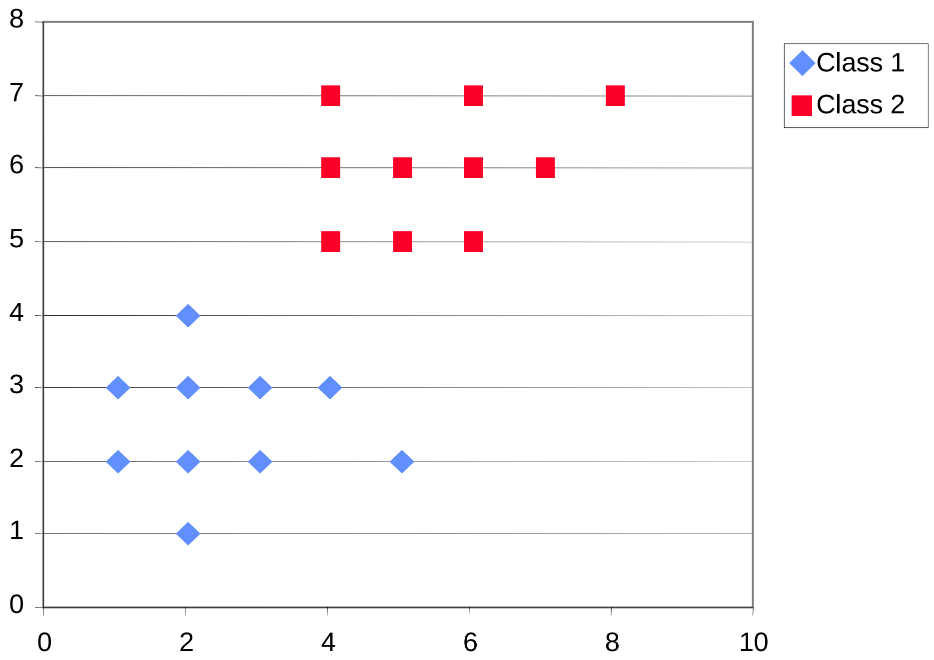
- Given the following samples of descriptors $\mathbf{x_1}$ and $\mathbf{x_2}$ of two classes:

Class 1

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

Class 2

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



- Determine (and graph) the equation of the curve that represents the decision frontier between the two classes according to the Euclidean distance.

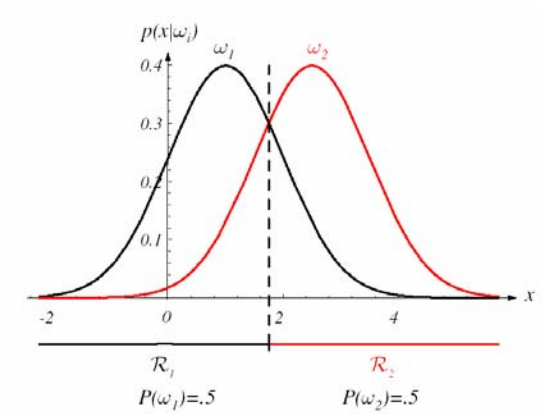
Euclidean distance

- Equivalent to Bayesian Classification when:
  - All classes are equally probable *a priori*

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

  - All descriptors have the same variance

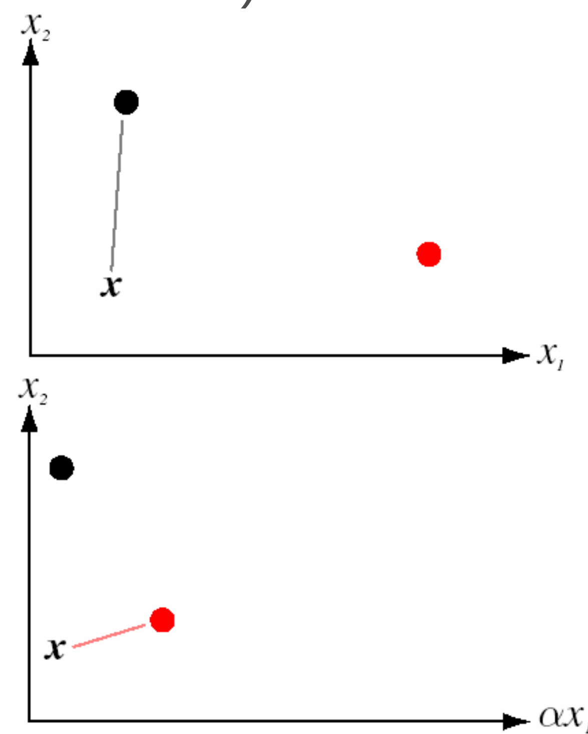
$$\sigma_{ik} = \sigma_{kl} = \sigma$$



- A class is always chosen; alternatively is it necessary to define an arbitrary maximal distance.

- Clear interpretation only if x_1 and x_2 have the same units
  - x_1 = perimeter
  - x_2 = area

- Sensitive to changes of scale (f.e. perimeter measured in *mm* or in *cm*).

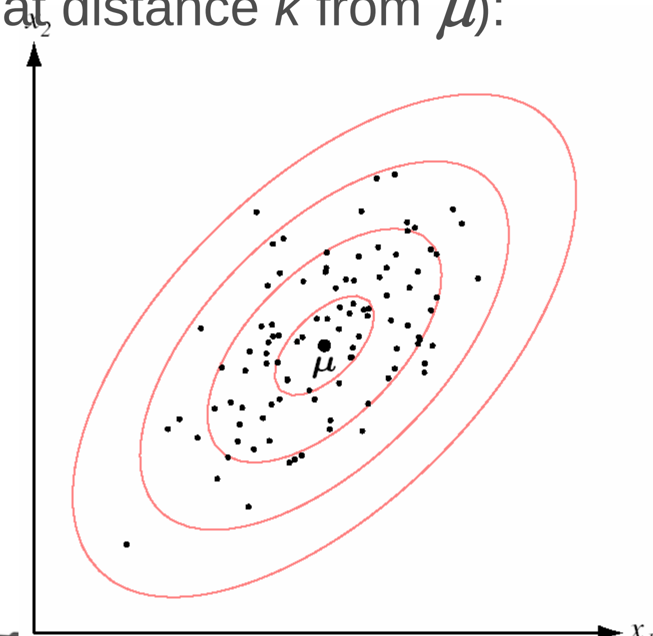


Minimal distances

- **Mahalanobis distance:**

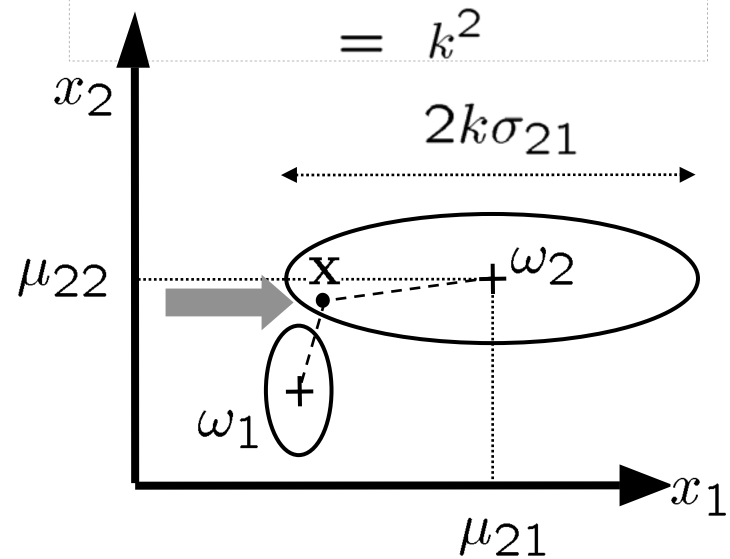
$$D^2(\mathbf{x}, \omega_i) = \sum_{j=1}^{m} \frac{(x_j - \mu_{ij})^2}{\sigma_{ij}^2}$$

- Adimensional distance; it considers the imprecision of the descriptor (the variance)
- General case: it defines a family of ellipses (equation of the points at distance *k* from 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 less precise than X_2 (has less influence in D^2).

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

- It can contradict the Euclidean distance.

Mahalanobis distance

- **Hypothesis test:**
  - μ_{ij}: expected value
  - σ_{ij}: expected deviation
  - x_j: observed value

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

- H_0: {$\mathbf{x}$ comes from ω_i}
- α: confidence level

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

Chi-squared tables

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

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

- Choose the class with smallest distance that passes the test.

A correct hypothesis will be Rejected with probability α (% of false negatives).

- Detection of false positives is more complex (it involves all classes).

Exercise

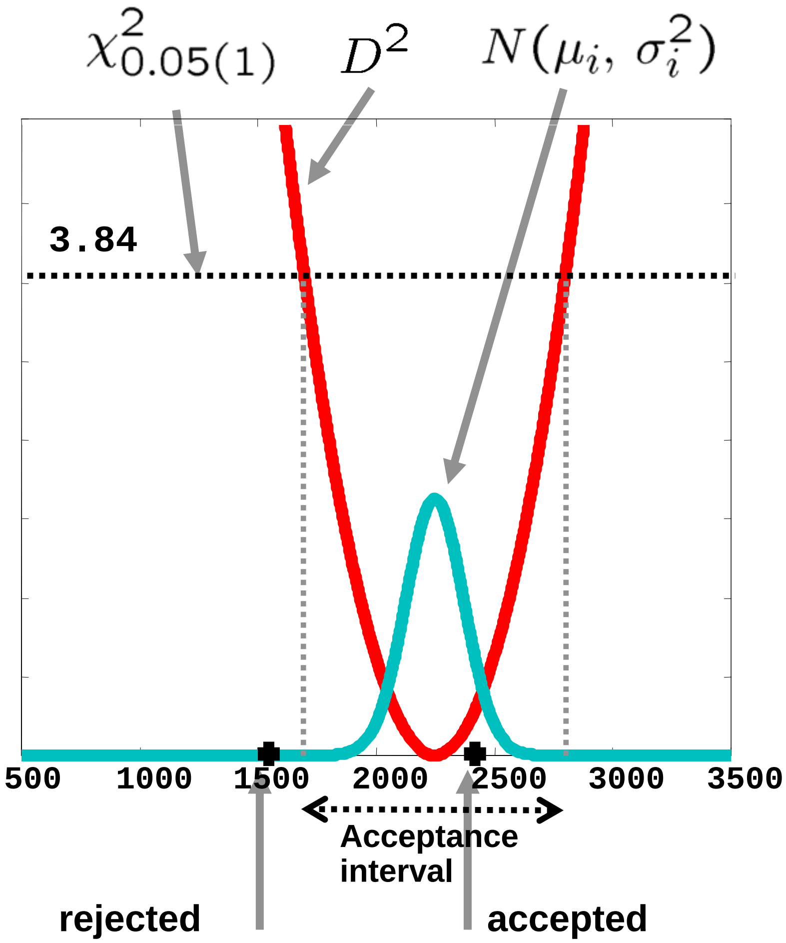
- Given an object characterized by *m* independent descriptors, and given a sample of each descriptor:

- Can each sample be statistically compatible with its corresponding descriptor, and nevertheless the samples be jointly incompatible with the object descriptors?

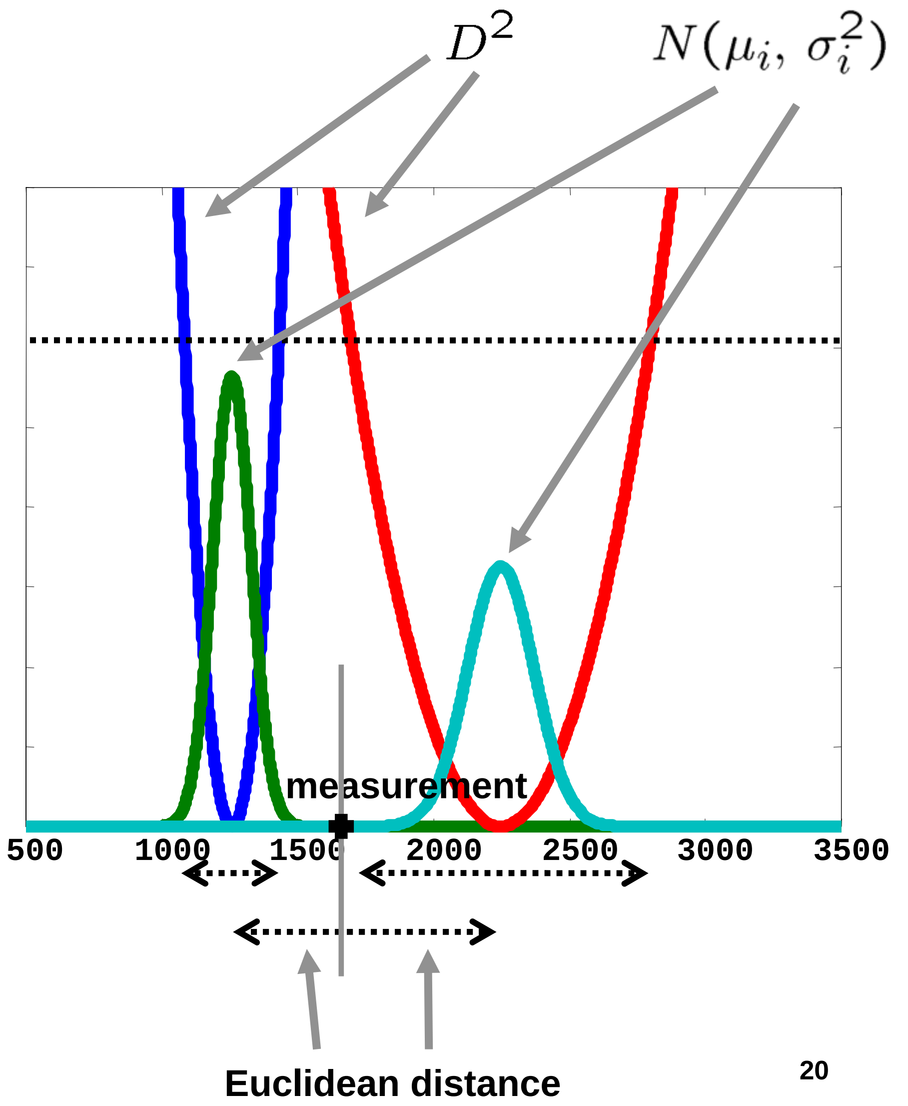
- Can the sample of some descriptors be statistically incompatible with their corresponding descriptors, and nevertheless the samples be jointly compatible with the object descriptors?

Euclidean .vs. Mahalanobis

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



- Euclidean .vs. Mahalanobis

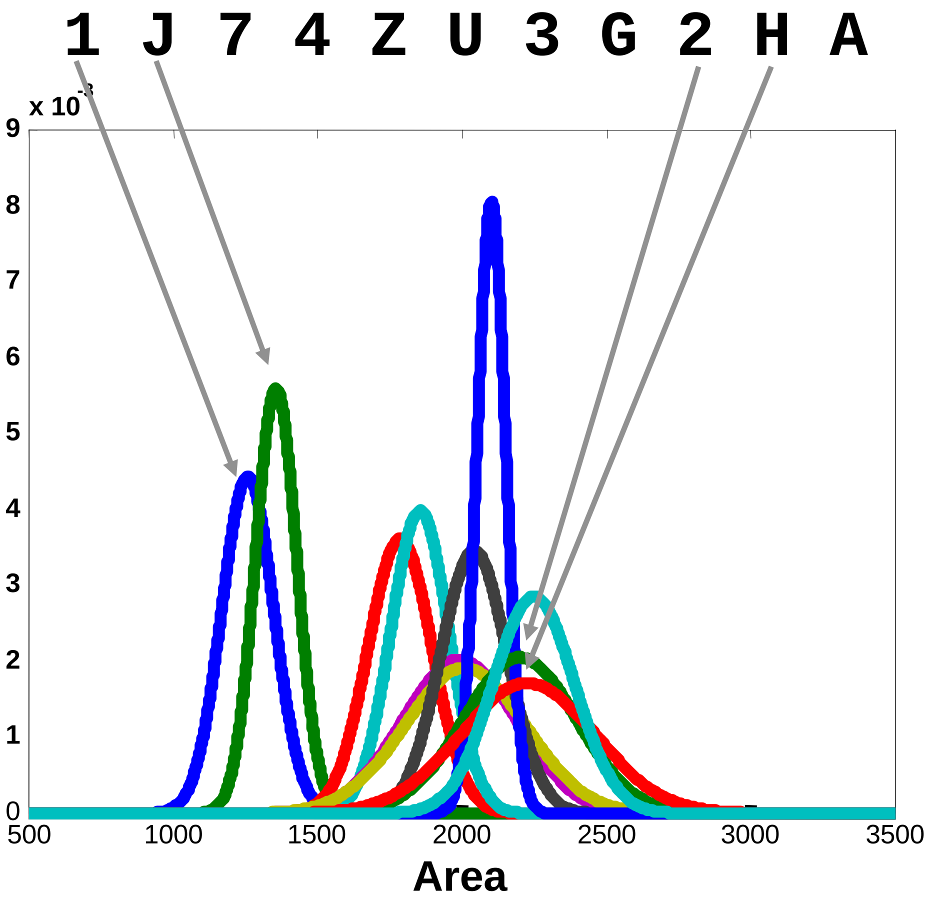


Discriminating between objects

- Distance between two classes:

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

- It allows to evaluate the potential of a set of descriptors



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

If the hypothesis is acceptable, classes ω_i and ω_k are indistinguishable

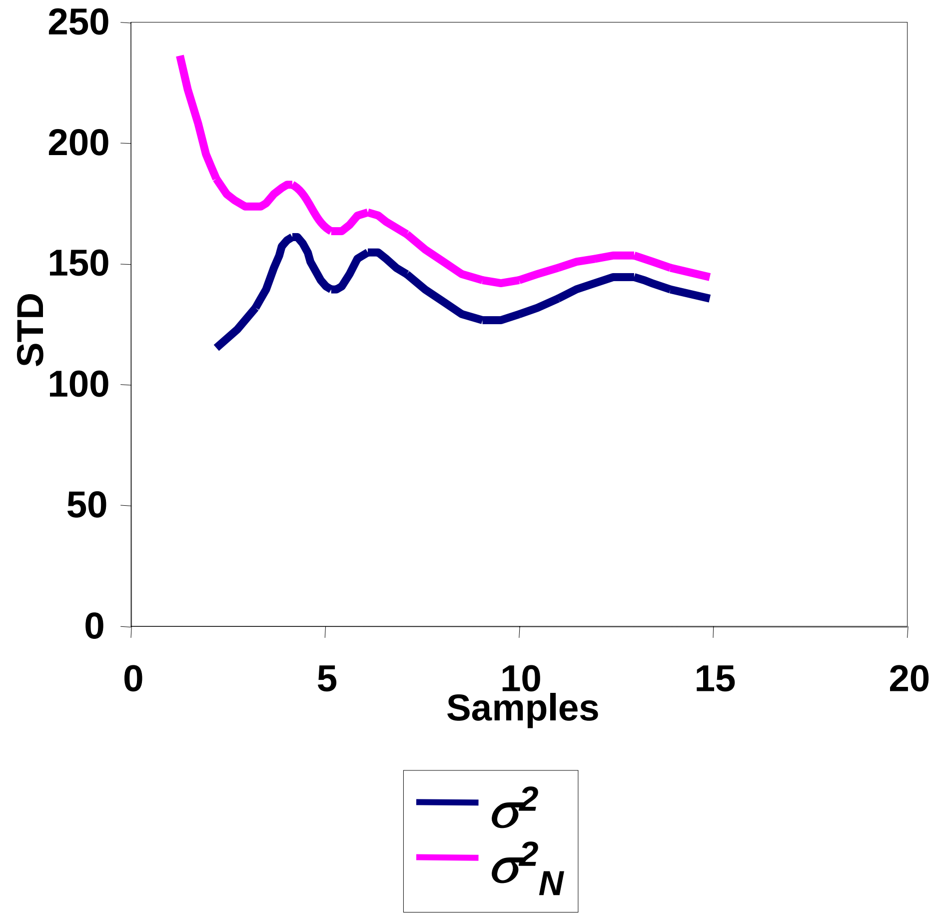
Number of samples

- Variance σ^2 is not defined when N=1
- If N is small, it tends to be optimistic (it underestimates the variance)

$$\sigma_N^2 = \frac{\sigma_{0N}^2 + \sum_{k=1}^{N}(x_k - \hat{\mu})^2}{N} = \frac{\sigma_{0N}^2}{N} + \frac{N-1}{N}\hat{\sigma}^2$$

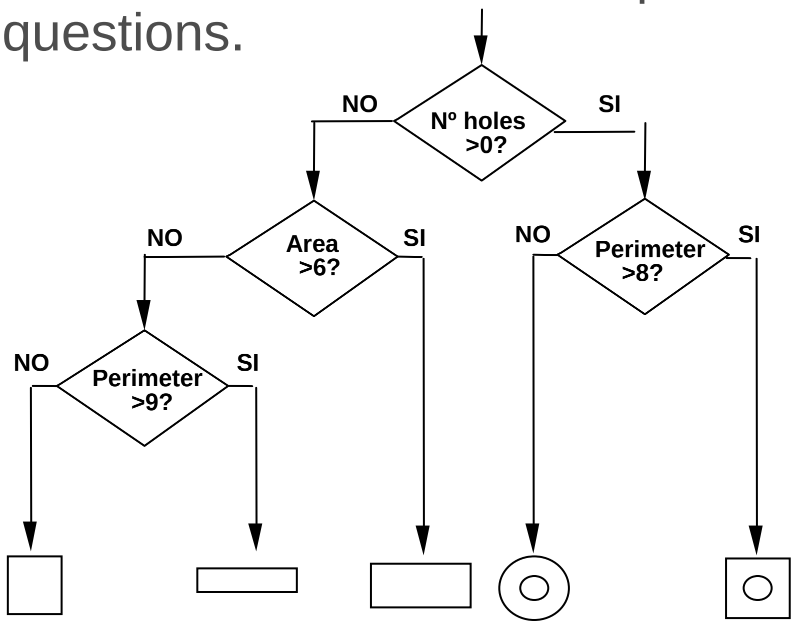
- σ_0^2: a priori estimation of variance (f. e. the square of 1% of the value of the descriptor)

- For a large N, if tends to be the classical variance



4. Decision trees

- Sequence of questions conditioned on answers to prior questions.



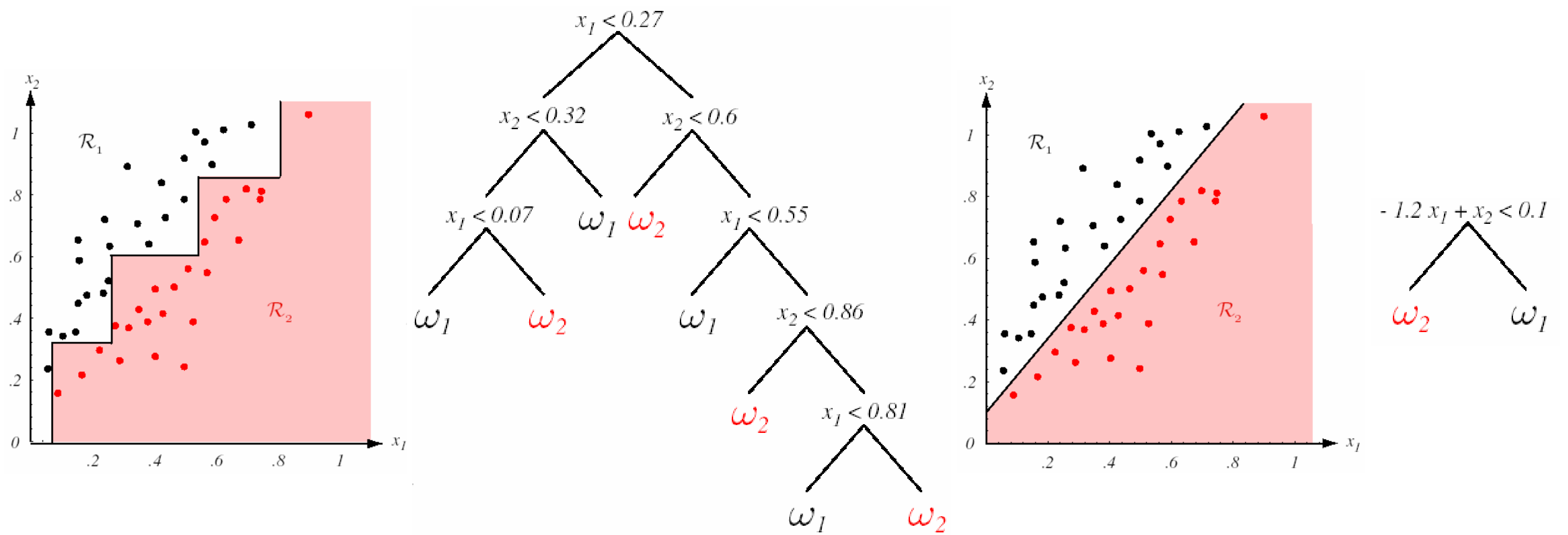
- **Non-metric method** (not based on measuring distances).
- **Advatages:**
  - Interpretability
  - Can include non numeric descriptors
  - Allows to incorporate expert knowledge

- **CART** (*Classification and Regression Trees*): Given a set of *m* samples of descriptors of *n* classes:
  - Recursive division of the samples according to a given descriptor.

1. Which factor to use for division?
2. Which descriptor to consult at each node?
3. When should division stop?
4. How to optimize a tree (making it smaller, simpler)?
5. What to do about *impure nodes*?
6. How to take into account unknown information?

Decision trees

- Importance of the selection of descriptors



- Disadvantages:
  - An error in one descriptor may preclude recognition of the object.
  - A class is always chosen (or you must introduce the unkown object in many places).

5. Other methods

- **Parametric methods** (all previous): estimation of parameters of function ***p(x/ω$_i$)*** characterizing each class and assumed known.

- **Non-parametric methods:**
  - Estimation of the distribution of probability:

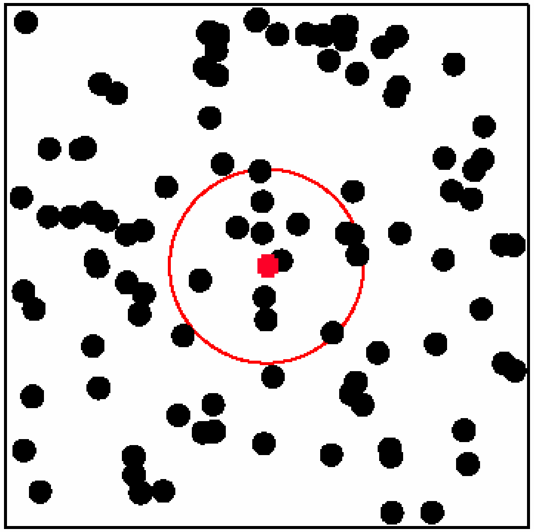
$$p(\mathbf{x}/\omega_i)?$$

  - Estimation of the *a posteriori* probability:

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

- ***A posteriori* probability:** Given a window *V* around ***x***, which includes *k* samples, k_i of which belong to ω_i:

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



Nearest Neighbor

- Given *N* samples of *m* descriptors belonging to *n* classes:

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

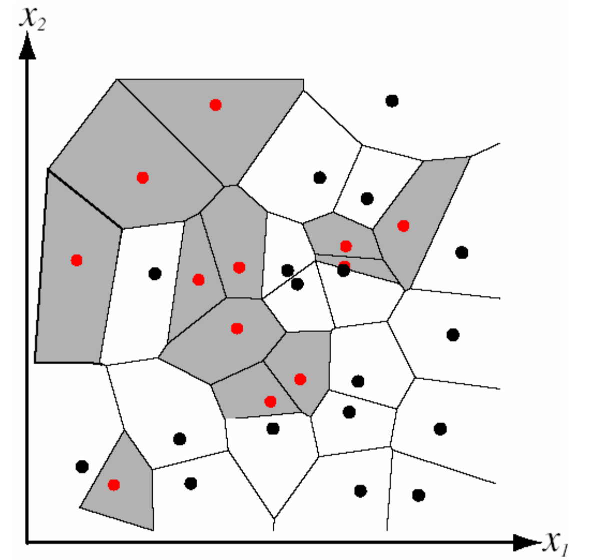
- Given *m* descriptors of the object to identify:

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

- Choose class ω_i such that:

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

- Use Voronoi tesellation

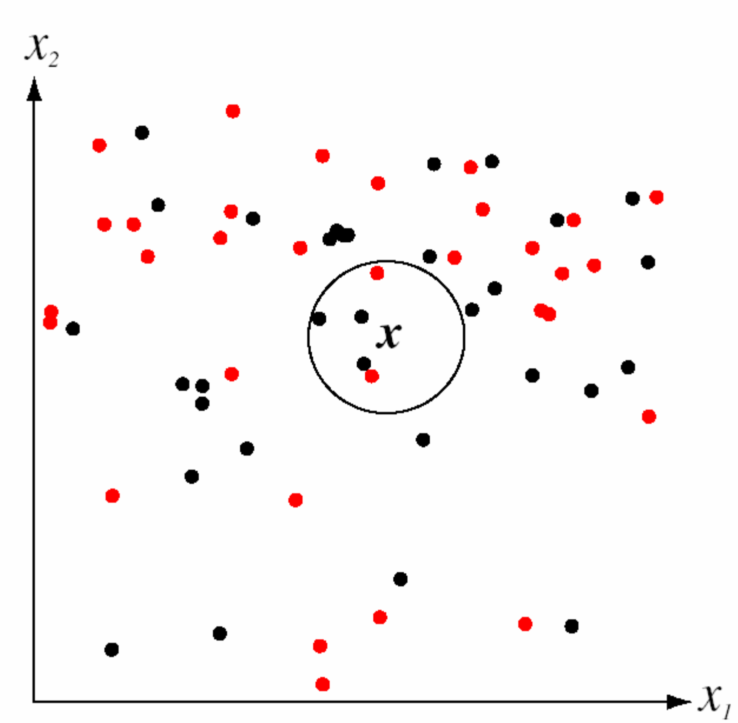


- **Suboptimal method**: error rate larger than Bayesian Classification.

¡The error rate doubles at most!

The *k* nearest neighbors

- Choose the class ω_i which is more frequent among the k nearest neighbors to $\mathbf{x}$.



- Ties can be avoided by incrementing *k* (*n*=2, *k* odd).

- For large *n*, *m* and *N*, it is computationally demanding.

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

- Partial distances:

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

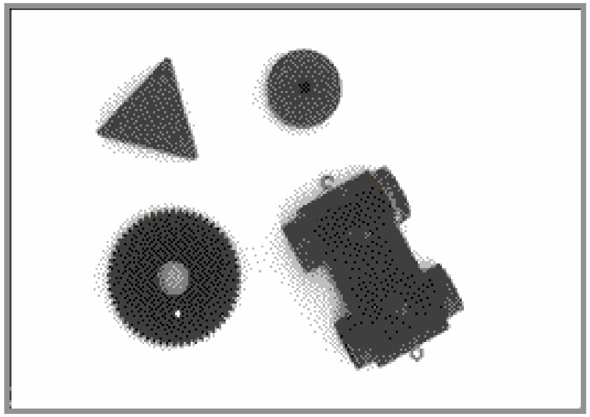
$$r \;<\; m$$

- The computation of a distance to a sample is abandoned when the partial distance is greater than the total distance to the *k*-th nearest neighbor.

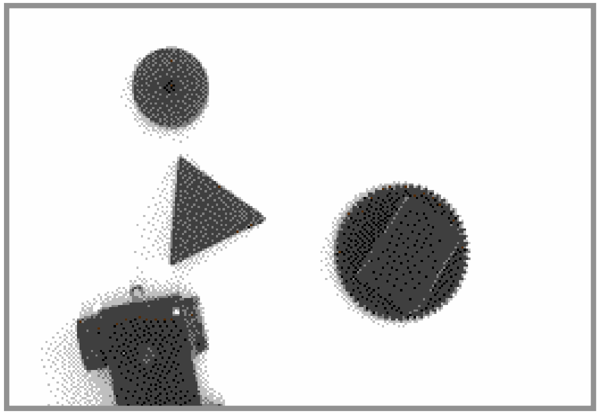
6. Conclusions

- Methods of **linear complexity**
- Descriptors are invariant in 2D

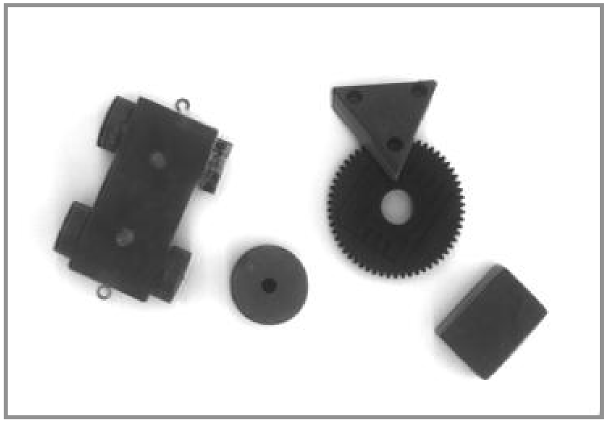
OK



Partially visible!



Overlap!



Geometric Recognition