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.

![Images of objects](image)

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

**Notation:**
- $n$ known objects (classes):
  $$ \omega = \{ \omega_1, \ldots, \omega_i, \ldots, \omega_n \} $$
- $m$ descriptors used:
  $$ x = \left( x_1, \ldots, x_j, \ldots, x_m \right)^T $$
- $N$ samples of descriptor $j$ of class $i$:
  $$ x_{ij} = \left( x_{ij1}, \ldots, x_{ijk}, \ldots, x_{ijn} \right)^T $$

**Given a sample $x$, which class $\omega_i$ does it correspond to?**

**Is it an unknown object?**

Different methods use different techniques to carry out this comparison.
2. Bayesian Classification

• Descriptors are considered random variables.

• Probabilistic model of each class:

\[ p(x/\omega_i) \]

• A priori probability of each class:

\[ P(\omega_i) \]

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

• Compute the probability that the sample comes from each class using Bayes' Rule.

\[ p(x/\omega_i) = p(x/\omega_j) \]

\[ P(\omega_i) = P(\omega_j) ? \]
Bayesian Classification

- **Bayes’ Rule:**

\[ P(\omega_i/x) = \frac{p(x/\omega_i)P(\omega_i)}{p(x)} \]

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

- It distributes the probability among all classes:

\[ \sum_{i=1}^{n} P(\omega_i/x) = 1 \]

- We decide \( \omega_i \) if:

\[ P(\omega_i/x) = \max_{j=1}^{n} P(\omega_j/x) \]
Bayesian Classification

• Influence of $P(\omega_i)$:

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

• Influence of $p(x|\omega_i)$:

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

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

• Example: area

\[ \hat{\mu} = 1990.7 \]
\[ \hat{\sigma} = 197.04 \]

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

\[ x \sim N(\mu, \sigma^2) \]
\[ p(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2} \left( \frac{x-\mu}{\sigma} \right)^2} \]

\[ Pr \left\{ \left| x - \mu \right| \leq 2\sigma \right\} \approx 0.95 \]
Descriptors

- **Sample mean**: representative value:

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

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

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

\[ x_{ij} \sim N(\hat{\mu}_{ij}, \hat{\sigma}^2_{ij}) \]
Independence of descriptors

- Independent:
  \[ \hat{\rho}_{ij} = 0 \]

- Linearly correlated:
  \[ \hat{\rho}_{ij} \approx -1 \]

Independent descriptors allow to discriminate more clearly between classes.
Bayesian Classification

• Assuming Gaussianity and statistical independence:

\[ p(x_i / \omega_i) = p(x_1 / \omega_i) p(x_2 / \omega_i) \cdots p(x_m / \omega_i) \]

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

• Iterative method:

\[ P^0(\omega_i) = P(\omega_i) \]

\[ P^d(\omega_i) = P(\omega_i / x_1, \ldots, x_d) \]

\[ = \frac{p(x_d / \omega_i) P^{d-1}(\omega_i)}{\sum_{j=1}^{m} p(x_d / \omega_j) P^{d-1}(\omega_j)} \]

• Iterate until the most probable object surpasses the threshold (computing all descriptors might not be necessary).
Sequential measurement of descriptors

• 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 classes $\omega_i$ and $\omega_j$:

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

• Example:

Can be computed a priori

<table>
<thead>
<tr>
<th></th>
<th>Círculo</th>
<th>GranCir</th>
<th>CirTruncad</th>
</tr>
</thead>
<tbody>
<tr>
<td>GranCir</td>
<td>63.47</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CirTruncad</td>
<td>1.94</td>
<td>102.64</td>
<td></td>
</tr>
<tr>
<td>CirSaliente</td>
<td>153.58</td>
<td>2.20</td>
<td>220.13</td>
</tr>
</tbody>
</table>

According to their J-divergence
Unkown object management

• Class $\omega_0$ representing the unknown object:

\[ p(x_j / \omega_0) = U(x_j^{\text{min}}, x_j^{\text{max}}) \]

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

\[
p(x_j / \omega_0) = \begin{cases} 
0, & x_j < x_j^{\text{min}} \\
\frac{1}{x_j^{\text{max}} - x_j^{\text{min}}}, & x_j^{\text{min}} \leq x_j \leq x_j^{\text{max}} \\
0, & x_j > x_j^{\text{max}}
\end{cases}
\]
3. Minimal distances

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

- **Euclidean distance:**
  
  \[ D^2(x, \omega_i) = \|x - \omega_i\|^2 = \sum_{j=1}^{m} (x_j - \mu_{i,j})^2 \]

- Choose the class of minimal distance (nearest neighbor).

- The decision frontier between two classes will be given:
  
  \[ H_{ij} : D^2(x, \omega_i) = D^2(x, \omega_j) \]
Exercise

- Given the following samples of descriptors $x_1$ and $x_2$ of two classes:

<table>
<thead>
<tr>
<th>Class 1</th>
<th>$x_1$</th>
<th>$x_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Class 2</th>
<th>$x_1$</th>
<th>$x_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>

- 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 it is 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:**
  \[ D^2(x, \omega_i) = \sum_{j=1}^{m} \frac{(x_j - \mu_{ij})^2}{\sigma_{ij}^2} \]

- 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 $\mu$):

\[
D^2(x, \omega_i) = \frac{(x_1 - \mu_{i1})^2}{\sigma_{i1}^2} + \frac{(x_2 - \mu_{i2})^2}{\sigma_{i2}^2} = k^2
\]

- $\omega_2$: $x_1$ 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:**
  - $\mu_{ij}$: expected value
  - $\sigma_{ij}$: expected deviation
  - $x_j$: observed value

  \[
  \frac{x_j - \mu_{ij}}{\sigma_{ij}} \sim N(0, 1)
  \]

  \[
  \frac{(x_j - \mu_{ij})^2}{\sigma_{ij}^2} \sim \chi^2_1
  \]

  \[
  D^2(x, \omega_i) \sim \chi^2_m
  \]

- $H_0$: $\{x$ comes from $\omega_i\}$
- $\alpha$: confidence level

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

### Chi-squared tables

<table>
<thead>
<tr>
<th>$m$</th>
<th>$\alpha$</th>
<th>0.05</th>
<th>0.025</th>
<th>0.01</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.84</td>
<td>5.02</td>
<td>6.64</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>5.99</td>
<td>7.38</td>
<td>9.22</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>7.82</td>
<td>9.36</td>
<td>11.32</td>
<td></td>
</tr>
</tbody>
</table>

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

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

- 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

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

- $\chi^2_{0.05(1)} = 3.84$

- Accepted/rejected

\[ \chi^2_{0.05(1)} \]

\[ 3.84 \]

\[ D^2 \]

\[ N(\mu_i, \sigma_i^2) \]

\[ 500 \quad 1000 \quad 1500 \quad 2000 \quad 2500 \quad 3000 \quad 3500 \]

- Euclidean .vs. Mahalanobis

\[ D^2 \]

\[ N(\mu_i, \sigma_i^2) \]

\[ \text{Euclidean distance} \]

\[ \text{Acceptance interval} \]

\[ \text{measurement} \]
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 \( \omega_i \) and \( \omega_k \) are indistinguishable.
Number of samples

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

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

$$
= \frac{\sigma^2_{0N}}{N} + \frac{N-1}{N} \sigma^2
$$

- $\sigma^2_0$: 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).

• **Advantages:**
  – 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
  – An error in one descriptor may preclude recognition of the object.
  – A class is always chosen (or you must introduce the unknown object in many places).
5. Other methods

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

- **Non-parametric methods:**
  - Estimation of the distribution of probability: $p(x/\omega_i)$.
  - Estimation of the *a posteriori* probability: $P(\omega_i/x)$.

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

$$P(\omega_i/x) \sim \frac{k_i}{k}$$
Nearest Neighbor

• Given $N$ samples of $m$ descriptors belonging to $n$ classes:
  \[ x_{i,k} = (x_{i,1}, \cdots, x_{i,N})^T \]

• Given $m$ descriptors of the object to identify:
  \[ x = (x_1, \cdots, x_m)^T \]

• Choose class $\omega_i$ such that:
  \[ x_{i,k} = \min_k D^2(x, x_{i,k}) \]

• Use Voronoi tessellation

¡The error rate doubles at most!

Suboptimal method: error rate larger than Bayesian Classification.
The $k$ nearest neighbors

- Choose the class $\omega_i$ which is more frequent among the $k$ nearest neighbors to $x$.

- 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:

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

  $$r < m$$

- 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