Introduction to Statistical Pattern Recognition

R.P.W. Duin
Pattern Recognition Group
Delft University of Technology
The Netherlands
prcourse@prtools.org

Pattern Recognition Problem
What is this? What occasion? Where are the faces? Who is who?

Pattern Recognition Problems
Which group?

Pattern Recognition Problems
To which class belongs an image
To which class (segment) belongs every pixel?
Where is an object of interest (detection);
What is it (classification)?

Pattern Recognition Books

Pattern Recognition: OCR
Kurzweil Reading Edge
Automatic text reading machine with speech synthesizer
Pattern Recognition: Pathology

Flow cytometry

Pattern Recognition: Seismics

Earthquakes

Pattern Recognition: Shape Recognition

Pattern Recognition is very often Shape Recognition:
- Images: B/W, grey value, color, 2D, 3D, 4D
- Time Signals
- Spectra

Pattern Recognition: Shapes

Examples of objects for different classes

Object of unknown class to be classified

A ? B

Pattern Recognition System

Sensor -> Representation -> Generalization

Feature Representation

Area
Perimeter

Pattern Recognition System

Sensor -> Representation -> Generalization

Pixel Representation

Pixel_1
Applications

- Biomedical: EEG, ECG, Röntgen, Nuclear, Tomography, Tissues, Cells, Chromosomes, Bio-informatics
- Speech Recognition, Speaker Identification.
- Character Recognition
- Signature Verification
- Remote Sensing,
- Meteorology
- Industrial Inspection
- Robot Vision
- Digital Microscopy

Notes: not always, but very often image based no strong models available

Possible Object Representations

- Measurement samples
- Feature vector
- Sets of segments or primitives
- Outline samples (shape)
- Symbolic structures
- (Attributed) graphs
- Dissimilarities

Requirements and Goals

- Set of Examples
- Physical Knowledge
- Representation
- Representation (Model)
- Pattern Recognition System
- Classification
- Minimize desired set of examples
- Minimize amount of explicit knowledge
- Minimize complexity of representation
- Minimize cost of recognition
- Minimize probability of classification error
Bayes decision rule, formal

\[ p(A|x) > p(B|x) \quad \Rightarrow \quad A \text{ else } B \]

Bayes:

\[
\frac{p(x|A) p(A)}{p(x)} > \frac{p(x|B) p(B)}{p(x)}
\]

\[ \Rightarrow \quad A \text{ else } B \]

2-class problems: \( S(x) = p(x|A) p(A) - p(x|B) p(B) > 0 \quad \Rightarrow \quad A \text{ else } B \)

n-class problems: Class(x) = \( \arg\max_{\omega} (p(x|\omega) p(\omega)) \)

Decisions based on densities

Length

\[ p\text{(length | male)} p\text{(male)} p\text{(length | female)} p\text{(female)} \]

What is the gender of somebody with this length?

Bayes:

\[
p\text{(female | length)} = \frac{p\text{(length | female)} p\text{(female)}}{p\text{(length)}} \]

\[
p\text{(male | length)} = \frac{p\text{(length | male)} p\text{(male)}}{p\text{(length)}} \]

Good features are discriminative.

Good features are have to be defined by experts

Sometimes to be found by statistics

e.g. Fisher Criterion:

\[ J_F(A, B) = \frac{\mu_A - \mu_B}{\sqrt{\frac{1}{2} \left( \frac{1}{\sigma_A^2} + \frac{1}{\sigma_B^2} \right)}} \]

Compactness

Representations of real world similar objects are close.

There is no ground for any generalization (induction) on representations that do not obey this demand.

(A.G. Arkedev and E.M. Braverman, Computers and Pattern Recognition, 1966.)

The compactness hypothesis is not sufficient for perfect classification as dissimilar objects may be close.

\( \Rightarrow \) class overlap

\( \Rightarrow \) probabilities
Distances and Densities

? to be classified as
B – because it is most close to an object A
A – because the local density of B is larger.

Distances: Scaling Problem

Before scaling: \( D(X,A) < D(X,B) \)
After scaling: \( D(X,A) > D(X,B) \)

How to Scale in Case of no Natural Features

Make variances equal:

\[
\frac{\text{color}}{\sqrt{\text{var(color)}}} = \frac{\text{perimeter}}{\sqrt{\text{var(perimeter)}}}
\]

Density Estimation:

What is the probability of finding an object of class A (B) on this place in the 2D space?
What is the probability of finding an object of class A (B) on this place in the 1D space?

The Gaussian distribution (3)

- Normal distribution = Gaussian distribution
- Standard normal distribution: \( \mu = 0, \sigma^2 = 1 \)
- 95% of data between \( [\mu - 2\sigma, \mu + 2\sigma] \) (in 1D!)

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

Multivariate Gaussians

\[
G = \begin{bmatrix} 3 & 1 \frac{1}{2} \\ 1 \frac{1}{2} & 2 \end{bmatrix}
\]

- \( k \)-dimensional density:
  \[
p(x) = \frac{1}{\sqrt{2\pi^k \det(G)}} \exp \left( -\frac{1}{2} (x-\mu)^T G^{-1}(x-\mu) \right)
\]
Introduction to Statistical Pattern Recognition

Multivariate Gaussians (2)

- \( G = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \)
- \( G = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \)
- \( G = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix} \)
- \( G = \begin{bmatrix} 3 & -1 \\ -1 & 1 \end{bmatrix} \)

Density estimation (1)

- The density is defined on the whole feature space.
- Around object \( x \), the density is defined as:
  \[
  p(x) = \frac{dP(x)}{dx} = \left( \frac{\text{fraction of objects}}{\text{volume}} \right)
  \]
- Given \( n \) measured objects, e.g. person’s height (m) how can we estimate \( p(x) \)?

Density estimation (2)

- **Parametric** estimation:
  - Assume a parameterized model, e.g. Gaussian
  - Estimate parameters from data
  - Resulting density is of the assumed form
- **Non parametric** estimation:
  - Assume no ‘formal’ structure/model, choose ‘approach’
  - Estimate density with chosen approach
  - Resulting density has no formal form

Parametric estimation

- Assume Gaussian model
- Estimate mean and covariance from data
  \[
  \hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i \quad G = \frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{\mu}) (x_i - \hat{\mu})
  \]

Nonparametric estimation

- **Histogram method**:
  1. Divide feature space in \( N \) bins
  2. Count the number of objects in each bin
  3. Normalize:
  \[
  \hat{p}(x) = \frac{\sum_{i,j=1}^{N_y} n_{ij} dx dy}{\sum_{i,j=1}^{N_y} n_{ij} dx dy}
  \]
- **Parzen density estimation** (1)
  - Fix volume of bin, vary positions of bins, add contribution of each bin
  - Define ‘bin’-shape (kernel):
  \[
  K(r) > 0 \quad \int K(r) dr = 1
  \]
  - For test object \( z \) sum all bins
  \[
  p(z) = \frac{1}{hn} \sum_{i} K \left( \frac{z - X_i}{h} \right)
  \]
Parzen density estimation (2)

- With Gaussian kernel:
  \[ K(x) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2\sigma^2}\right) \]

Parametric vs. Nonparametric

- **Parametric** estimation, based on some model:
  - Model parameters to estimate
  - More samples required than parameters
  - Model assumption could be incorrect resulting in erroneous conclusions

- **Non parametric** estimation, hangs on data directly:
  - Assume no "formal" structure/model
  - Almost no parameters to estimate
  - Erroneous estimates are less likely