BIM488 Introduction to Pattern Recognition

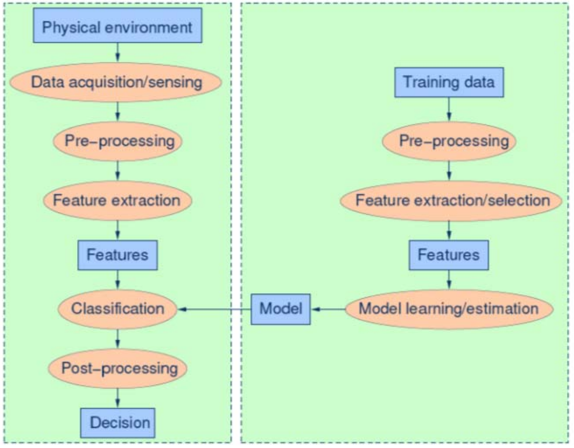
Classification Algorithms – Part I

Outline

- Introduction
- Bayes Decision Theory
- Bayesian Classifier
- Minimum Distance Classifiers
- Naive Bayes Classifier
- Nearest Neighbor (NN) Classifier

Introduction

Pattern Recognition System



Introduction

- There exist numerous classification algorithms.
- We are going to describe some of those classifiers in this course.

Bayes Decision Theory

- This chapter discusses classification techniques inspired by Bayes decision theory.
- In a classification task, we are given a pattern and the task is to classify it into one out of *M* classes.
- The number of classes is assumed to be known a priori.
- Each pattern is represented by a set of feature values which make up *l* dimensional feature vector, x.

$$\underline{x} = [x_1, x_2, \ldots, x_l]^T$$

- Each pattern is represented uniquely by a single feature vector and that it can belong to only one class.

Bayes Decision Theory

- Assign the pattern represented by feature vector x to the most probable of the available classes

$$\omega_1, \omega_2, \ldots, \omega_M$$

That is,

$$\underline{x} \rightarrow \omega_i : \underset{\text{maximum}}{P(\omega_i | \underline{x})}$$

- Probability that unknown pattern belongs to the respective class wi, given that corresponding feature vector takes the value x.

Bayes Decision Theory

Recall the Bayes rule (2-class case)

Posterior probability of class wi given x

Class conditional pdf of x given wi

$$p(\underline{x})P(\omega_i|\underline{x}) = p(\underline{x}|\omega_i)P(\omega_i) \Rightarrow$$

$$P(\omega_i|\underline{x}) = \frac{p(\underline{x}|\omega_i)P(\omega_i)}{p(\underline{x})}$$

Pdf of x

Prior probability of class wi

where

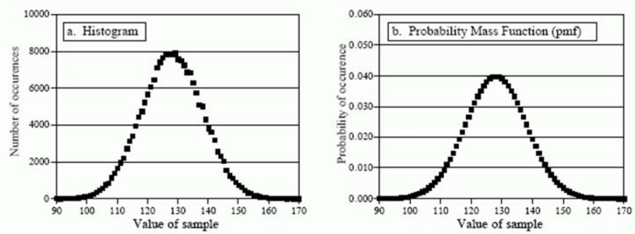
$$p(\underline{x}) = \sum_{i=1}^{2} p(\underline{x}|\omega_i)P(\omega_i)$$

Pdf: Probability Density Function

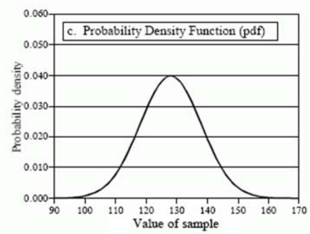
Bayes Decision Theory

- Probability P(.)
  - prior knowledge of how likely is to get a pattern
- Probability density function p(x)
  - how frequently we will measure a pattern with feature value x

Bayes Decision Theory



The relationship between (a) the histogram, (b) the probability mass function (pmf), and (c) the probability density function (pdf). The histogram is calculated from a finite number of samples. The pmf describes the probabilities of the underlying process. The pdf is similar to the pmf, but is used with continuous rather than discrete signals. Even though the vertical axis of (b) and (c) have the same values (0 to 0.06), this is only a coincidence of this example. The amplitude of these three curves is determined by: (a) the sum of the values in the histogram being equal to the number of samples in the signal; (b) the sum of the values in the pmf being equal to one, and (c) the area under the pdf curve being equal to one.



Bayesian Classifier

- The Bayesian classification rule:
  - Given $\underline{x}$ classify it to ω_i if:

$$P(\omega_i|x) > P(\omega_j|x), \quad \forall j \neq i$$

  - Since p(x) is the same for all classes,

$$p(x|\omega_i)P(\omega_i) > p(x|\omega_j)P(\omega_j), \quad \forall j \neq i$$

Bayesian Classifier

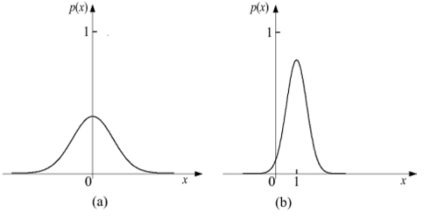
- Gaussian (Normal) pdf is extensively used in pattern recognition
- N(μ, Σ) notation is used to describe normal distribution
- In one dimensional case (single feature vector):

  μ = Mean value

  $\Sigma = \sigma^2$ = Variance
- In multidimensional case (multiple feature vectors):

  μ = Mean vector

  Σ = Covariance matrix



Bayesian Classifier

- The one-dimensional case:

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

- The Multivariate (Multidimensional) case:

$$p(\underline{x}) = \frac{1}{(2\pi)^{\frac{\ell}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} (\underline{x} - \underline{\mu})^T \Sigma^{-1} (\underline{x} - \underline{\mu}) \right)$$

Bayesian Classifier

Example . Compute the value of a Gaussian pdf, $\mathcal{N}(m, S)$, at $x_1 = [0.2, 1.3]^T$ and $x_2 = [2.2, -1.3]^T$, where

$$m = [0,\ 1]^T,\ \ S = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Solution.

```
m=[0 1]'; S=eye(2);
x1=[0.2 1.3]'; x2=[2.2 -1.3]';
pg1=comp_gauss_dens_val(m,S,x1);
pg2=comp_gauss_dens_val(m,S,x2);
```

The resulting values for $pg1$ and $pg2$ are 0.1491 and 0.001, respectively.

Bayesian Classifier

Example. Consider a 2-class classification task in the 2-dimensional space, where the data in both classes, ω_1, ω_2, are distributed according to the Gaussian distributions $\mathcal{N}(m_1, S_1)$ and $\mathcal{N}(m_2, S_2)$, respectively. Let

$$m_1 = [1,\ 1]^T, \quad m_2 = [3,\ 3]^T, \quad S_1 = S_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Assuming that $P(\omega_1) = P(\omega_2) = 1/2$, classify $x = [1.8,\ 1.8]^T$ into ω_1 or ω_2.

Solution.

```
P1=0.5;
P2=0.5;
m1=[1 1]'; m2=[3 3]'; S=eye(2); x=[1.8 1.8]';
p1=P1*comp_gauss_dens_val(m1,S,x);
p2=P2*comp_gauss_dens_val(m2,S,x);
```

The resulting values for $p1$ and $p2$ are 0.042 and 0.0189, respectively, and x is classified to ω_1 according to the Bayesian classifier.

Minimum Distance Classifiers

1. The Euclidean Distance Classifier
2. The Mahalanobis Distance Classifier

Minimum Distance Classifiers

The optimal Bayesian classifier is significantly simplified under the following assumptions:

- The classes are equiprobable.
- The data in *all* classes follow Gaussian distributions.
- The covariance matrix is the *same* for all classes.
- The covariance matrix is diagonal and *all* elements across the diagonal are *equal*. That is, $S = \sigma^2 I$, where I is the identity matrix.

Minimum Distance Classifiers

- Under these assumptions, it turns out that the optimal Bayesian classifier is equivalent to the minimum Euclidean distance classifier.

- That is, given an unknown *x*, assign it to class ωi if

$$||x - m_i|| \equiv \sqrt{(x - m_i)^T (x - m_i)} < ||x - m_j||, \quad \forall i \neq j$$

$m1 = mean\ of\ c1,$
$m2 = mean\ of\ c2$
$x = ?$

$d1 = (x - m1)^T (x - m1)$

$d2 = (x - m2)^T (x - m2)$

$if\ d1 > d2$
$\quad x\ belongs\ to\ c1$
$else$
$\quad x\ belongs\ to\ c2$

Minimum Distance Classifiers

- If one relaxes the assumptions required by the Euclidean classifier and removes the last one, the one requiring the covariance matrix to be diagonal and with equal elements, the optimal Bayesian classifier becomes equivalent to the minimum Mahalanobis distance classifier.

- That is, given an unknown *x*, it is assigned to class *ωi* if

$$\sqrt{(x-m_i)^T S^{-1}(x-m_i)} < \sqrt{(x-m_j)^T S^{-1}(x-m_j)}, \quad \forall j \neq i$$

- where *S* is the common covariance matrix.

Minimum Distance Classifiers

Example. Consider a 2-class classification task in the 3-dimensional space, where the two classes, ω_1 and ω_2, are modeled by Gaussian distributions with means $m_1 = [0, 0, 0]^T$ and $m_2 = [0.5, 0.5, 0.5]^T$, respectively. Assume the two classes to be equiprobable. The covariance matrix for both distributions is

$$S = \begin{bmatrix} 0.8 & 0.01 & 0.01 \\ 0.01 & 0.2 & 0.01 \\ 0.01 & 0.01 & 0.2 \end{bmatrix}$$

Given the point $x = [0.1, 0.5, 0.1]^T$, classify x (1) according to the Euclidean distance classifier and (2) according to the Mahalanobis distance classifier. Comment on the results.

Minimum Distance Classifiers

Solution. Take the following steps:

Step 1. Use the function *euclidean_classifier* by typing

```
x=[0.1 0.5 0.1]';
m1=[0 0 0]'; m2=[0.5 0.5 0.5]';
m=[m1 m2];
z=euclidean_classifier(m,x)
```

The answer is $z = 1$; that is, the point is classified to the ω_1 class.

Minimum Distance Classifiers

Step 2. Use the function *mahalanobis_classifier* by typing

```
x=[0.1 0.5 0.1]';
m1=[0 0 0]'; m2=[0.5 0.5 0.5]';
m=[m1 m2];
S=[0.8 0.01 0.01;0.01 0.2 0.01; 0.01 0.01 0.2];
z=mahalanobis_classifier(m,S,x);
```

This time, the answer is $z = 2$, meaning the point is classified to the second class. For this case, the optimal Bayesian classifier is realized by the Mahalanobis distance classifier. The point is assigned to class ω_2 in spite of the fact that it lies closer to m_1 according to the Euclidean norm.

Naive Bayes Classifier

- In the Naive Bayes classification scheme, the required estimate of the pdf at a point *x* is computed by

$$p(x) = \prod_{j=1}^{l} p(x(j))$$

- That is, the components (features) of the feature vector *x* are assumed to be *statistically independent*.
- For example, a fruit may be considered to be an apple if it is red, round, and about 4" in diameter. Even if these features depend on each other or upon the existence of the other features, a naive Bayes classifier considers all of these properties to independently contribute to the probability that this fruit is an apple.

Naive Bayes Classifier

- Decision rule (similar to Bayes classification):

$$p(x|\omega_i)P(\omega_i) > p(x|\omega_j)P(\omega_j), \quad \forall j \neq i$$

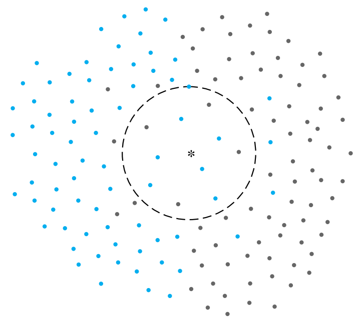
Nearest Neighbor (NN) Classifier

- Nearest neighbor (NN) is one of the most popular classification rules.
- We are given c classes, ωi , $i = 1, 2, . . . , c$, and a point x, and N *training* points, xi , $i = 1, 2, . . .,N$, in the l-dimensional space, with the corresponding class labels.
- Given a point, x, whose class label is unknown, the task is to classify x in one of the c classes. The rule consists of the following steps:

Nearest Neighbor (NN) Classifier

1. Among the N training points, search for the k neighbors closest to x using a distance measure (e.g., Euclidean, Mahalanobis). The parameter k is user-defined. Note that it should not be a multiple of c. That is, for two classes k should be an odd number.
2. Out of the k-closest neighbors, identify the number, k_i, of the points that belong to class ω_i.
3. Assign x to class ω_i, for which $k_i > k_j$, $j = i$. In other words, x is assigned to the class in which the majority of the k-closest neighbors belong.

Nearest Neighbor (NN) Classifier



Example: For k=11 → 11-NN Classification

Summary

- Introduction
- Bayes Decision Theory
- Bayesian Classifier
- Minimum Distance Classifiers
- Naive Bayes Classifier
- Nearest Neighbor (NN) Classifier

References

- S. Theodoridis, A. Pikrakis, K. Koutroumbas, D. Cavouras, *Introduction to Pattern Recognition: A MATLAB Approach,* Academic Press, 2010.
- S. Theodoridis and K. Koutroumbas, *Pattern Recognition* (4th Edition), Academic Press, 2009.