ELE888/EE8209 Intelligent Systems (2015)
Lab 1: Bayesian Decision Theory

Objective

To understand and implement Bayes decision rule for performing simple classifications

Background

Bayesian decision theory is a fundamental statistical approach to the problem of classification. It makes the assumption that the decision problem is posed in probabilistic terms, and that all the relevant probability values are known. The Bayes formula states that the posterior probability of a state of nature or class (j) given the observation or feature (x) can be computed using the prior probability of the class (ω_j) and the class-conditional probability of feature (x) given the class (ω_j). According to the Bayes decision rule, the feature (x) is then assigned to the class(ω_j) with the highest posterior probability. The formal definition of the Bayes formula is given by the Eq. 1 as follows:

Bayes Formula

$$P(\omega_j \mid x) = \frac{p(x \mid \omega_j) \cdot P(\omega_j)}{p(x)} \tag{1}$$

where in case of c classes

$$p(x) = \sum_{j=1}^{c} p(x \mid \omega_j) \cdot P(\omega_j) \tag{2}$$

$P(\omega_j \mid x)$ = Posterior probability of class (ω_j) given the feature (x).
$p(x \mid \omega_j)$ = Class-conditional prob. density of feature (x) given the class (ω_j).
$P(\omega_j)$ = Prior probability of class (ω_j).
$p(x)$ = Probability density function of the feature (x).

In general for a multi-dimensional feature input $\mathbf{x} = [x_1\ , x_2\ , ..., x_k\ ]$, the Bayes formula could be written as:

$$P(\omega_j \mid \mathbf{x}) = \frac{p(\mathbf{x} \mid \omega_j) \cdot P(\omega_j)}{p(\mathbf{x})} \tag{3}$$

Laboratory Exercises

Part I

The *lab1.zip* file contains this handout, data *irisdata.mat*, a MATLAB script file *runlab1.m* and a skeleton MATLAB program *lab1.m*. Open *runlab1.m* with the MATLAB editor and run it. This script loads 2 matrix variables `irisdata_features` and `irisdata_labels` to the MATLAB workspace, corresponding to the IRIS data set investigated in Lab0.

The matrix `irisdata_features` is a double array of size 150 × 4 (i.e. 150 rows and 4 columns), where each row represents an individual sample, and each column represents a particular feature. The features for the IRIS data set (columns 1-4 respectively) are: [x_1 = *Sepal Length (cm);* x_2 = *Sepal Width (cm);* x_3 = *Petal Length (cm);* x_4 = *Petal Width (cm)*]. The matrix `irisdata_labels` is a vector of strings, size 150 × 1 (stored in cell array format). Effectively, each row in `irisdata_labels` supplies the label (class name) for the corresponding row (sample) in `irisdata_features`.

In this lab, we will constrain our attention to only 2 of these classes [ω_1 = *Iris Setosa*; ω_2 = *Iris Versicolour*].

runlab1.m preprocesses the IRIS labels to generate a set of *numeric* labels (stored in `numericLabels`). The script produces a figure that shows histogram (distribution) of feature x_1, for each of these two classes. A new matrix `trainingSet`, is then constructed which omits the 3rd class and merges the numeric labels with the features. Please familiarise yourself with the techniques used to preprocess this data.

The class-conditional probability densities of each x_i are assumed to be Gaussian. The general from of a univariate Gaussian density is given by:

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \tag{4}$$

where μ and σ[1] are mean and standard deviation of the density function.

1. Using a single discriminant function g(x_2), design a 2 class minimum-error-rate classifier(dichotomizer) from the given data, to classify IRIS samples into either *Iris Setosa* or *Iris Versicolour*, according to the feature: *Sepal Width.*

2. Using the shell program *lab1.m*, write a program that will take an individual sample value as the input and will return the posterior probabilities and the value of g(x_2)

3. Identify the class labels for the feature values using your program, and indicate their respective posterior probabilities and discriminant function values: x_1 = [3.3, 4.4, 5.0, 5.7, 6.3]

4. Arrive at a optimal threshold (T_{h1}) that separates classes ω_1 and ω_2 (theoretically or experimentally. Justify your result.

5. Suggest how T_{h1} would be affected if a higher penalty is associated with classifying class ω_2 as class ω_1 - show with experiment.

6. Adjust your program to accept *Sepal Length* as the discriminating feature g(x_1). Suggest which of the two features (x_1,x_2) might be a better choice for separating the two classes ω_1 and ω_2. Justify.

[1] In MatLab coding, use the default Bessel's correction command "std(X)", use "Help std" for detailed information.

Part II (Optional)

This is a bonus part and is used to recover marks lost in Part I (TA will decide). Students are encouraged to do this part to get a better understanding of multivariate data. For this part of the lab we will use both features x_1 and x_2.

The class-conditional probability densities of x are now assumed to be multivariate Gaussian. The general form of a multivariate Gaussian density is given by:

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|^{0.5}} \exp\left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right] \tag{5}$$

where x is a d component vector, $\boldsymbol{\mu}$ is the d component mean vector, Σ is the d by d covariance matrix, and $|\Sigma|$ and Σ^{-1} are its determinant and inverse.

1. Using both the features [x_1 , x_2] and a single discriminant function $g(x)$, design a 2 class minimum-error-rate classifier(dichotomizer) from the given data.
2. Write a program that will take the feature vector as input and will return the posterior probabilities and the value of $g(x)$.
3. Plot the feature space using the two features.
4. Identify and verify the class labels for the following feature vectors: [2 6]; [4.4 3]; [5 3.5]; [5.3 2]; [5.5 2.5]; [6.6 3.5]; [4.5 6.1]

What to Submit?

1. Documented MATLAB code for Part I of the lab exercise. The programs should be demonstrated to the TA. (5 marks)
2. A report including the answers to the questions asked in Part I, your observations, plots, and conclusions. (5 marks)