Name:

STN:

In The Name of Almighty
ECE Dept. Isfahan University of Technology
Statistical Pattern Recognition
Computer Assignment #1

Exercise 1:

Computer Exercise 2.5 of text book (page 85)

Exercise 2:

In this exercise, you are to design and test the BME classifier for three different data sets. These datasets referred at as *Iris*, *Liquid* and *Normal*. These data sets are described in the following:

Iris Data: This data is for three species of iris. Each iris sample is characterized by four features, Petal length, Sepal length, Sepal width, and Petal width. There are 50 samples per class for a total of 150 samples. The database is arranged as one sample per line for a total of 150 lines. The first 50 lines are samples from class 1, the next 50 are from class 2 and the last 50 are from class 3. Each row contains five entries for: Class, Petal Length, Sepal Length, Sepal Width and Petal Width.

Liquid Data: The Liquid database is the result of a chemical analysis performed on Liquids produced in a factory. This database contains 178 6-dimensional features belonging to three classes. The data in the first column is the class type and the rest of 6 columns are 6 features representing its chemical. The numbers of samples in the three classes are 59, 71, and 48, respectively.

Normal Data: This data is simulated from two bivariate Normal densities with the following parameters:

$$\boldsymbol{\mu}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \boldsymbol{\Sigma}_1 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \qquad \boldsymbol{\mu}_2 = \begin{bmatrix} 4 \\ 0 \end{bmatrix}, \boldsymbol{\Sigma}_2 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

Two files are provided for this database. One file is for training and the other one is for testing. Each file contains a total of 1000 samples with 500 samples from each class, respectively. For each row, the data is column 1 is the class and the following two numbers are the two features.

For this assignment, assume that the feature vectors of each class have a multivariate normal pdf with unknown mean and covariance matrices. Use the maximum likelihood estimate for estimation of the parameters of the pdf. Also assume that the a priori probabilities of all classes are equal. For all the cases, you are to design the Bayes Minimum Error (BME) classifier.

a) For the Iris Data, evaluate the performance of the BME classifier using the leave-one-out method. This means that you should design 150 different classifiers and each

time use the available 149 training samples to estimate the required parameters. Then the one left out sample is to be classified by this designed classifier.

For this study provide the following information:

- The estimates of the mean and covariance matrices of all three classes for the first classifier, i.e. when the first sample in the database is left out.
- In the leave-one-out procedure, provide the following information for each misclassified sample: (1) sample number (out of 50), (2) actual class, (3) class to which it is misclassified, (4) value of discriminant function for actual class, (5) value of discriminant function for the class it is classifying to. For instance: Sample no.7 from class 1 incorrectly classified to class 3 with g_1= … and g_3= ….
- Provide the confusion matrix for the classifier after all 150 samples are tested. Label the row and column number of this matrix. Compute error rates for each class.

b) Repeat ***a*** for the Liquid database. In this case, you will be building 178 different BME classifiers. Provide the same kind of information as in 1.

c) For the Normal database, design the BME classifier using the Training file and then test the designed classifier on the Testing file.

- Provide a two-dimensional graph showing all the Testing and Training samples using two different symbols to show samples of class1 and 2, respectively. Show the misclassified Testing samples from each class with a different color.
- List all the misclassified samples
- Find the empirical error rate for each class and the overall error rate
- Compute the theoretical error rate and compare it with the empirical rate.