**CS 557 STATISTICAL PATTERN RECOGNITION AND LEARNING**

**FALL 2014**

**ASSIGNMENT 4**

**DUE: Saturday, 18th October, 2014.**

**PROBLEM**

1. Read the dataset trainSynthetic.txt and testSynthetic.txt. The last columns in both files are the labels. The first two columns are the attributes. This is synthetic data generated in Matlab by using a non-linear function to map the data points to classes.

2. Make a scatter plot of all the points of the dataset. Make sure you use a different symbol for class 1 and a different symbol for class 2 so that you can clearly see the class boundary.

3. Classify all the points comprising the testSynthetic.txt using the K-nearest neighbor classifier.

Run your simulation for different values of K (interval size) and fill out the following table. Also, show the points with errors in your plot for one value of K. So you need to add two plots in your report. One for step 2 and one for errors for one value of K.

Compute the average error rate as

average ((total errors class 1)/(total of class 1), (total errors class 2)/(total of class 2))

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Method** | **K** | **Total errors of class 1** | **Total errors of class 2** | **Average error rate** |
| Nearest Neighbor | 1 |  |  |  |
|  | 3 |  |  |  |
|  | 9 |  |  |  |
|  | 15 |  |  |  |
|  | 21 |  |  |  |
|  | 101 |  |  |  |
|  | 2 values of K your choice |  |  |  |

**TO SUBMIT**

1. Make a folder with your roll number as folder name. Put Matlab’s source code in it and place it in the ‘submit assign4’ folder on xeon. PLEASE DO NOT EMAIL
2. **Hard** **copy** of a report which is **not more than one page** long that describes your experiments and your observations. It must contain the table above and the plot for the dataset of the two classes and one plot of errors for one value of K. YOUR comments and conclusions about this method.