Assignment 1

Note:

This assignment should be done by each student individually. You can discuss it in general terms with other students; however, the files you hand in, e.g., the report and codes should be your own. If I find your reports/codes are the same as or similar to the others, both of you cannot get the scores for this assignment.

1. Problem Statement

In this assignment, you have one data set with 3000 samples. You can randomly select 2500 samples as training data, and 500 samples as test data. Then, you are required to classify the test data using Bayesian Theorem. To estimate the densities, you need to use one parameter estimation (e.g., MLE or Bayesian estimation) and one non-parameter estimation method (e.g., KNN or Parzen window), respectively (Select one you like). For each model, you should give the detail that you have to consider, e.g., how do you select the parameters in the model. Finally, you should list and analyze the results of each model including the test errors and the impact of the parameters, etc.

By Xin-Shun Xu *November 14, 2013*

2. Submissions

Finally, you are required to submit the following files:

- A report which describes the details of your implementation
- The codes that are written with good style. I suggest to use Matlab; but if you like, you can also write them in other programming languages, e.g., Python, R, Java, C etc.

3. About the data

The file **data1.mat** contains all the samples. In which, each sample is 3 dimensions, and the first two variables comprise the input x, and the last one is its label.

4. Deadline

All files should be submitted before Dec. 1, 2013.