**CS 557 STATISTICAL PATTERN RECOGNITION AND LEARNING**

**FALL 2015**

**ASSIGNMENT 3**

**DUE: Saturday, 03 October, 2015.**

**PROBLEM**

1. Read the dataset sylvester.mat in Matlab. This is the ecology dataset and modified from the unsupervised and transfer learning challenge. Please see:

<http://www.causality.inf.ethz.ch/unsupervised-learning.php>

2. The labels of the training points and test points are given. Classify each point in the test file using the following three methods based on MAP:

* Naive density estimate
* Kernel density estimate
* K-nearest neighbor algorithm

3. Note: You will estimate the density using the training points and their labels only. The labels of the test set are provided so that you can compute BER.

4. For density estimation, make a plot of the balanced error rate for different values of h (at least 5) and plot the BER for various h values. For K-nearest neighbor plot the BER for different K values.

5. Note: For kernel density if you get NaN or Inf then take the log for probability values to avoid underflow/overflow errors.

**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 assign3’ folder on xeon. PLEASE DO NOT EMAIL
2. **Hard** **copy** of a report which is **not more than two pages** long that describes your experiments and your results along with graphs. It must contain the table of results as well the graph and YOUR comments and conclusions about the three methods.