August-December 2016 Semester
CS6690: Pattern Recognition
Section 1: B.Tech. and Dual Degree Students
Programming Assignment I

Date: 14th September, 2016
Note: Each batch of students must use the datasets identified for that batch. The total data of a class is to be divided into 3 parts as follows: 70% as training data , 15% as validation data, and the remaining 15% as test data.

Dataset 1: 2-dimensional artificial data of 3 or 4 classes:
(a) Linearly separable data set for static pattern classification
(b) Nonlinearly separable data set for static pattern classification
(c) Overlapping data set for static pattern classification
(d) Trajectory data set for sequential pattern classification

Dataset 2: Real world data sets:
(a) Image classification data set for static pattern classification
(b) Speaker identification and verification data set for varying length pattern (Set of local feature vectors representation) classification
(c) Isolated word recognition data set for sequential pattern classification

Classifiers to be built for static pattern datasets (a), (b) and (c) in Dataset 1 :

1. K-Nearest Neighbour classifier
2. Naive-Bayes classifier
   a. Covariance matrix for all the classes is the same and is $\sigma^2 I$
   b. Covariance matrix for all the classes is the same and is C
   c. Covariance matrix for each class is different
3. Bayes classifier
   a. Covariance matrix for all the classes is the same and is C
   b. Covariance matrix for each class is different

Classifiers to be built for datasets (a) and (b) in Dataset 2:

1. Naive-Bayes classifier
2. Bayes classifier

Classifiers to be built for dataset (d) in Dataset 1 and dataset (c) in Dataset 2 :

1. Discrete HMM based classifier
2. Continuous density HMM based classifier

Report should include the results of studies presented in the following forms for each classifier and for each dataset:

1. Classification accuracy on test data
2. Confusion matrix based on the performance for test data
3. Decision region plots with the training data superposed for datasets (a), (b) and (c) in Dataset 1
4. FRR vs FAR plots for speaker verification data set

Report should also include your observations about the performance and the nature of decision surface for each classifier, and for each dataset.

Deadline for submission of report: **5.00PM, Thursday, 13th October 2016**