

**July-November 2018 Semester**  
**CS5691: Pattern recognition and Machine Learning**  
**Programming Assignment II**

Date: **October 11, 2018**

Deadline for submission of PDF file of report: **Wednesday, October 31, 2018**

**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

**Dataset 2:** Real world data sets:

- (a) Data set for static pattern classification
- (b) Image data set for varying length pattern (Set of local feature vectors representation) classification

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

1. K-Nearest Neighbour classifier
2. Naive-Bayes classifier with a Gaussian distribution for each class
  - 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 with a Gaussian distribution for each class
  - a. Covariance matrix for all the classes is the same and is  $C$
  - b. Covariance matrix for each class is different
4. Naive-Bayes classifier with a GMM for each class
5. Bayes classifier with a GMM for each class
6. Bayes classifier with K-nearest neighbours method for estimation of class-conditional probability density function

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

1. Naive-Bayes classifier with a GMM for each class
2. Bayes classifier with a GMM for each class
3. Bayes classifier with K-nearest neighbours method for estimation of class-conditional probability density function

**Use the cross-validation method to choose the best values of hyperparameters.**

**Report should include the following for each classifier and for each dataset:**

1. Table of classification accuracies of model on training data and validation data for different values of hyperparameter
2. Classification accuracy of the best configuration of the model on test data
3. Confusion matrix for the best configuration of model, on training data and test data
4. Decision region plots for the best configuration of the model, for datasets (a), (b) and (c) in **Dataset 1**. Superpose the training data on the decision region plot.

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