

February-May 2021 Semester
CS5691: Pattern recognition and Machine Learning
Programming Assignment II

Date: **27th March, 2021**

Deadline for submission of PDF file of report: **Sunday, 18th April, 2021**

Dataset 1: 2-dimensional artificial data:

- (a) Linearly separable data set for static pattern classification
- (b) Nonlinearly separable data set for static pattern classification

Dataset 2: Real world data sets:

- (a) Image 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 Dataset 1(a) :

1. K-nearest neighbours classifier, for $K=1$, $K=7$ and $K=15$
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

Classifiers to be built for Dataset 1(b) :

1. K-nearest neighbours classifier, for $K=1$, $K=7$ and $K=15$
2. Bayes classifier with a GMM for each class, using full covariance matrices
3. Bayes classifier with a GMM for each class, using diagonal covariance matrices
4. Bayes classifier with K-nearest neighbours method for estimation of class-conditional probability density function, for $K=10$ and $K=20$

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

1. Bayes classifier with a GMM for each class, using full covariance matrices
2. Bayes classifier with a GMM for each class, using diagonal covariance matrices

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 the 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 the model, on training data and test data
4. Decision region plots for the best configuration of the model, for Datasets 1(a) and 1(b). Superpose the training data on the decision region plot. For the Bayes classifiers using Gaussian distributions or GMMs, superpose the plots of level curves on the training data.

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