

July-November 2023 Semester
CS5691: Pattern recognition and Machine Learning
Programming Assignment II

Date: **5th October, 2023**

Deadline for submission: **Thursday, 26th October, 2023**

Dataset 1: 2-dimensional artificial data:

- (a) Linearly separable data set for 2 classes
- (b) Nonlinearly separable data set for 2 classes

Dataset 2: Image data set for 5 classes. (Dimension of feature vector is 81)

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 every class
 - a. Covariance matrix for both the classes is the same
 - b. Covariance matrices for two classes are different
3. Bayes classifier with hypersphere based Parzen window method for density estimation.
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 Dataset 1(b) and Dataset 2:

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

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

Report should include the following for every classifier and for every 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.