July-November 2025 Semester
CS5691: Pattern recognition and Machine Learning
Programming Assignment II

Date:**26th September, 2025**

Deadline for submission of report: **5PM on Wednesday, 1st October, 2025**

Dataset 1: 2-dimensional data: Linearly separable data set for 3 classes
Dataset 2: 2-dimensional data: Nonlinearly separable data set for 2 classes
Dataset 3: Image data set (Dimension of feature vector: 36) for 5 classes

Classifiers for Dataset 1:

1. K-nearest neighbours classifier, for K=1, K=5 and K=9
2. Bayes classifier with a Gaussian distribution for every class
    a. Covariance matrices for all the classes are the same
    b. Covariance matrices are different

Classifiers for Dataset 2:

1. K-nearest neighbours classifier, for K=1, K=5 and K=9
2. Bayes classifier with a Gaussian distribution for every class
    a. Covariance matrices for all the classes are the same
    b. Covariance matrices are different
3. Naive-Bayes classifier with a Gaussian distribution for every class
    a. Covariance matrices for all the classes are the same
    b. Covariance matrices are different
4. GMM based classifier (Q = 4, 6, 8, 10) using
    a. Full covariance matrices
    b. Diagonal covariance matrices

Classifiers for Dataset 3:

1. K-nearest neighbours classifier, for K=1, K=9 and K=15
2. Bayes classifier with a Gaussian distribution for every class
3. Naive-Bayes classifier with a Gaussian distribution for every class
4. GMM based classifier (Q = 2, 3, 4, 5) using
    a. Full covariance matrices
    b. Diagonal covariance matrices

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

Report should include the following:

1. Table of classification accuracies on training data, validation data and test data, for each of the datasets.
2. Confusion matrix for the best configuration of the model, on training data and test data, for each of the datasets.
3. Precision, recall and F1 score measures for each of the classes, and the average precision, average recall and average F1 score, for the best configuration of the model and for each of the datasets.
4. Decision region plots for the best configuration of the model for each of the classifiers for Datasets 1 and 2. Superpose the training data on the decision region plot. For the Bayes classifiers, Naïve-Bayes classifiers, and GMM based classifiers, superpose the plots of level curves on the training data.