July-November 2024 Semester
CS5691: Pattern recognition and Machine Learning
Programming Assignment IV

Date:**30th October, 2024**
Deadline for submission of report: **11.59PM on Saturday, 2nd November, 2024**

Dataset 1: 2-d data: Linearly separable data (Same as Dataset 1 of Assignment 2)
Dataset 2: 2-d data: Nonlinearly separable data for 2 classes (Same as Dataset 2 of Assignment 2)
Dataset 3: Image data (Dimension of feature vector: 35) for 5 classes (Same as Dataset 3 of Assignment 2)

Exercises:

1. Classifier for Dataset 1 using linear kernel based SVMs

2. Classifier for Dataset 2 using (a) polynomial kernel based SVM and (b) Gaussian kernel based SVM. (Use 4 different values of kernel hyperparameter : (a) Degree for polynomial kernel and (b) Kernel width for Gaussian kernel). (Use values of C as 1, 10, 100)

3. Classifier for Dataset 3 using (a) polynomial kernel based SVMs and (b) Gaussian kernel based SVMs. (Use 4 different values of kernel hyperparameter: (a) Degree for polynomial kernel and (b) Kernel width for Gaussian kernel). (Use values of C as 1, 10, 100)

4. Perform PCA for Dataset 3 and identify a suitable value of the reduced dimension l.

5. Classifier for Dataset 3 with reduced dimension representation obtained in Exercise 4, as the input to the classifier. The classification models are as follows: (a) GMM, (b) MLFFNN and (c) SVM. The configurations of GMM and MLFFNN are the same as in Assignment 3. The SVM is the best performing model in Exercise 3.

Report should include the following:
1. Datasets 1, 2 and 3: A table of classification accuracies for the training data, validation data and the test data, for each of the classifiers.
2. Datasets 1, 2 and 3: Confusion matrices for training data and test data for each of the classifiers in Exercises 1 and 5, and for the best models in Exercises 2 and 3.
3. Decision region plot for classifier in Exercise 1**.** Superpose the training data on the decision region plot. Mark the support vectors.
4. Decision region plots for Dataset 2: Plots for the best performing model of (a) polynomial kernel based SVM, and (b) Gaussian kernel based SVM. Superpose the training data on the decision region plots. Mark the bounded and unbounded support vectors.
5. Dataset 3: A table of percentage of bounded and unbounded support vectors out of the training examples, for the best performing model of (a) polynomial kernel based SVM, and (b) Gaussian kernel based SVM, in Exercises 3 and 5.
6. Exercise 4: Plot of Cumulative Variance (σ_{cl}^2) vs l