

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

Date: **November 7, 2018**

Deadline for submission of PDF file of report: **Wednesday, December 5, 2018**

Dataset 1: Data for sequential pattern classification

- (a) On-line handwritten character data
- (b) Spoken digit data

Dataset 2: 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

Dataset 3: Image data set

Classifier to be built for Dataset 1 : Discrete HMM for Datasets 1(a) and 1(b) (Use vector quantization method for converting a sequence of feature vectors to a sequence of cluster indices.)

Classifiers to be built for Dataset 2 and Dataset 3:

1. Multi-class logistic regression based classifier using polynomial basis functions and Gaussian basis functions.
2. Perceptron for Dataset 2(a) only
3. Multilayer feedforward neural network (MLFFNN) with 2 hidden layers. Use the cross-entropy error function. Use the stochastic gradient descent method with the generalized delta rule for weight update.
4. Linear kernel based C-SVM for Dataset 2(a) only
5. Nonlinear kernel based C-SVM using polynomial kernel and Gaussian kernel.

Optional exercise that carries bonus marks : C-SVM for Dataset 1(a)

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. Outputs of two nodes in the first hidden layer, two nodes in the second layers and all the nodes in the output layer of MLFFNN after epochs 1, 2, 10, 50 and final epoch for Dataset 2(b)
5. Decision region plots for the best configuration of the model, for datasets 2(a) and 2(b) in **Dataset 2**. Superpose the training data on the decision region plot. In the plots for SVM based classifiers, indicate the support vectors (unbounded and bounded) among the training examples.

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