January-May 2017 Semester
CS6011: Kernel Methods for Pattern Analysis
Programming Assignment II

Date: February 23, 2017
Deadline for submission of PDF file of report: **Wednesday, March 22, 2017**

Datasets:
Dataset 1: Bivariate data: (a) Linearly separable classes, (b) Nonlinearly separable classes,
  (c) Overlapping classes
Dataset 2: Black and white image data
Dataset 3: Multivariate data

Classification Models:
1. Logisitic regression model with polynomial basis functions for Datasets 1(a), 1(b), 1(c) and Dataset 3
2. MLFFNN with 2 hidden layers using the cross-entropy error function for all the datasets
3. MLFFNN with 3-level stacked autoencoder based pre-training for Dataset2
4. MLFFNN with 3-level stacked RBM based pre-training for Dataset2
5. Convolutional neural network (CNN) for Dataset2
6. Linear kernel based SVM for Dataset 1(a)
7. Polynomial kernel based C-SVM for Datasets 1(a), 1(b) and 1(c)
8. Gaussian kernel based C-SVM for all the datasets

Presentation of Results:
1. Decision region plots for Datasets 1(a), 1(b) and 1(c) and for different models specified above. Superpose the training examples on the decision region plots. In the decision region plots for SVMs, mark the unbounded and bounded support vectors.
2. Confusion matrix for the test data of each dataset and for different models with the best performance.
3. Plots of outputs for each node in the second hidden layer and for each of the output nodes, for Dataset 1(b) after 1, 2, 10, 50, 100 epochs and after training is complete, for the best performing configuration of MLFFNN in Model 2.
4. Display of features extracted by different levels of stacked autoencoder in Model 3 and stacked RBM in Model 4, at the end of pre-training.
5. Display of feature maps after training of CNN
6. Images of kernel gram matrices to justify the choice of kernel parameters for the SVMs with best performance, for each kernel and for each dataset as specified above.

Selection of model complexity in Models 1, 2, 3, 4 and 5 is to be done using the cross-validation method.

Selection of kernel parameters and the hyperparameter C in Models 7 and 8 is to be done using the cross-validation method.

A single report by a team should also include the details of the models used and the observations about the results of studies.