CS6011: Kernel Methods for Pattern Analysis
Programming Assignment III

Date: April 4, 2016 **Deadline for submission of report**: 4PM, Thursday, May 12, 2016

Task1 Regression :

Dataset 1: 1-dimensional input data of curve fitting
Dataset 2: 2-dimensional (Bivariate) input data

Model : ν -SVR using Gaussian kernel

Presentation of Results :

1. Plots of underlying function, ε -tube, target output and approximated function obtained for Dataset 1. Mark the unbounded and bounded support vectors.
2. Plots of Mean Squared Error (MSE) vs ν on training data, validation data and test data, for Datasets 1 and 2
3. Plots of model output and target output for training data, validation data and test data, for Dataset 2
4. Scatter plot for training data, validation data and test data, for Datasets 1 and 2
5. Comparison of performance of SVR with that of Linear Model for Regression, RBF and MLFFNN for Datasets 1 and 2 including the surfaces realized.

Task2 Novelty detection :

Dataset 3: 2-dimensional input data of overlapping classes
Dataset 4: Multivariate input data

Model : ν -SVDD using Gaussian kernel

Presentation of Results :

1. Mark the bounded and unbounded support vectors for Dataset 3 and plot decision regions.
2. Percentage of true positives and false alarms for test data in Datasets 3 and 4.

Task3 Clustering :

Dataset 5 : 2-dimensional data of nonlinearly separable classes

Model : Kernel K-means clustering using Gaussian kernel

Presentation of Results :

1. Decision region plots for K-means clustering and Kernel K-means clustering after initialization, 2^{nd} iteration , after convergence and after an intermediate iteration.

Task4 Semisupervised learning :

Dataset 6: 2-dimensional input data
Dataset 7: UCI dataset

Models :

a. Self- training with ν -SVM
b. Graph-based semi-supervised method using Label propagation
c. Semi-supervised SVM

Presentation of Results :

1. Comparison of classification accuracy on test data for these three methods and supervised ν SVM for Dataset 6 and Dataset 7.
2. Decision region plots for three methods and supervised ν -SVM for Dataset 6 and Dataset 7.
3. Comparison of classification accuracy on test data for varying sizes of labeled data for Dataset 7.

Task5- Classification/Clustering using kernels for structured data

Choose a type of structured data (text, string, graph, tree) and perform classification or clustering using a kernel method, and present the results.

In this part, you have to choose the **task** you want to perform, the **dataset** you want to use and the **kernel** as well. You have to mail this by April 18, 4pm to the TA email id.

Note: Report should also include your observations about the results.