

**January-May 2016 Semester**  
**CS6011: Kernel Methods for Pattern Analysis**  
**Programming Assignment I**

**Date:** February 11, 2016

**Deadline for submission of PDF file of report:** 4PM, Thursday, March 03, 2016

**Classification tasks**

**Datasets:**

**Dataset 1:** 2-dimensional input data: (a) Linearly separable classes, (b) Nonlinearly separable classes, (c) Overlapping classes

**Dataset 2:** Image data

**Model:** MLFFNN with 2 hidden layers

**Presentation of Results:**

1. Decision region plots for Datasets 1(a), 1(b) and 1(c).
2. Confusion matrix for each dataset
3. Plots of outputs for each of the hidden nodes and 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.

**Regression tasks**

**Datasets**

**Dataset 1:** 1-dimensional (Univariate) input data

**Dataset 2:** 2-dimensional (Bivariate) input data

**Models:**

1. Polynomial curve fitting for Dataset 1
2. Linear model for regression using Gaussian basis functions for Dataset 2. Use the centers of clusters as the centers of Gaussian basis functions. Clusters may be formed using the K-means clustering method on the training data. Use the same value of width parameter for all the basis functions.
3. MLFFNN with 1 hidden layer for Dataset 1
4. MLFFNN with 2 hidden layers for Dataset 2
5. Generalized RBF model for Datasets 1 and 2

**Presentation of Results:**

- Dataset 1 and Model 1: Plot of the approximated functions obtained using training datasets of different sizes, for different model complexities and for different values of regularization parameter. (Similar to Figures 1.4, 1.6 and 1.7 of Bishop's book).
- Dataset 2 and Model 2: Plot of target output and model output for training data, for different model complexities and for different values of regularization parameter.
- Plots of model output and target output for training data, validation data and test data. (For Datasets 1 and 2) (All 5 models)
- Scatter plot with target output on x-axis and model output on y-axis, for training data, validation data and test data. (For Datasets 1 and 2) (All 5 models)
- Dataset 2 and Model 4: Plots of outputs for each of the hidden nodes and output nodes after 1, 2, 10, 50, 100 epochs and after training is complete, for the best performing configuration of MLFFNN.

**Selection of model complexity and regularization parameter 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.**