February-May 2021 Semester
CS5691: Pattern recognition and Machine Learning
Programming Assignment 1

Teacher: C. Chandra Sekhar

Date: 1st March, 2021

Deadline for submission: **Wednesday, 17th March, 2021**

Datasets:

Dataset 1: 1-dimensional (Univariate) input data
Dataset 2: 2-dimensional (Bivariate) input data
Dataset 3: Multivariate real world data

Regression Tasks:

Task 1. Polynomial curve fitting for Dataset 1
Task 2. Linear model for regression using polynomial basis functions for Dataset 2
Task 3. Linear model for regression using Gaussian basis functions for Datasets 2 and 3

Regularization methods:

1. No regularization in Tasks 1, 2 and 3
2. Quadratic regularization in Tasks 1, 2 and 3
3. Tikhonov regularization in Task 3

Presentation of Results:

- For Task1: Plot of the approximated functions obtained using training datasets of different sizes (10 and 200), for different model complexities (degrees 2, 3, 6 and 9) and for different values of λ. (Similar to Figures 1.4, 1.6 and 1.7 of Bishop's book).

- For Task2: Plots of the surfaces of the approximated function obtained using training datasets of different sizes (50, 200 and 500), for different model complexities (degrees 2, 3 and 6) and for different values of λ. The training data points need to be superposed on the surface.

- For Tasks 2 and 3: Scatter plots with target output t_n on *x*-axis and model output $(\boldsymbol{x}_n, \boldsymbol{w})$ on *y*-axis for the best performing model, for training data and test data.

- Tasks 1, 2 and 3: Tables showing the E_{RMS} on the training data, the validation data and the test data, for different models

The best performing model is to be selected using the cross-validation method.

Report: A single report by a team should include the details of the models used and the observations about the results of studies. The report should be properly formatted. The sections should be numbered. The figures and tables should be numbered and should have proper captions.