July-November 2024 Semester
CS5691: Pattern recognition and Machine Learning
Programming Assignment 1

Date: 30th August, 2024

Datasets
Dataset 1: 1-dimensional (Univariate) input data
— Training Dataset 1(a): 10 examples, Training Dataset 1(b): 50 examples
Dataset 2: 2-dimensional (Bivariate) input data
— Training Dataset 2(a): 25 examples, Training Dataset 2(b): 100 examples
Dataset 3: Multivariate data

Regression Model: Linear model for regression using polynomial basis functions

Regularization method: Quadratic regularization

Hyperparameters:

Dataset	Degree of Polynomial M	Regularization Coefficient λ
Dataset 1	3, 6, 9	0.001, 0.1, 1
Dataset 2	2, 4, 6	0.001, 0.1, 1
Dataset 3	2, 3	0.000001, 0.0001, 0.1

Presentation of Results:

• For Dataset 1: Plots of the approximated functions (curves) obtained using training datasets of different sizes (10 and 50), for different model complexities with no regularization, and for different values of λ with model complexity as 9. The training data points need to be superposed on the curve.

• For Dataset 2: Plots of the surfaces of the approximated function obtained using training datasets of different sizes (25 and 100), for different model complexities with no regularization, and for different values of λ with model complexity as 6. The training data points need to be superposed on the surface.

• For Datasets 1(a), 1(b), 2(a), 2(b), 3: Scatter plots with target output t_n on x-axis and model output $y(x_n, w)$ on y-axis for the best performing model, for training data and test data.

• For Datasets 1(a), 1(b), 2(a), 2(b), 3: Tables showing the E_{RMS} on the training data, the validation data and the test data, for models without and with regularization

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