July-November 2025 Semester
CS5691: Pattern recognition and Machine Learning
UG Section
Programming Assignment 1

Date: 29th August, 2025

Deadline for submission of report: 11.59PM on Wednesday, 3rd September, 2025.

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 with 3 input variables and 3 output variables

Regression Models:
1. Linear model for regression using polynomial basis functions
2. Linear model for regression using Gaussian basis functions

Regularization method: Quadratic regularization

Regularization should be used only when there is overfitting

Hyperparameters:

Regression model using polynomial basis functions:

Dataset	Degree of Polynomial M	Regularization Coefficient λ
Dataset 1	3, 5, 7, 9	0.001, 0.1, 1
Dataset 2	2, 4, 6, 8	0.001, 0.1, 1
Dataset 3	2, 3, 4	0.000001, 0.0001, 0.1

Regression model using Gaussian basis functions:
- Number of basis functions should be between 5% to 10% of the number of training examples.
- The width parameter is to be chosen such that the values of Gaussian basis functions are spread over the range of 0.0 to 1.0.

Presentation of Results for each of the Regression Models:

- 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 in the cases of overfitting. 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 in the cases of overfitting. 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 $y(\boldsymbol{x}_n, \boldsymbol{w})$ 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.