

**July-November 2018 Semester**  
**CS5691: Pattern recognition and Machine Learning**  
**Programming Assignment I**

**Date: August 30, 2018**

**Deadline for submission of PDF file of report: Thursday, September 20, 2018**

**Datasets**

- Dataset 1:** 64x64 pixel, gray level image data
- Dataset 2:** 1-dimensional (Univariate) input data
- Dataset 3:** 2-dimensional (Bivariate) input data
- Dataset 4:** Multivariate real world data

**Task 1: Eigenanalysis**

Perform eigenanalysis on the covariance matrix for pixel representations of images in Dataset 1. Use eigenvectors corresponding to significant eigenvalues for reconstruction of images.

**Task 2: Regression Models**

1. Polynomial curve fitting for Dataset 2
2. Linear model for regression using polynomial basis functions for Dataset 3
3. Linear model for regression using Gaussian basis functions for Datasets 3 and 4

**Regularization methods for regression:**

1. Quadratic regularization for all regression models
2. Measure of roughness of hypersurface for models using the Gaussian basis functions

**Presentation of Results:**

- Dataset 1: Compare the original images and reconstructed images for 5 randomly chosen images in Dataset 1, for different number (1, 10, 20, 40, 80, 160, 320 and 640) of significant eigenvalues and their eigenvectors.
- Dataset 2, Model 1: Plot of the approximated functions obtained using training datasets of different sizes, for different model complexities and for different values of  $\lambda$ . (Similar to Figures 1.4, 1.6 and 1.7 of Bishop's book).
- Dataset 3, Models 2 & 3: Plot of the approximated functions obtained using training datasets of different sizes, for different model complexities and for different values of  $\lambda$ .
- Datasets 2, 3 and 4, Models 1, 2 and 3: Scatter plot with target output  $t_n$  on x-axis and model output  $y(\mathbf{x}_n, \mathbf{w})$  on y-axis for the best performing model, for training data and test data.
- Datasets 2, 3 and 4, Models 1, 2 and 3: A table showing the  $E_{RMS}$  on the training data and the test data.

Model selection 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. The report should be properly formatted. The sections should be numbered. The figures should be numbered and should have proper captions.