

**August-December 2024 Semester**  
**CS616: Statistical Pattern Recognition**  
**CS612: Statistical Pattern Recognition Laboratory**  
**Programming Assignment 3**

**Date: 14<sup>th</sup> October 2024**

**Deadline for submission of code and report: Sunday, Nov. 03, 2024, 10:00 PM**

**Datasets:**

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

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

For Dataset 1 and Dataset 2, divide the data into training, and test data. Train, and test split should be 70% and 30% respectively.

**Models:**

1. Polynomial curve fitting for Dataset 1
2. Linear model for regression using Gaussian basis functions with the centers of clusters as the centers of Gaussian basis functions for Datasets 2. Clusters may be formed using the K-means clustering method for the training data.

**Selection of model complexity and regularization parameter is to be done using the cross-validation method.**

**Presentation of Results:**

1. Plot of the approximated functions (For Dataset 1):
  - a. Training datasets of size 10, 50, 100 and complete training set.
  - b. Model complexity (degree of polynomial): 2 to 9
  - c. Regularization to be used for model complexity that leads to over-fitting. Plots to be given for different values of regularization parameter.
  - d. Give the weight values for each case (before and after the regularization).
2. Plot of the approximated functions (For Dataset 2):
  - a. Model complexity (Number of basis functions): 2, 4, 8, 16, 32, 128 and 256
  - b. Regularization to be used for model complexity that leads to over-fitting. Plots to be given for different values of regularization parameter.
3. Plots of the values of mean squared error (MSE) on training data, and test data, for different model complexities and for different values of regularization parameter. (For Datasets 1 and 2)
4. Plots of model output and target output for training data, and test data. (For Datasets 1 and 2). Give the plots for the best model complexity after cross-validation.
5. Give the plots of model output and target output for training data, and test data for the best regularization parameter on model complexity equal to 9 (For Dataset 1) and 256 (For Dataset 2).
6. Scatter plot with target output on  $x$ -axis and model output on  $y$ -axis, for training data, and test data. (For Datasets 1 and 2). Give the plots for the best model complexity after cross-validation.

**Report by a team should include the plots and observations about the results of studies.**

Each group of students must use the dataset identified for that group only.

The expectation of the assignment is to implement from scratch using Python or MATLAB or any other programming language.

**Note:** You are not supposed to use libraries of polynomial regression, K-means clustering etc.

Report should be in **PDF** form and report by a team should also include the observations about the results of studies.

***Instruction:***

**Upload in Moodle all your codes in a single zip file.**

- **Give the name of the code folder as** Group<number>\_Assignment4\_code  
Example: Group01\_Assignment4\_code.
- **Give the name of the zip file as** Group<number>\_Assignment4\_code.zip  
Example: Group01\_Assignment4\_code.zip

**Upload the report as PDF file.**

- **Give the name to the report file as** Group<number>\_Assignment4\_report.pdf  
Example: Group01\_Assignment4\_report.pdf

**We will not accept the submission if you don't follow the above instructions.**