National Institute of Technology Calicut

Department of Computer Science and Engineering

CS6155D Topics in Data Analytics - Winter 2022-23 - M. Tech

Assignment | Due by 10PM, 10-04-2023 | Total Marks: 20

Points to note:-

- You can use either Python or R to complete the assignment
- Submission details will be announced later

Q.1: Create 500 points of the function given below :

$$f(x) = x^3 + 5x^2 + 7x + 2$$

Assume x takes values in the range (-4, 2). Plot the points generated.

Q.2: Add random noise to 30% of the above 500 points, selected at random. Assume that random noise follow a Gaussian distribution. Plot the points after adding the noise.

Divide the points in Q.2 in 80:20 way to obtain Train set and Test set.

Q.3: Fit a Linear regression model for the points in the Train set. Create you own function without using built-in function of Linear regression. Use the gradient descent method.

Q.4: Report the Train error and Test error. Plot the fitted model also.

Q.5: Fit a quadratic and cubic models and compare the Test error rates with that in Q.4. Which model is the best one? Is your best model close to the original equation used for creating the data in Q.1?

Q.6: Assume that all the points with x value less than 0 belongs to Class 1 and rest of the points belongs to Class 2. Let Blue colour represents Class 1 points and Red colour represents Class 2 points. Plot the points.

Q.7: Build a Logistic regression model of classification for the above case and draw the decision boundary also. Report the confusion matrix for the Test set. Create your own function for Logistic regression.

Q.8: Plot the ROC curve and find the AUC also.