

CS5011 – Machine Learning

SOFTWARE ASSIGNMENT - 1

1. Correlation Assignment:

The assignment is to measure the correlation, and produce a scatterplot, that shows the relationship between any two variables. The attached “Q1_data_xx.Rda” file contains the predictors ($x_1, x_2, \dots$) and the outcome (y). Use R and perform experiments to:

- i) Calculate the correlation between the predictors and also between the predictor and the outcome.
- ii) Generate the scatterplot matrix.
- iii) Based on the correlation values, discuss about the influence of predictors ($x_1, x_2, \dots$) on y .
- iv) Fit linear model on the data; Based on the coefficient of the predictors, identify the significant predictors.

File Names (Links to download files are given in the webpage):

Q1_data_01.Rda

Q1_data_02.Rda

(Refer to Table 1 for your assigned dataset)

2. Regression - Polynomial Fitting:

Consider the problem of fitting one-dimensional data with a polynomial. Write an R code to:

- i) Plot function y given in “Q2_fun_xx”.
- ii) Randomly extract 100 points from the function and add normally distributed noise to the data points to get “noisy data”, $\hat{y}$.
- iii) Fit polynomial of degree d (values given in the table 1 below) to the noisy data.
- iv) Compute the bias and variance for the models fitted.
- v) Plot the bias-variance plot.

Functions: (Code for both functions)

$$\text{Q2_fun_01: } y = e^{-5(x-0.3)^2} + 0.5 e^{-100(x-0.5)^2} + 0.5 e^{-100(x-0.75)^2}$$

$$\text{Q2_fun_02: } y = 2 - 3x + 10x^4 - 5x^9 + 6x^{14}$$

Deadline : 25/02/2020

Table 1:

S.No	Roll number	Q1 Data	d values		
1	CS19M502	Q1_data_01.Rda	1	9	18
2	CS19M503	Q1_data_02.Rda	2	10	19
3	CS19M504	Q1_data_01.Rda	3	11	20
4	CS19M505	Q1_data_02.Rda	4	12	21
5	CS19M506	Q1_data_01.Rda	5	13	22
6	CS19M507	Q1_data_02.Rda	6	14	23
7	CS19M508	Q1_data_01.Rda	7	15	24
8	CS19M509	Q1_data_02.Rda	8	16	25
9	CS19M510	Q1_data_01.Rda	1	17	18
10	CS19M511	Q1_data_02.Rda	2	9	19
11	CS19M512	Q1_data_01.Rda	3	10	20
12	CS19M513	Q1_data_02.Rda	4	11	21
13	CS19M514	Q1_data_01.Rda	5	12	22
14	CS19M515	Q1_data_02.Rda	6	13	23
15	CS19M516	Q1_data_01.Rda	7	14	24
16	CS19M517	Q1_data_02.Rda	8	15	25
17	CS19M518	Q1_data_01.Rda	1	16	18
18	CS19M519	Q1_data_02.Rda	2	17	19
19	CS19M520	Q1_data_01.Rda	3	9	20
20	CS19M521	Q1_data_02.Rda	4	10	21
21	CS19M522	Q1_data_01.Rda	5	11	22
22	CS19M523	Q1_data_02.Rda	6	12	23
23	CS19M524	Q1_data_01.Rda	7	13	24
24	CS19M525	Q1_data_02.Rda	8	14	25
25	CS19M526	Q1_data_01.Rda	1	15	18
26	CS19M527	Q1_data_02.Rda	2	16	19
27	CS19M528	Q1_data_01.Rda	3	17	20
28	CS18M504	Q1_data_02.Rda	4	9	21
29	CS18M505	Q1_data_01.Rda	5	10	22
30	CS19M501	Q1_data_02.Rda	6	11	23
31	CS18M518	Q1_data_01.Rda	7	12	24