

CS6464 - CSLT

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

2. Central Limit Theorem:

Suppose X is a random variable whose probability distribution is specified by “Q2_dist_xx”. Given an iid sample of size m such that $S_m = \sum_{i=1}^m X_i$. According to CLT, for large m , S_m can be approximated by normal distribution. Prove it by plotting the density functions of S_m and the normal distribution, with parameters obtained using CLT, for different m values (given in the table 1 below).

Distributions: (Code for both distributions)

Q2_dist_01: $X \sim \text{Gamma}(\alpha, 1)$, where α is the shape parameter.

Q2_dist_02: $X \sim \text{Binomial}(n, p)$, where $n \in \mathbb{N}$ and $p \in [0, 1]$ are the parameters.

3. 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 “Q3_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)

$$\text{Q3_fun_01: } y = e^{-5(x-0.3)^2} + 0.5 e^{-100(x-0.5)^2} + 0.5 e^{-100(x-0.75)^2}$$

$$\text{Q3_fun_02: } y = 2 - 3x + 10x^4 - 5x^9 + 6x^{14}$$

Table 1:

S.No	Name	Q1 Data	<i>m</i> values	<i>d</i> values
1	Krishnachaitanya Gogineni	Q1_data_01.Rda	1,5,10,50,100,500	1,8,15,22
2	Prateep Bhattacharjee	Q1_data_02.Rda	2,6,20,60,200,600	2,9,16,23
3	Saptakatha Adak	Q1_data_01.Rda	3,7,30,70,300,700	3,10,17,24
4	Avishek Bhattacharjee	Q1_data_02.Rda	4,8,40,80,400,800	4,11,18,25
5	Sayanti Bardhan	Q1_data_01.Rda	5,9,50,90,500,900	5,12,19,26
6	Sandeep Narayan P	Q1_data_02.Rda	6,10,60,100,600,1000	6,13,20,27
7	Vinod Karmenghe	Q1_data_01.Rda	7,11,70,110,700,1100	7,14,21,28