24.02.2023

CS6464W - CSLT
SOFTWARE ASSIGNMENT - 1

Release Date: 24th Feb 2023
Submission Date: 23rd March 2023

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, ...) 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, ...) 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)
   Q2_fun_01: $y = e^{-5(x-0.3)^2} + 0.5\,e^{-100(x-0.5)^2} + 0.5\,e^{-100(x-0.75)^2}$
   Q2_fun_02: $y = 2 - 3x + 10x^4 - 5x^9 + 6x^{14}$

3. 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: $X \sim Gamma(\alpha, 1)$, where α is the shape parameter.

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

Deadline : 24/03/2023

Table 1:

S.No	Roll number	Q1 Data	Q2: d values	Q3: m values
1	CS21M501	Q1_data_01.Rda	1,9,25	1,5,10,50,100,500
2	CS21M504	Q1_data_02.Rda	2,10,26	2,6,20,60,200,600
3	CS21M509	Q1_data_01.Rda	3,11,27	3,7,30,70,300,700
4	CS21M010	Q1_data_02.Rda	4,12,28	4,8,40,80,400,800
5	CS21M014	Q1_data_01.Rda	5,13,29	5,9,50,90,500,900
6	CS21M015	Q1_data_02.Rda	6,14,30	6,10,60,100,600,1000
7	CS21M017	Q1_data_01.Rda	7,15,23	7,11,70,110,700,1100
8	CS20M502	Q1_data_02.Rda	8,16,24	8,12,120,800,1200
9	Nagarajan	Q1_data_02.Rda	8,16,24	8,12,120,800,1200