BACHELOR OF ARTS IN ECONOMICS EXAMINATION, 2019

(2nd Year, 2nd Semester)

COMPUTER APPLICATION - II

Time :Two Hours | Full Marks : 30

Answer Question number 1 and any two from the rest.

1. Find the correct answer of any 5 questions with explanation 2x5 =10

A. Suppose you have the following data with single real-valued input-output variables. What is leave-one out cross validation mean square error in case of linear regression (Y = bX+c)?

X(independent variable)	Y(dependent variable)
0	2
2	2
3	1

a) 10/27 b) 20/27 c) 50/27 d) 49/27

B. Which of the following is/are true about "Maximum Likelihood estimate (MLE)"?

1. MLE may not always exist
2. MLE always exists
3. If MLE exist, it (they) may not be unique
4. If MLE exist, it (they) must be unique

a) 1 and 4 b) 2 and 3 c) 1 and 3 d) 2 and 4

C. To test linear relationship of y(dependent) and x(independent) continuous variables, which of the following plot is best suited?

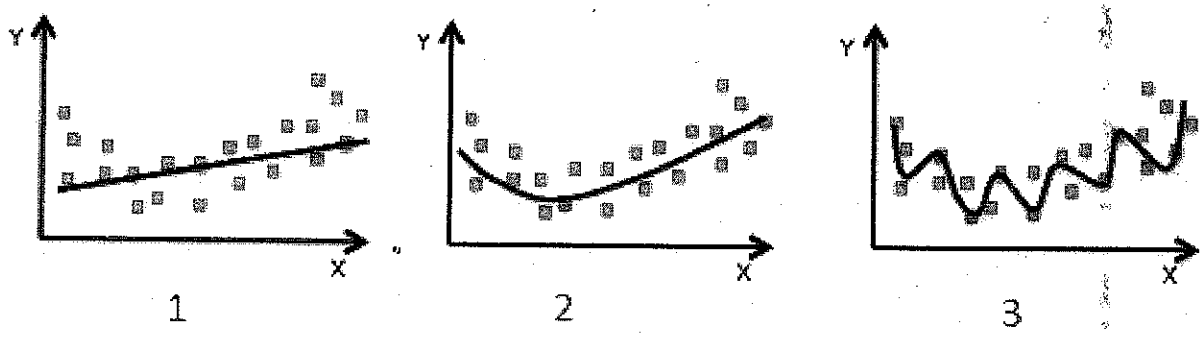
a) Scatter plot b) Barchart c) Histograms d) None of these

D. Which of the following statement(s) can be true post adding a variable in a linear regression model?

1. R-Squared and Adjusted R-squared both increase
2. R-Squared increases and Adjusted R-squared decreases
3. R-Squared decreases and Adjusted R-squared decreases
4. R-Squared decreases and Adjusted R-squared increases

a) 1 and 2 b) 1 and 3 c) 2 and 4 d) None of the above

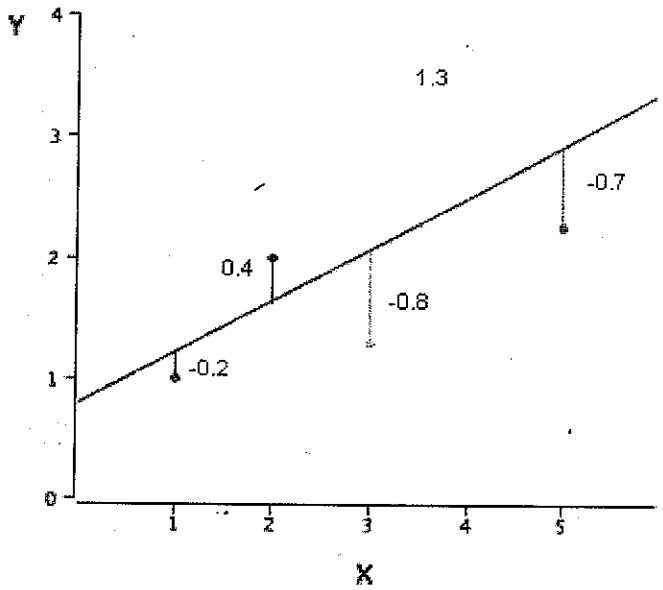
E. The following visualization shows the fit of three different models (in blue line) on same training data. What can you conclude from these visualizations?



1. The training error in first model is higher when compared to second and third model.
2. The best model for this regression problem is the last (third) model, because it has minimum training error.
3. The second model is more robust than first and third because it will perform better on unseen data.
4. The third model is overfitting data as compared to first and second model.
5. All models will perform same because we have not seen the test data.

a) 1 and 3 b) 1 and 3 c) 1, 3 and 4 d) Only 5

F. The graph below represents a regression line predicting Y from X. The values on the graph shows the residuals for each predictions value. Use this information to compute the SSE.



a) 3.02 b) 0.75 c) 1.01 d) None of these

2. Write a Step by step method to fit a data to Multiple Linear Regression. How will you compare two models in R? What is *Cross Validation*? Write a short R code to 3-fold cross validation (Using DAAG package).

5+2+1+2

3. Let us predict weight of a person when his height is known; height is stored in x and weight in y as given below

```
x <-c(151,174,138,186,128,136,179,163,152,131)
y <-c(63,81,56,91,47,57,76,72,62,48)
```

Write all the R code and print result step by step and predict the weight for a person with height 170.

10

4. Let x represents a sample collected from a normal population with unknown mean and standard deviation. We want to test if the population mean is equal to 9, at significance level 5%.

Now if the sample is stored in x, such that

```
X <- c(6.2, 6.6, 7.1, 7.4, 7.6, 7.9, 8, 8.3, 8.4, 8.5, 8.6,8.8,
8.8, 9.1, 9.2, 9.4, 9.4, 9.7, 9.9, 10.2, 10.4, 10.8, 11.3, 11.9)
```

Write an R code preform a t test and interpret the result.

10

----------X----------