For More Questions ***Click Here***

Q1. Which of the following step / assumption in regression modeling impacts the trade-off between under-fitting and over-fitting the most.

A. The polynomial degree

B. Whether we learn the weights by matrix inversion or gradient descent

C. The use of a constant-term

Solution: A

Choosing the right degree of polynomial plays a critical role in fit of regression. If we choose higher degree of polynomial, chances of overfit increase significantly.

Q5. In a linear regression problem, we are using "R-squared" to measure goodness-of-fit. We add a feature in linear regression model and retrain the same model.

Which of the following option is true?

A. If R Squared increases, this variable is significant.

B. If R Squared decreases, this variable is not significant.

C. Individually R squared cannot tell about variable importance. We can't say anything about it right now.

D. None of these.

Solution: C

"R squared" individually can't tell whether a variable is significant or not because each time when we add a feature, "R squared" can either increase or stay constant. But, it is not true in case of "Adjusted R squared" (increases when features found to be significant).

Q6. Which one of the statement is true regarding residuals in regression analysis?

A. Mean of residuals is always zero

B. Mean of residuals is always less than zero

C. Mean of residuals is always greater than zero

D. There is no such rule for residuals.

Solution: A

Sum of residual in regression is always zero. It the sum of residuals is zero, the 'Mean' will also be zero.

Q7. Which of the one is true about Heteroskedasticity?

A. Linear Regression with varying error terms

B. Linear Regression with constant error terms

C. Linear Regression with zero error terms

D. None of these

Solution: A

The presence of non-constant variance in the error terms results in heteroskedasticity. Generally, non-constant variance arises because of presence of outliers or extreme leverage values.

Q8. Which of the following indicates a fairly strong relationship between X and Y?

A. Correlation coefficient = 0.9

B. The p-value for the null hypothesis Beta coefficient =0 is 0.0001

C. The t-statistic for the null hypothesis Beta coefficient=0 is 30

D. None of these

Solution: A

Correlation between variables is 0.9. It signifies that the relationship between variables is fairly strong.

On the other hand, p-value and t-statistics merely measure how strong is the evidence that there is non zero association. Even a weak effect can be extremely significant given enough data.

Q9. Which of the following assumptions do we make while deriving linear regression parameters?

1. **The true relationship between dependent y and predictor x is linear**
2. **The model errors are statistically independent**

3. **The errors are normally distributed with a 0 mean and constant standard deviation**
4. **The predictor x is non-stochastic and is measured error-free**

A. 1,2 and 3.

B. 1,3 and 4.

C. 1 and 3.

D. All of above.

Solution: D

When deriving regression parameters, we make all the four assumptions mentioned above. If any of the assumptions is violated, the model would be misleading.

Q10. To test linear relationship of y(dependent) and x(independent) continuous variables, which of the following plot best suited?

A. Scatter plot

B. Barchart

C. Histograms

D. None of these

Solution: A

To test the linear relationship between continuous variables Scatter plot is a good option. We can find out how one variable is changing w.r.t. another variable. A scatter plot displays the relationship between two quantitative variables.