Name: Dharnidhar Deshmukh

Ref. Code: IIPRSS5306

Assignment (Linear Regression)

1. **Explain the equation of simple linear regression. Which terms contribute the most for predicting y?**
   In simple Linear Regression we use 'y = mx + c' or ' y= wx +b' equation. Here 'y' is a Dependant variable and m is the estimated slope, X is a Independent variable or predictor, and b is the estimated intercept. In predicting Y, Independent variable contribute most.

2. **Do we need to perform feature scaling in Linear Regression?**
   Yes, We need to perform Feature Scaling when we are dealing with Gradient Descent Based algorithms (Linear and Logistic Regression, Neural Network) and Distance-based algorithms (KNN, K-means, SVM) as these are very sensitive to the range of the data points.

3. **What are the assumptions of simple Linear Regression?**
   1) Linear relationship: Linear regression needs linear relationship between independent (x) and dependent variable. LR is very sensitive to outliers, it is very important to check outliers before performing regression analysis.
   2) Normality- Linear regression requires all variables to be normally distributed. the error terms also should be normaly distributed.
   3) Multicollinearity- There should be very little or no multicollinearity in the data multicollin. multicollinearity happens when the independent variables are highly correlated(Linearly correlated) with each other.
   4) Homoscedasticity- The error terms must have a constant variance with respect to the regression line This phenomenon is known as homoscedasticity. The presence of non-constant variance is refer to as Heteroscedasticity.
   5) Autocorrelation- There should be no correlation between the residual(error) terms. absence of this phenomenon is called as autocorrelation.

4. **How will the output be impacted, if the check for multicollinearity is overlooked?**
   Multicollinearity reduces the precision of the estimated coefficients, which weakens the statistical power of your regression model. You might not be able to trust the p-values to identify independent variables that are statistically significant.

5. **In the regression analysis sheet, what does the adjusted r square denote?**
   The adjusted R-squared is a modified version of R-squared that adjusts for predictors that are not significant in a regression model. Compared to a model with additional input variables, a lower adjusted R-squared indicates that the additional input variables are not adding value to the model.

6. **Give the null and alternate hypothesis of linear regression?**
   the chief null hypothesis is H0 : $\beta 1 = 0$, and the corresponding alternative hypothesis is H1 : $\beta 1 = 0$ If this null hypothesis is true, then, from $E(Y) = \beta 0 + \beta 1x$ we can see that the population mean of Y is $\beta 0$ for every x value, which tells us that x has no effect on Y.

7. **Is linear regression impacted due to the curse of dimensionality?**
   linear regression model will not suffer from Curse of Diamensionality. This is because the number of parameters in the OLS will NOT increase exponentially with respect to the number of features / independent variables / columns.

8. **Why do we take the squared differences and simply not differences?**
   Squaring always gives a non-negative value, so the sum will always be zero or higher.

9. **What is the significance of an F-test in a linear model?**
   The F-Test of overall significance in regression is a test of whether or not your linear regression model provides a better fit to a dataset than a model with no predictor variables.

10. **What are the disadvantages of the linear regression?**
    1) Linear Regression Is Limited to Linear Relationships.
    2) Linear Regression Only Looks at the Mean of the Dependent Variable.
    3) Linear Regression Is Sensitive to Outliers.