**Assignment 4**

1. What is the difference between a univariate and a multivariate function?

Ans:

In univariate, the there is only one variable present in the data. This doesn’t give any relationship with other features/variables.

Example can be just height of students in a class.

This type of data can be utilized for finding the central tendencies (Mean, Median and Mode), spread of the data (Quartiles, Variance, and standard deviation)

In Multivariate, there are more than one variable, hence some relationship with other features/variables can be identified for predictive modelling.

2. What are the various ways to check the applicability of a particular regression model on a dataset?

Ans:

For applying regression on a dataset following conditions must be met.

* Linearity: Linear relationship between input features and target variable.
* Homoscedasticity: In residual plot, the variance should be constant. Should not follow any specific pattern.
* Multi-collinearity: Features should not be highly correlated. Can be found out by calculating VIF, Eigen values or Correlation coefficient.
* Independent features: Features selected should be independent.

3. What are the basic assumptions of the Linear Regression Algorithm?

Ans:

Assumptions of Linear Regression:

* Linearity: Linear relationship between input features and target variable.
* Homoscedasticity: In residual plot, the variance should be constant. Should not follow any specific pattern.
* Multi-collinearity: Features should not be highly correlated. Can be found out by calculating VIF, Eigen values or Correlation coefficient.
* Independent features: Features selected should be independent.

4. What are the different Evaluation metrics used in Linear Regression?

Ans:

* Mean Absolute Error(MAE)
* Mean Squared error (MSE)
* Root Mean Squared Error (RSME)
* R2 Score

5. Why do we square the residuals instead of the Modulus?

Ans:

For ease of calculating its derivative the residuals are squared. For maximizing or minimizing a particular function we will generally have to calculate its derivative and equal it to zero. And squared function are easy to calculate derivatives.

6. Which evaluation metric should you prefer if you have a lot of outliers present in it?

Ans:

Mean Absolute Error (MAE) should be used. Formula is

MAE = 1/n \* ∑ |Yactual – Ypredicted|