MACHINE LEARNING- WORKSHEET 3

1. In the linear regression equation y = θ0 + θ1x, θ0 is the:

Ans- **C) y intercept**

2. True or False: Linear Regression is a supervised learning algorithm.

Ans- A) True

3. In regression analysis, the variable that is being predicted is:

Ans- B) the dependent variable

4. Generally, which of the following method(s) is used for predicting continuous dependent variables?

Ans- B) Linear Regression

5. The coefficient of determination is:

Ans- C) the correlation coefficient squared

6. If the slope of the regression equation is positive, then:

Ans- B) y increases as x increases

7. Linear Regression works best for:

Ans- A) linear data

8. The coefficient of determination can be in the range of:

Ans- A) 0 to 1

9. Which of the following evaluation metrics can be used for linear regression?

Ans- **B) RMSE**

D) MAE

10. Which of the following is true for linear regression?

Ans- A) Linear regression is a supervised learning algorithm.

11. Which of the following regularizations can be applied to linear regression?

Ans- A) Ridge B) Lasso D) Elastic Net

12. Linear regression performs better for:

Ans- A) Large amount of training samples with small number of features.

13. Which of the following assumptions are true for linear regression?

Ans- A) Linearity B) Homoscedasticity D) Normality

14. Explain Linear Regression?

Ans- **Linear Regression is a machine learning algorithm based on supervised learning. It performs a regression task. Regression models describe the relationship between variables by fitting a line to the observed data. Linear regression models use a straight line, while logistic and nonlinear regression models use a curved line. Regression allows you to estimate how a dependent variable changes as the independent variable(s) change. There are two types of linear regression - Simple and Multiple.**

Since linear regression shows the linear relationship, which means it finds how the value of the dependent variable is changing according to the value of the independent variable. Linear regression makes predictions for continuous/real or numeric variables such as sales, salary, age, product price, etc.

The theory of linear regression is based on certain statistical assumptions. It is crucial to check these regression assumptions before modelling the data using the linear regression approach.

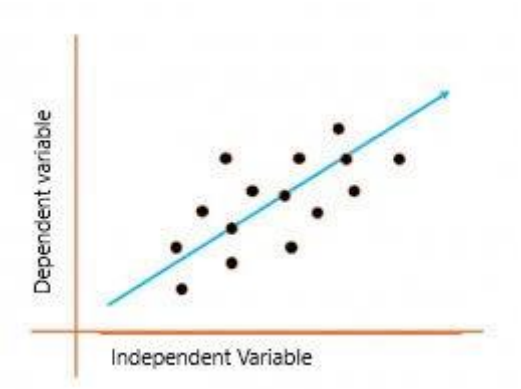
Mainly there are 7 assumptions taken while using Linear Regression:

1. **Linear Model**
2. **No Multicollinearity in the data**
3. **Homoscedasticity of Residuals or Equal Variances**
4. **No Autocorrelation in residuals**
5. **Number of observations Greater than the number of predictors**
6. **Each observation is unique**
7. **Predictors are distributed Normally**

15. What is difference between simple linear and multiple linear regression?

Ans-

Simple Linear Regression: Require single feature to model a linear relationship with a target variable, i.e. if there is a single input variable X (independent variable), such linear egression is called *simple linear regression*.



The above graph presents the linear relationship between the output(y) variable and predictor(X) variables. The blue line is referred to as the best fit straight line. Based on the given data points, we attempt to plot a line that fits the points the best.

To calculate best-fit line linear regression uses a traditional slope-intercept form which is:

$Y = B_0 + B_1 X$.

Where,

- Y is the dependent variable
- X is the independent variable
- B1 is the slope, which determines the angle of the line
- B0 is the intercept, which is a constant determining the value of y when x is 0

The goal of the linear regression algorithm is to get the best values for B0 and B1 to find the best fit line. The best fit line is a line that has the least error which means the error between predicted values and actual values should be minimum.

Multiple Linear Regression: Uses multiple features to model a linear relationship with a target variable. When working with multiple independent variables, we're still trying to find a relationship between features and the target variables. The only difference is that there's more features we need to deal with.

The equation for a multiple linear regression is shown below.

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \ldots + \hat{\beta}_n x_n$$

If we look at the first half of the equation, it's the exact same as the simple linear regression equation! What does the other half of the equation mean? Well they're just added features! If we add more features, our equation becomes bigger.