**MACHINE LEARNING – WORKSHEET 11**

**(LINEAR REGRESSION)**

**(Solutions)**

**ANS 1)** B) always increases

**ANS 2)** B) SST = SSR + SSE

**ANS 3)** A) difference between the actual value and the predicted value.

**ANS 4)** C) By its slope

**ANS 5)** B) can be either -1 or 1

**ANS 6)** A) Scatter plot

**ANS 7)** B) f-statistics

**ANS 8)** C) Ridge

**ANS 9)**

A) It shows the causal relationship between dependent and independent variables

B) It shows the positive or negative relation between dependent and independent variables

D) It is a straight line that is the best approximation of the given data sets

**ANS 10)** C) Automatic feature selection

**ANS 11)** A) Normal Equation

**ANS 12)**

R-squared, also known as the coefficient determination, defines the degree to which the variance in the dependent variable (or target) can be explained by the independent variable (features).

Similar to R-squared, the Adjusted R-squared measures the variation in the dependent variable (or target), explained by only the features which are helpful in making predictions. Unlike R-squared, the Adjusted R-squared would penalize you for adding features which are not useful for predicting the target.

**ANS 13)**

In ML, cost functions are used to estimate how badly models are performing. Put simply, **a cost function is a measure of how wrong the model is in terms of its ability to estimate the relationship between X and y.** This is typically expressed as a difference or distance between the predicted value and the actual value. The cost function (you may also see this referred to as loss or error.) can be estimated by iteratively running the model to compare estimated predictions against “ground truth” — the known values of y.

The **MSE** in the case of Linear Regression is the cost function. It is simply the **mean** **of the squared differences between predicted y and actual y (i.e. the residuals).**

**ANS 14)**

**SST**: The **sum of squares total**, denoted **SST**, is the squared differences between the observed dependent variable and its **mean.**

**SSR**: The second term is the **sum of squares due to regression**, or **SSR**. It is the sum of the differences between the predicted value and the **mean** of the dependent variable.

**SSE**: The last term is the **sum of squares error**, or **SSE**. The error is the difference between the observed value and the predicted value.

**ANS 15)**

* Mean Squared Error (MSE)

## Mean Absolute Error (MAE)

## R-squared or Coefficient of Determination

## Root Mean Squared Error (RMSE)