**MACHINE LEARNING AISC1003 – (M07 Group 1)**

**Case Study 1** – Linear regression

**Name** : Palvai Saikiran Reddy.

**ID number** : 500197829.

**Linear Regression :**

In general, the term ‘Regression’ is used to find the relationship between variables. Linear regression uses the relationship between the data points to draw a straight line through all of them. Basically, linear regression is a linear model. When there is a single input variable the method is termed as simple linear regression and when there are multiple input variables, then it refers to the method as multiple linear regression.

1. Simple Linear Regression : while using basic linear regression with a single input, statistics such as means, standard deviations, correlations and covariance can be used to estimate the coefficients.
2. Ordinary Least Squares : When we have more than one input, we can use Ordinary Least Squares to estimate the values of the coefficients. This procedure seeks to minimize the sum of the squared residuals. This means that given a regression line through the data we calculate the distance from each data point to the regression line, square it, and sum all the squared errors together. This is the quantity that ordinary least squares seek to minimize, and the data is used as matrix and uses linear algebra operations to estimate the optimal values for the coefficients.
3. Gradient Descent: In this method, when there are one or more inputs, we can utilise an iterative method to optimise the coefficient values by reducing the model’s error on the training data.
4. Regularization: Regularization methods (Regularization methods are extensions of the linear model training process) seek to both minimize the sum of the squared error of the model on the training data (using ordinary least squares) but also to reduce the complexity of the model (like the number or absolute size of the sum of all coefficients in the model).

The two popular examples of regularization procedures for linear regression are:

Lasso Regression and Ridge Regression.

**Linear Regression Model Representation:** It is a linear equation that combines a specific set of input values (x) the solution to which is the predicted output for that set of input values (y). So, both the input values (x) and the output value are numeric.

The three main metrics for model evaluation in regression are :

* R Square/Adjusted R Square.
* Mean Square Error (MSE)/Root Mean Square Error (RMSE).
* Mean Absolute Error (MAE).