**Linear Regression**

## Aaryan

## CO21BTECH11001

Let number of features of dataset = n

Let number of sets of features = m

Data consists of matrices X and y where column of X represents the feature of dataset and element of y represents the value of variable dependent on set of features listed in row of X.

Linear regression is a type of regression algorithm where it assumes a linear relationship between dependent (X) and independent (y) variables.

Let

where is known as parameter.

We define a hypothesis function as follows –

where

We will calculate a value of which best fits the approximation –

To do this, we will define a cost function as follows –

We can see from here that when , our assumption is satisfied.

**Objective –** Minimize or Converge the cost function.

There are two approaches to do this –

# **Gradient Descent Algorithm –**

# **Normal Equation Method –**

If then will get minimized.

Solving for that, we get –

After getting optimal , we can get the value corresponding to a new data as

**Questions –**

1. Is Linear Regression a regression algorithm or classification algorithm?

**Ans.** Regression algorithm

1. Why do we need to take ?

**Ans.** Because in the hypothesis function there is a constant term apart from the linear combination of and , which is , so the multiplier of can be any value. For simplicity, we take it as 1.

1. How can we increase the accuracy of linear regression?

**Ans.** We can increase the accuracy of linear regression by outlier treatment, i.e. we have to remove the values of sparse feature which are usually less than 1 percentile or more than 99 percentile.

1. What are the disadvantages of linear regression model?

**Ans.** It is sensitive and dependent on outliers, which effect the overall accuracy of the model.

1. What are the fields where linear regression is usually used?

**Ans.** It is usually used in Businesses, Statistics, Medical Science etc.