**Logistic 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.

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

In a binary classification problem, y consists of only two values, usually 0 and 1.

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 –

The above approximation is only for a binary classification problem. In any other classification problem, we can similarly fix landmarks for

Now, we will define a function which is a measure of probability of accuracy of hypothesis function, which is known as log-likelihood function.

**Objective –** Minimize or Converge the log-likelihood function.

There are two approaches to do this –

# **Gradient Ascent Algorithm –**

# **Newton’s Algorithm of Classification –**

Algorithm –

*Stop*

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

**Questions –**

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

**Ans.** Regression algorithm

1. What is the type of decision surface in Logistic Regression algorithm?

**Ans.** A linear curve (straight line)

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. What is the range of values of hypothesis function?

**Ans.** (0,1)

1. Name three methods by which we can increase the accuracy of logistic regression?

**Ans.** Removal of incomplete dataset , Feature Scaling/ Normalization, Removal of outliers of sparse features.

1. What are the disadvantages of linear regression model?

**Ans.** It constructs linear boundaries which is not as accurate in non-linear problems.