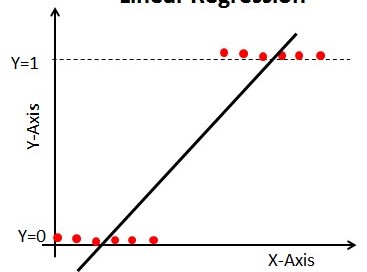
Logistic Regression

Logistic Regression is the popular Supervised Machine learning algorithm for classification. Logistic regression is a statistical technique used to predict the outcome of a dependent variable based on previous observations. It's a type of regression analysis and is a commonly used algorithm for solving classification problems. It calculates the probability that a given value belongs to a specific class. If the probability is more than 50%, it assigns the value in that particular class else if the probability is less than 50%, the value is assigned to the other class. Therefore, we can say that logistic regression acts as a binary classifier. The outcome of Binary Logistic regression is Dichotomous meaning it will have only two possible outcomes.  Let’s now understand the concept of classification with an example. Suppose you want to classify batsmen into two categories (inform/not in form). We can take the average of their last 10 matches scores as a feature for classification. Using average as a training feature, we can build and train a classification model.

Why Logistic regression is named as Regression even though it is a classification problem?

You must be wondering if it is a classification algorithm why it is called regression. Let’s consider a classroom and you want to classify the heights of the students and categorize them into tall or short. Now we will try to solve this problem using Linear regression. In linear regression we will draw a hyperplane for the given

data.



Considering

x-axis=Heights

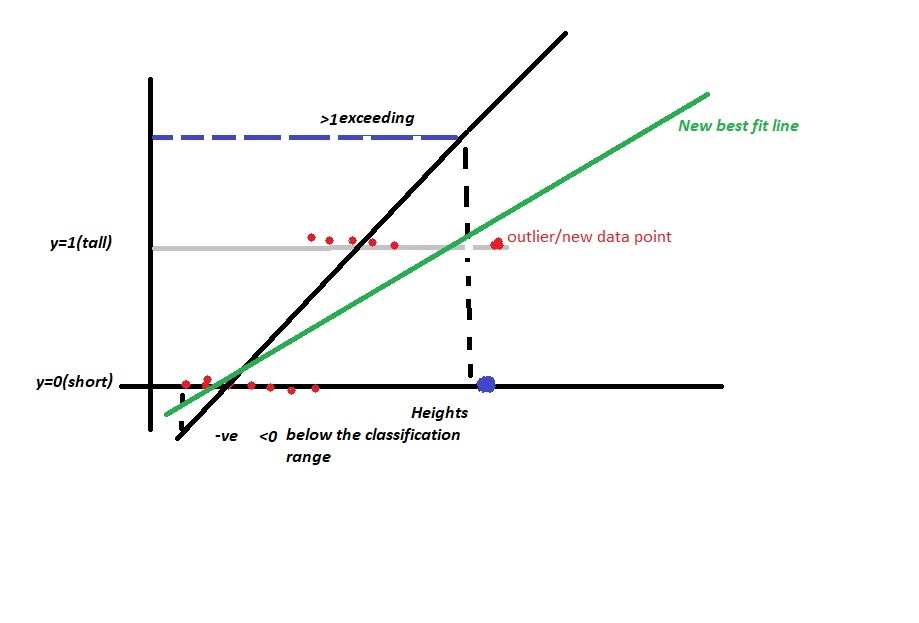
y-axis->1=tall

y-axis->0=short

if h(x)>=4 we can conclude that person as tall(ie y=1)

if h(x)<4we can conclude that person as short(ie y=0)

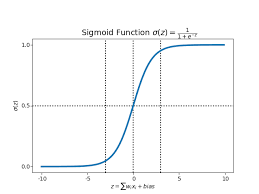
If we can classify this problem using linear regression then why exactly we need logistic regression

Suppose if we consider a new data point or an outliner then we will get a new best fit line overriding the previous dataplane. 

From the above diagram we can say that by adding a new outliner a new best fit line has created. This new plane will result in a change of previous classification values. This will create a high error rate. Every time we need to create a new best fit line as the data points will be keep coming. There is another glitch by using linear regression, some inputs will generate the outcome which will exceed the classification range. And classification will take a wrong path. So, we precisely can’t use the linear regression for these kinds of problems. Logistic Function/Sigmoid function is used to avoid all these misconceptions created by linear regression. Logistic Regression is called as regression because the underneath concept used is linear regression but the cost function used is sigmoid or logistic function.

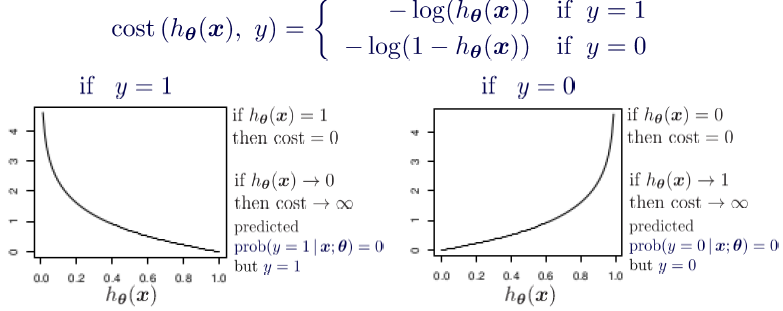
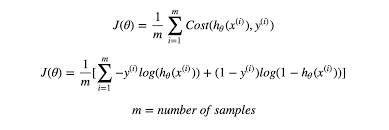
Logistic Regression Cost function

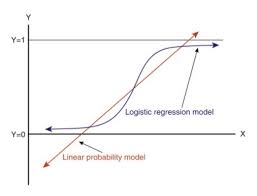
Before moving to Logistic regression cost function let us take a look at sigmoid function.



Sigmoid function values lie between 0 and 1. To prevent the outliners impact we will use the sigmoid function.

Cost function





Applications of Logistic Regression

-> Logistic regression is usually applied to linearly separable data.

->Classifies whether the tissue is malignant or Benign.

->To predict whether email is spam or not.

->Marketing

->finance

Advantages

->No assumptions are involved.

->Easy to implement

->less prone to over-fitting.

->Easily extendable to multinominal regression.

->very fast, good accuracy.

Disadvantages

->Easily outperformed by the many other machine learning algorithms.

->Logistic Regression algorithm begins to act weak if there is to much of missing or large amount of data.

->Curve Overfitting if the data is large or inconsistent.

->In logistic regression the independent variables are linearly related to the logarithmic odds (log(p/1-p))

Types of Logistic Regression

Binomial

Two outcomes

Multinomial

In multinomial logistic regression the dependent variables or the classified classes will be 3 or more.

->uses maximum likelihood estimation for classifying.