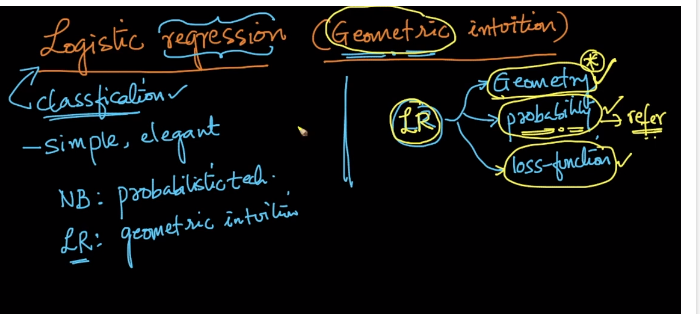
**Geometric Intuition of Logistic Regression:**

Logistic regression is one of the most wonderful classification algorithm.

It may have regression in its name but it is an classification algorithm and it can be derived using multiple methods.

1. Geometric representation
2. Probability
3. Loss function

So all the methods reach to same result and we will see Geometric representation and Loss function in full detail and will touch the surface for Probability in this course.



Currently we will start with Geometric representation.

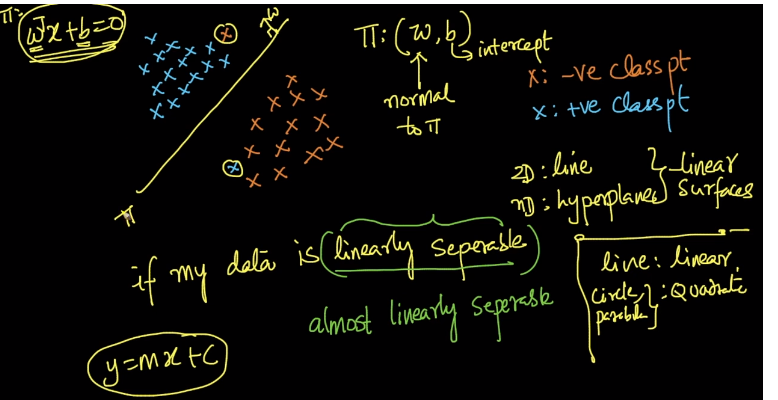
So we have studied about equation of plane and line in linear algebra and as we know that we can use those equation to separate out data points.

There are two terms called “Linearly Separable” which means that our data can be separated using plane or line without making any errors but as we can see in below image

We have one +ive point below our line and one –ive point above our line and this sort of data which can be separated using line or plane but there will be some miss classified points are called as “Almost Linearly Separable”.

In equation of plane we have two terms i.e. w and b where w in an vector normal to plane and b in an intercept(if line passes through origin it is 0).

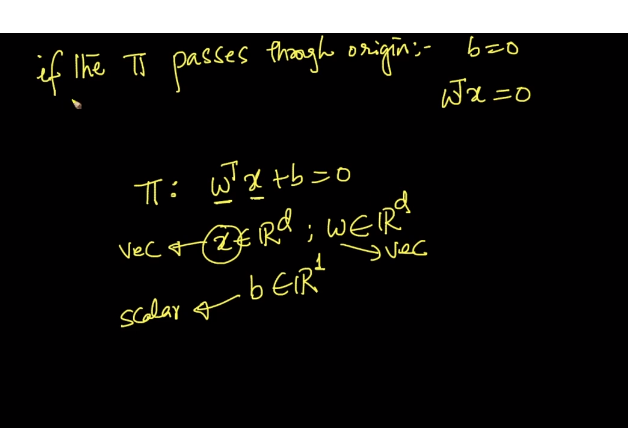
We are separating data through a line or plane than we are considering one side of the plane as positive and another side as negative.



If a line passes through origin the intercept b will become 0 and we assume it passing through origin to minimize our calculation for now.

Here x is and d dimensional vector which we get form our data and w is also an d dimensional which we need to find so that it can separate our data in best possible manner.

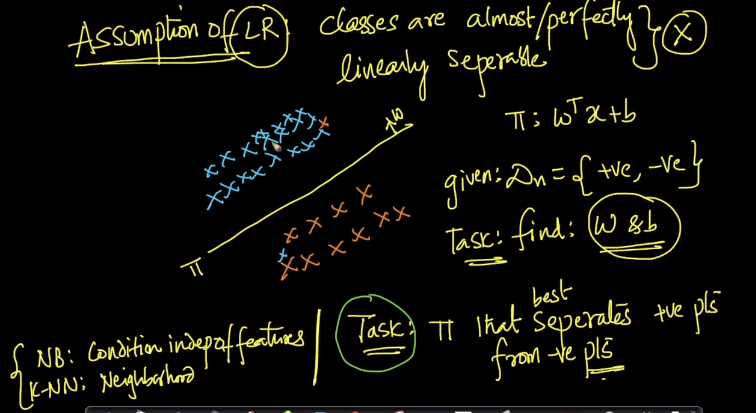
And b is obviously and scalar data.



**As in NB the assumption was that all the features are independent and so in linear regression it has its assumption as the classes are almost or say perfectly linearly separable.**

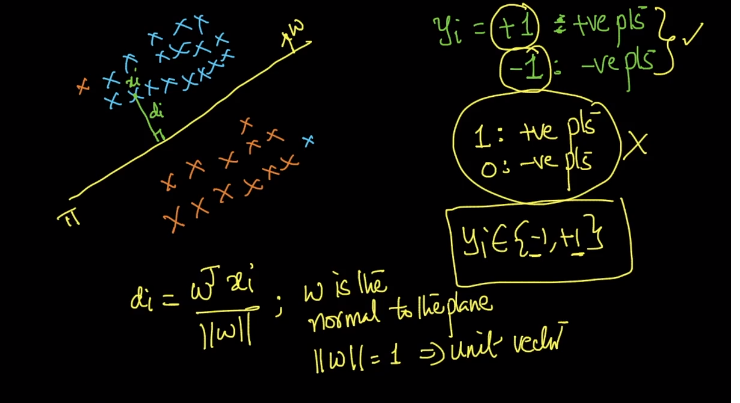
Which means it will consider that the data is linearly separable and try to find best w which will separate the data.

SO the whole task of linear regression is to find best values for w and b so it separates the given data in best possible manner.



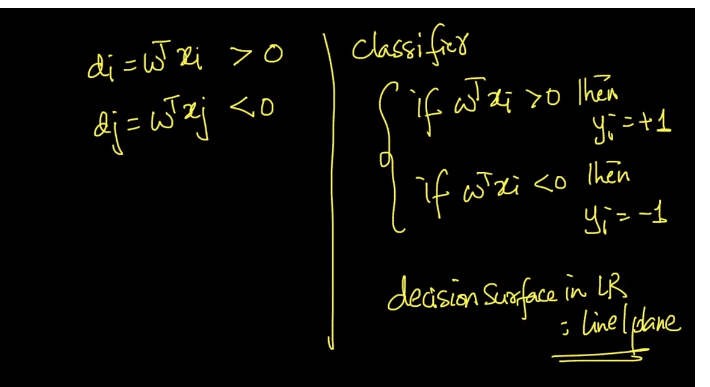
We will now consider our positive point as +1 and negative points as -1 instead of (1,0).

The reason behind this is when we trying to find distance of any given point from our plane we are multiplying the value of W(Transpose) to Xi and if we consider our Xi as 0 so it will make no sense.



SO now when we are calculating W(Transpose) \* Xi and if we get it > 0 then we get point as +1 i.e. positive and if we get it <0 then we get point as -1 i.e. negative.

And in this manner we divide our points.



Now the question is how do we know, can we improve our model more or is this the best model we have got?

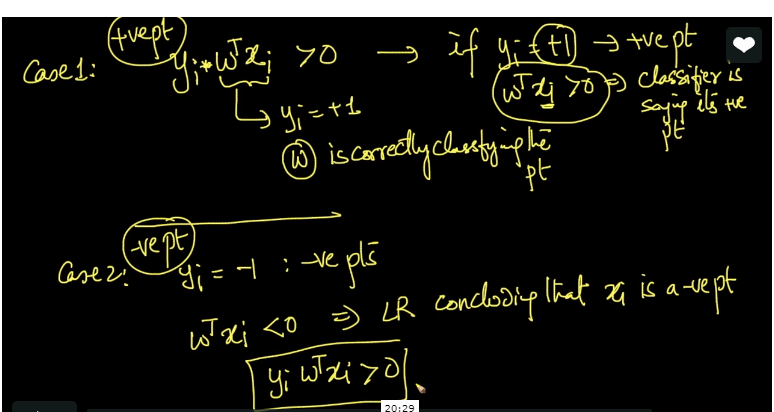
So to answer this question we need to see multiple cases.

Case 1: Let’s say we have got our point as positive and when we multiply it by Yi i.e. 0 or 1 and if we get positive than we can say that the point is correctly classified.

For example we have choose a point to be positive and if it actually is positive so on multiplying + \* + we will get a + value which means it is correctly classified.

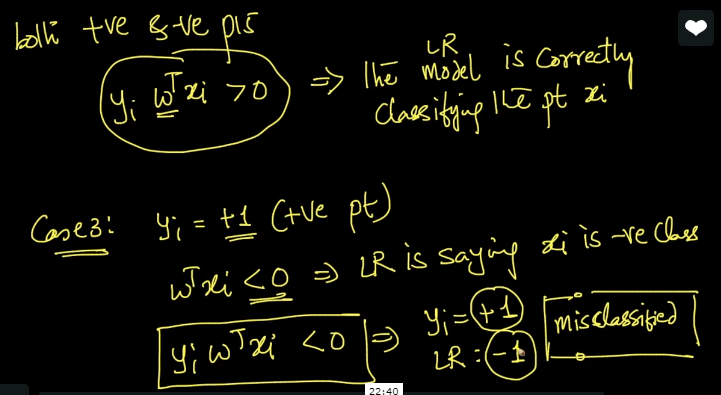
And if we choose a point to be negative and it actually is negative so on multiplying - \* - again gives us +ive value which means it is correctly classified.

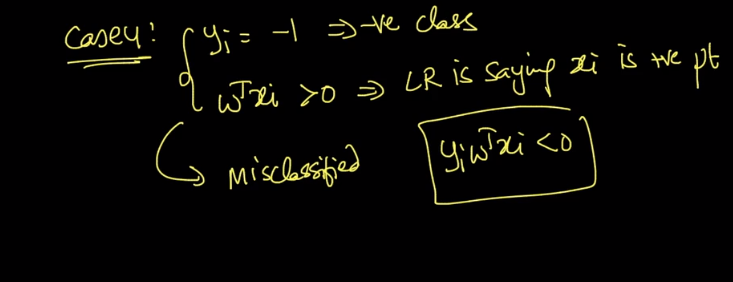
And so this is what case 1 and case 2 are.



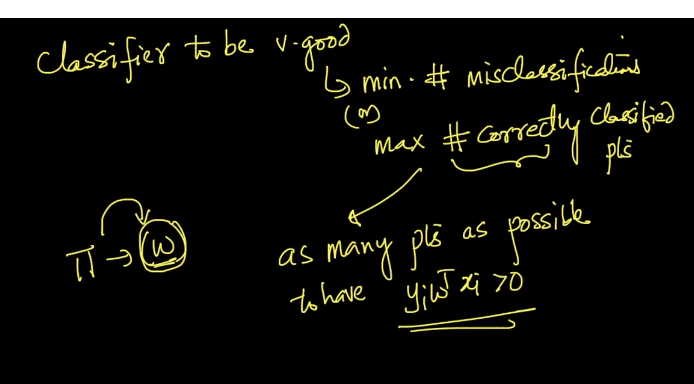
Now lets see case 3 and case 4 :

Suppose we choose a point to be negative and it actually is positive or vice versa, than in both the case on multiplying the values that is Yi \* W(transpose\*X) we will get a negative value which shows that the point is not correctly classified or misclassified.





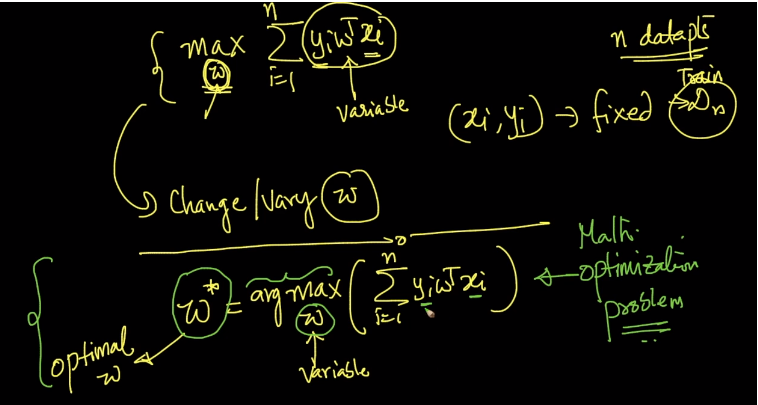
Our aim is to find the most optimal values of w and b so it reduces the misclassification and maximize the correctly classified points.



SO how can we formulate this is shown in below image.

We want to maximize the sum of the multiplication value of Yi and W(transpose)\*Xi and if most of the values in this multiplication are positive so it will result in maximum sum and if more values are negative than it will reduce the sum or even give us a negative value as sum.

But since we have value of X and Y as fixed because those are our data points so we have got w which can be varied in such way to maximize the sum.

.

Comments:

