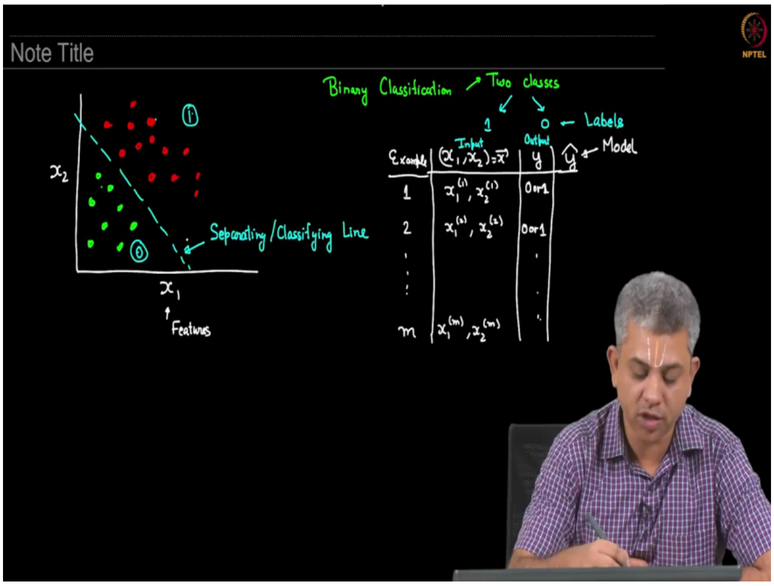
Machine Learning for Engineering and Science Applications
Professor Dr Balaji Srinivasan
Indian Institute of Technology, Madras
Department of Mechanical Engineering
Logistic Regression

(Refer Slide Time: 0:27)

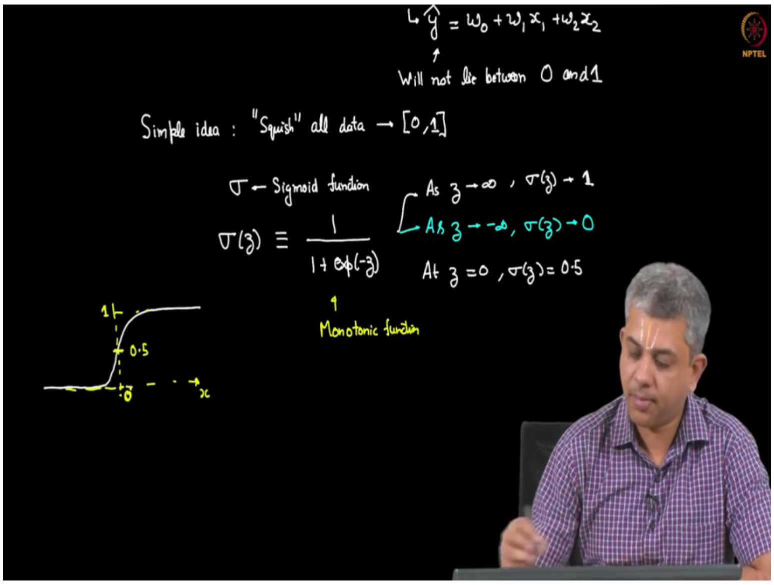


In this video we will be looking at our 1st classification algorithm called logistic regression, note that this is a even though the name regression is sitting there, it is actually a classification algorithm. So in the previous video we saw linear regression which tries to fit an optimal usually least-squares fit to some given data using weights which are linear. So now in this video we will look at the idea of trying to classify 2 different sets of data for binary classification. By binary classification I mean there are 2 possible classes, we label these classes simply 1 and 0, these are the labels that we are giving our classes.

So suppose you collect some bunch of example and once you plot them suppose X1 and X2 are two features of your data and you are classifying your data recording to this okay. Now what you see when you plot the figure is that they are nicely clustered to one side or the other side and intuitively we can draw a line separating the two, separating or the classifying line. Now can we use our linear regression idea in order to achieve some sort of classification of this sort okay? So here is the problem, you are given X1 and X2 which is your input X factor, you have a bunch of examples, each of data point is an example, so as we use with linear regression we will use the same notation so on and so forth.

Once again we will use m examples; this is our input and our output as before is Y. Now unlike linear regression, remember that our output has to be 1 or 2 classes, either in this case the red class or the green class which is class 1 or class 0. So our label Y should always be either 0 or 1 always and we have to make, our job is to get a predictive model which does the same. It does the same, ideally it should classify this as 0 whenever this is 0 and as 1 whenever it is 1 and of course we would like to do it for new points as well okay for some extra new point which has been now introduced, so how can we do this?

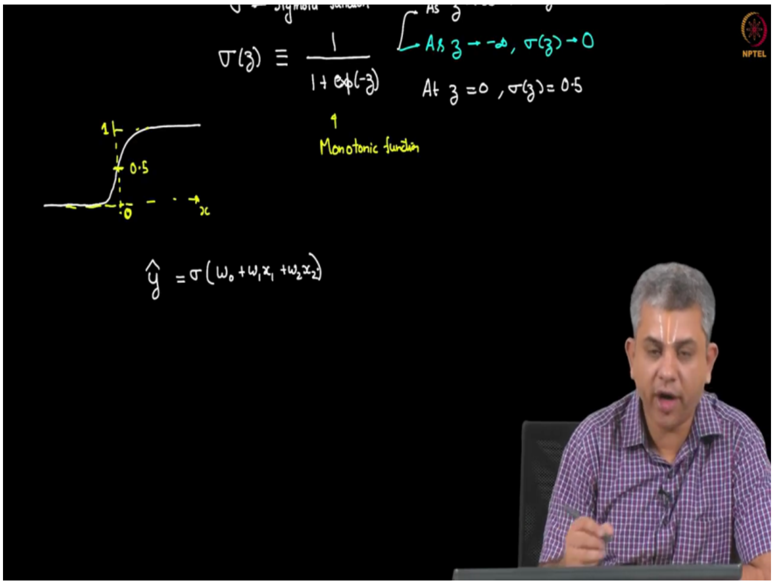
(Refer Slide Time: 4:11)

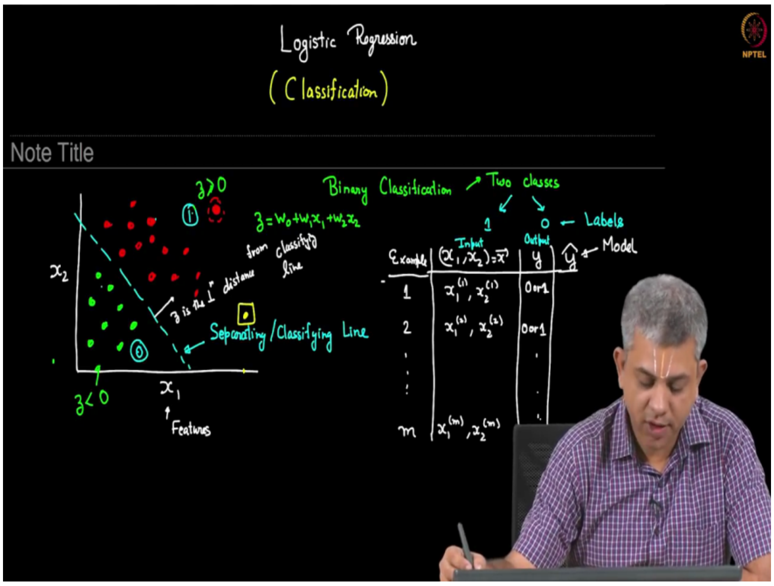


Suppose we try simple linear regression, there would be a problem, what is that problem? So suppose I introduce a new point somewhere here ok, if I had fit a simple linear regression line whatever model I use, if I use a linear model Y hat equal to let us say W0 plus W1X1 plus W2 X2 regardless of your weights W1 and W2 if I pick up a far enough point, this value will come very high it is not going to lie between 0 and 1. Similarly if I choose a point at the extreme, there is no way for you to ensure that a simple linear model will always give values between 0 and 1, so which is why we use a very simple idea which is too squish all data to the range 0 1 which is what we require as our output okay. How do we do this?

Once again there is a very simple idea for this, we use something called the Sigma function. So Sigma function is defined as follows, 1 by 1 plus E to the power minus Z, what does it look like? You can see its limits as Z tends to infinity, Sigma of Z tends to 1 because E to the power minus Z tends to 0. As Z tends to minus infinity, Sigma of Z tends to 0 because this tends to infinity. Now exactly at Z equal to 0, Sigma of Z is equal to 0.5, notice also that this is a monotonic function, so if you plot the Sigma curve it looks somewhat like this. At X equal to 0 the Sigma value is 0.5 and as it tends to infinity it is going to be 1 and as X tends to minus infinity it is going to tend to 0.

(Refer Slide Time: 7:58)

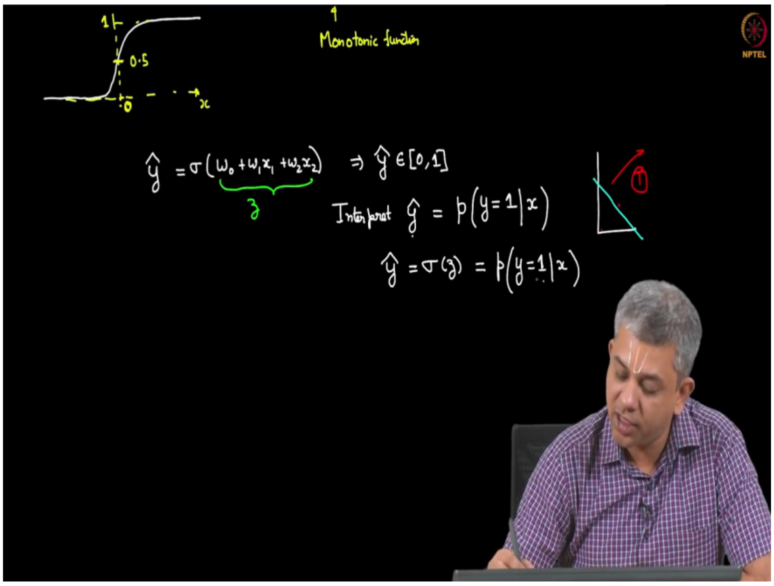




So what this tells us that if I predict Y hat as instead of using what I had for linear regression. If I use Sigma of linear regression, this would tell me that Y hat will always lie between 0 and 1. Now this has an additional advantage which is now we can interpret Y hat as the probability that the output belongs to class 1 even the input X, let me explain. Now let us look at this line here, we want this value let us call this something Z this value Z which is equal to W0 plus some weights W1X1 plus W2 X2 such that if it has to lie in class 1 then Z has to be really high ok. If it has to belong to class 0 we know that Z has to be really low ok.

Now suppose I look at a point which is somewhere here, let us say this is a new point ok, what this tells me is that this point is almost certain to lie in class 1. Just looking at this data the further away we get away from this classifying line then more and more uncertain we are that this point belongs to class 1. Similarly, the further away we get to decide of the line the further and further closer certain we are that we belong to class 0 okay. So what we can think of Z as is Z is the perpendicular distance from the classifying line.

(Refer Slide Time: 11:40)



How does that help us? The classifying line then is the line Z equal to 0 which means Sigma of Z is equal to 0.5, if I come to this side Sigma of Z becomes close to 1, if I come to this side Sigma of Z becomes close to 0, the closer I am to this line the more uncertain we are about where it lies, whether it lies on class 1 or whether it lies on class 0. Therefore it is easy to now interpret Y hat which was equal to Sigma of Z as the probability that you belong to class 1 okay for example, if your Sigma value is close to 0.5 okay then what it means is you are not really certain about how close it is to class 1, it means probability is approximately 0.5 that it is class 1.

Let us see Sigma is close to 0.99 okay then we know that it is really far away and that we are very-very certain that it lies in class 1. Suppose Z the Sigma value is 0.01 then we know that the probability that it lies in class I is actually pretty low it is equal to 0.01. So this is the simple idea behind logistic regression, we will see how to compute classifying lines using logistic regression in later videos.