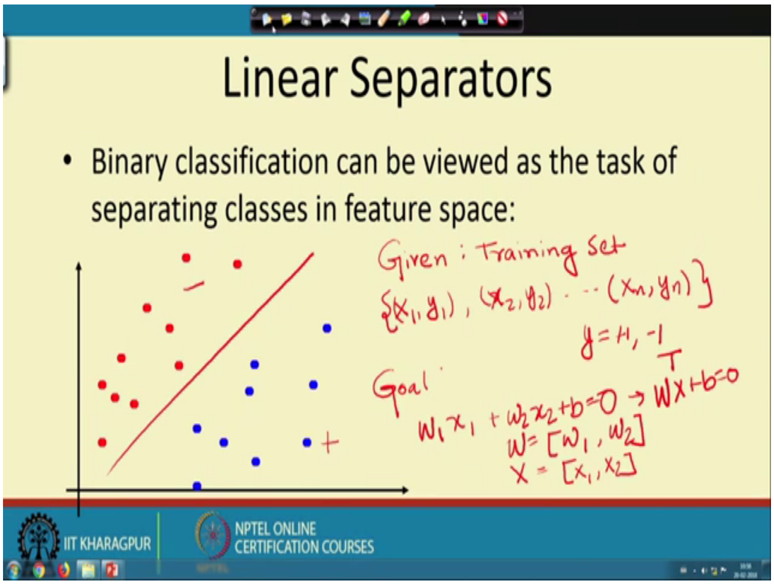
Data Mining
Prof. Pabitra Mitra
Department of Computer Science & Engineering
Indian Institute of Technology, Kharagpur

Lecture - 24
Support Vector Machine – III

Welcome. We continue our discussion on the Support Vector Machine classifier. Let me quickly review what we have studied so far.

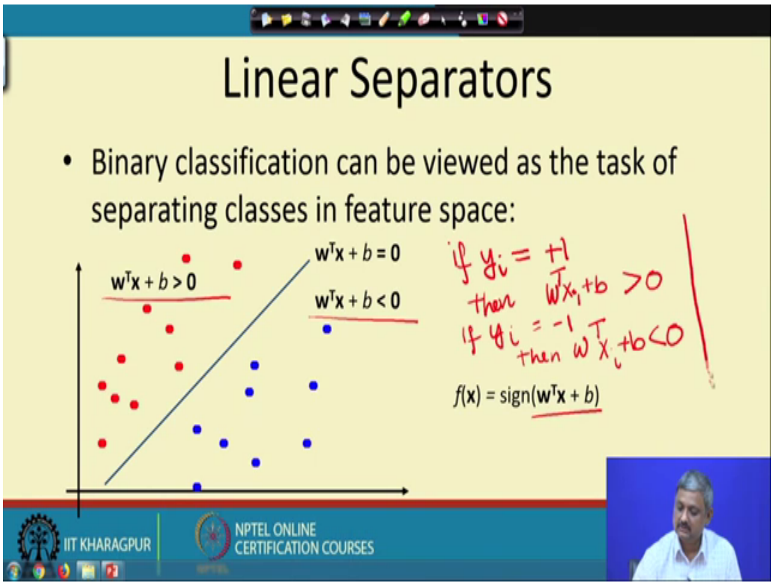
(Refer Slide Time: 00:26)



So, our job was the following given the training set which is of this form a set of vectors in this case 2 dimensional vectors and their class levels 2 y 2 a set of vectors and their class levels. So, in this case we consider a two class problem where y can be either plus one positive class or minus one negative class. So, given this training example our goal is to find a hyper plane in 2 dimensional line of the form w 1 X 1 plus w 2 X 2 plus b equals 0. Find a good linear discriminant which will classify the point. So, in this case I have two points; the red points and the blue points. The blue points are all positive examples and the red points are all negative class examples.

So, I want to find the values of w 1, w 2 and b corresponding to a good linear discriminant. So, if I consider w as the vector of these weights and X as the attribute vector or the feature vector I can rewrite this as in this form w transpose w is a vector X vector plus b equals 0. So, you can extend this beyond 2 dimensions also alright. So, our goal is to find w and b.

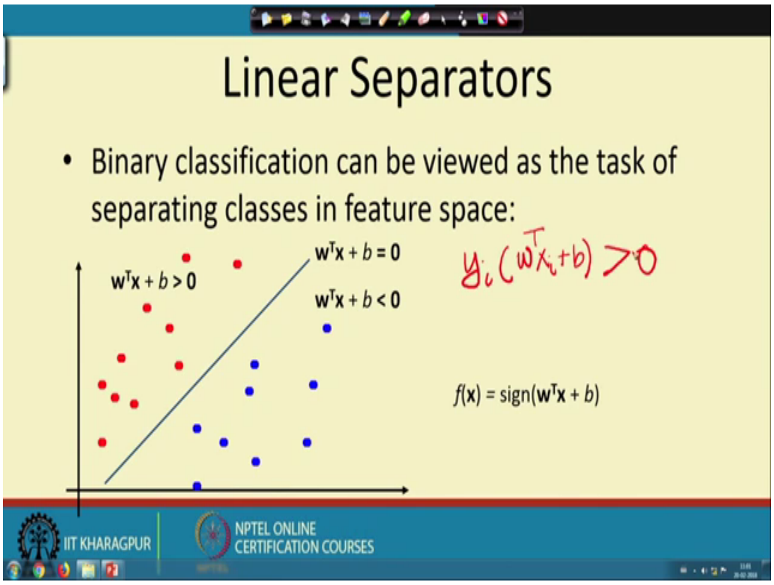
(Refer Slide Time: 02:57)



So, we observed that in order to find a good w and b the first thing you need to do is to ensure that the line corresponding to the w and b puts all the plus points in one side that the non origin side the side in which the origin does not lie and all the minus points the red points in this case in the other side. So, this will successfully do this for all the entire training set, we saw that.

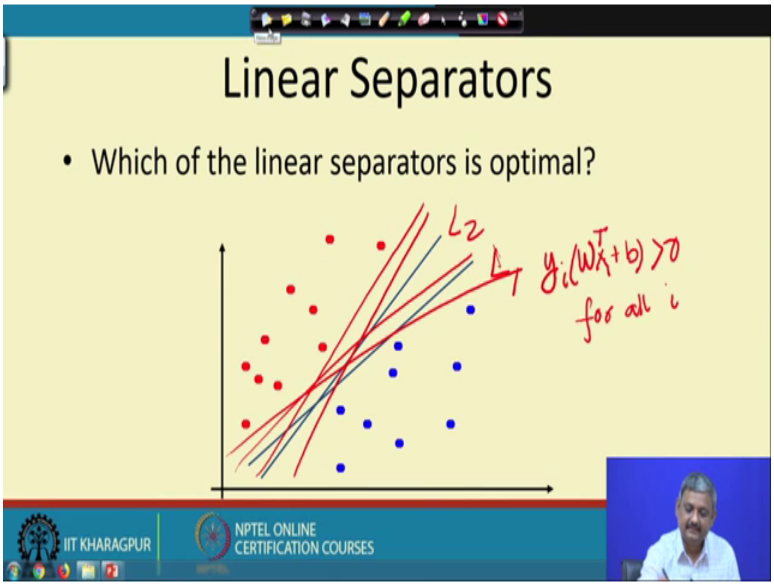
The first condition we need is this finding a good linear separator and then once we find that what we can do is to actually use the sign of this quantity which will decide which side a point X is in origin or non origin and to accordingly from the class. So, in this side it is actually less than 0. So, it is the origin side actually and this side it is greater than 0 and on the line it is equal to 0. So, we observed that to satisfy. So, what we want w and b to be is that if y i for a training set x i y i is plus 1 then w T x i plus b x i is the ith training instant should be greater than 0. If y i equal to minus 1 then. So, this I am basically writing down the condition for correct classification. If y i is plus 1 positive plus x i would lie on the origin side greater than 0 side if y i is minus one x i should as lie on the negative side and what we did was to combine these 2 equations combined into the equations in a single equation which says that.

(Refer Slide Time: 05:11)



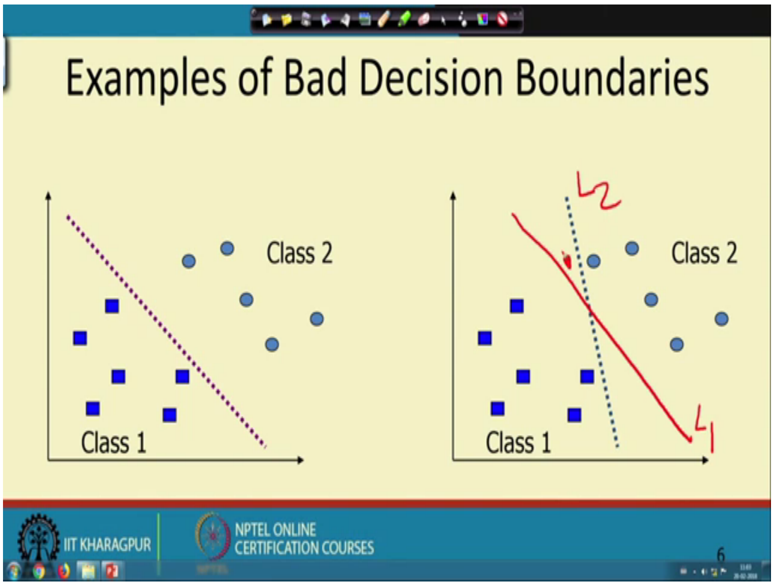
The sign of y i and the sign of this quantity should be the same if y i is negative this would be negative is y i is positive this should positive. So, in other words their sign is same means their product is greater than 0, first quantity. So, I guess this is clear. So, this is the first condition that we need from our linear classifier, alright.

(Refer Slide Time: 05:55)



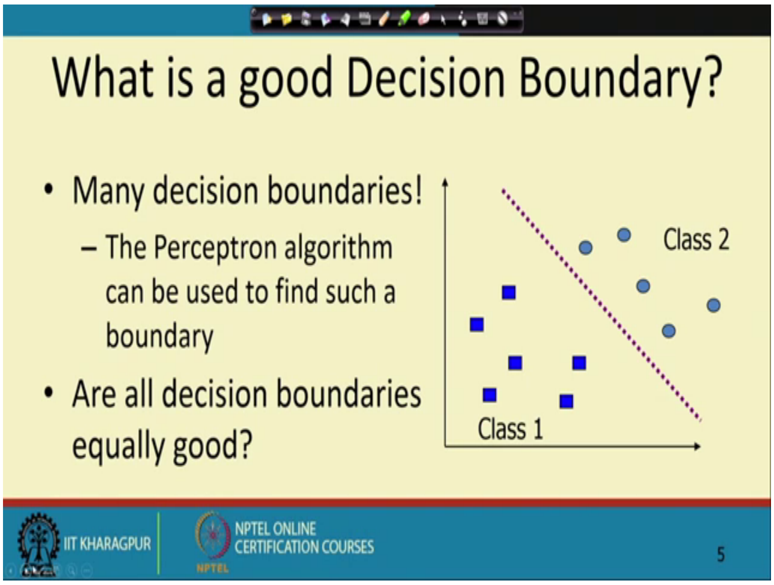
So, let us see we further observe that there are actually many such line which satisfy this condition. So, both these these lines L 1 and L 2 satisfy the condition that w T x i plus b greater than 0, all i; i is the training set x i makes to x i they satisfy. In fact, if the 2 classes are separable we actually showed that there are infinite number of such lines I can draw as many line as I want which will satisfy this condition and what we decided is to sort of not take any line, but take the central line.

(Refer Slide Time: 07:03)



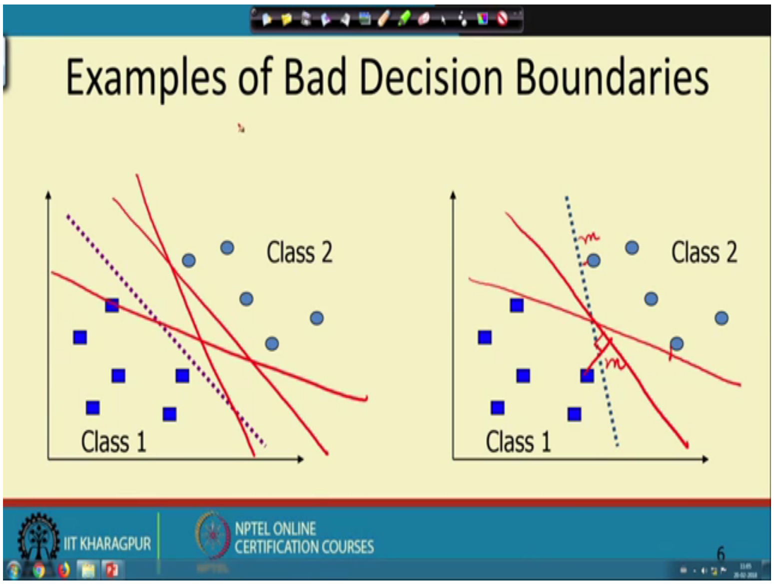
Yes. So, take the central line because of the fact that maybe the lines which are kind of cornered they are not good. So, these 2 are not a good line whereas, at sorry these 2 is not a good line whereas, and the one passing to the centre is probably a good line.

(Refer Slide Time: 07:27)



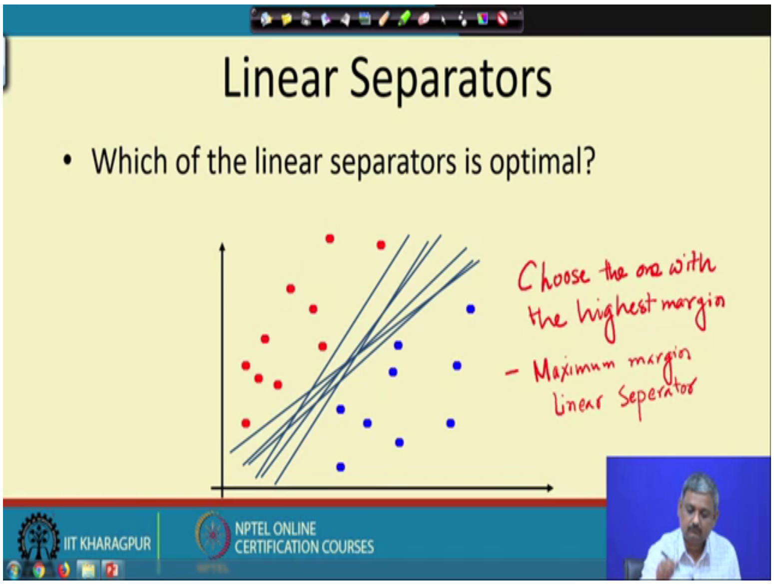
The reason we are good in support of the central line rather than this line is that the performance of this central line let me call it as L 1 whereas, let me call this as L 2 on a new point the generalization performance is better. For example, I have a point here the central line would still put it into class 2 whereas; the corner line L 2 would put it in erroneously as class one. So, generalization is better and we had a mean of quantifying this corner and the central thing. So, we said that is one quantity which is higher in this corner compared to the higher in the centre compared to the corner line.

(Refer Slide Time: 08:32)



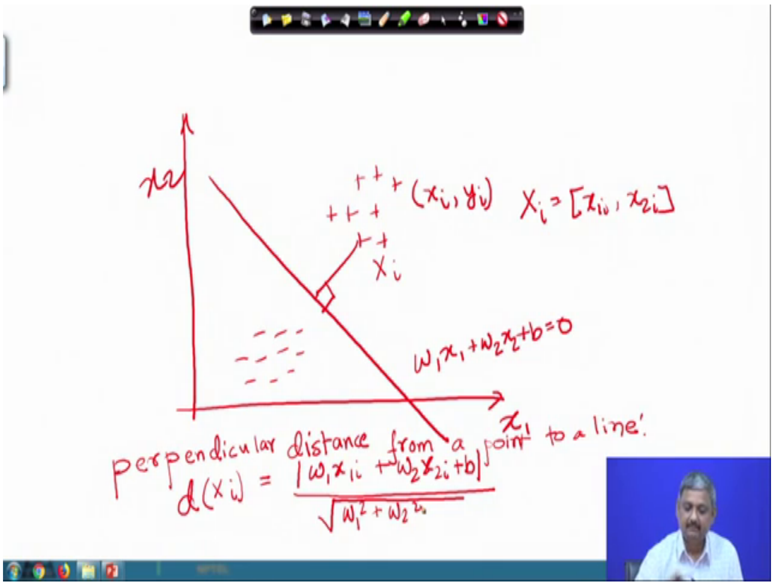
What is the quantity? What we did with among all these training points we find out the closest to the line which one is the closest to the line. For example, this is the closest to this red line, so, I drop a perpendicular from the closest point and I call that perpendicular distance as the margin of that line with respect to these points and then we saw that similarly the margin here is sorry not this only this much. So, the margin is only this much for this line. We saw the central line indeed as the highest margin among them all. So, if you take another line like this till also have a smaller margin draw perpendicular that distance. So, we decided that among all these correctly classifying lines on the training set, we will take the one having the highest margin. So, let me come back to this example.

(Refer Slide Time: 10:17)



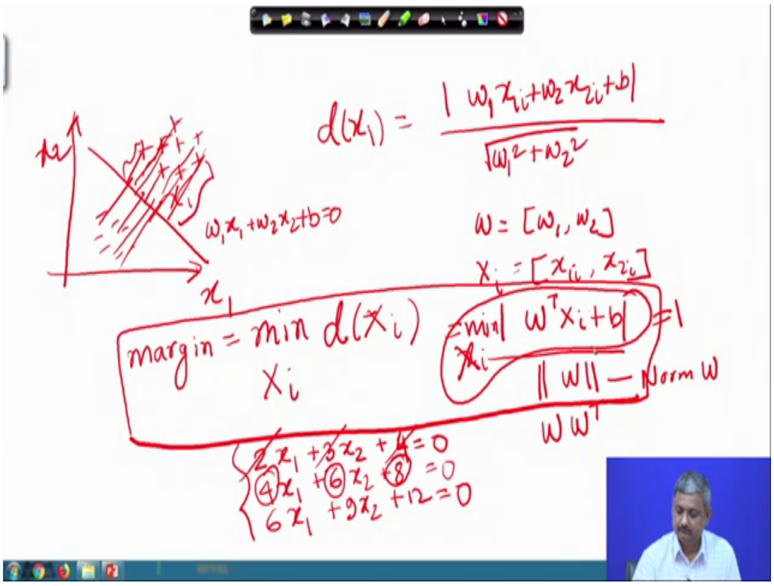
So, among all these lines all this possible line, the highest margin. In other words, you get a maximum margin linear separator all these are linear separators, sorry. We want to take the highest margin linear separator let us see how to do that let me draw an example and explain you.

(Refer Slide Time: 11:27)



So, these are my training points each of the training point is a pair x i and that class level y i. In this particular example X i is a vector having 2 components x 1 i and x 2 i. In general, there can be more components to dimensional in this case. So, if I consider a line of the form the perpendicular distance from certain x i, let me call this as x i to this line. What is the perpendicular distance of a point to a line you remember from your class 12 coordinate geometry? Perpendicular distance of a point x 1 y 1 to a line of this form, let me write down d x i is nothing, but the positive value absolute value of w 1 x 1 i plus w 2 x 2 i just plug in these values remember we are looking at the sign of this quantities now I looking at this magnitude. W 1 x 1 i w 2 x 2 i divided by root over w 1 square plus w 2 square, this is the distance from a point to a line. Let me introduce again the vector notation for this that will be. So, let me draw it.

(Refer Slide Time: 14:14)



Take a point x i. So, d x i to this line is absolute value of w 1 x 1 i plus w 2 x 2 i plus b. So, let me write it in a slightly different vector notation. So, as you remember I have previously defined w vector to be this pair w 1, w 2 in general in higher dimension it will be w 1, w 2, w 3 in 3 dimension and so on and x i to be this vector x 1 i, x 2 i then I can write down d x of i as absolute value of w vector transpose x i the same format written earlier divided by. Note that this denominator is nothing, but the magnitude of the w vector the amplitude of the w vector it is the distance of the w vector w is this vector. So, in notation we introduce for that is this the norm of W, we call it as norm W. In fact, you can write it in a slightly different form also.

In fact, if you take w vector and multiply by it is own transpose you take the w vector multiplied by it is own transpose then you get norm of w square. So, sorry I missed one thing I missed the b here, please do not mind this mistake, I missed the b here. So, anyway to sum up this is my distance from a point to a line and what is my margin. So, distance is this distance take any point draw perpendicular take any point draw perpendicular. These are the distance d i, d j, d 1, d 2 and so on, these are the distances.

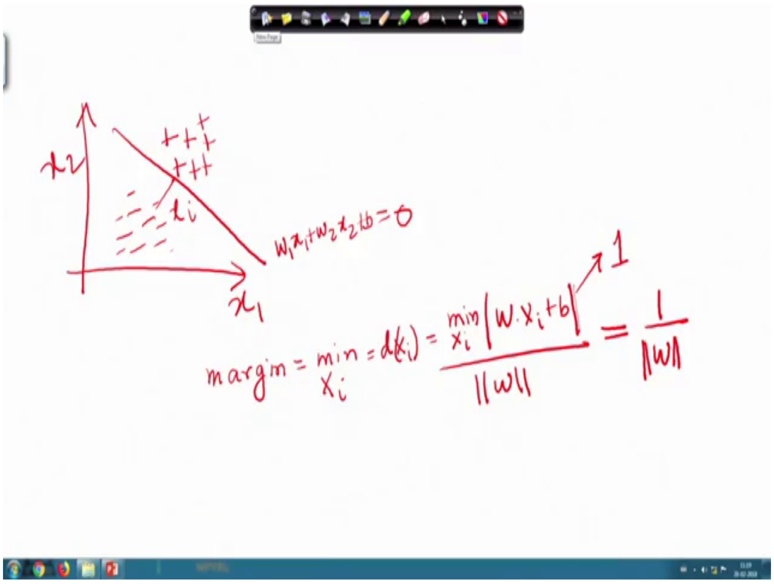
So, can you tell, what is the margin? The margin now you can see is nothing, but the smallest of these values the closest point the perpendicular the closest point you will have the smallest perpendicular distance. So, margin is nothing, but the minimum of these distances the smallest of these distances over all the X i's. So, that is my definition of the margin this is my note down this is a very important quantity in the margin and as I have said before. All I want to do is to is to maximize this margin find w b which for a given training set will maximize this margin which kind of a max min problem maximize the minimum distance.

Now, I will do a small kind of trick you look at these 2 equations of the lines. This is 4. Note the, if you examine this 3 lines they have different values of w 1, w 2 b different values of w 1, w 2 and b, but are they really different lines. No. If you plot them they will turn out to be the same line and they of course, if they are the same line they will have the same margin. So, which one should I took which value of w and b should I took. So, what I will do is the following I will scale w 1, w 2 and b by multiplying by some factor.

So, the smallest of this value of the numerator of this margin for the numerator the smallest value of the numerator, minimum of these over x i. Sorry, capital X i minimum of this value. So, see this numerator for every point there will be some value for every X i there will x 1 there will be some value x 2 there will be some value one of them will be the smallest. So, what I do I multiply w and b by a constant such that this smallest value turns out to be 1. This is a critical point you try to understand it may appear a bit tricky, but you can actually always do a very simple thing that we are doing. What I doing is that I am just choosing a value of w 2, W; I can multiply w and b by the same quantity what I basically done here is that w and b are multiplied by the same constant and got different values, but they are the same line.

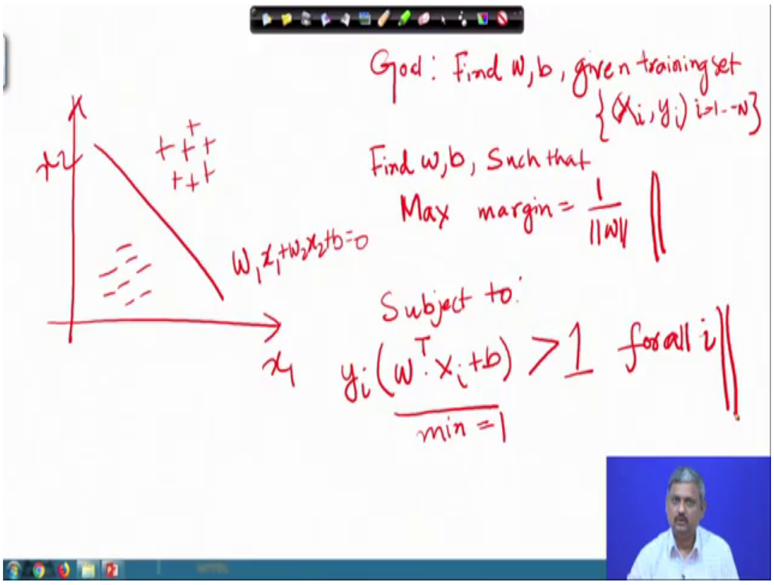
So, I can multiply w and b both by some quantity. So, that the smallest this value overall X i's becomes 1. I can always do that I can always do that you think a little bit you take any line take some point take some examples you will see you can always do that. So, if that is so, then things become very simple for me, then what I can do is that I maybe fear of the things a little bit.

(Refer Slide Time: 22:13)



The numerator of this quantity is independent of X i. So, the minimization I can apply only over the numerator and I have scaled w and b to make this equal to one this entire minimum thing. So, I can rewrite margin as this. So, this seems a pretty nice formula, 1 by the slope of the line, 1 by the coefficient of the line absolute value of the coefficient of the line that seems a nice quantity. So, let me summarize, let me again write down the equation of the final way by which we will compute this line.

(Refer Slide Time: 24:03)



Even the training set, find w and b. So, what I did I said that it has to satisfy 2 conditions maximize margin and linearly separate. So, I write that find w b such that subject to the constant the earlier consent that the sign of y and the sign of this quantity should be same that I had written earlier should be greater than 0 for all i, this is the first condition we had derived that all points should be correctly less positive points on plus side negative points on negative side. In fact, I will modify this equation a little bit if you remember the definition of the margin we said. So, this y is either plus 1 and minus 1.

So, it only changes the sign it changes the value of this product. This quantity the smallest value we have said is the numerator of the margin which we have said it can go down to one the smallest value is 1. So, this entire quantity will be a positive quantity earlier you said it is a positive quantity. Now, we say it is not just a positive quantity it has a value it is positive and it is value is above 1, because this quantity the minimum of this quantity. The minimum of this much this much quantity is 1 by scaling w and b you have already got that that is why you got the numerator to be 1 here.

So, this is an optimization problem maximize something subject to some constant. We will solve this optimization problem to get value of w and b in my next lecture. So, up to this is clear, optimization problems maximize margin subject to separability. So, in my next talk, I end here today. In my next talk I will explain how to solve this optimization problem.

Thank you.