Essential Mathematics for Machine Learning

Prof. S. K. Gupta
Department of Mathematics
Indian Institute of Technology, Roorkee

Lecture - 37
Error Minimizing LPP

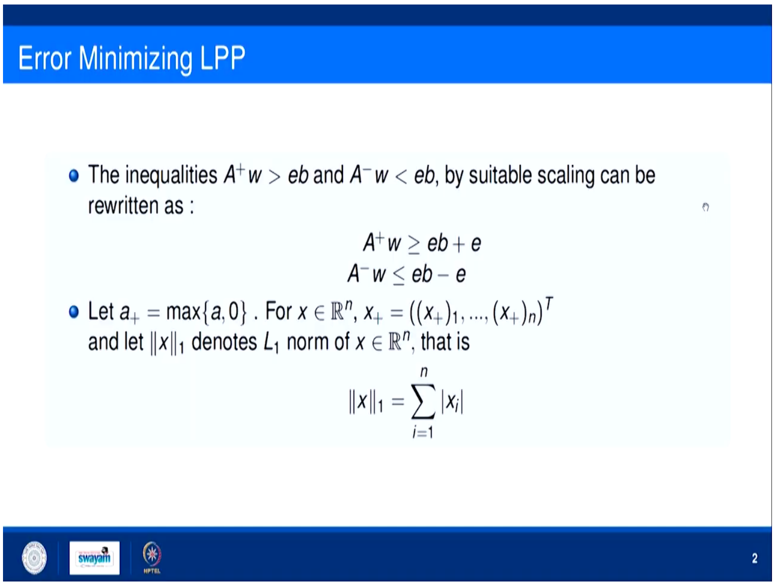
Hello friends. So, welcome to lecture series on Essential Mathematics for Machine Learning. In the last lecture we have see the basic concepts of SVM. We have seen that if you are having two class of data points and, if they are we are able to find a hyperplane which can classify the two classes then they are linearly separable, otherwise they are not linearly separable means they are they will be non-linearly separable.

Also, we have seen one result; that means, if two if we are having a two sets and their convex hulls are disjoint then; that means, they are linearly separable. So, this condition is if an only if so; that means, if the convex hulls are not disjoint, they will not be linearly separable.

So, now the question arises that if is there any mathematical formulation of such type of problems that seeing a mathematical result we can analyse whether a given set of data points are linearly separable or not. So, the answer is yes, and that is error minimising LPP.

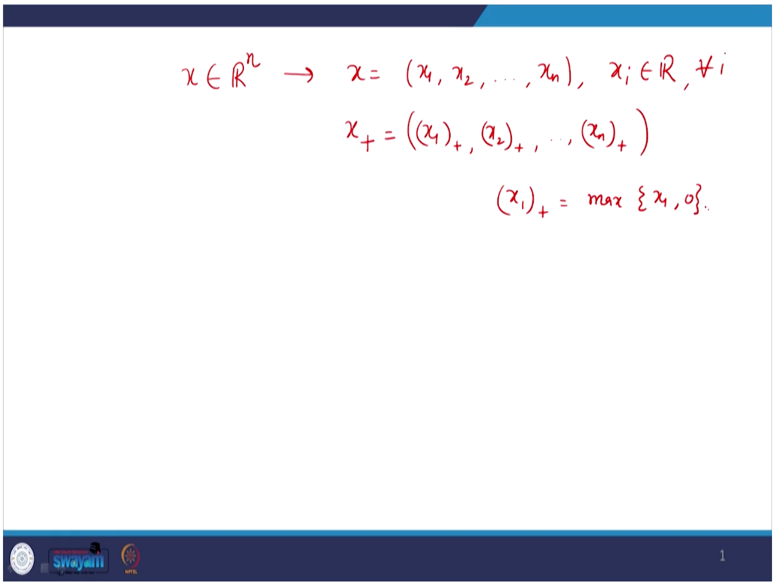
So, what do we mean by and what is how we can formulate error minimising LPP? By seeing that problem we can see that whether the given class of two class of data point is linearly separable or not.

(Refer Slide Time: 01:48)



So, how can we formulate? So, this we have already seen that by a suitable scaling these two inequalities these two inequalities can be convert into these two inequalities ok. We need greater than equal to type or less than equal to type to deal a to deal with a LPP, linear programming problems. How can we solve the problem? Let us suppose A plus is equal to maximum of A or 0.

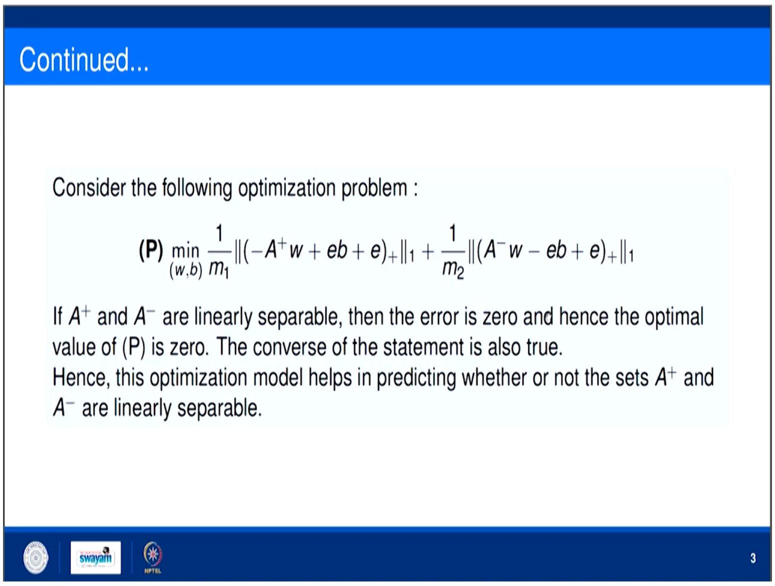
(Refer Slide Time: 02:33)



If it is if A is positive it will be A otherwise, if A is negative it will be 0. Now, if x is in Rn what do you mean by x plus x plus in that case x plus means plus of 1, I mean this x c if x is in Rn such that; that means, x is equal to x 1 x 2 up to x n such that each x i belongs to R for all i.

Now, if I am telling x plus so; that means, x 1 plus x 2 plus and xn plus and plus means? Plus means if I am taking x 1 plus; that means, maximum of x 1 or 0. Now, here I am taking L 1 norm. What do you mean by L 1 norm of x? L 1 norm means simple means summation of mod of x i. x is in Rn so, this L 1 norm here means i from 1 to n mod of x i. Now, consider the following optimization problem.

(Refer Slide Time: 03:37)



See, what optimization problem I am taking see, here this A plus w, I am taking in the right hand side. So, this A plus w when it go right hand side then it will be minus A plus w plus eb plus e less than equal to 0 and similarly here I am bringing these two terms in the left hand side. So, it is A minus w minus eb plus e less than equal to 0.

So, here we are having that term. So, it is there are m 1 points here. So, this m 1 plus m 1 points plus norm 1 and here also m 2 points and norm 1 ok. So, plus means what? Plus means maximum of maximum of, if I am taking a plus that mean maximum of A and 0.

Now, here each term is less than equal to 0 it is less than equal to 0. It is less than equal to 0, if this value entire value comes out to be 0 then this means they are linearly separable why it is so? See, there are two terms here and both the terms are non negative, because norm 1 is there.

Norm 1 means sum of mode values sum of m 1 values here, sum of m 2 values here and if error is if this is comes out to be 0; that means, each term is equal to 0 and if this is 0 this is 0 and this plus means maximum of A and 0. So, if it is 0; that means, none of the term is positive all are 0 then only it will be 0.

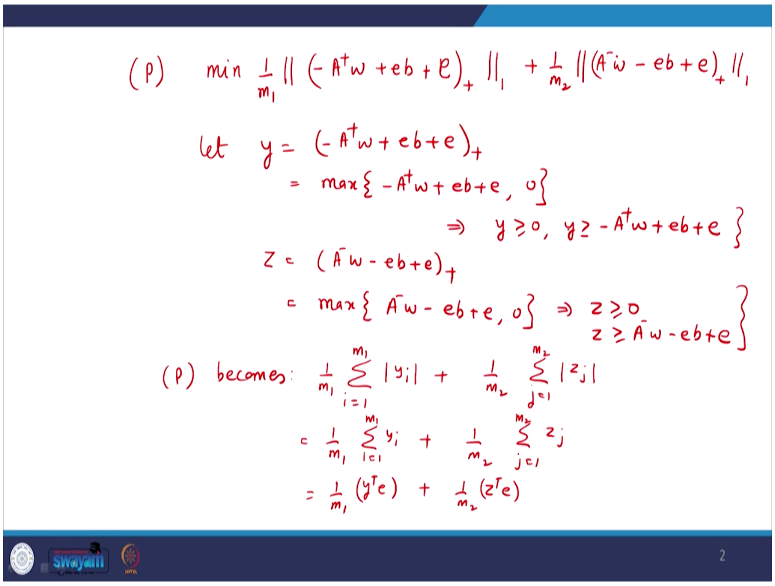
Similarly, here also all the term will be 0 ok. So, if; that means, this inequality is true and here also this inequality is true. So, if this inequality is true and this inequality is true; that means, we are able to find a hyperplane satisfying these two inequalities and; that means, the classes are linearly separable and if this value is not 0, if this value is not 0; that means, there will exist at least one point for at least one point for which this inequality is not holding, does not satisfy, at least one so; that means, the classes are not linearly separable ok.

So, here the same thing is there that if A plus and A minus are linearly separable, then the error, we are calling this is error then the error is 0 and hence, the optimal value of P is 0. The converse of the statement is also true, if it is 0 then the error will be 0.

Hence, this optimization model helps in predicting whether or not the sets A plus and A minus are linearly separable. Now, the question is that it is not a LPP, it is a non-linear problem.

How can we model it in a LPP or a simplified way so that we can predict we can solve it and we can find whether a given LPP is whether a given problem is linearly separable or not. So, to simplify this so let us understand it here.

(Refer Slide Time: 07:17)



So, what is problem P? Problem P is minimum of norm of minus A plus w plus e b minus e norm plus norm 1 plus it is 1 by m 1 times and here 1 by m 2 times norm of, here norm of minus and plus here. So, it is also plus and it is A minus w minus eb plus e norm plus norm 1. So, this is the problem and if this error comes out to be 0; that means, the given data points are linearly separable.

So, now how can we convert this? So, let to minus A plus w plus eb plus e plus. So, this means it is maximum of minus A plus w plus eb plus e or 0. So, if it is maximum of these two this implies; first of all this implies y is greater than equal to 0 and y is greater than equal to minus A plus w plus e b plus e ok.

Similarly, let z equal to; z equal to A minus w minus e b plus e plus. So, that further means maximum of A minus w minus e b plus e or 0. So, this implies z is greater than equal to 0 and z is greater than equal to A minus w minus e b plus e. So, this point and this point.

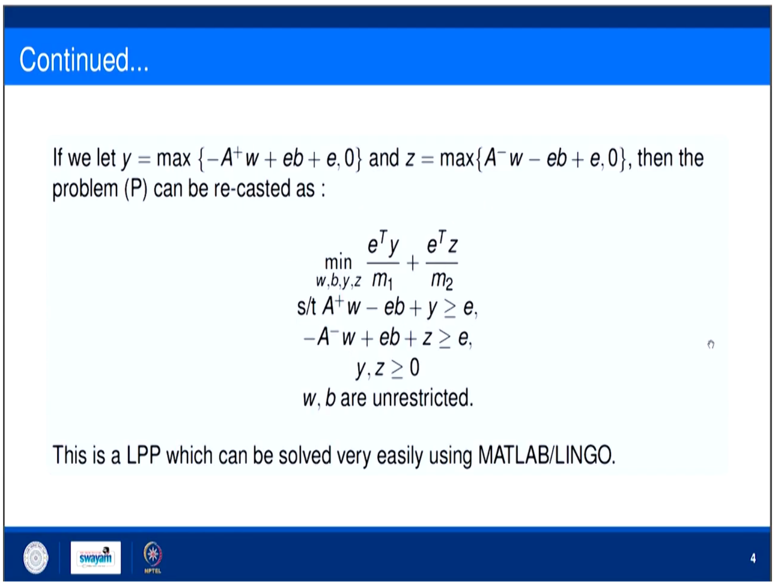
Now, let us put y and z here in P. So, P problem becomes P becomes; see if we put y here, so now, it is nothing, but norm of y and since, y cannot be negative, because y is either 0 or positive from this.

So, hence y is cannot be negative so mod of y will be itself and we are summing it, because by the L 1 norm this is nothing, but sum of see this is nothing, but sum of i from 1 to m 1 times mod of y i times 1 by m 1 plus 1 by m 2 summation j from 1 to m 2 mod of z mod of z j, subject to these conditions, these four conditions.

Now, since y i cannot be negative, because it is maximum of this and this either it is positive or 0 similarly z is either positive or 0. So, it cannot be negative. So, mod will be itself. So, this will be equal to 1 by m 1 summation i from 1 to m 1 y i plus 1 by m 2 summation j from 1 to m 2 m 2 points zj and this can be further written as m 1.

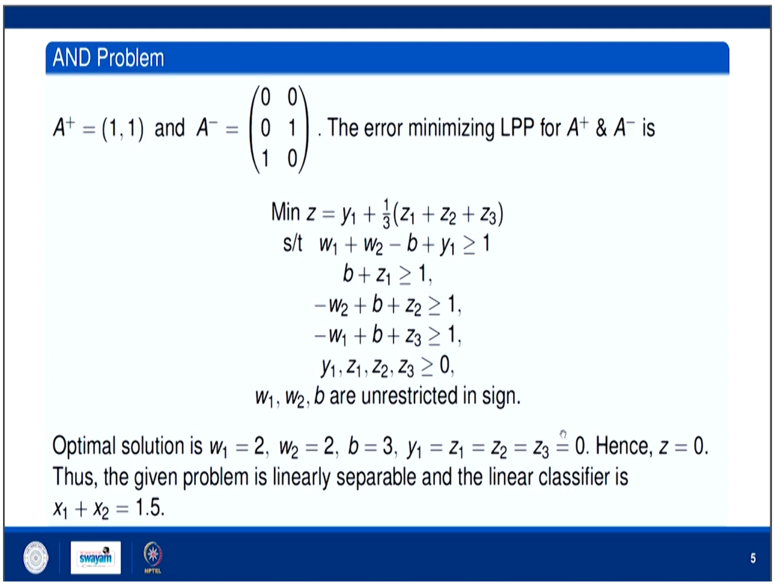
So, it is what? It is y 1 plus y 2 plus y 3 up to m 1 times y m 1, it is z 1 plus z 2 plus z z 3 up to z m 2. So, this can be written as y transpose e, e is vector of ones y 1 plus y 2 plus y 3 up to y m 1 and this is 1 by m 2 times z transpose e z 1 plus z 2 plus z 3 up to z m 2. So, this will be the objective function subject to these constraints.

(Refer Slide Time: 11:15)



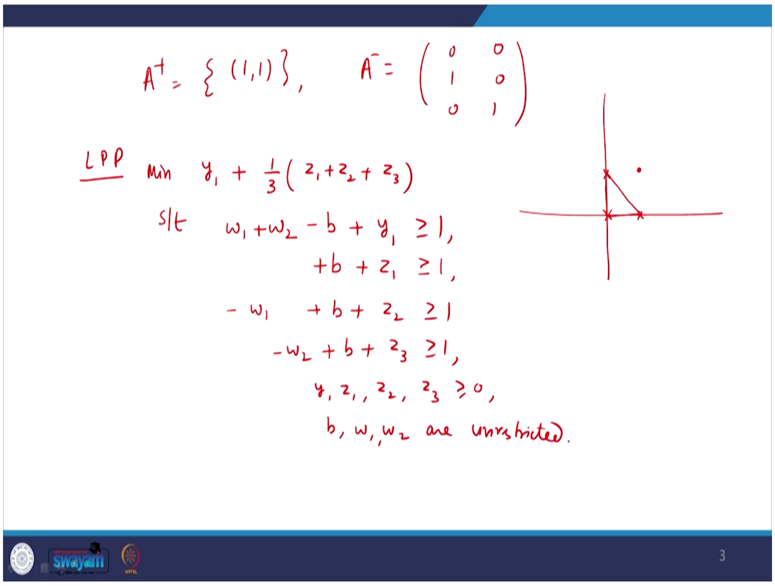
Now, the problem will be what? The problem can be recast into e transpose y or y transpose e both are same upon m 1 plus e transpose z upon m 2 subject to these conditions where w and b are unrestricted in sign. So, this is a simple linear programming problem and we can solved it very easily by using any LPP solver. If this value of the objective function comes out to be 0; that means, the problem is, the classification problem is linearly separable.

(Refer Slide Time: 11:53)



So, let us discuss an example based on this. So, example is we are this is an and problem where A plus having a one data point and A minus having a three data points.

(Refer Slide Time: 12:08)



So, what is A plus and what is A minus? A plus contain only one data point 1 1 and A minus contain three data points 0 0 1 0 0 1. These are three data points in. So, if you see by geometry then A minus is 0 0 1 0 and 0 1 these are three points and A plus having a one point only 1.11. So, of course, the convex hull of these three are only this and the convex hull of these this point is itself.

So, they are linearly separable, but let us verify this result by formulating error minimising LPP. So, how can you formulate the error minimising LPP? So, let us see. How many points here? Only one. So, there will be only one, y say y 1. How many points here? Three points. So, there will be 3 variables here, z 1 z 2 z 3.

So, what be the what will be the objective function? The LPP will be, objective function will be it is y 1 plus 1 by 3 times z 1 plus z 2 plus z 3. It is 1 by m 1 m 1 is 1, because it is only one point 1 by m 2 m 2 are there are three points 1 by m 2. So, minimum of this subject to, what about other conditions?

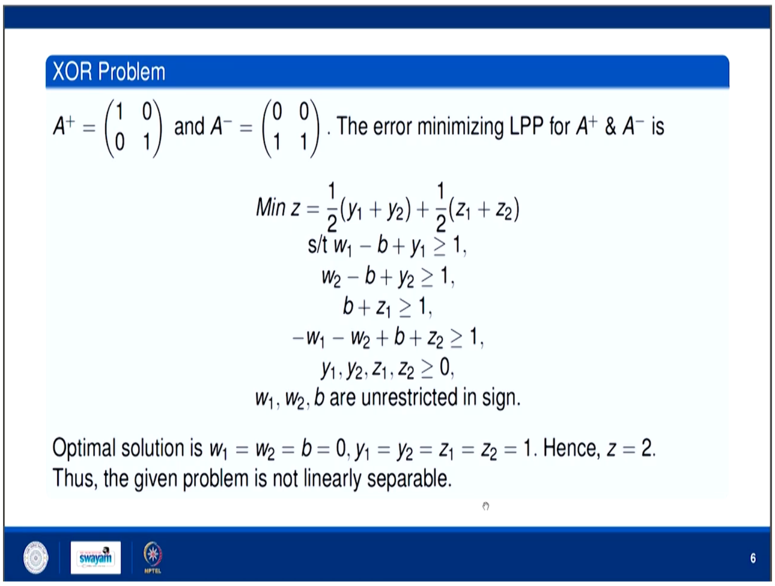
So, let us see the first condition the first condition is A transpose w minus e b plus y greater than equal to e, A plus w. Now, A plus w, A plus means 1 1; that means, A plus w 1 plus w 2, w is not a single w it is in R 2. So, w 1 plus w 2 then it is minus eb minus b plus y 1 greater than equal to 1.

Next is minus A minus A minus w. So, minus A minus w means minus now, when you multiply this first point with w 1 w 2, it will be 0 ok. So, that will not come. So, that will be plus e b plus z plus eb; that means, b plus z 1 greater than equal to 1.

Next third point is again minus of w 1 plus b plus z 2 greater than equal to 1 minus of w 2 plus b plus z 2 z 3 greater than equal to 1 y and z 1 z 2 z 3 are greater than equal to 0 and b w 1 w 2 are unrestricted. So, this will be the error minimising LPP, which we have formulated here or the same we have formulated here also.

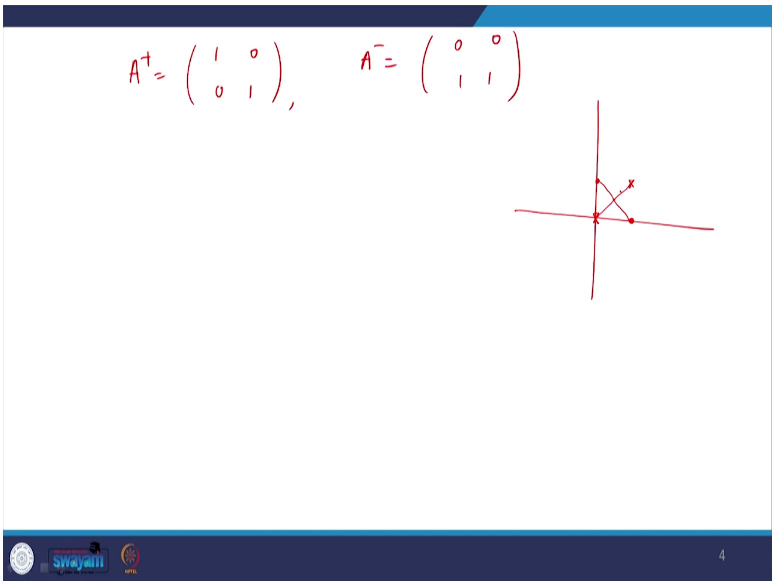
Now, now if you solve if you solve this LPP by any LPP solver so what we obtain? We obtain that w 1 equal to 2 w 2 equal to 2 b equal to 3 and y 1 z 1 z 2 z 3 all are 0 so; that means, the value of the objective function comes out to be 0, because all are 0 and since objective function come comes out to be 0; that means, problem is linearly separable and what is the linear classifier? Linear classifier will be w 1 x 1 plus w 2 x 2 equal to b, w 1 is 2 w 2 is 2 and b is 3 so; that means, x 1 plus x 2 equal to 1.5 it is a classifier ok.

(Refer Slide Time: 16:02)



So, similarly if you are having XOP problem ok. So, what is XOR problem? XOR problem is a plus A plus R 1 0 0 1 and A minus is 0 0 1 1 ok.

(Refer Slide Time: 16:09)

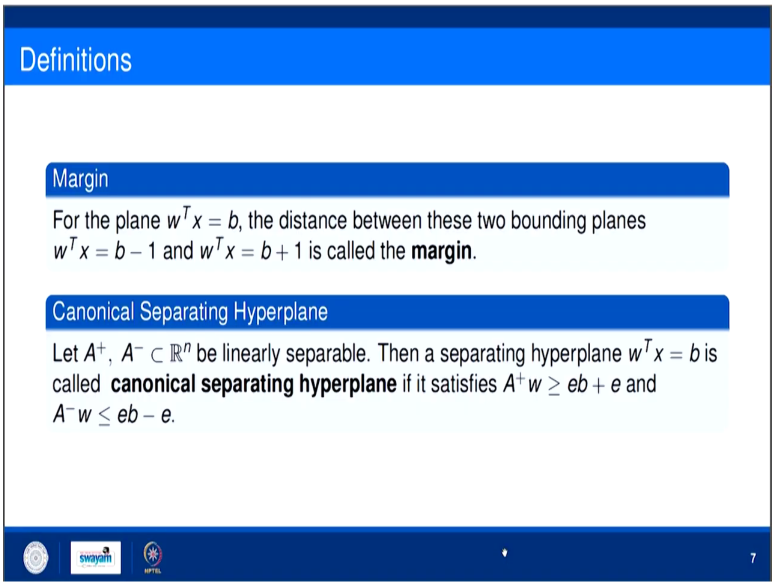


So, if you see here 1 0 1 0 is something here 0 1 is something here 0 0 is something here and 1 1 is something here. So, the convex hull of these two point is this line and the convex hull of these two point is this line. So, they are not disjoint so; that means, they are not linearly separable. So, the same thing we can also verify it by formulating error minimising LPP of the in the same way ok.

So, if we formulate a error minimising LPP. So, we will get this LPP and by solving this we can find the optimal solution as w 1 equal w 2 plus equal to b equal to 0 and y 1 y 2 z 1 z 2 all are 1 so; that means, that means the value of the objective functions comes out to be 2.

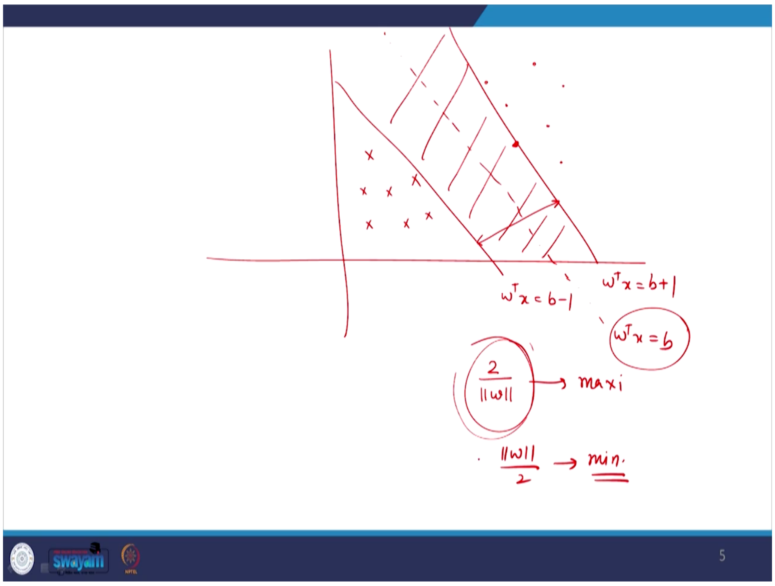
So, since the value of z is not 0; that means, that means the problem is not linearly separable ok. In this way we can analyse whether a given, when whether a binary class classification problem which we are having by formulating a simple LPP, we can find out whether a whether a given classification problem is linearly separable or not. Now, we have some basic definitions, what are that? What are definitions? The first definition is margin. What do you mean by margin?

(Refer Slide Time: 17:50)



For a plane w transpose x equal to b, the distance between these two bounding planes between these two bounding plane w transpose x equal to b minus 1 and w transpose x equal to b plus 1 is called the margin, the distance between these two bounding planes. Now, you are having this type of structure.

(Refer Slide Time: 18:10)

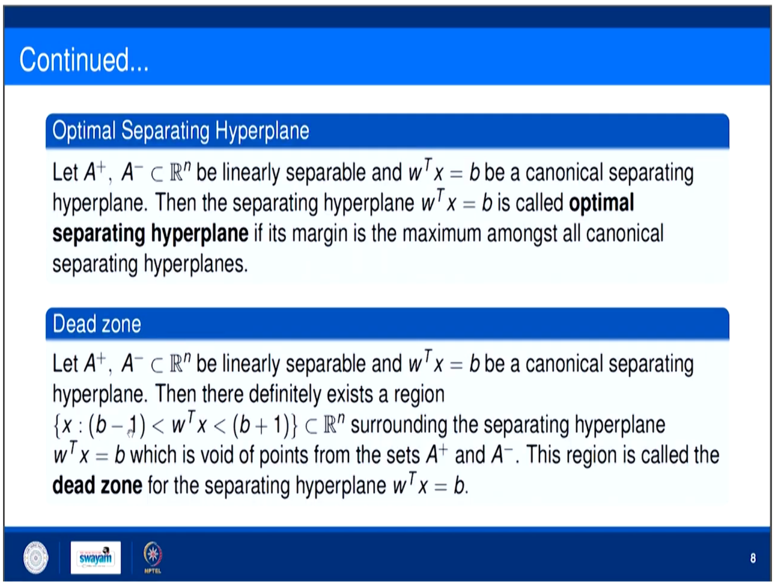


You are having plus 1 class you are having minus 1 class. Suppose, this is suppose this is the w transpose x equal to b minus 1 and this is w transpose x equal to b plus 1. This is on the point w transpose x equal to b plus 1 and this is some plane which is w transpose x equal to b. So, distance between these two bounding planes is called margin.

So, what is the distance? What is the distance between these two planes? So, the distances clearly between these two plane is 2 upon norm of w. They are distance between these two hyperplanes ok. You find out distance from origin to this plane, origin to this plane and the difference between the two is nothing, but 2 upon norm of w that will be the distance between these two hyperplanes.

Next is canonical separating hyperplanes. Let A plus and A minus be linearly separable. We are assuming. Then a separating hyperplane w transpose x equal to b is called canonical hyper canonical separating hyperplane, if it satisfy this and this inequality ok. So, this plane will be called canonical hyperplane, if we are able if this and this inequality holds.

(Refer Slide Time: 19:44)



Now, optimal separating hyperplane. Now, in this if you see, these are linearly separable, but there are infinite number of so many hyperplane which can be found which can separate these two classes, but what will be an optimal separating hyperplane?

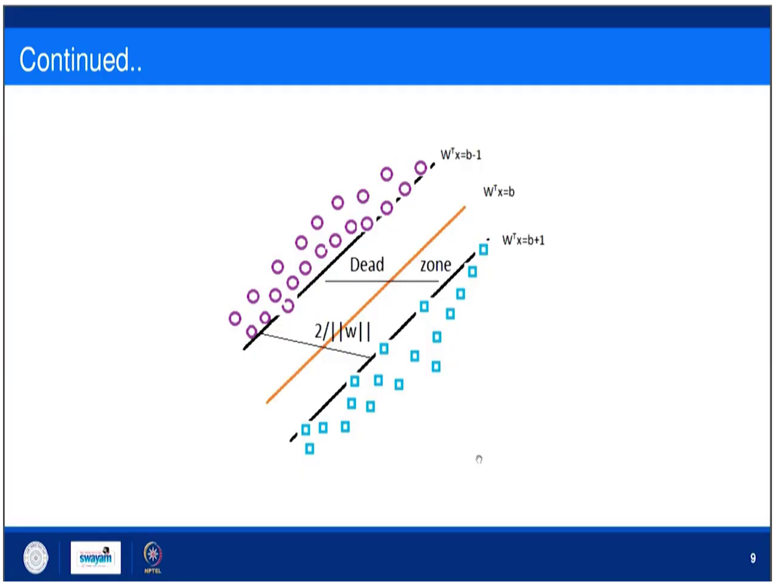
So, how do you find optimal separating hyperplane? So, suppose this is an canonical separating hyperplane, then the separating hyperplane, this is called optimal hyperplane if the margin between these between the bounding plane is maximum and; that means, if this distance, if this distance maximum which is a distance between the two bounding plane, then we say that this plane is called optimal separating hyperplane.

Now, what is the dead zone. So, this here, here this region is called dead zone. So, dead zone basically means all those region where w transpose x is less than b plus 1 greater than b minus 1 this region, where there is no point from A plus or A minus that region is called dead zone.

So, now what will be our aim, next aim? We have seen that whether a two classes are linearly separable or not that we have seen mathematically also, geometrically also, but if we are interested to know that what will be the optimal separating hyperplane; that means, we have to we have to maximise this distance or this to maximise or we have to minimise this distance.

Now, the question is, if we are minimising this distance, then how can we solve this? How we can find out the optimal separating hyperplane?

(Refer Slide Time: 21:40)



So, that thing; so, this I have explained by the geometry. So, that thing we will explain in the next class that if we are having a, if you are interested to find out an optimal separating hyperplane given that it is a linearly separable, then how can we formulate an optimization model for this.

So, thank you very much.