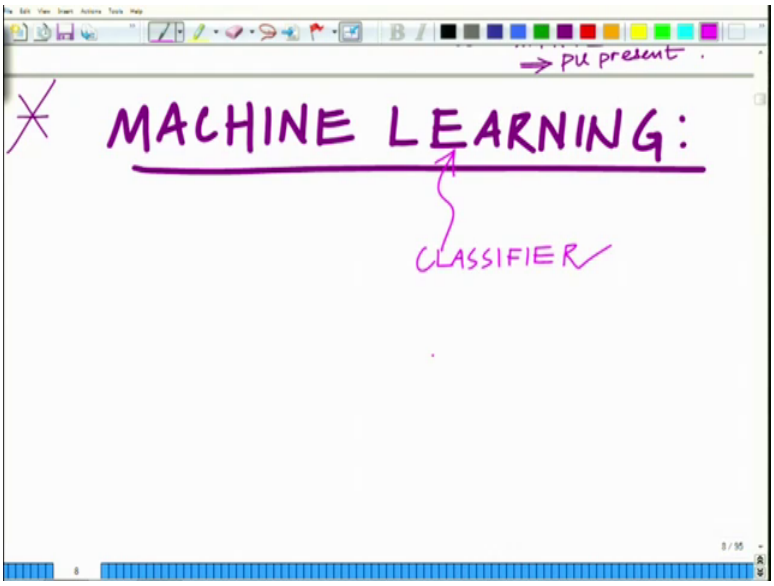
Applied Optimization for Wireless, Machine Learning, Big Data
Prof. Aditya K. Jagannatham
Department of Electrical Engineering
Indian Institute of Technology, Kanpur

Lecture – 61
Practical Application: Linear Classifier (Support Vector Machine) Design

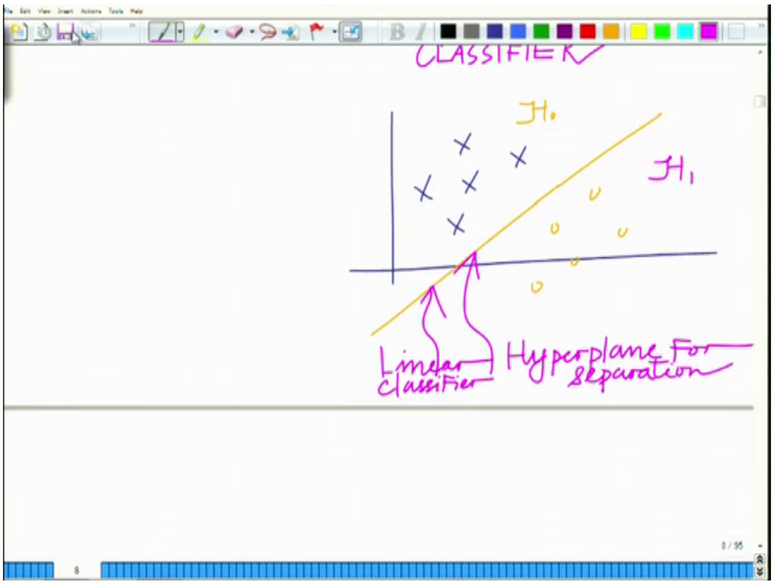
Hello welcome to another module in this massive open online course. So, we are looking at convex optimization and its application for machine learning alright and let us continue our discussion.

(Refer Slide Time: 00:23)



So, what we want to look at is we are looking at applications of convex optimization or machine learning. And, well in this and in particular we are looking at classification or building the optimal classifier and we have seen that this can be done as follows.

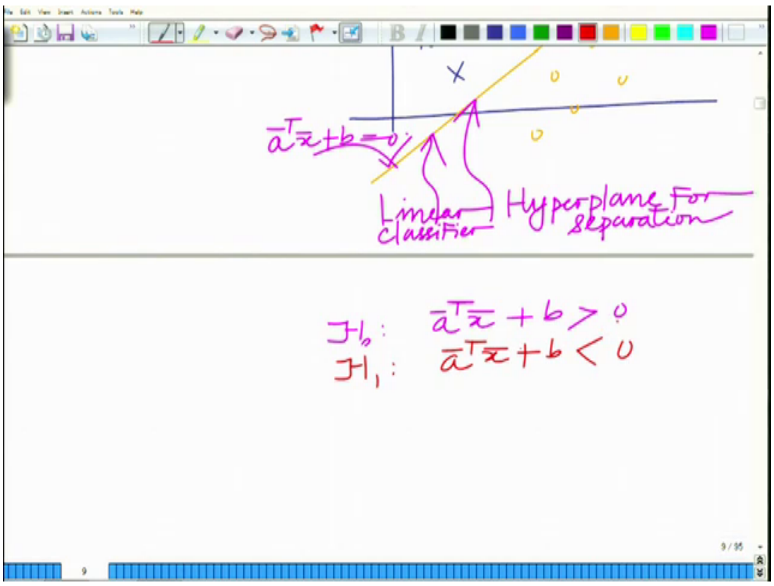
(Refer Slide Time: 00:55)



If you have a set of points test data set two sets of points corresponding to hypothesis 0 and hypothesis 1. So, this is hypothesis 0 this is hypothesis 1 this is absence of hypothesis 0 is absence of primary user hypothesis 1 is presence of primary user.

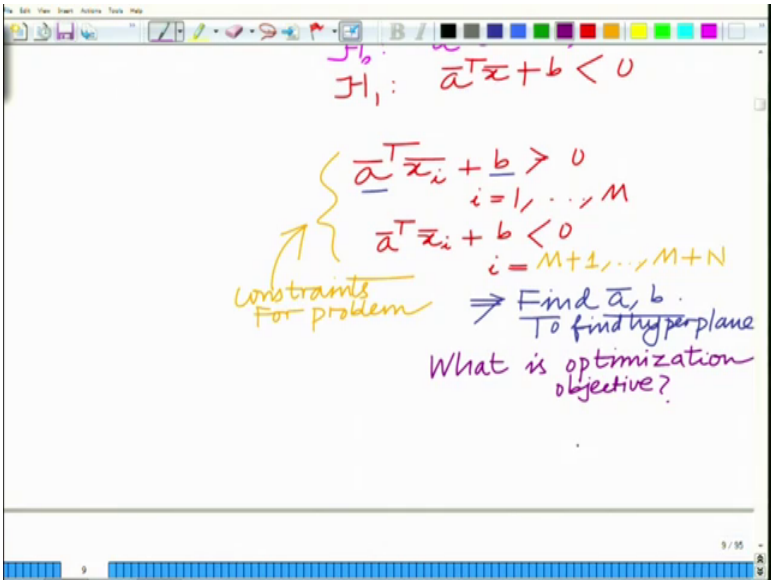
And this is the hyperplane that is separating them, this is the hyperplane for separation. And since this is linear alright you can also call this linear separation or linear classification you can also call this as a linear classifier.

(Refer Slide Time: 01:53)



And we said for this to be true if this is your hyperplane a bar transpose x bar plus b equals 0 all the points corresponding to H 0 must satisfy a bar transpose x bar plus b greater than 0. And points corresponding to H 1 a bar transpose x bar plus b less than 0 alright and the way we build this is we have this x bar i is M plus N points.

(Refer Slide Time: 02:24)

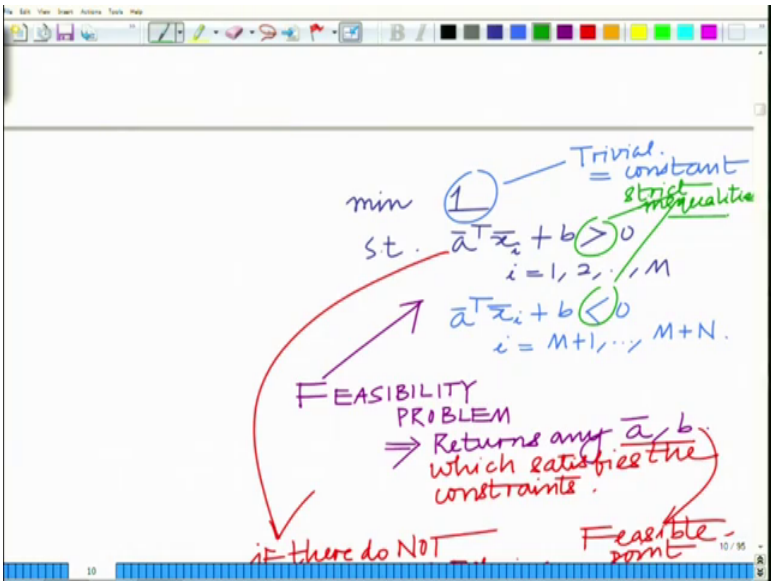


So, we want to design this find the hyperplane such that a bar transpose x bar i plus b greater than 0 for i equals 1 up to M and a bar transpose x bar i plus b is less than 0 for i plus 1 or i equals M plus 1 for that matter i equals M plus 1 up to M plus N and you can say these are our constraints. So, we have constraints for our problem constraints for our problem of the to design the optimal classifier alright. That can separate these two sets of points in the training data set or the test data set, now the problem is the first point is now these are the constraints.

Now, what is so we have to find the some optimal hyperplane implies we have to find a bar comma b to find the hyperplane. Now what is the optimization objective here the constraints find: what is the optimization? What is optimization objective and you will see that we do not have an optimization objective alright. Now, how do you formulate the optimization problem in this context and you will realize something interesting is that given this optimization paradigm or given this constraint we do not need an optimization objective.

There is any hyperplane or any combination of a bar and b which satisfies these constraints that is for all points x bar i i equal 1 2 M a bar transpose x bar i plus b greater than 0 and i equal to M plus 1 2 M plus N a bar transpose x bar i plus b less than 0. Any a bar and b satisfying this set of constraints is fine with us which means we can formulate a trivial, we can formulate an optimization problem with a trivial optimization objective and that is as follows.

(Refer Slide Time: 05:01)

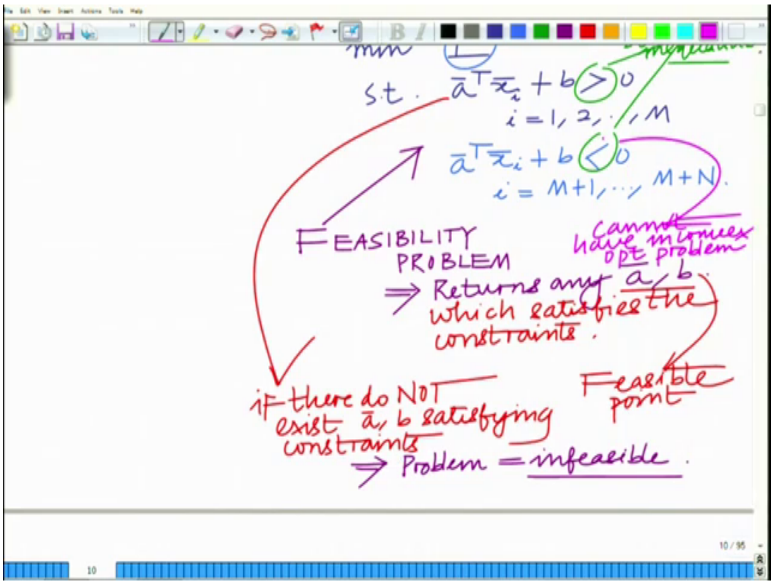


So, I can simply set the optimization objective to 1 alright any constant does not matter minimize 1 such that a bar transpose x bar x i plus b greater than 0 i equals 1 2 up to M. And a transpose x bar i plus b is less than 0 i equal to 1 or i equal to M plus 1 up to. So, this optimization objective is a trivial optimization objective, the objective is constant which means it cannot be minimized any further alright the objective is constant at 1. So, what will this return this will return any feasible point, any feasible point in the sense any a bar and b which satisfy the set of constraints alright which are able to separate these two sets of points.

And in fact, that is what we are looking for and such a problem so this will result so this will yield a feasible a bar and b and this type of optimization problem is a trivial objective is termed as the feasibility problem. Because we are only interested in saying is this problem feasible are they linearly separable, if they are linearly separable what is any hyperplane that linearly separates these two sets of points.

So, this is a feasibility problem so this is a trivial objective. So, this is also termed as the this is also termed as the feasibility problem implies returns any a bar comma b which satisfies returns any a bar comma b which satisfies the constraints above. So, we are only interested in finding a feasible point so this is basically a bar and b.

(Refer Slide Time: 07:28)



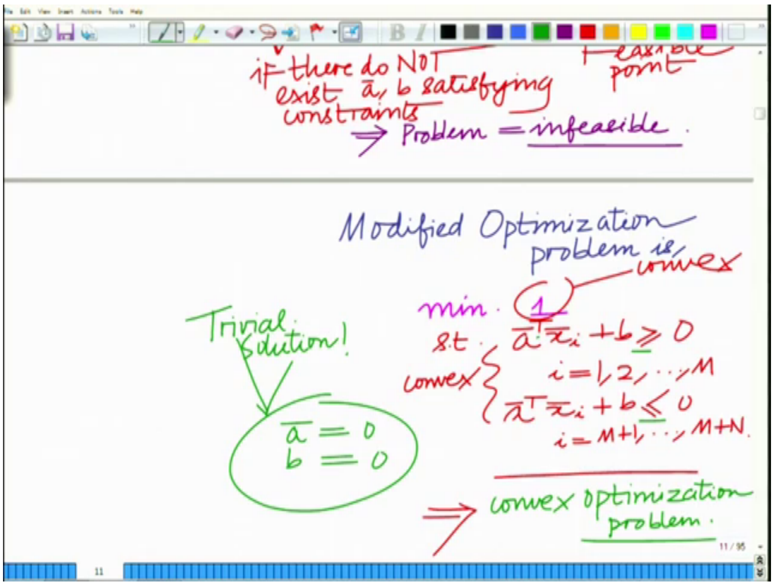
So, we are asking the question is this problem feasible does there exist any point which satisfies the constraints. If there is exist any point which satisfies the constraint the problem is feasible otherwise the problem is infeasible alright.

If there do not exist a bar and a bar comma b satisfying constraints implies problem is, the problem is infeasible. So, this is simply feasibility problem we are trying to check if the problem is feasible and that returns the a bar and b which characterize the hyperplane and that can be used for linear separation alright. Once you have the a bar and b you have a new point x bar plug it into this a bar transpose x bar plus b, if it is greater than 0 implies hypothesis H 0 is less than 0 implies hypothesis H 1 which means primary user is present.

Now, we still have a problem what is the problem if you look at this carefully you will observe that these constraints are strict inequalities. So, these are strict inequalities and you cannot have these strict inequalities in a convex optimization problem.

Because you have to include the boundaries cannot have in a cannot have the strict inequalities in the convex optimization problem alright, this is not a convex out where you have the strict inequality alright. You cannot have it is the optimization resulting optimization problem is non convex which means the inequalities cannot restrict alright; not a problem we can modify the strict inequalities we can simply modify the optimization problem as follows.

(Refer Slide Time: 09:58)



So, our modified optimization problem is the followings we have the same objective trivial objective minimize 1 subject to a bar transpose x bar i plus b greater than or equal to 0. Now from strict inequality I have simply made it greater than or equal to and a bar transpose x bar i plus b less than or equal to 0 i equal to M plus 1 up to M plus N very easy alright.

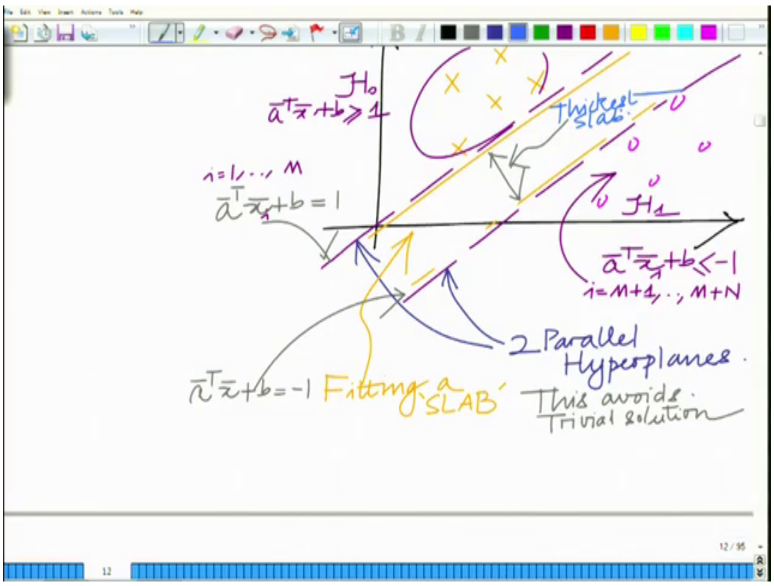
Now, you do not have strict inequalities anymore these are half spaces a bar transpose x bar i plus b greater than equal to 0 each constraint represents a half space alright. So, these are a fine and therefore, the constraints, so these are convex functions alright each function is of the constraints is a convex function the objective is constant in fact; this is trivially convex.

So, constraints are convex and this is a convex optimization process everything is fine now everything seems fine, but there is a problem I am going to point it out. This is a convex optimization problem objective is convex constraints are convex so this is a convex optimization problem where then is the problem in this. The problem rise in the fact that the moment you specify this and remember it is a feasibility problem we are simply interested in finding a feasible a bar and b if you observe closely you will see that you have this greater than equal to 0 less than equal to 0.

If you set a bar equal to 0 b equal to 0 correct that trivially satisfies this problem a bar transpose x bar i plus b equal to 0 a bar transpose x bar i plus b equal to 0 for i equal to in fact, 1 up to M plus N. So, this trivial solution so now the problem is this feasibility problem will always have the trivial solution. So, this would simply yield a bar equal to 0 b equal to 0 alright even if the points are not separable it will simply yield a bar equal to 0 b equal to 0 alright. So, you are stuck now with the series of trivial previously you had a non convex optimization problem the moment you are relaxed the strict inequalities make them inequalities already you are stuck with the trivial solution a bar equal to 0 b equal to 0 and that is precisely the problem alright.

So, you will have to work out another approach which does not yield the trivial solution, but yields is actually a hyperplane that separates and a convex optimization problem alright. We want a convex optimization problem that yields hyperplane that separates these two subsets of points and that is what we want to do next. And to do that we will now further modify this optimization problem as follows what we are going to do, so what we want to do is.

(Refer Slide Time: 13:40)



Now, we have these two sets of points hypothesis H 1 and hypothesis H 0 and hypothesis H 1. Now, we will modify this problem to design not one, but to design two hyperplanes and that is the novel solution. So, we want to modify it to design not one, but separate them by 2 parallel hyperplanes. In fact, what you are doing is you are fitting a slab if you look at this what you are doing is you are fitting a slab not just a hyperplane.

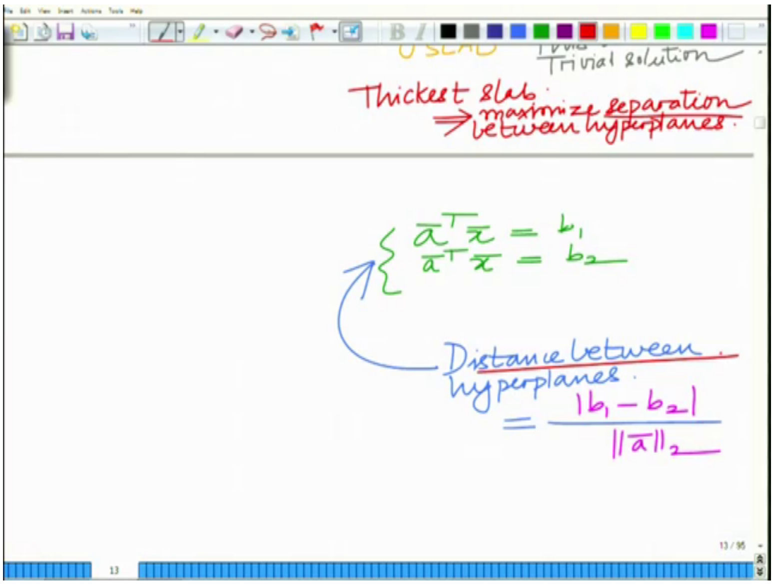
But of course, this is a continuous slab you are fitting a slab this hyperplane is characterized by let us say a bar transpose x bar plus remember they are parallel. So, a is going to be same only the constant is going to change, so a bar transpose x bar plus b equal to minus 1. And this hyperplane is characterized by a bar transpose x bar plus b equal to 1 I am sorry this one is a bar transpose x bar plus b equal to 1 this other one is a bar transpose x bar plus b equal to minus 1.

And therefore, all the points in hypothesis H 0 will satisfy a bar transpose x bar plus b greater than or equal to 1 and all these points in hypothesis H 1 will satisfy a bar transpose x bar plus b less than or equal to minus 1. This is for this is for i equal to 1 up to M and this is for i equal to M plus 1 up to M plus N.

Now, we are fitting a slab between these two sets of points in the training data set now how do we want to design that slab alright. So, now, we want to fit a slab ok, so now, we want to fit a slab. Now further if you observe so now, that problem is solved so we want we do not know so we avoided that problem the previous trivial solution by designing two hyperplanes not a single hyperplane two hyperplanes. So, this avoids the trivial solution by the way you cannot now simply have a bar b bar equal to 0 alright.

Because if a bar and b are equal to 0 then left hand side will be 0 it cannot be greater than equal to 1 and less than equal to minus 1. So, this avoids the trivial solution this avoids the trivial solution. Now further what else can we do we can do something interesting we can not only fit a slab but we want to fit the thickest slab that is we want to maximize the separation between the hyperplanes to make it robust.

(Refer Slide Time: 17:29)

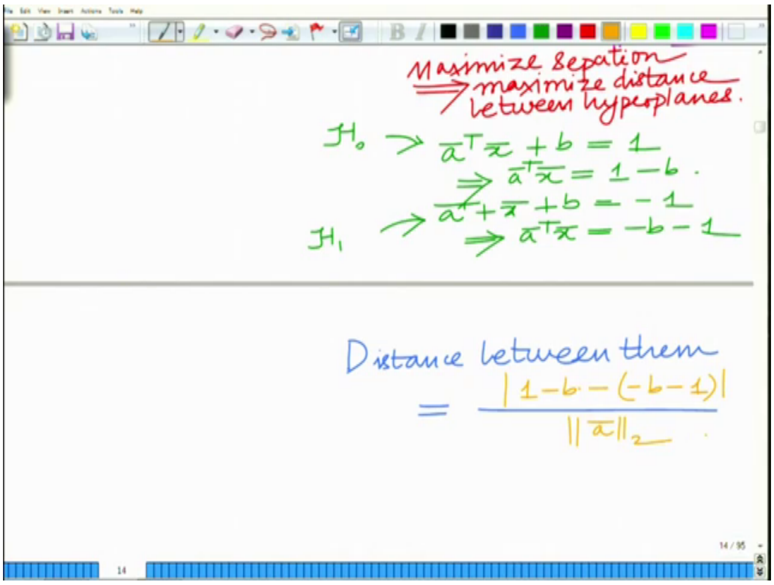


We want to find the thickest slab implies the thickest slab implies maximize the separation maximize the separation between the hyperplanes only do you want to not just fit any slab, but do you want to fit the thickest slab. So, want to maximize the separation between the two hyperplanes how do we do that.

Now, we know that and we must have seen this in one of the problems before that if we have to parallel hyperplanes a bar transpose x bar equal to b 1 a bar transpose x bar equals b 2. The distance between these two hyperplanes if you look at the distance between these two hyperplanes this is equal to b 1 minus b 2 magnitude b 1 minus b 2 by norm of a.

In the two norm of a this is the distance between the two hyperplanes therefore, to maximize the slab maximize the separation implies we have to maximize the distance between the hyperplanes.

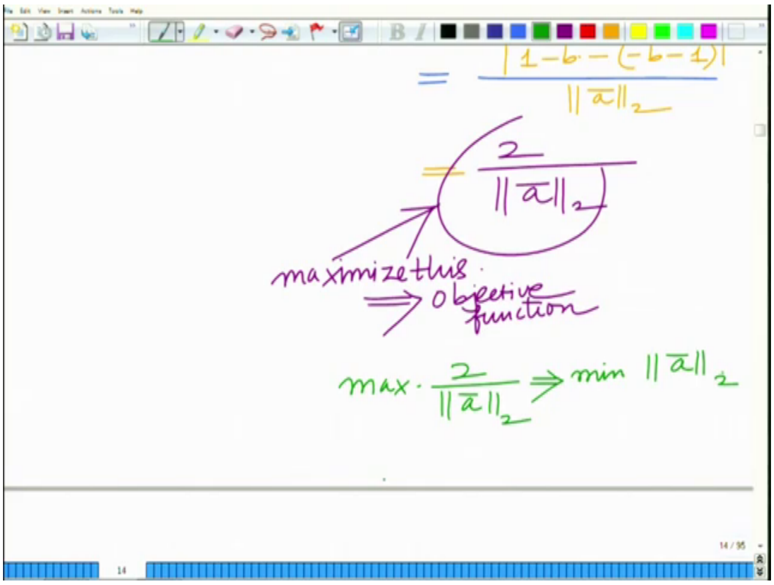
(Refer Slide Time: 19:05)



So, maximize the separation implies maximize the distance between the two hyperplanes. Now, what are the two hyperplanes remember the two hyperplanes are a bar transpose x bar plus b equals 1 implies a bar transpose x bar equals 1 minus b. This is the H 0 hyper hyperplane and a bar transpose x bar plus b equals minus 1 implies a bar transpose x bar equals minus b minus 1 this is the H 1 hyperplane.

And the distance between them distance between hyperplanes this is simply magnitude 1 minus b minus minus b minus 1 divided by norm a bar this is the 2 norm of a bar which is equal to now you can see 2 divided by the 2 norm of a bar this is the distance between the hyperplanes. So, we have to maximize this separation implies this is our objective function or this is basically our cost function.

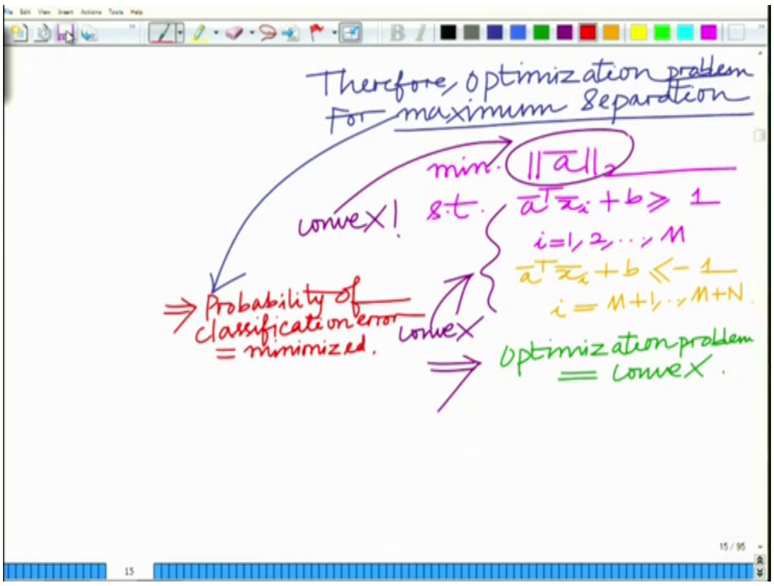
(Refer Slide Time: 20:38)



We have to maximize is to maximize the distance of separation between hyperplanes we have to maximize the distance between the hyperplanes which is nothing, but 2 divided by norm a bar.

Now maximize 2 is a constant norm a bar is positive maximize 2 divided by norm a bar implies you can equivalently minimize norm a bar the 2 norm all these are 2 norms. Because, maximizing 1 over norm is 2 over norm a bar is maximized when norm a bar is minimized and that is now our optimization from.

(Refer Slide Time: 21:49)



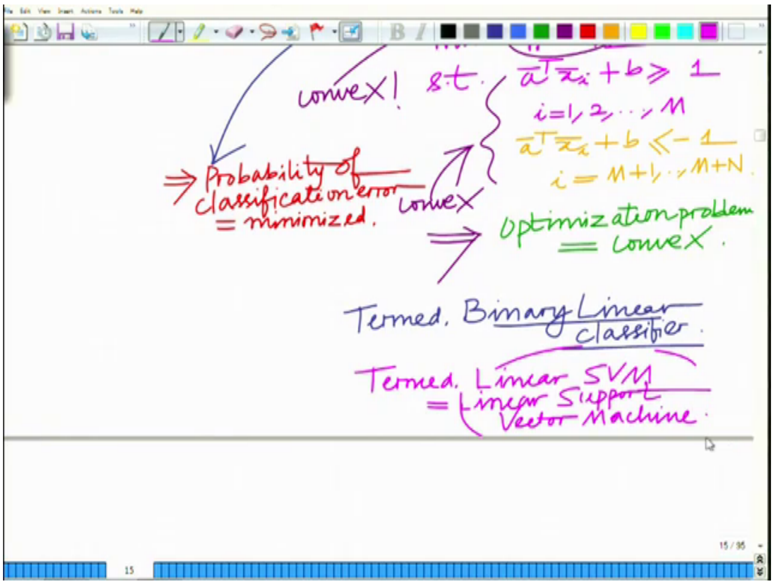
Therefore the optimization problem for maximum which you can also say is a robust separation problem is minimize norm a bar such that sorry subject to a bar transpose x bar i plus b greater than or equal to 1, for i equals 1 2 up to M. And a bar transpose x bar i plus b less than or equal to minus 1 for i equal to M plus 1 up to M plus N.

And now if you see the norm this is convex norm is convex these are half spaces, these are convex implies optimization problem equals convex. And more importantly there is no trivial solution for this a bar equal to 0 b equal to 0 does not satisfy this alright. So, the trivial solution has been avoided, we have a convex optimization problem avoided to the trivial solution and we are finding the hyperplanes such that you are fitting the thickest possible slab or you have you have the set of hyperplanes with the maximum possible separation between them separating these two sets of points alright.

So, there is the probability the error the probability of classification error therefore, is going to be because remember as the separation becomes smaller and smaller there is a high chance that because of noise you might have point from one set crossing over into another set alright, because of noise you might have the point being misclassified. So, the moment you are maximizing the separation between two hyperplanes the probability of error also becomes minimum alright.

So, when you maximize the separation this implies the probability of classification error is automatically minimized, the probability of classification error is minimized. So, this is the linear classifier and linear classification into two sets this is termed binary.

(Refer Slide Time: 25:04)



This is termed binary linear classifier and is also termed as a linear SVM where SVM stands for support; this is one of the cutting edge in this is cutting edge in classification linear support not supporting; I am sorry support vector linear support vector machine linear SVM or linear support vector machine ok. So this is in fact, the most you can say one not very latest, but this is definitely the cutting edge or one of the a very popular alright. And we do a very popular and very efficient mechanism or a very efficient tool for linear separation. In fact, it can be extended also fairly easily to non-linear separation alright.

But in this current from it is simply a linear SVM that is a linear support vector machine which is employed for which can be employed as a binary release in linear classifier to classify two sets of points. We have seen a simple example in a cognitive radio scenario you have once you sense the spectrum, you have a measurement you would like to classify if it belongs if it belong a if it corresponds to either hypothesis 0, that is the primary user is absent or hypothesis 1 primary user are present alright.

So, all such binary classification problems so or several such binary classification a broad class of binary such classification problems can be handled by the linear SVM. And in fact, because this is a convex optimization problem it can be solved fairly efficiently to come up with a linear classifier alright. So, we will stop here and continue in the subsequent modules.

Thank you very much.