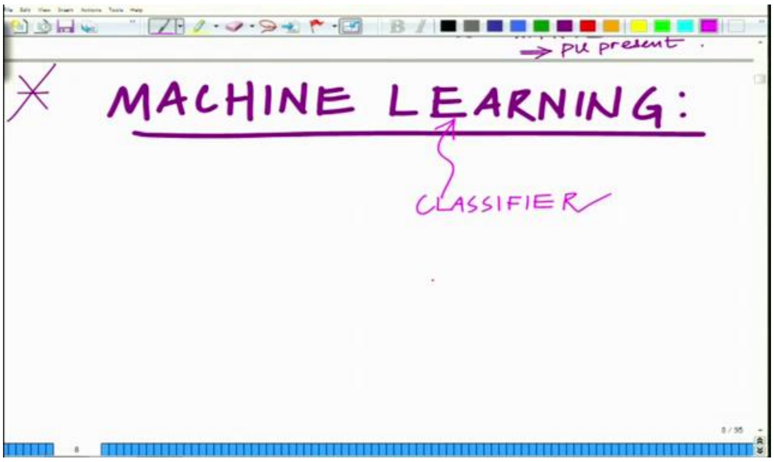
Applied Optimization for Wireless, Machine Learning, Big Data
Prof. Aditya K. Jagannatham
Department of Electrical Engineering
Indian Institute of Technology, Kanpur

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

Keywords: *Linear Classifier, Support Vector Machine*

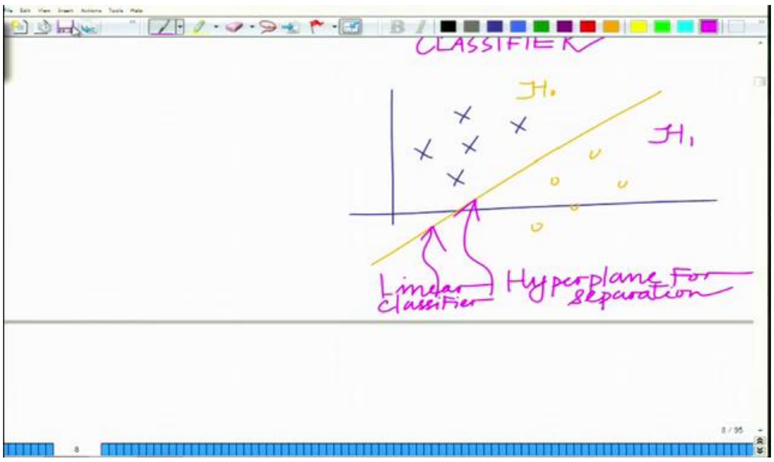
Hello welcome to another module in this massive open online course. So we are looking at convex optimization and its application for machine learning and let us continue our discussion.

(Refer Slide Time: 00:23)



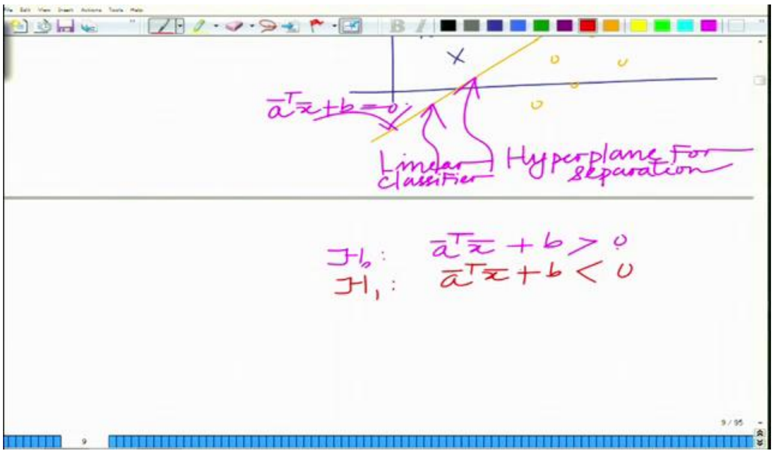
So we are looking at applications of convex optimization or machine learning and in particular we are looking at building the optimal classifier.

(Refer Slide Time: 00:55)



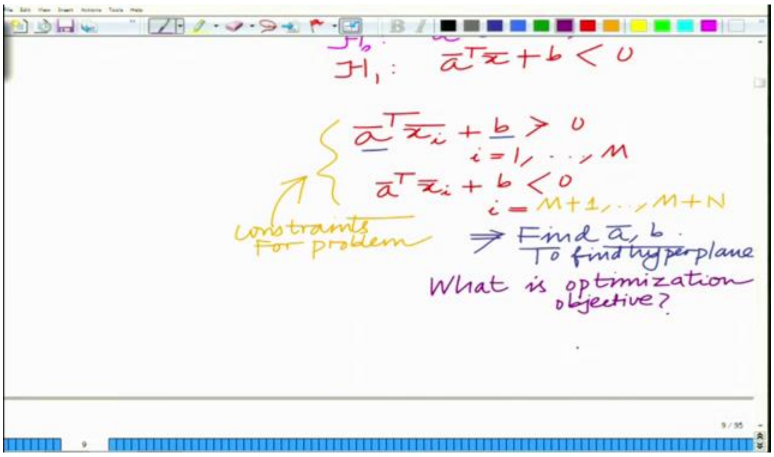
If you have two sets of points corresponding to hypothesis 0 and hypothesis 1 and this is the hyperplane that is separating them and since this is linear you can also call this as a linear classifier.

(Refer Slide Time: 01:53)



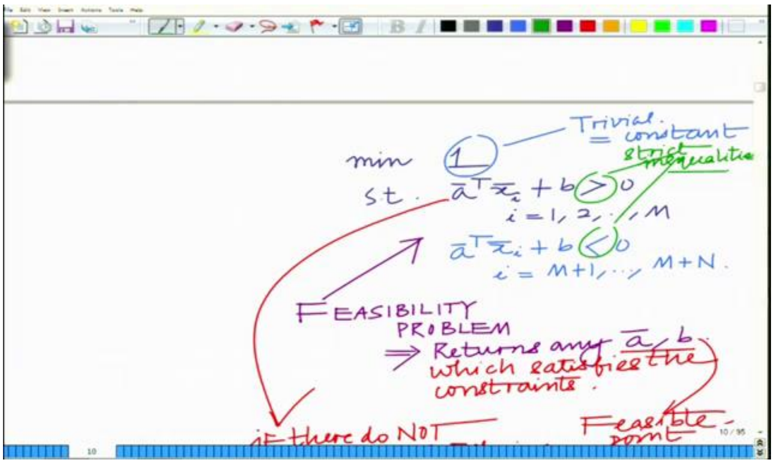
So we have $\begin{matrix} H_o : \bar{a}^T + b > 0 \\ H_1 : \bar{a}^T + b < 0 \end{matrix}$.

(Refer Slide Time: 02:24)



So we have the constraints and we do not have an optimization objective. Now, how do you formulate the optimization problem in this context and you will realize something interesting that given this constraint we do not need an optimization objective. Any $\bar{a}$ and b satisfying this set of constraints is fine which means we can formulate an optimization problem with a trivial optimization objective and that is as follows.

(Refer Slide Time: 05:01)

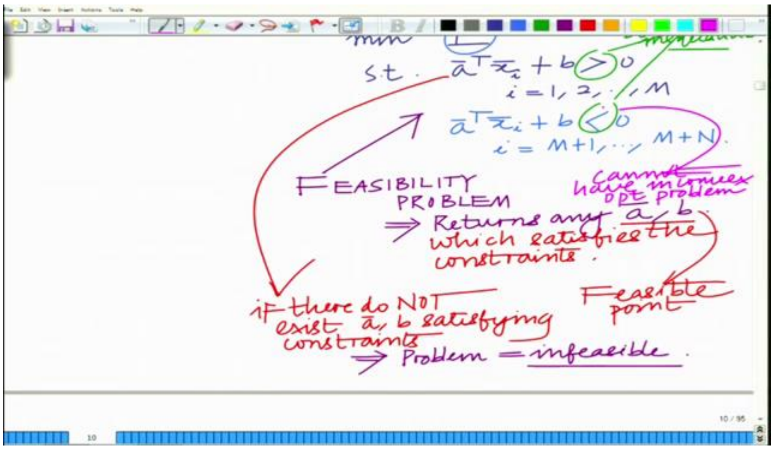


So I can simply set the optimization objective to 1, any constant does not matter so we have

$$\begin{aligned} &\min \quad 1 \\ &s.t \quad \bar{a}^T \bar{x}_i + b > 0 \\ &\qquad i = 1, 2, \ldots, M \\ &\qquad \bar{a}^T \bar{x}_i + b < 0 \\ &\qquad i = M+1, \ldots, M+N \end{aligned}$$

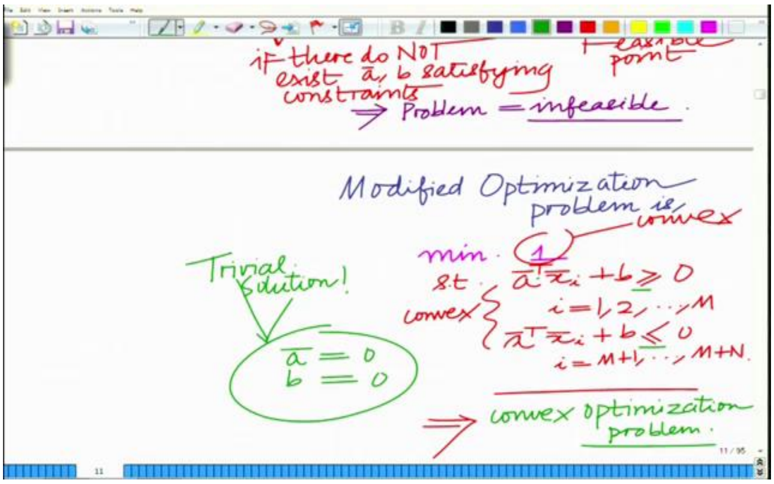
. So this is a trivial optimization objective, the objective is constant which means it cannot be minimized any further. So this will return any feasible point in the sense any $\bar{a}$ and b which satisfy the set of constraints and which are able to separate these two sets of points. This type of optimization problem is termed as the feasibility problem.

(Refer Slide Time: 07:28)



So if there is any point which satisfies the constraint, the problem is feasible otherwise the problem is infeasible. So we are trying to check if the problem is feasible and once you have the $\bar{a}$ and b you can build the classifier and thereby solve this. Now the other problem is that these constraints are strict inequalities. So you cannot have these strict inequalities in a convex optimization problem. So we can modify the optimization problem as follows.

(Refer Slide Time: 09:58)

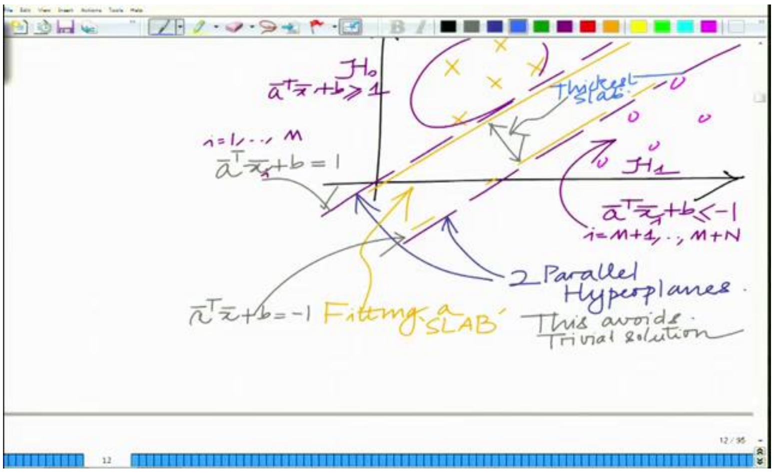


So our modified optimization problem is the following, we have

$$\begin{aligned} &\min \quad 1 \\ &s.t \quad \bar{a}^T \bar{x}_i + b \geq 0 \\ &\qquad i = 1, 2, .., M \\ &\qquad \bar{a}^T \bar{x}_i + b \leq 0 \\ &\qquad i = M+1, .., M+N \end{aligned}.$$

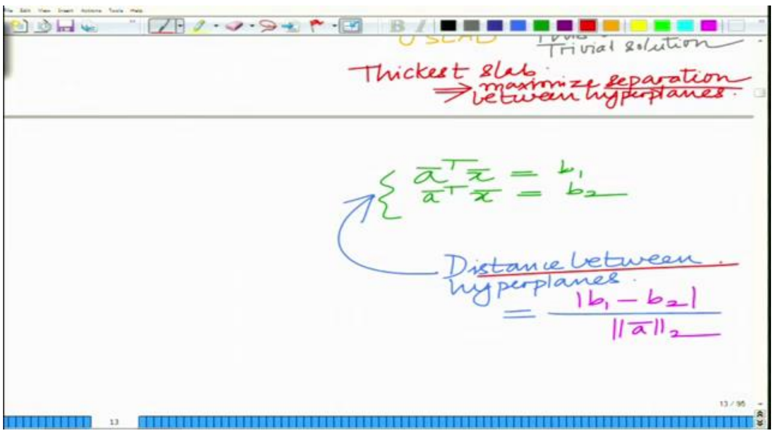
Now you do not have strict inequalities anymore, these are half spaces. So these are affine and therefore, these are convex functions. So this is a convex optimization problem. The problem arises in the fact that it is a feasibility problem, we do not have strict inequalities and if you set $\bar{a} = 0, b = 0$ that trivially satisfies this problem. So now this feasibility problem will always have the trivial solution. So even if the points are not separable it will simply yield $\bar{a} = 0, b = 0$. So you will have to work out another approach which does not yield the trivial solution, but yields actually a hyperplane that separates these two sets of points. And to do that we will now further modify this optimization problem as follows.

(Refer Slide Time: 13:40)



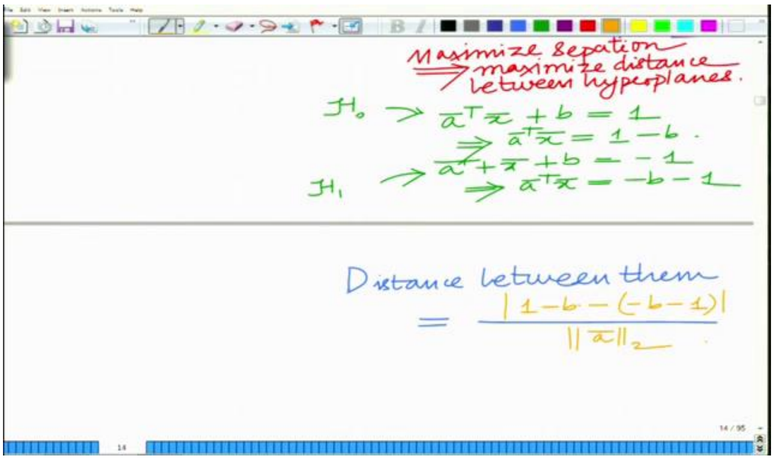
Now, we have these two sets of points, hypothesis H_1 and hypothesis H_0 and we design two hyperplanes and that is the novel solution. So we want to design two parallel hyperplanes. In fact, what you are doing is you are fitting a slab, not just a hyperplane, but this is a continuous slab. So this hyperplane is characterized by $\bar{a}^T\bar{x}+b=1$ and the other one is $\bar{a}^T\bar{x}+b=-1$. All the points in hypothesis H_0 will satisfy $\bar{a}^T\bar{x}+b\geq 1$ and all these points in hypothesis H_1 will satisfy $\bar{a}^T\bar{x}+b\leq -1$. Now, we are fitting a slab between these two sets of points in the training data set. So this avoids the trivial solution. Now 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)



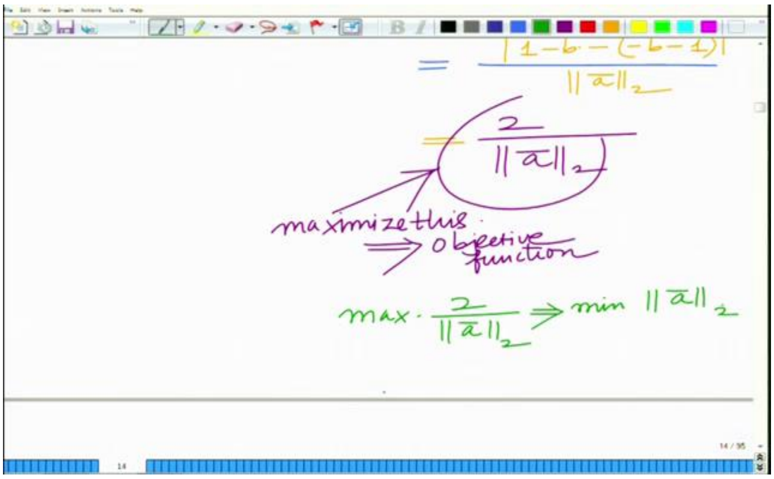
Now, we know that if we have two parallel hyperplanes $\bar{a}^T \bar{x} = b_1$ and $\bar{a}^T \bar{x} = b_2$, the distance between these two hyperplanes is $\frac{|b_1 - b_2|}{\|\bar{a}\|_2}$ and therefore, to maximize the separation implies we have to maximize the distance between the hyperplanes.

(Refer Slide Time: 19:05)

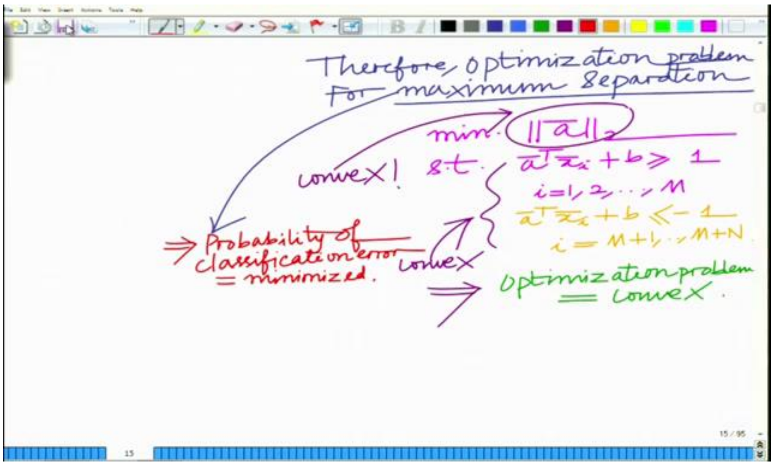


So this is done as shown in these slides.

(Refer Slide Time: 20:38)



(Refer Slide Time: 21:49)

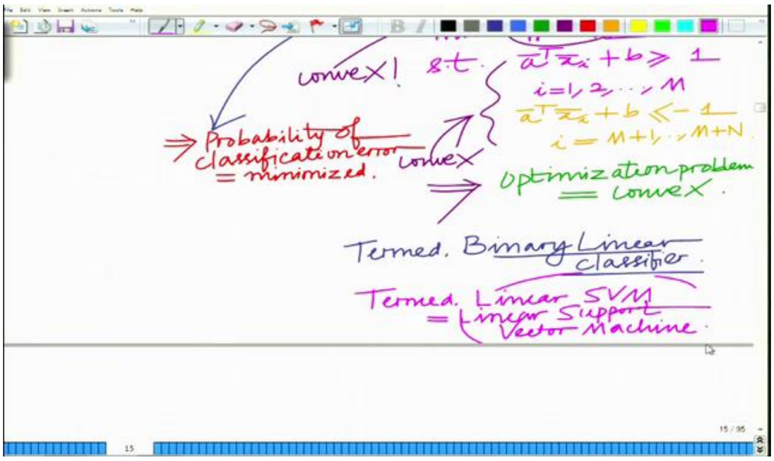


Therefore, the optimization problem for maximum which is a robust separation problem is

$$\begin{aligned} &\min \; \left\|\bar{a}\right\|_2 \\ &s.t \quad \bar{a}^T \bar{x} + b \geq 1 \\ &\qquad i = 1, 2, .., M \\ &\qquad \bar{a}^T \bar{x} + b \leq -1 \\ &\qquad i = M + 1, .., M + N \end{aligned}$$

. And this is a convex optimization problem. And more importantly there is no trivial solution. So we have a convex optimization problem, avoided the trivial solution and we are finding the hyperplanes such that we are fitting the thickest possible slab or you have the set of hyperplanes with the maximum possible separation between them separating these two sets of points. So as the separation becomes smaller and smaller there is a high chance that because of noise you might have points from one set crossing over into another set. So the moment you are maximizing the separation between two hyperplanes the probability of error becomes minimum. So this is the linear classifier and linear classification into two sets is termed binary linear classifier and is also termed as a linear SVM where SVM stands for support vector.

(Refer Slide Time: 25:04)



So this is in fact the cutting edge or one of the efficient mechanisms for linear separation. But in this current form, it is simply a linear SVM which can be employed as a binary linear classifier to classify two sets of points. All such binary classification problems can be handled by the linear SVM. So we will stop here and continue in the subsequent modules. Thank you very much.