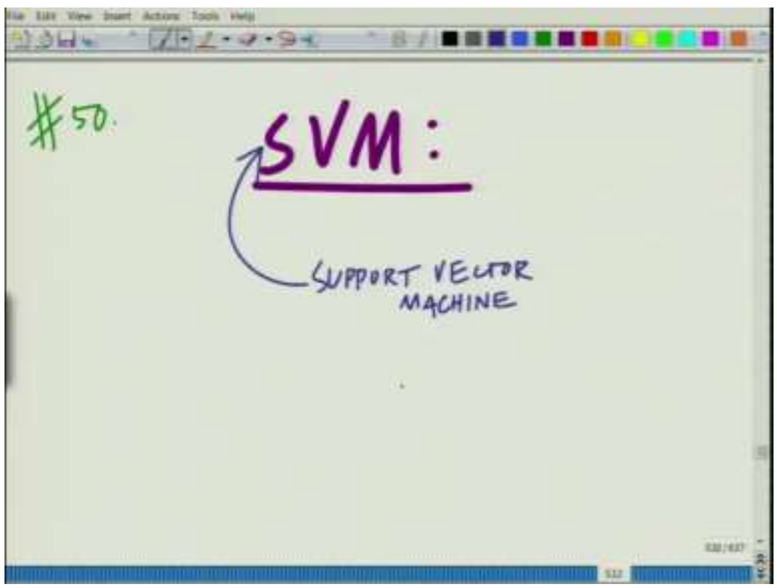
Applied Linear Algebra for Signal Processing, Data Analytics and Machine Learning
Professor Aaditya K Jagannatham
Department of Electrical Engineering
Indian Institute of Technology Kanpur
Lecture 50
Support Vector Machines (SVM): Problem Formulation via maximum hyperplane separation

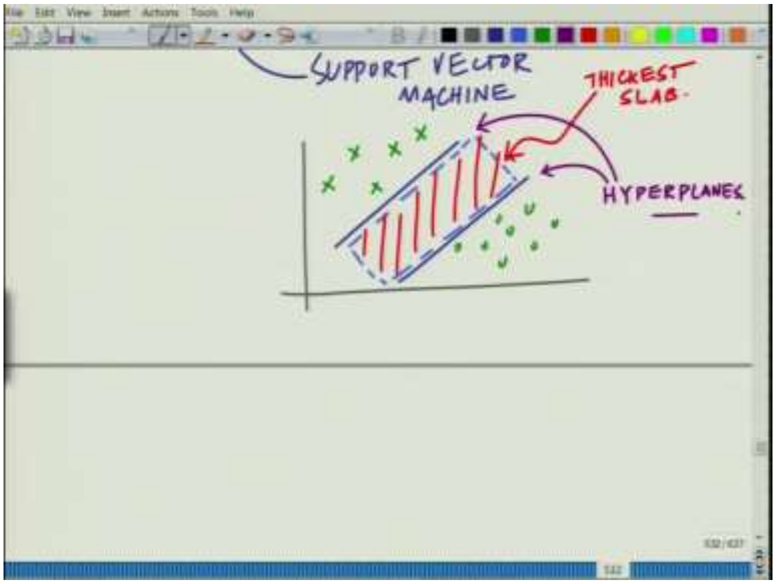
Hello! Welcome to an another module in this massive open online course. So we are discussing about SVMs that is Support Vector Machines and their application in machine learning and essentially how to design these support vector machines to solve a classification problem in the area of machine learning.

(Refer Slide Time: 00:36)



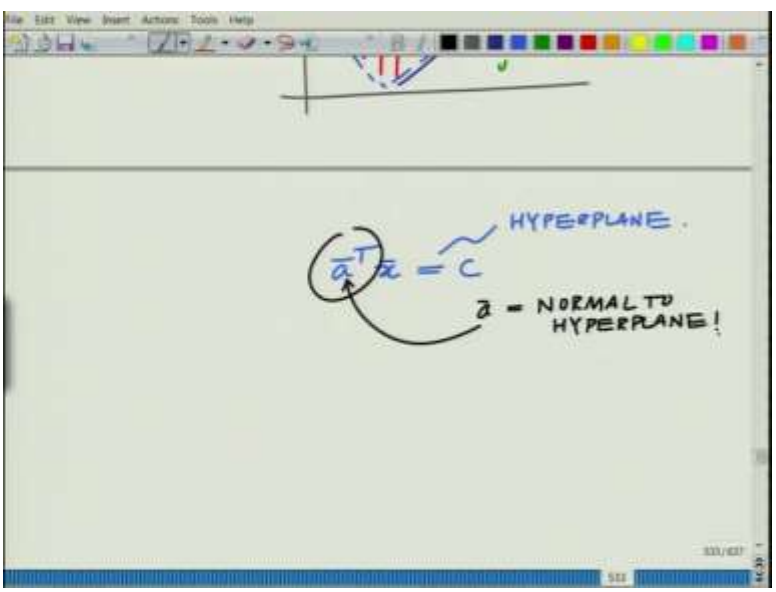
So let us continue our discussion. So we are looking at SVM or what we also call as the support vector machine, which is essentially a machine to classify two sets of points, right? And we said the central philosophy in the support vector machine is essentially now to design two hyper-planes.

(Refer Slide Time: 01:07)



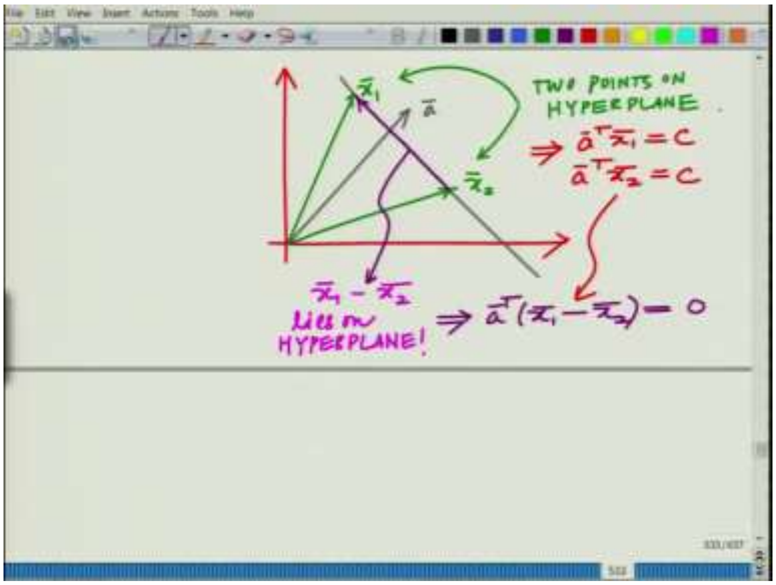
In fact these are two parallel hyper-planes such that, let us say you have one set of points over here, another set of points over here, then you fit the thickest possible slab. What are we trying to do is basically trying to maximize this separation so that you fit the thickest possible slab between these two classes. So that is essentially your separate support vector machine and this is essentially, these are essentially your hyper-planes in n dimensions. Now, how do we start this problem?

(Refer Slide Time: 02:10)



Now, let us go back and take a look at our hyper-plane, remember we said the hyper-plane satisfies the equation a bar transpose x bar equals to c. So we have the hyper-plane a bar transpose x bar equals to c, remember this is our hyper-plane and the point is, now it is not difficult to see that this vector a bar is the normal to the hyper-plane.

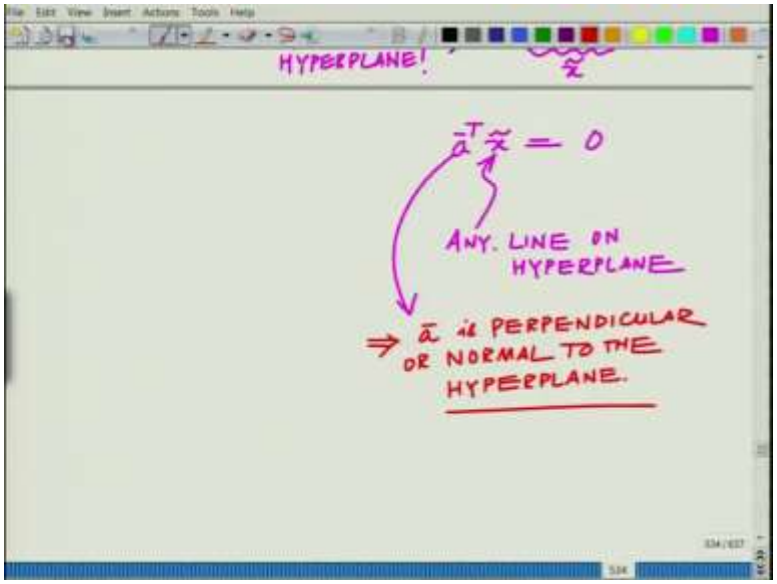
(Refer Slide Time: 02:56)



That is not very difficult to see because, let us again draw a picture over here to illustrate that you have this hyper-plane and you have this vector a bar. I am drawing it as a normal but you can quickly see that essentially we are going to prove these facts. So let us say we have two points x 1 bar on the hyper-plane and we have another point x 2 bar. Therefore these are the two points on the hyper-plane which means we must have, this implies that these both must satisfy the hyper-plane equation. That is a bar transpose.

That is we must have a bar transpose x 1 bar equal to c and a bar transpose x 2 bar equal to c and both of these together, these imply that a bar transpose you subtract 1 minus the other, one from the other, then you have a bar transpose x 1 bar minus x 2 bar equal to 0 and remember x 1 bar minus x 2 bar is the line on the hyper-plane, lies on the hyper-plane. So this is your x 1 bar minus x 2 bar which essentially lies on the hyper-plane. So this implies that any line on the hyper-plane x 1 bar minus x 2 bar that essentially if you treat this as your x tilde.

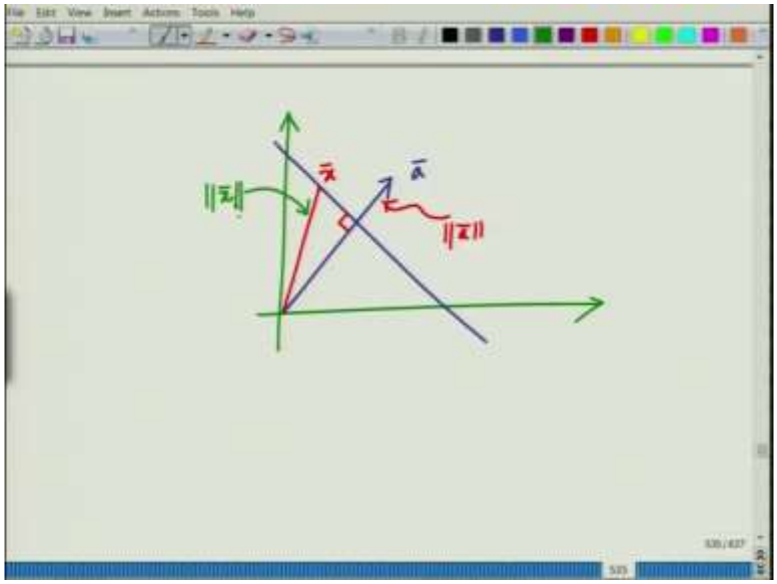
(Refer Slide Time: 05:02)



So a bar transpose x tilde equal to 0 where x tilde lies, and now you can see this is any line or line segment or any line on the hyper-plane. So a bar is perpendicular to every line on the hyper-plane. So a bar essentially this implies that a bar is perpendicular or basically is normal, normal again means the same thing, normal to the hyper-plane.

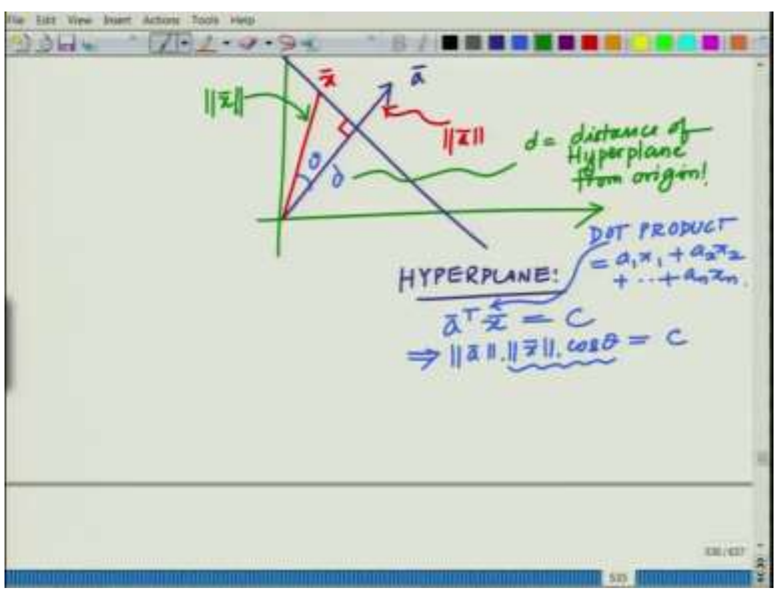
So a bar is perpendicular to the hyper-plane. So we have established that a bar times x tilde equals to 0 where x tilde is any line on the hyper-plane that essentially implies that a bar is the normal to the hyper-plane. Now, look at this interesting thing.

(Refer Slide Time: 06:22)



Now let us go back, take a look once again at the diagram. What that tells us is the following, now let us look at our hyper-plane. There is an abstraction, the hyper-plane will be in n dimensions. I am just showing the representation in two dimensions. So this is your a bar which is the normal and let us say you have any other points which is your x bar. This is basically 90 degrees and now you have the vector a bar and this length of this is norm a bar. Length of this x bar is obviously norm x bar.

(Refer Slide Time: 07:14)

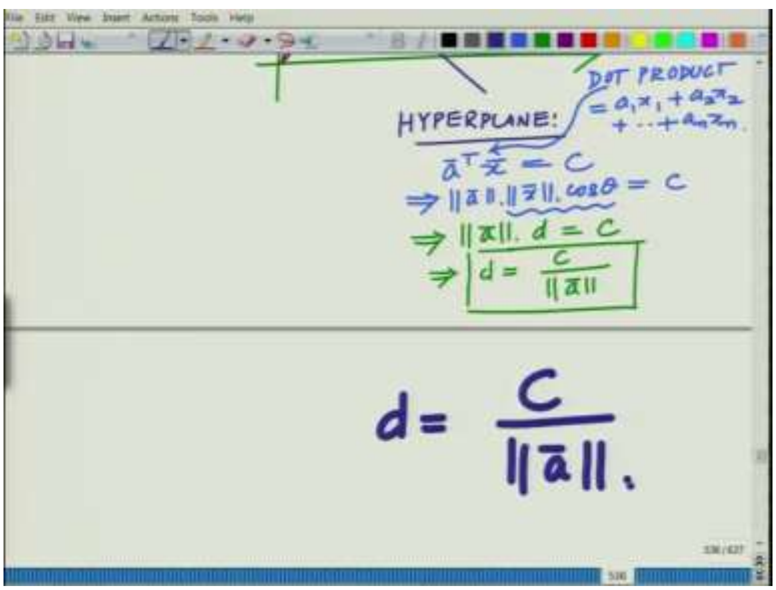


Now, look at this we have the equation of hyper-plane a bar, the hyper-plane satisfies, remember the fundamental equation of the hyper-plane is a bar transpose x bar equal to c. Let us say this is the equation of the hyper-plane a bar transpose x bar equal to c. Now what this says is that basically this means, now look at this we know what is the dot product a bar transpose x bar.

This is essentially norm of the vector a times norm of the vector x times cosine theta where theta is the angle between the vectors a bar and the x bar. So if you call this angle as theta, so this is essentially this implies that a bar transpose x bar which is essentially nothing but the dot product. This is the dot product.

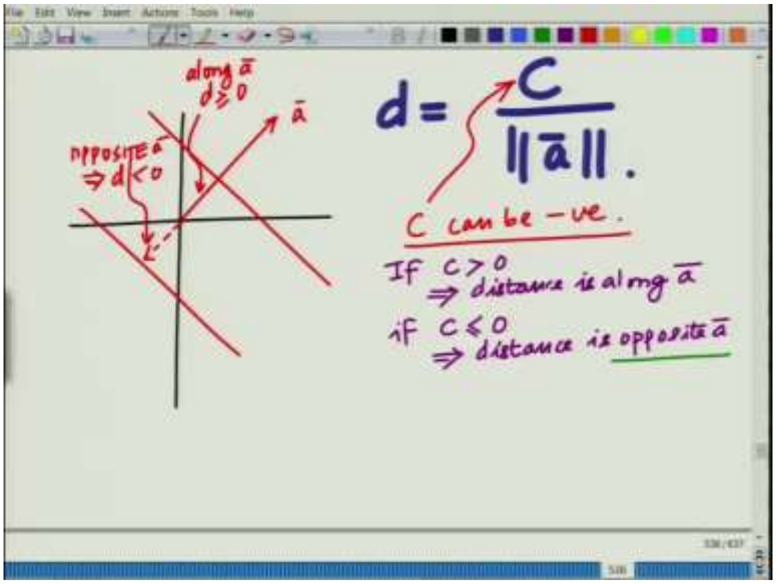
That is basically this is a 1 times x 1 plus a 2 times x 2 plus so on, a n times x n. So this is essentially norm of a bar times norm of x bar times cosine theta is equal to c, but if you look at this quantity norm of x bar times cosine theta that is nothing but d, the distance of the hyper-plane from the origin. So if you look at this d, d equals distance of hyper-plane from the origin.

(Refer Slide Time: 09:28)



So essentially we have, this implies that norm of a bar times d equal to c which essentially implies that d equal to c divided by norm of a bar. That is essentially the interesting relation. So d is the essentially distance of the hyper-plane from the origin and we have this interesting relation that is if you look at this, what we have just derived is essentially that d equals c divided by norm a bar.

(Refer Slide Time: 10:14)



And if you think about this, this quantity c, this c can be negative. Here, this is, although it is a distance it can be negative. The reason being if c is greater than 0 implies distance is along the vector, the normal vector a bar. On the other hand if c less than equal to 0 this implies the distance is opposite that of a bar.

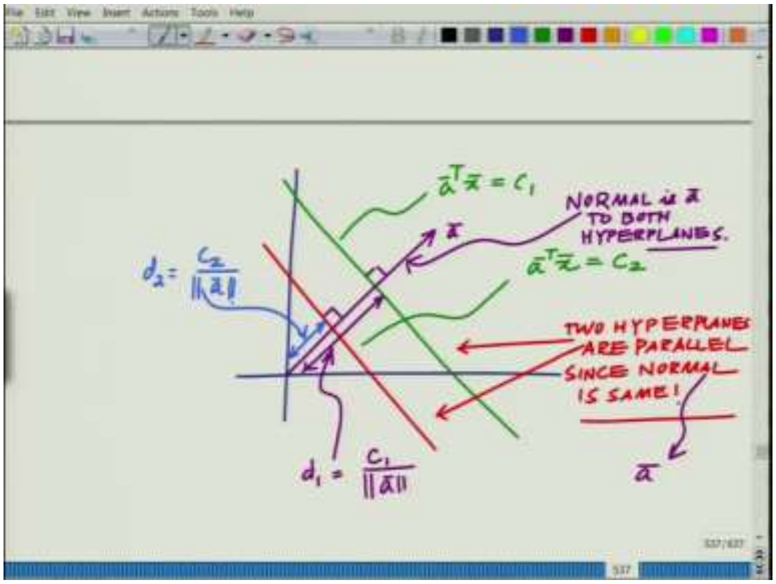
That is the distance is along the direction that is opposite a bar, that is you have two situations here. So essentially if you look at this, what you will have is that, you will have this hyper-plane, this hyper-plane both are at the same distance.

This is the normal let us say a bar. So this is essentially where your d will be greater than or equal to 0 and this, so this is basically along a bar and this is basically what is this? This is opposite. This implies that d is less than 0 or you can say less than or equal to 0. It will be 0 if that hyper-plane passes through the origin that is c equal to 0.

So that also basically checks the formula. If you have hyper-plane such that a bar transpose x bar equal to 0 that is your constant c 0 then essentially the hyper-plane is passing through 0. So the distance from the origin is 0.

Further, if you look at two particular hyper-planes which only differ in the constant. That is you have a bar transpose x bar equals to c 1, a bar transpose x bar equals to c 2. You can clearly see that this two hyper-planes are parallel because the normal is the same.

(Refer Slide Time: 12:46)



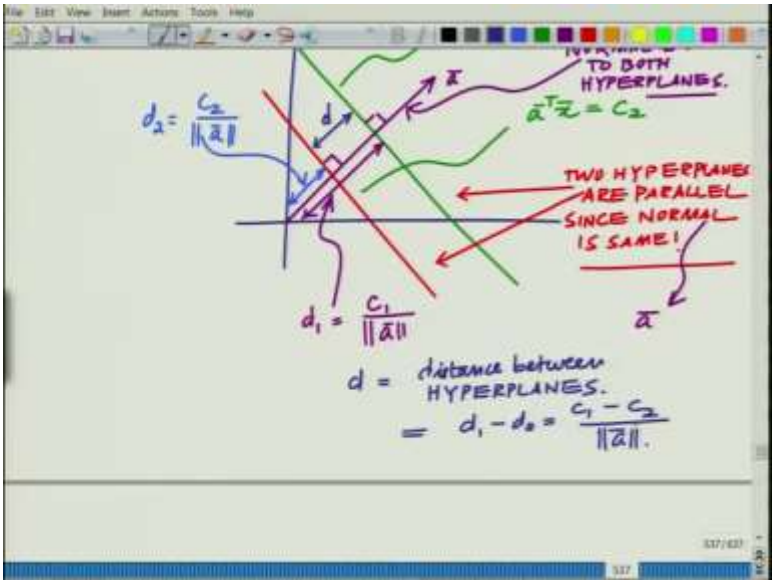
So if you have two hyper-planes, now the point here is, again all these are simple principles that you might have already learnt in your high school. That is if you have these two hyper-planes, let us draw these two hyper-planes. So you have these two hyper-planes.

This is a bar transpose x bar equal to c 1, this is a bar transpose x bar equal to c 2. These two hyper-planes will be parallel, so these two hyper-planes are parallel since the normal is the same, what is the normal? Normal is basically nothing but the normal vector, is basically nothing but a bar, that is if you look at this, this is your a bar and a bar is perpendicular to both.

So normal vector to both, the normal vector is a bar to both the hyper-planes, normal vector to both these hyper-planes is essentially your vector a bar. Now therefore what, you can ask what is the distance between these two hyper-planes? That brings us to the distance between these two hyper-planes and now you can naturally see the distance to between these two hyper-planes is basically d 1 minus d 2.

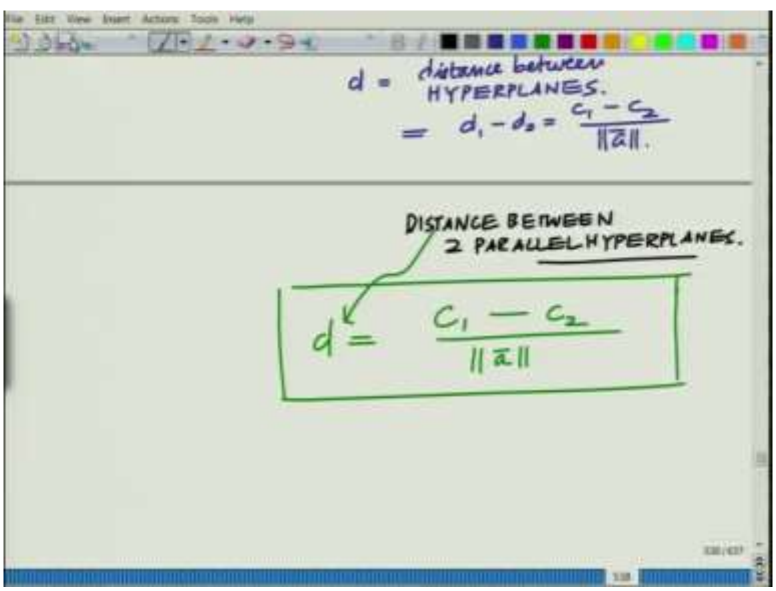
Where d 1 is the distance of the first one from the origin, d 2 is the distance of the second from one from the origin. So if you look at d 1 we already know d 1 equal to c 1 divided by norm a bar and if you look at the distance to the second hyper-plane, that is essentially your d 2 which is equal to c 2 divided by norm a bar.

(Refer Slide Time: 15:37)



And therefore the distance between the two hyper-planes naturally, if you look at this if you call it as d, d equal to now distance between the hyper-planes. This is equal to d 1 minus d 2 equals basically c 1 minus c 2 divided by norm of a bar.

(Refer Slide Time: 16:12)

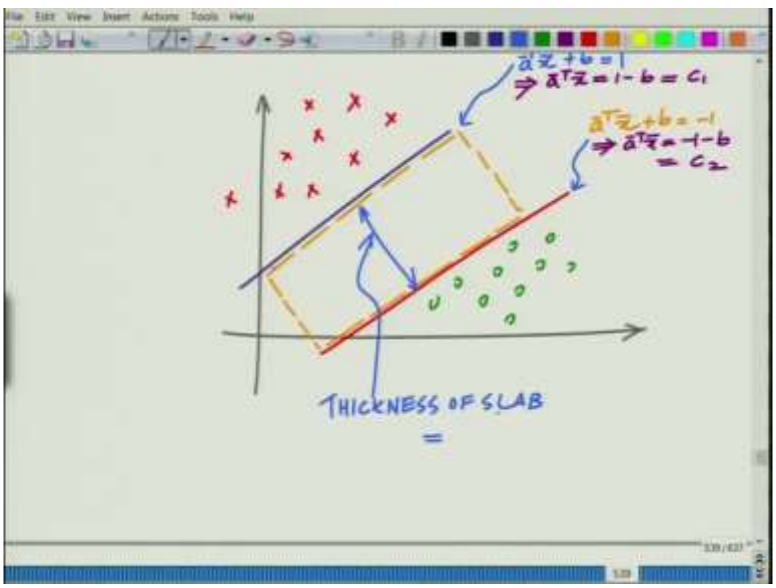


So that is the interesting, that d, distance between the hyper-planes d equals c 1 minus c 2 by norm a bar and remember this is the distance between the two parallel hyper-planes. Otherwise they are going to intersect and in which case the distance is, of course it is not constant and the minimum distance will be 0.

So this is the distance between two parallel hyper-planes, so this is the distance between the two parallel hyper-planes. Therefore we have essentially, now what we have done is we have found distance between the two parallel hyper-planes. Now let us go back to our support vector machine problem and try to see what is the problem over there.

Now if you go back to the support vector machine problem you will realize that we have exactly the same problem, we have these two hyper-planes which are trying to insert the slab. Remember the thickest possible slab between these two sets of points that have to be classified, so essentially where we stand now is that you have these two sets of hyper-planes.

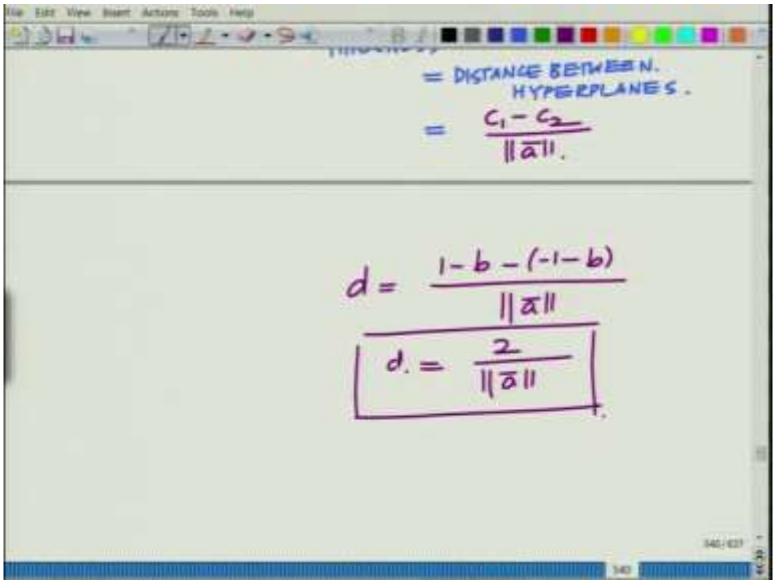
(Refer Slide Time: 17:40)



So essentially where we stand now is you have these two hyper-planes, one of these is basically the hyper-plane. Remember a bar transpose x bar so you have these two points, two sets of points and you have, so this is your a bar transpose x bar plus b equal to 1 and this is your a bar transpose x bar plus b equal to minus 1 and if you look at the distance between these two.

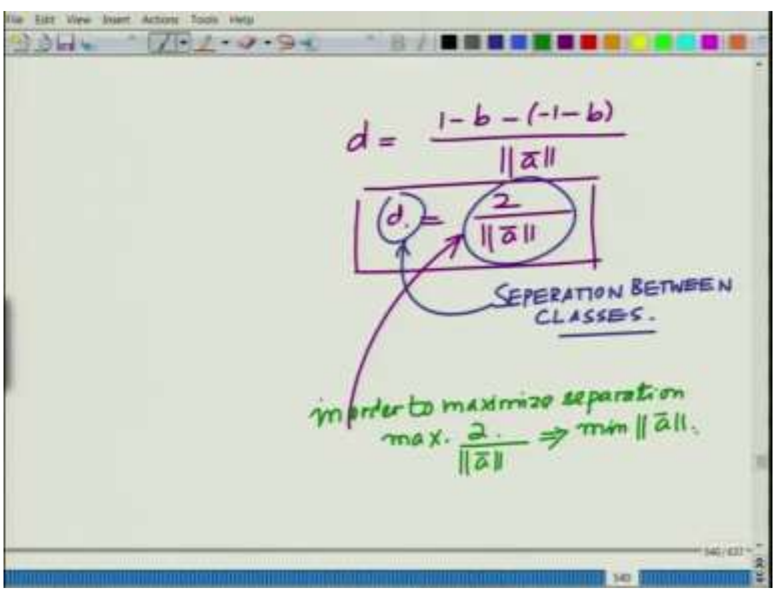
Now look at this implies that a bar transpose x bar equals 1 minus b that is you can call that as c 1 and this implies a bar transpose x bar equals to minus 1 minus b, which you can call as c 2 and therefore. If you look at the distance that is the thickness of this slab, now once again we come to the problem our problem of the thickest slab. The thickness of this slab, if you think about that is basically going to be equal to.

(Refer Slide Time: 19:43)



So the thickness of this slab equal to distance between hyper-planes equals c 1 minus c 2 divided by norm a bar which in this equal to d equal to c 1 that is basically you have 1 minus b minus c 2 minus 1 minus b divided by norm of a bar. Which is essentially if you look at this is equal to essentially 2 by norm of a bar, so this is essentially what we are calling as the separation between the points, separation between the classes.

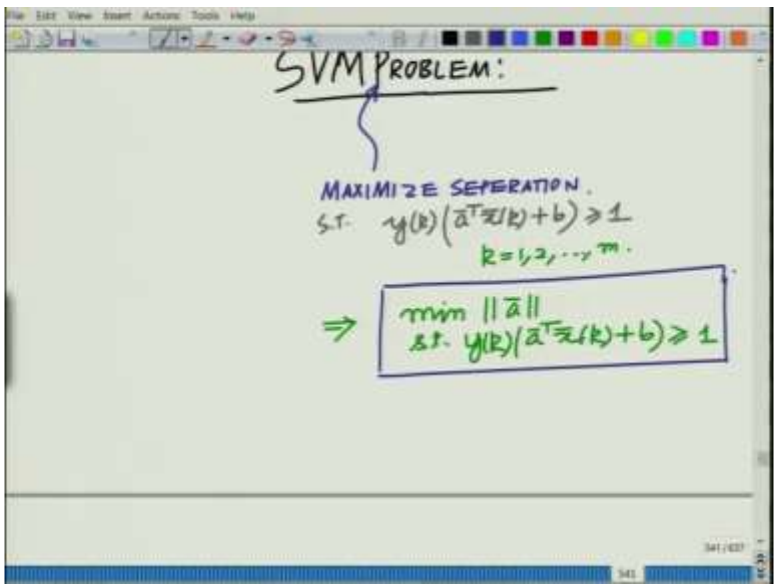
(Refer Slide Time: 20:42)

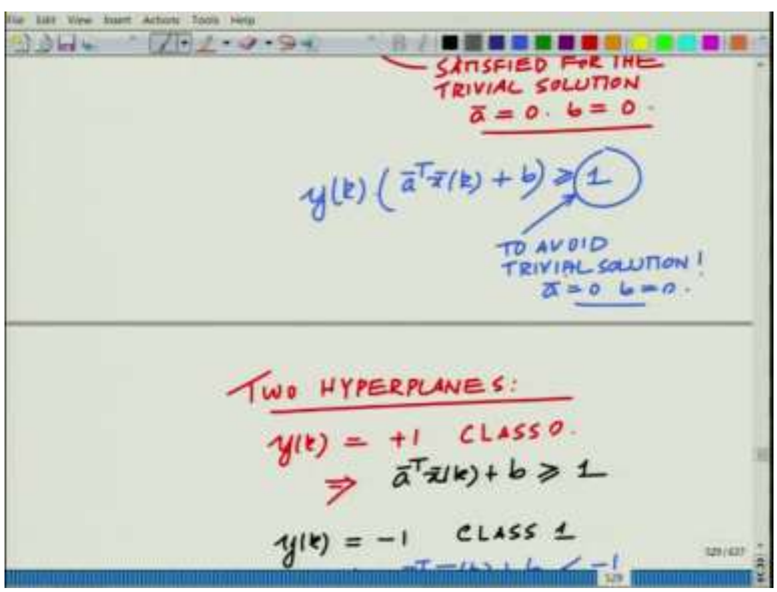


This is the thickness of the slab. Which is essentially the separation between the classes, and therefore, now we have to find the hyper-planes or the slab which maximizes the separation between the classes.

In other words we have to maximize the distance between the hyper-planes or maximize the quantity 2 divided by norm of a. So to maximize the separation therefore it naturally follows, in order to maximize separation we have to maximize 2 divided by norm a bar that means minimize, take the reciprocal that is minimize norm a bar.

(Refer Slide Time: 21:57)

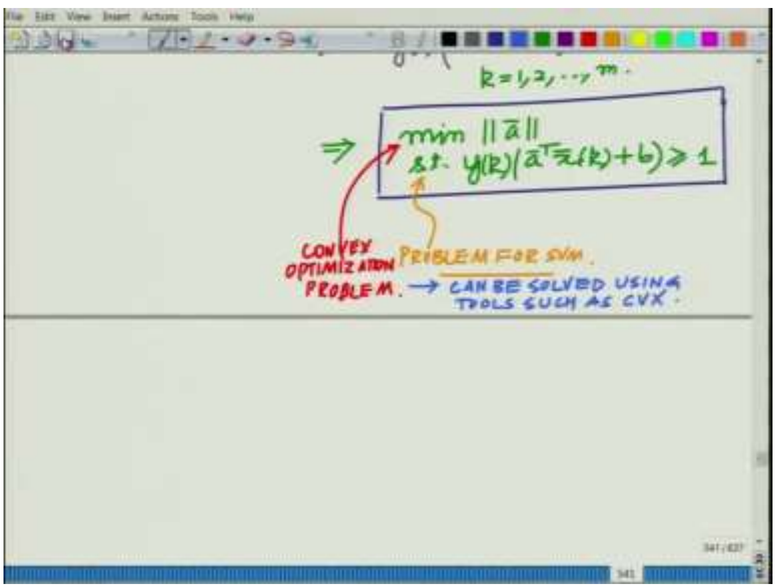




And therefore finally our SVM problem can be formulated as follows. Maximize the separation this is essentially, remember we have the objective, which is to essentially maximize the separation. Subject to the constraint remember y k times, you have a bar transpose x bar, that is if you go back and take a look at this, what we have over here that is the constraint is that a bar transpose x bar k plus b greater than equal to 1.

So that is, essentially what that means is, we have the constraints a bar transpose x bar k plus b greater than or equal to 1 for k equal to the m points 1, 2 up to m. This implies the net problem will be minimized; this is what we have shown, to maximize the separation we have to minimize norm a bar, subject to the constraint that y of k times a bar transpose x bar of k plus b is greater than or equal to 1, this is the problem for the support vector machine.

(Refer Slide Time: 24:06)



So this is essentially the problem for the support vector machine, and this is what is known as a convex optimization problem, this is known as you have the objective function, you have the constraints, this is what is known as a convex optimization problem and this can be solved efficiently using several software.

So this is a convex optimization problem, so this can be solved efficiently using tools or rather computational tools, such as for instance CVX, one can readily solve this using several computational tools software, such as CVX and that essentially shows how this principles of linear algebra, and in fact we have used a lot of these principles of linear algebra in geometry.

That is hyper-planes, equation of a hyper-plane, the inner product, distance of the hyper-plane from the origin, parallel hyper-planes, distance between these hyper-planes and then eventually designing the classifier or the support vector machine which essentially maximizes the separation between these two classes of points that is essentially your support vector machine.

In fact this is one of the most important, one of the prominent tools that has been used as, in fact also currently being used and one of the most attractive features about the support vector machines. As you can see is the simplicity of the analysis because it is linear in nature, it simply builds based on hyper-planes, designing the hyper-planes, such that you choose the set of hyper-planes, parallel hyper-planes with the maximum distance between them and that can be posed as a convex optimization problem which can be solved rather efficiently.

So that is an application, interesting application of the principles of linear algebra in the context of machine learning. So we will conclude this discussion here and continue with other aspects in the subsequent modules. Thank you very much.