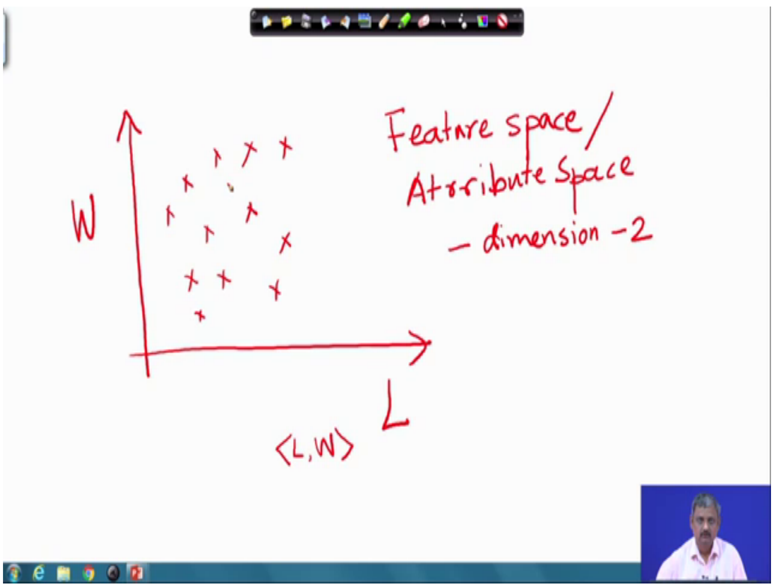
Data Mining
Prof. Pabitra Mitra
Department of Computer Science & Engineering
Indian Institute of Technology, Kharagpur

Lecture - 22
Support Vector Machine - I

We continue our discussion on the classification algorithms. So, so far we have studied 3 classification algorithms, the decision tree which splits up the attribute space into rectangles, boxes. Then we studied the Bayesian classifier where what we do is that, we try to figure out the probability of a point of an instance belonging to a class and whichever class has the highest probability we put it into that. And then we also saw the nearest neighbor classifier where, which is an approximation of the Bayes classifier where we look at the neighbors and then we assign to the majority class among it is neighbors.

So, now we extend our methods to one more algorithm for classification. The idea is like this, I will explain you with an figure is that, what I want to do is to find off show you a yes show you a representation like this. So, what do you want to do?

(Refer Slide Time: 01:43)

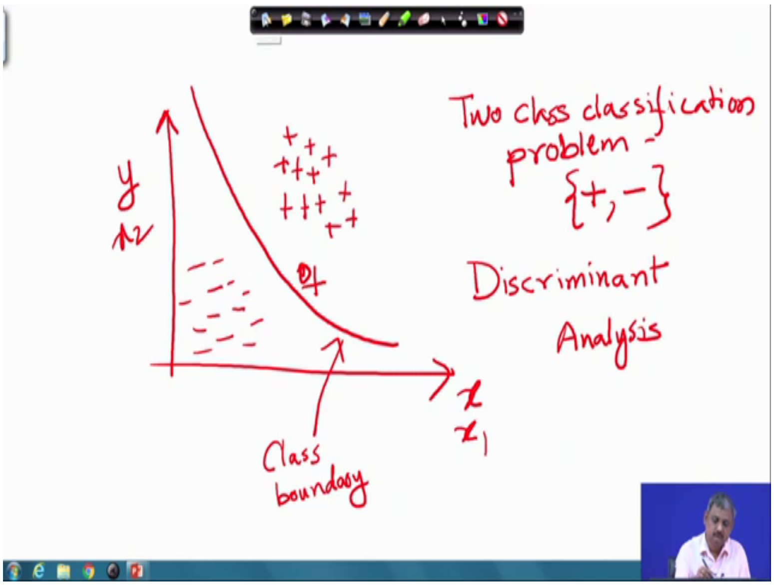


Suppose, I have a Feature space say length of some object and say the weight of some object.

And one thing you must have realized by now is that any particular object which is a pair of say in this case, length and weight values can be represented as a point in this Feature space. So, this I call as a yes or Attribute space. In this case, the dimension is 2, 2 features are there. In general it can be higher dimension. And so now, we have points that any representation as points in such spaces ok.

So now, we can look at things like this.

(Refer Slide Time: 03:06)



What we can look at is that, let me look at a feature space, say instead of l w, let me call it as x and y now or maybe x 1 and x 2. And now suppose we are considering a 2 class problem and let me problem, I can name these 2 classes are as anything, but let me for, so that I can pictorially explain you properly. Let me call one of them as a positive class and other as a negative class ok.

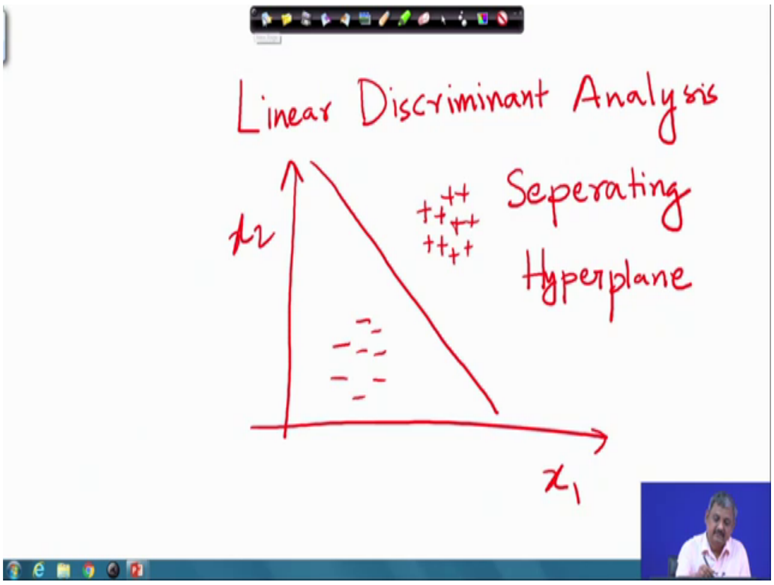
And now, I draw all my positive class points, positive class points in the feature space and also I draw all my negative class points. And what I really want to do is to find a rule which will sort of, if I get a new point it will put either in the positive class or in the negative class; it will classify the new point.

So, one way of doing this is called a dis. So, we have seen other methods previously, is called a Discriminant Analysis. So, what that Discriminant Analysis does is to draw a some boundary which. So, basically draw what is called a Class boundary, Class boundary. So, how does it help me get a rule? What I will do is that, when I get a new point, I will just check which side of the discriminant it is. For a specifically for this 2 class problem, I will check whether the new point is in positive side or it is in the negative side and based on that finding out which side it is I will assign it the class accordingly. So, in this case it is the positive class ok.

So that is clear. So what I do basically, 2 things you have to understand. One is I am representing points objects as points in some 3 dimensional space. Second is I am drawing some kind of a boundary to separate the classes, points from each class.

In general, it could have been a multi class problem also.

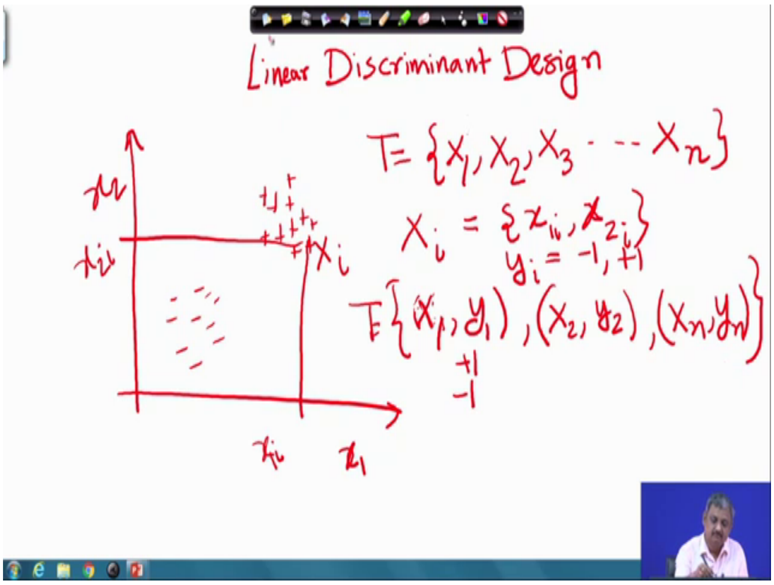
(Refer Slide Time: 06:55)



Let me consider a particular case of this problem, which I will discuss in this lecture, is the case of Linear Discriminants; that means, if I have 2 features and 2 classes as well, I want to draw a line, because it is linear discriminant I want to draw a line which will separate this ok.

So it is a, in 2 dimension it will be a line, in higher dimension it 3 dimensional it will be a plane, in higher dimensions it will be so called hyperplane rather I would call it a Separating hyperplane. So, what I want to do today is to find a method for designing such a hyperplane, separating hyperplane. Let me formally write down the, formally write down some mathematical notation that we will need.

(Refer Slide Time: 08:42)



So, my job is; for explaining I am considering 2 attributes in general there can be more attributes.

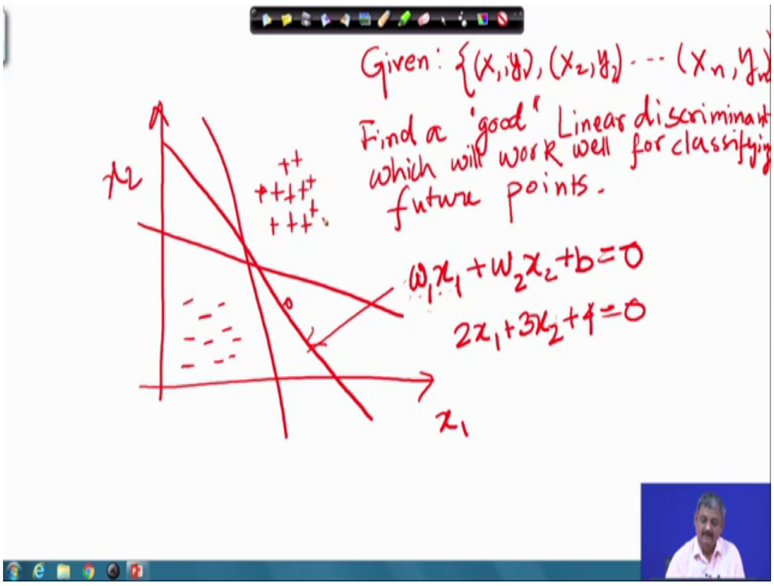
So, what I have is a set of points from 2 classes. This set I will call as a training set t, for which I know the class levels. So, what will training set consists of? It will consist of certain number of vectors; I write them as x 1 or rather let me write it as a subscript itself. I am writing in this in capital letters, where each of these vectors X i has 2 components; small x 1 i and small x 2 y which are the 2 feature values. So, suppose this is my x I; this is small x 1 i and this is small x 2 i ok.

So, I have some points given like. I am just writing down the notation. And note on, my training set is not yet complete because I know the levels, I know the class values for each of these endpoints. So, what I have along actually my training set consists of not just x 1, x 2 values, but a sorry capital letters let me write it properly let me write it proper notation. So, I have capital x 1 and the y 1 value, y 1 here I denote as the class level. So, y 1 can take a value positive or negative class, it can take a value plus 1 or minus 1; it can take on value either plus 1 and minus. So, this is one data point ok.

So, similarly I have n number of data points. So, x 1 is like this, y i is either minus 1 or plus 1. So, I have n number of such training point that is given to me. I know before end, a set of endpoints and their class levels. So, what do I do with these end points? This has become little cluttered; let me draw it again.

So, what I do with these end points is to use them to find out a good linear discriminant.

(Refer Slide Time: 12:18)



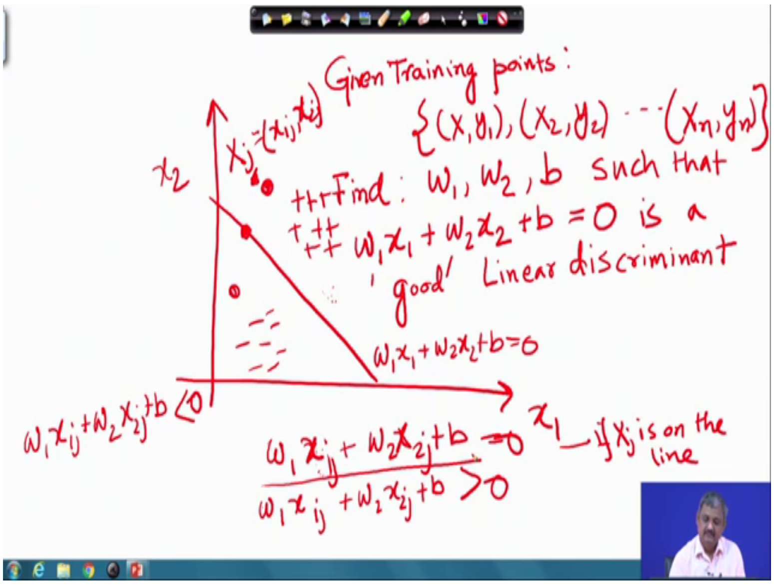
So, given discriminant which will work well classifying future points? So, I get a new point it should correctly classify ok.

So let me see, what is the equation of this linear discriminant we are talking about, equation of the linear discriminant; we are talking about. So, you think a little bit and you will appreciate that I can write down a linear discriminant as a equation of a straight line sorry mistake sorry, w 1 x 1 plus w 2 x 2 plus some constant b equals 0. So, x 1 is the x coordinate, x 2 is the y coordinate, w 1 w 2 are the slopes. So, I can use w 1, w 2 to find the slopes, plus b.

So, if you note this is nothing but the equation of a straight line. So, if w 1 is 2, w 2 is 3, w 2 is 3, then I can and b is 4 this is nothing but the equation straight line in 2 dimensions. So, I can give values of w 1, w 2, w b and get different-different straight lines. So, b sort of determines the offset. If b is 0, the straight line will pass through the origin.

So, what is my job now? My job now is to, given these training points, given this training let me draw again.

(Refer Slide Time: 16:08)



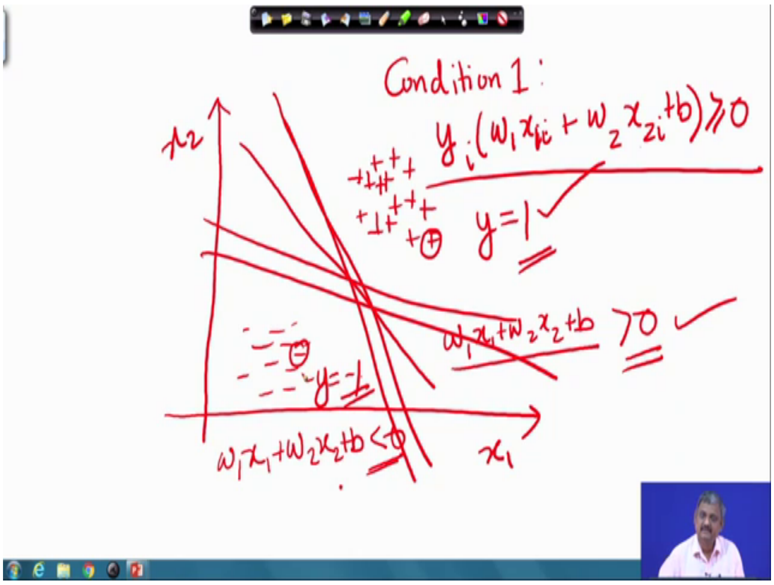
Find, such that, linear discriminant. So in other words, design a linear discriminant. So, given points form by that class, find w 1 x 1 plus 0. So, that is my job, that is what is a data miner my job is, finding this classification boundary ok.

So, now notice certain point, certain thing. Suppose I take a point, say I call it x j not here, maybe I take a point here; suppose I take a point x j which has 2 coordinates x 1 j, x 2 j, small x 2 j. So, what is the value of this quantity, x sorry x 1 j, given some w, given some line; horrible handwriting let me correct it x 1 j.

What is the, what can I say about this quantity for different points x j? If the point is on this line, if the point falls exactly on this line, I can, you can actually figure out that this quantity if I just plug in x I j, x 2 j, x 1 j, x 2 j will be equal to 0. If x j is on the line right, what happens if x j is here then; that means, if x j is in the non origin side, non origin side which I call as the positive side then this quantity will be greater than 0. So, all points on this side plus side, non origin side will have value of this quantity greater than 0.

Similarly, all points on this side, we will have a value less than 0; we will have a value less than 0. So, one thing you notice is that, what I am actually doing is that I am kind of, by looking at sign of this quantity, I am looking at sign of this quantity that I am examining. So, how does that help me; it helps in the following way.

(Refer Slide Time: 21:09)



So, suppose this is my b, if I pick up a point whose y value is plus 1, the sign of this quantity should be greater than 0. So, if it is plus 1, this will be greater than 0. Similarly I if I pick up a 1, if the sign of y is minus 1, then this quantity, this point should lie on the origin side. So, basically what I am saying all positive points, positively labeled points should lie on non origin side and all negatively level points should lie on the origin side.

So, here 0; so, if I am trying to draw, my job we remember is to draw a good hyperplane. One property this good hyperplane, should satisfy is that it should put all plus points on the origin side, non origin side and all minus points in the origin side, no no sorry origin side.

Another way of telling is that, the sign of this quantity and this quantity should be same and the sign of this quantity and this quantity should be same. If y is negative, this value should be negative for a point. If y is positive, this value should be positive all right. So, that is my first condition x 1 i 2 x 2, this means if both of them same sign their product will always be positive. So, that is the first condition I need from w 1, w 2, w b, you understand this. So, in the next exposition what I will explain is that, we need one more condition, condition 2, just this condition of separating 2 classes and 2 side is not enough. We need one more condition. So, I will stop in this lecture here. In the next class I will start with explaining that condition.

Thank you.