Deep Learning

Prof. Prabir Kumar Biswas
Department of Electronics and Electrical Communication Engineering
Indian institute of Technology, Kharagpur

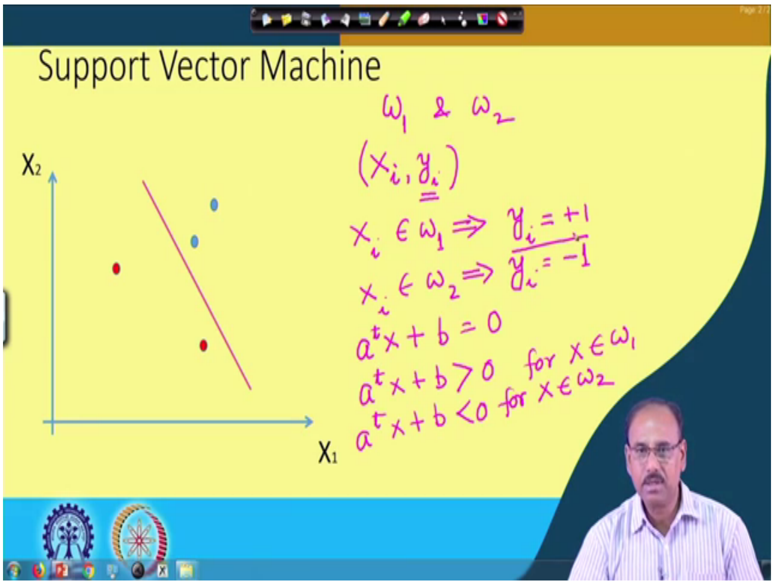
Lecture - 12
Support Vector Machine - II

Hello, welcome to the NPTEL Online Certification course on Deep Learning. You remember in the previous class we started our discussion on the Support Vector Machine. So, in today's lecture, we will continue with the same discussion.

(Refer Slide Time: 00:42)

So, in the previous class we have just introduced or gave a brief introduction of what the Support Vector Machine is and today we are going to talk about what should be the design approach of a Support Vector Machine.

(Refer Slide Time: 00:59).

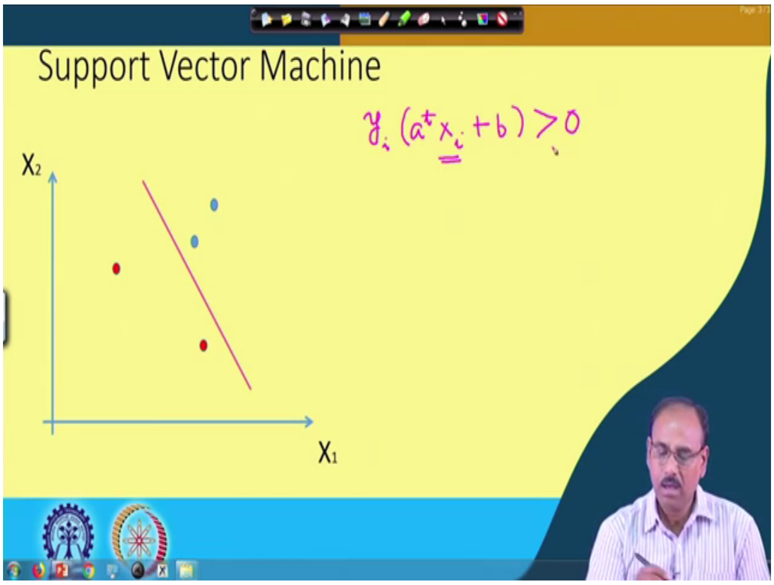


So, we have seen that in our case we will assume again a two class problem.

So, we have the feature vectors given from two classes; omega 1 and omega 2 and all the training vectors we assume that are given as leveled pair in the sense that attaining vector X i which is the i th training vector will be given as a pair X i y i where this y i indicates the level. So, if the training vector y i is taken from class omega 1 that is if y i belongs to class omega 1, then we will set sorry X y belongs to class omega 1, then we will set the level y i to be plus 1 and if X i the training vector X i is taken from class omega 2, then we will set y i the level to be equal to minus 1.

So, that indicates that given a separating plane with an equation a transpose X plus b is equal to 0 if this is the separating plane between the feature vectors belonging to class omega 1 and class omega 2, then our classification rule was a transpose X plus b greater than 0 for X taken from class omega 1 or if X belongs to class omega 1 and a transpose X plus b will be less than 0 for X taken from class omega 2. Now by introduction of these levels that is y i is equal to plus 1 for if X i belongs to class omega 1 and y i equal to minus 1 if X i is taken from class omega 2 then I have a uniform classification rule.

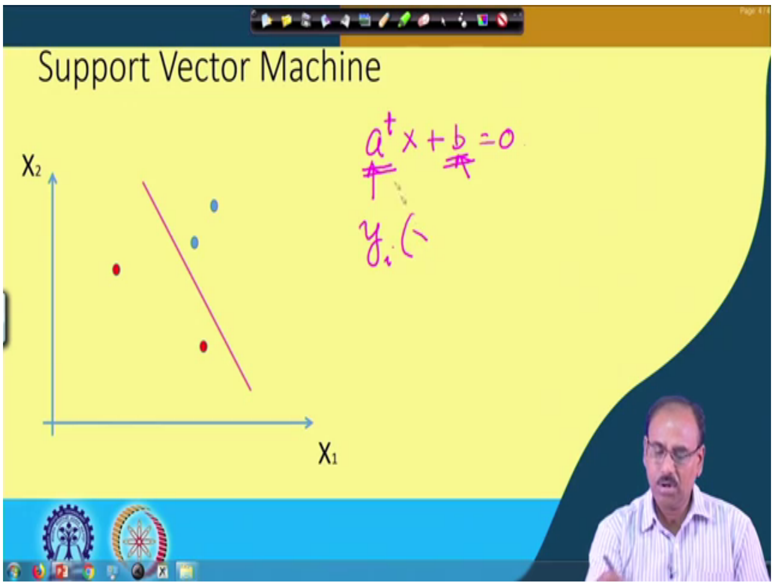
(Refer Slide Time: 03:19)



That is I can write y i a transpose X i plus b will be greater than 0.

If X i is correctly classified by the separating plane a transpose X plus b equal to 0 and this will be less than 0 if X i is misclassified by the separating plane a transpose X plus b equal to 0.

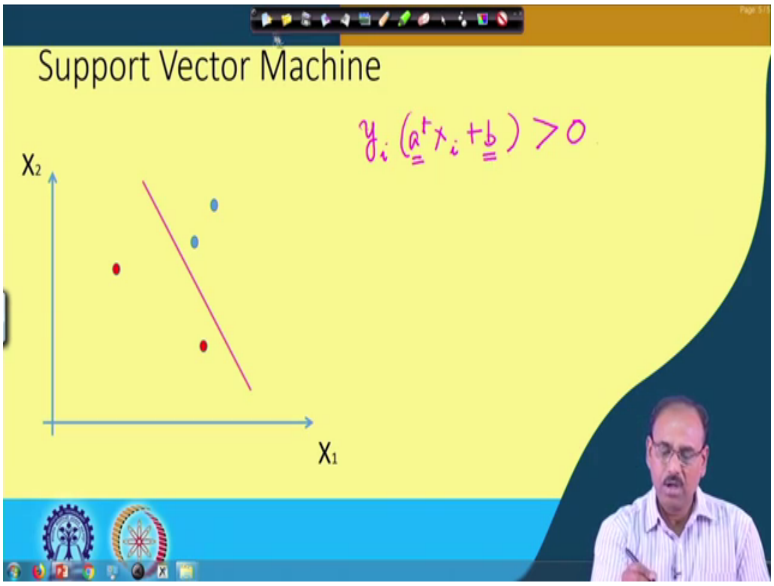
(Refer Slide Time: 03:57)



And we have also seen as in this equation a transpose X plus b equal to 0, a is a vector which is orthogonal to the separating plane and b is a bias which indicates what is the position or location of the separating plane. So, as a is orthogonal to X if I modify a, that means the orientation of the separating plane will be different whereas, if I modify b, then the position or location of the separating plane will be different in a feature space.

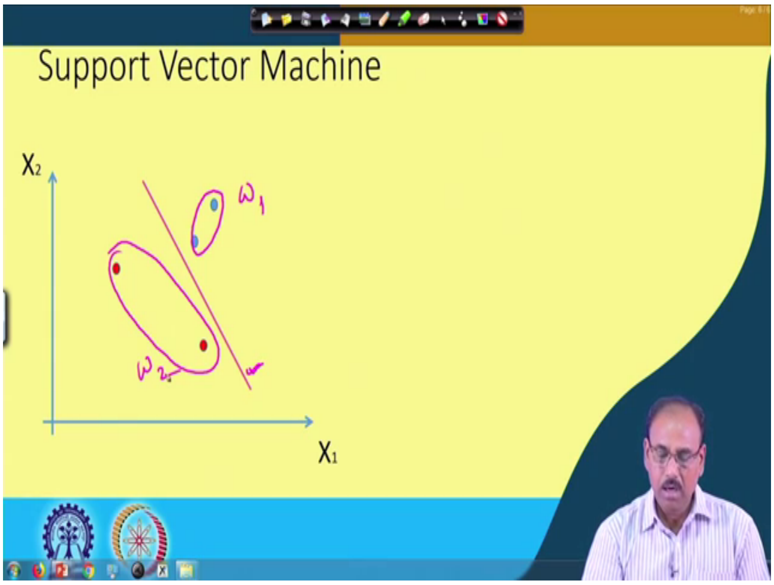
So, for different values of a and b, I have got I can obtain different separating planes and maybe many of those departing planes will satisfy the same condition that is a i y i into a transpose X i plus b to be greater than 0.

(Refer Slide Time: 04:55)



Now for different values of a the vector a and for different values of the bias b, I get different such planes, but for each such plane I will have the different margins or different confidence level of classification. So, what is that?

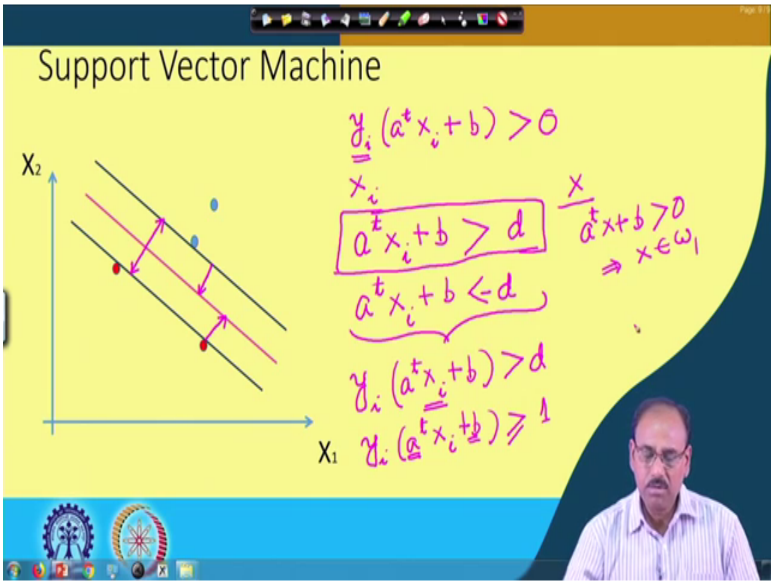
(Refer Slide Time: 05:26)



So, here I take this particular separating plane which separates between this set of set of feature vectors which belong to class omega 1 and these two feature vectors which belong to class omega 2.

Now, given this you find that if I take this particular separating plane, this separating plane gives me a margin which is given by this, so that the distance between these two planes gives me the margin or what is the confidence level of the confidence level given by this particular classifier. Similarly if I take another separating plane set, this one here again you find that the margin is given by this much ok. So obviously the margin given in this option is less than the margin given in the previous option.

(Refer Slide Time: 06:44)



To continue further if I take this separating plane, then again the margin is given by this. So, out of so many options which one should be preferred and that is the scope of the Support Vector Machine that is what the Support Vector Machine does. The Support Vector Machine tries to get a separating plane which maximizes the margin and for such a separating plane the separating plane should be at a maximal distance from the vectors belonging to both the classes. That means, the vectors belonging to class omega 1 should try to maximize the distance of the separating plane from the vectors belonging to class omega 1 and it also tried to maximize the distance from the vectors belonging to class omega 2 right.

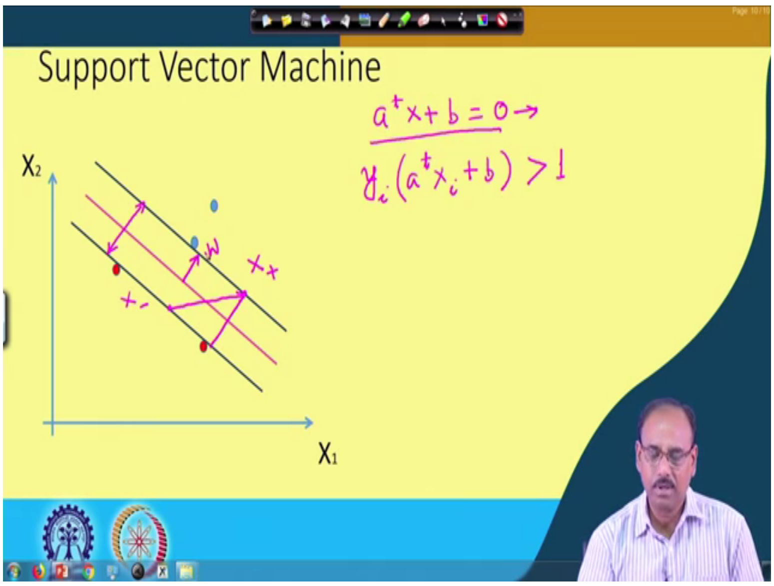
So, I should get that particular separating plane. I should try to obtain that particular separating plane which maximizes this margin and for classification my rule is that I must have y i times a i a transpose X i plus b. That should be greater than 0. This is for the classification, but as I am talking about the margin I want that for correct classification of a reliable classification for every X i, the distance from the separating plane must be more than a certain threshold. So, that distance as we said earlier that a measure of the distance is given by a i transpose X i plus b.

So, if a i transpose X i plus b equal to 0, that means X i falls on the separating plane in which case the distance of X i from the separating plane is 0. For any non-zero value if X i is taken from class omega 1, then I must have a transpose X i plus b to be greater than certain threshold say d and if X i is taken from class omega 2, then I should have a transpose X i plus b should be less than minus d and this should be true for all the training samples whether the training samples are taken from class omega 1 or the training samples are taken from plus omega 2.

So, if X i is taken from class omega 1, then this should be satisfied that is a transpose X i plus b should be greater than d and if the training sample X i is taken from class omega 2, then this one should be satisfied that is a i a transpose X i plus b must be less than d less than minus d and by taking this particular option I have an uniform criteria that is a y i a transpose X i plus b should always be greater than d irrespective of from whichever class this training sample X I has been obtained. What I can do is, I can always normalize this expression.

So, while designing I can have the condition that y i a transpose X i plus b should be greater than or equal to 1 and I will use this approach while designing the classifier or while choosing the separating plane, but for classification my rule will be once I fix what should be a and what should b after designing the separating plane or choosing the separating plane using the training vectors, then for any unknown X my classification rule can be that a transpose X plus b greater than 0 indicates that X belongs to class omega 1 or if a transpose X plus b becomes less than 0, then my decision will be that X should be classified to class omega 2.

(Refer Slide Time: 11:24)

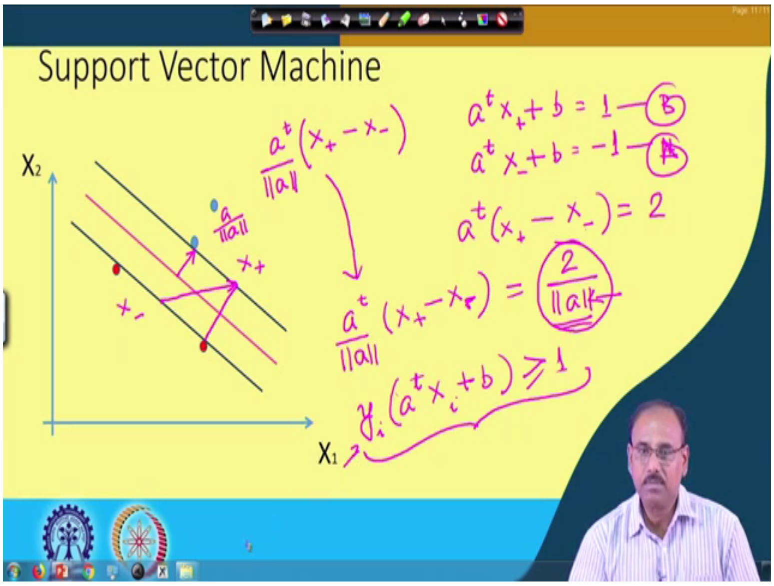


So, right now our aim is that I should choose this separating plane a transpose X plus b equal to 0 which satisfies the condition that y i a transpose X i plus b must be greater than 1. So, that is after normalization. So, how I can do that?

So, what I am saying is that in this particular equation of this particular separating plane I should take that particular separating plane which maximizes this margin. So, how I can obtain this margin and how I can maximize this margin? So, for that let us take one vector on this margin which is say X plus and I have take I will take another vector on this margin which is say X minus. So, X plus is taken within the class omega 1 region and X minus is taken within omega 2 region.

So, a vector X plus minus X minus is a vector drawn from X minus 2 X minus 2 X plus and once I have this vector, then from here you find that I can obtain that margin which is given by this as a dot product of the vector X plus minus X minus with the unit vector in that direction of w, right.

(Refer Slide Time: 13:24)



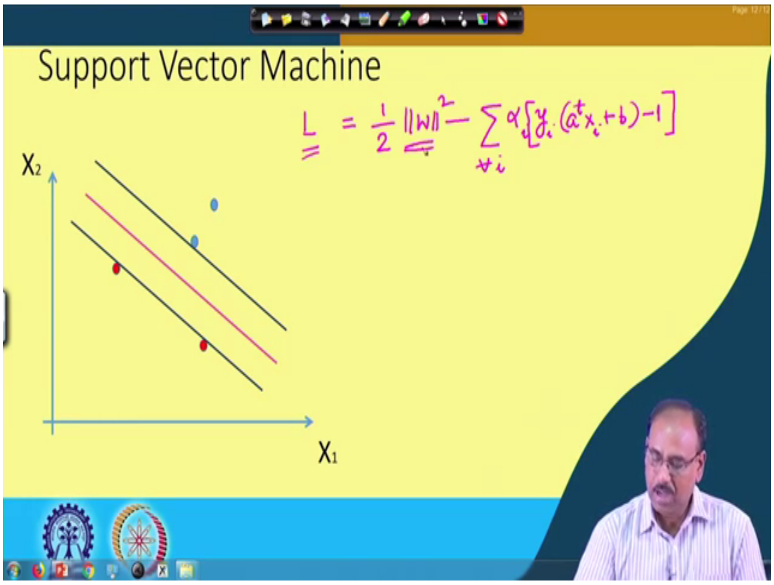
So, the situation that I have over here is I have taken a vector X plus in omega 1 region, I have taken a vector X minus in omega 2 region, drawn a vector from X minus 2 X plus and then from this I have to find out the margin which is nothing, but dot product of the vector drawn from X minus 2 X plus with the unit vector in the direction of w which is nothing, but orthogonal to the separating plane and the unit vector in this direction is given by a upon mod of a. So, the margin that you get is X plus minus X minus take the dot product of this or a transpose or into X plus minus X minus upon mod of a. This is what is the margin given by this particular separating plane and now you remember that we had the situation because this X plus is on the margin.

So, I have a transpose X plus b is equal to 1 and because X minus is on the margin on the negative side or on the margin into omega 2 side. So, I have this particular equation a transpose X minus plus b is equal to minus 1. So, from here you find that a transpose X plus minus X minus just subtracting if I call it equation A. From equation B I get a transpose X plus minus X minus is equal to 2. So, by using this you find that the margin a transpose upon mod a into X plus minus X minus sorry is given by 2 upon mod of a. So, as we said earlier that I should choose or I aim to choose that particular separating plane which maximizes the margin and the margin comes out to be 2 upon mod of a.

So, I should choose that particular a which maximizes this and here you find that obviously as mod of a comes in the denominator, I can maximize this term indefinitely by making a smaller and smaller, but that is not the solution because the A and B that I choose also must satisfy the requirement that y i a transpose X i plus b that has to be greater than or equal to 1. So, I have to minimize a subject to the constraint that a transpose y i into a transpose X i plus b have to be greater than or equal to 1. So, it becomes a constrained optimization problem and as you know that to solve accountant optimization problem, we have to make use of Lagrangian.

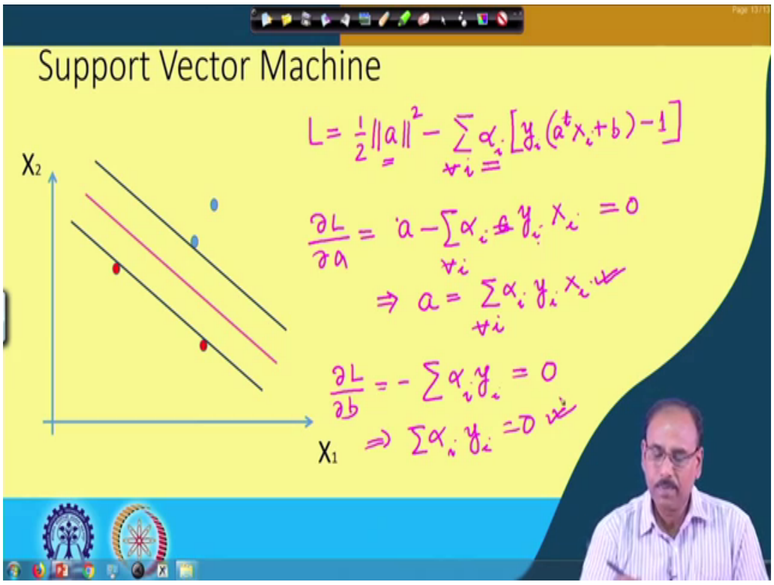
So, here what I have to do is, I have to form a Lagrangian using this particular constant.

(Refer Slide Time: 17:44)



So, the Lagrangian can be formed like this. I form l as I have to minimize mod of w. So, the Lagrangian that I form there I write half of mod of w square. Why I am taking at taking this as half of mod of w square will be clear very soon and then minus alpha i y i times a transpose X i plus b minus 1 take the sum of this over all i. So, this becomes my Lagrangian for constant optimization problem. So, this l of the Lagrangian has to be minimized with respect to sorry I am using the term a not w.

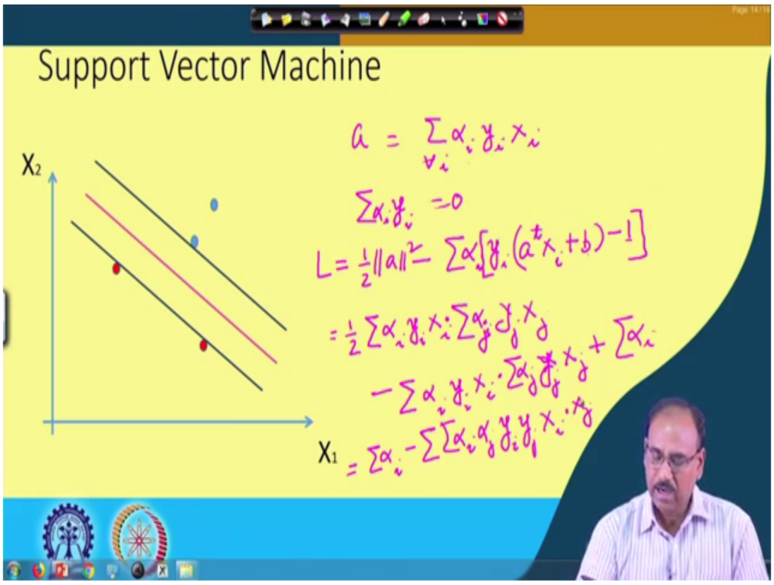
(Refer Slide Time: 18:52)



So, let me put it like this that my lagrangian l will be half of mod of a square minus sum of alpha i y i times a transpose X i plus b minus y and this summation has to be taken over all i. So, this lagrangian has to be minimized with respect to a and it has to be maximized with respect to our lagrangian multipliers which are alpha i. So, first for this optimization problem as you know that we have to make use of the differential operators. So, first let us try to differentiate l with respect to a and if I do that, it simply becomes a minus, it becomes alpha i a transpose sorry it simply becomes alpha i y i X i sum of this over all i. So, when I differentiate l with respect to a, it becomes a minus sum of alpha i y i X i and that has to be equated to 0 which gives me the solution vector a or the orientation of the separating plane to be equal to sum of alpha i y i X i summation has to be taken. Over all that is all the training vectors which are given for designing the Support Vector Machine in the same manner.

If I take the differential of l with respect to b what do I get ? The first term because there is no b over here. This becomes 0 over here, it becomes minus sum of alpha i y i yeah. So, it is sum of alpha i y i b. So, if I differentiate this with respect to b, it simply becomes sum of alpha i y i and that i if I equate to 0, this simply gives me that sum of alpha i y i has to be equal to 0.

(Refer Slide Time: 22:03)

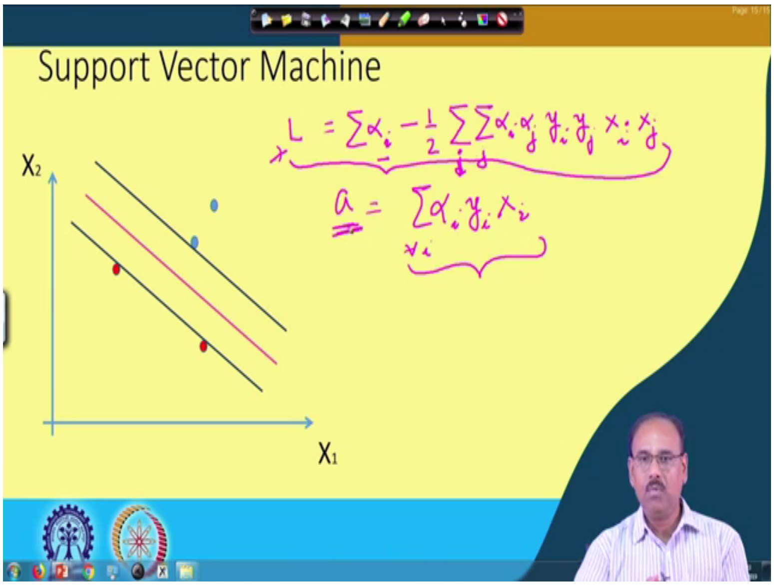


So, I get two intermediate solutions that is a is equal to sum of alpha i y i X i summation over all i and the other i get is sum of y i alpha i y i that is equal to 0.

So, now let us see what Lagrangian that we had. We had Lagrangian equal to half of mod a square minus sum of alpha i y i a transpose X i plus b minus 1. This was the Lagrangian and over here a is nothing, but sum of alpha i y i X i. So, putting that in this expression it simply becomes half of alpha i y i X i into I can write the other y as alpha i or alpha j y j X j minus what I have over here. So, I will put a transpose a is nothing, but a dot a. So, let us put it as a dot product.

So, over here again it becomes sum of alpha i y i X i again dotted with sum of alpha j y j X j. So, that takes care of alpha i y i a transpose X i plus or minus b times sum of alpha i y i and sum of alpha i y i equal to 0 and then I get plus sum of alpha i right and this simply gets gives me sum of alpha i minus double summation alpha i alpha j y i y j X i dotted with X j.

(Refer Slide Time: 24:59)



So, the final Lagrangian that i have is l is equal to sum of alpha i minus half alpha i alpha j y i y j, then X i dot X j or X i transpose X j.

So, this is the final form of Lagrangian and you find that under what we get as a. A is equal to sum of alpha i y i X i summation over all i and here it has to be summation over all j and summation over all i. So, the solution vector a is given by this expression alpha i y i X i taken summation over all i and what should be the values of alpha. The values of alpha will be should be those alphas which maximizes this expression of this Lagrangian. So, now you can make use of any of the optimization tool to optimize l with respect to alphas and the state of such alphas that you get which maximizes this. l can give you what is my solution vector a.

And once you have the solution vector a, you get your separating plane and this is the separating plane which maximizes the margin or in other words, this separating plane will give you our robust linear classifier.

So, today what we have done is, we have tried to find out a linear boundary between the feature vectors taken from two different classes; omega 1 and omega 2 and using Support Vector Machine we have tried to find out one such linear separator i will plane between the two separating planes in such a manner that this separator maximizes the margin between the vectors belonging to class omega 1 and the vectors belonging to class omega 2.

So, so far whatever we have discussed, whether it is a linear discriminator or a Support Vector Machine we have considered a problem which is only two class problem. So, next we will generalize this and try to find out that how we can obtain or how we can extend similar concepts to multi class problems. With this I stop here today.

Thank you.