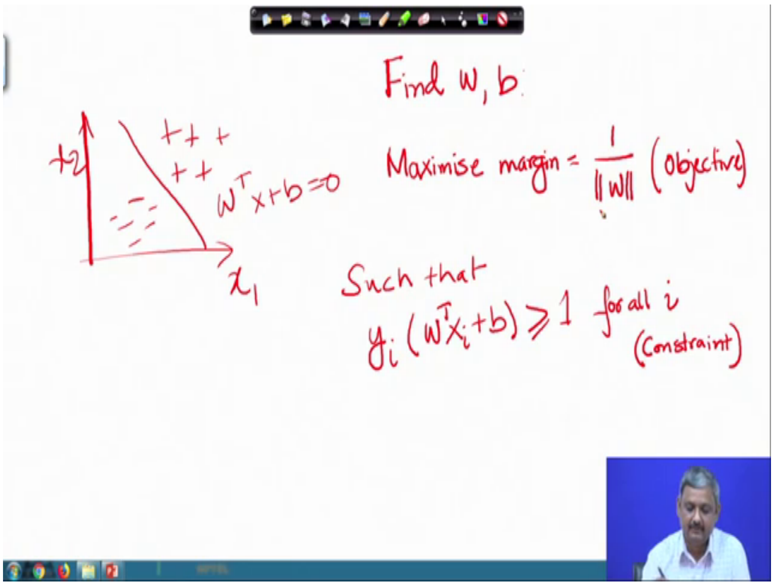
Data Mining
Prof. Pabitra Mitra
Department of Computer Science & Engineering
Indian Institute of Technology, Kharagpur

Lecture – 25
Support Vector Machine – IV

We continue our discussion on the support vector machine.

(Refer Slide Time: 00:21)

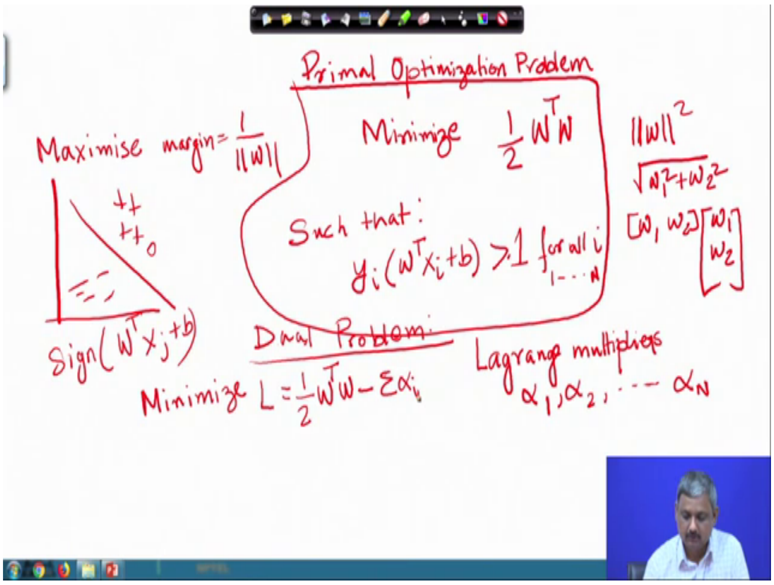


To quickly recollect we would like to form a decision boundary between the plus 2 class initially plus and minus classes of this form, of the form W transpose x plus b equals 0 and we want the 1 with the highest margin.

So, we saw that we can actually solve a optimization problem to get the values of W and b. Find W and b; there are the optimization problem is the following maximize margin, which we found to be 1 by norm of w, such that a constant y i W transpose xi plus b greater than equal to greater than 1 sorry not greater than equal to greater than 1 for all i.

So, this is a standard optimization problem, where we have a objective function which we want to minimize in a constant. So, what we do is slightly change the objective function; see since this is the magnitude this will always be positive. So, I can actually rewrite the optimization problem as the following.

(Refer Slide Time: 02:26)



I am rewriting it again instead of maximize margin which was equal to 1 by norm of w, I write down minimize W norm of w. In fact, since W is norm of W is positive, I can actually minimize the square of norm of w. So, instead of norm of W I am doing the square of it and I showed that the square off because norm of W remember it is W 1 square plus W 2 square. So, square of that is W transpose W into. So, W transpose W is of this form is W you do this multiplication you will see will get square of this quantity.

So, sorry we minimize this quantity. In fact, not this I put some up or some reason such that, y i W transpose x i plus b greater than 1 for all i. So, this is a optimization problem I want to solve; what are the free note that this x i y i these are the training set these are given for i equal to 1 to n these are given, so the free variables are W and p.

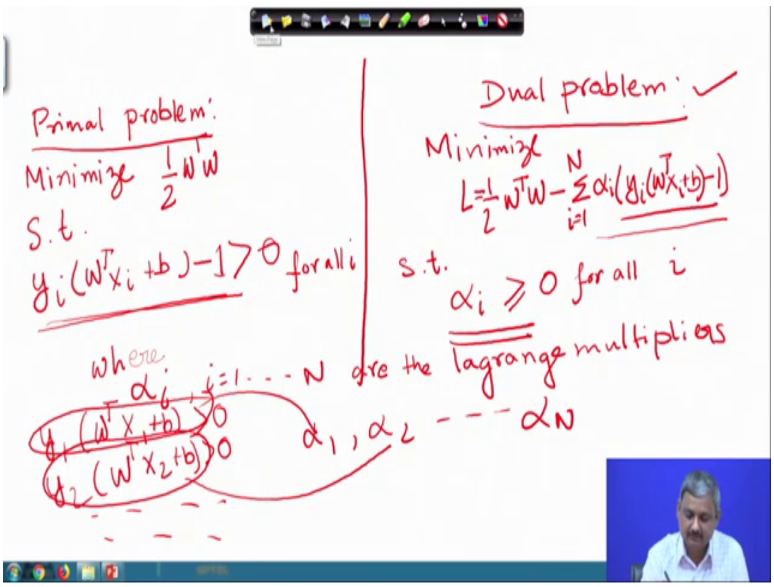
This has an objective function which we are minimizing and some constant. So, this problem I will call as the reason, I am giving some name it is called a primal optimization problem so this is clear. So, this is a clean and nice optimization problem W and some constant W transpose W and some constant W and b you have to find out and once we know W and b I can use that to draw my hyper plane. So, obtain my hypothec and when a new point comes I check which side of the hyper plane, it is I just check the sign of W transpose x j let me call it, we just take the sign if it is positive plus class negative minus class ok.

So, the reason I call it primal is that I will convert to an equivalent problem which we will call as dual problem, I am not going into the optimization theory it turns out that this transform problem the dual problem has exactly the same solution as the primal problem they have identical solution. So, if I solve that dual problem I still get my value of W and b, what is the dual problem? What the dual problem does is that it introduces a new set of free variables known as Lagrange multipliers.

New set of free variables not just 1 introduces many what it does? So, you have n training points for each and every training point I introduce a Lagrange multiplier alpha 1 alpha 2 dot up to alpha N. So, basically n new free variables I introduced, these 3 variables I call the Lagrange multiplier I denote them by the symbol alpha 1 alpha 2 alpha N, and the dual problem will take this form it will take minimize; what the dual problem will do basically I am saying is that it will convert this complex constant into a simpler constant, by moving the constant to the objective function.

I introduce a term called lagrangian which is times let me write it down here. Sorry let me explain.

(Refer Slide Time: 10:18)

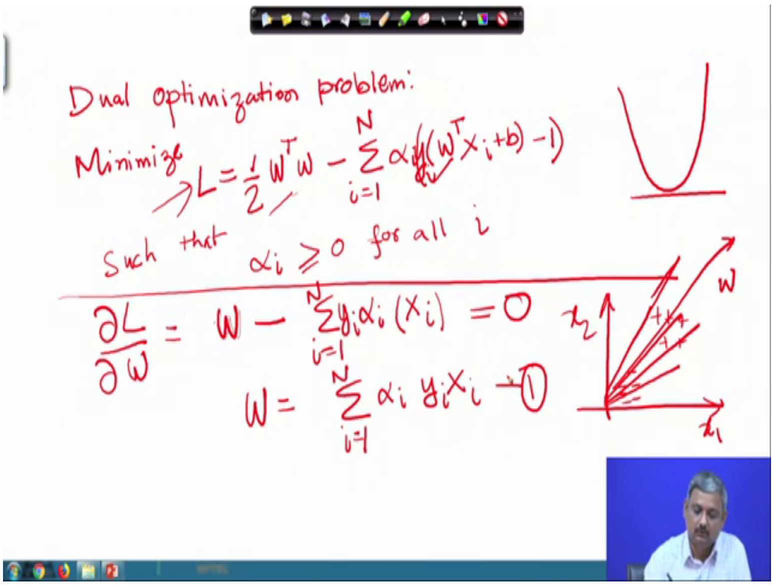


So, what I have done see this was the constant, I can actually I could have rewritten the constant this way, I could have rewritten the constant as; so what I do this constant I take here multiplied by alpha i, one thing you note that this is not just a single constant for each and every I have this constant; so that means, if I put some value of i 1, 2, 3, 4 I get a set of constants ok.

So, actually this means a set of constant like this and this is actually a set of constant so each and every constant I move here each constant corresponding Lagrange multiplier alpha 1 alpha 2 alpha n I multiply. So, I take this I multiply by alpha 1 I take this, I multiply by alpha 2 and add them all up that becomes my new objective function l and what is the constant? Constant is just that each of the Lagrange multipliers should be positive or 0 positive or 0.

So, the problem is see now we have more free variables wb and so many Lagrange multipliers, but still it is easier to solve I will show it soon. So, these are the 2 problems using optimization theory, you can actually show that they have identical solutions; they have the same 1 and the same solution. So, what I will do is that, I will solve this dual problem let me see how to solve it.

(Refer Slide Time: 12:58)

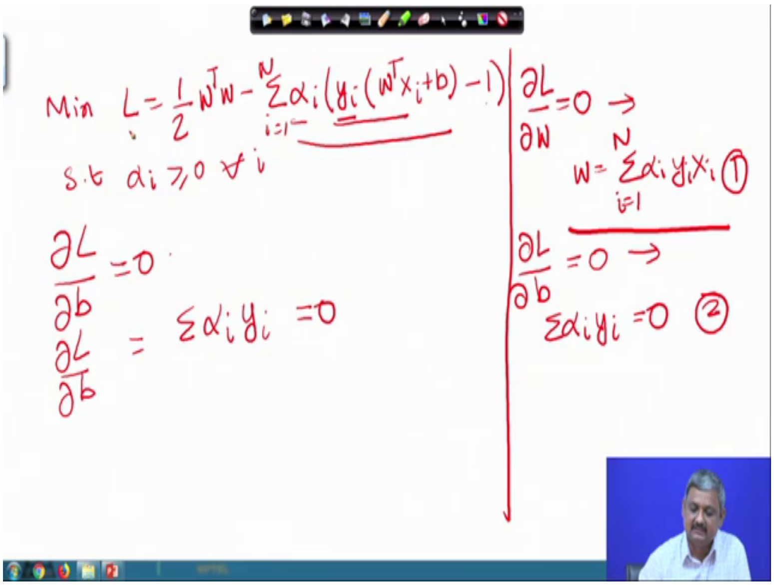


I am no longer writing the primal problem, I want to solve this problem right and once I solve it I will get value of W and b ok.

So, you know for minimization problem one thing holds that, if you take any function and at the minima take the derivative it becomes 0. So, the derivatives in this case the partial derivatives will vanish at the minima, let us see so first take partial derivative with respect to W del L del w. So, this I will take partial derivative with W you can work out it is just like plain plus 12 calculus let us see what it becomes. So, this is like half W square so derivative of half W square is w, these W goes this becomes W summation alpha i x i the second term alpha I b is independent of W so 0 summation alpha is 0.

So, this is the partial derivative at minima this will be 0. So, what we have is oh sorry I missed something very important is. So, there was a y i you go back to the previous thing I missed the y i, so there should be y i here. Simple derivative I missed a derivative please correct me in the previous thing. So, it actually means that the weight vector I call it as a weight vector is summation our y i x i alpha i your all the 10 on points, what it means? You take each and every x i multiply it by alpha i, reverse it is sign depending on the sign of yi add them all up, resultant vector is W here note this down.

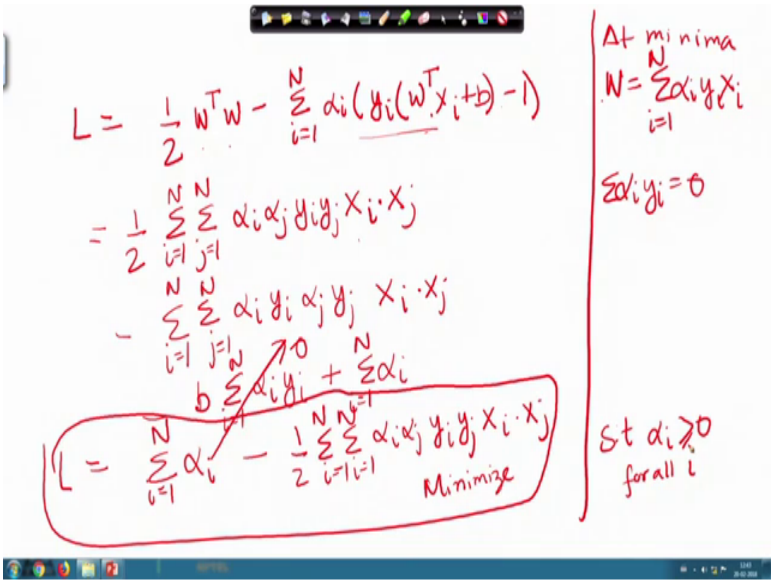
(Refer Slide Time: 17:24)



Now, wait is a symbol for all, now let us equate del L del b to 0, I equated del L del W earlier let us see; this term is free of b, only the second term b, comes outside, so summation wi so this term this term is free of. So, summation alpha I yi b comes outside take derivative with b goes away this is the third term is also free of b so 0 partial availability.

Let me write it down, see this is very simple algebra you have to just follow it. So, at minima we have these 2 conditions. So, I do not may take the derivative with alpha is now let me plug in this value of W in this equation of l plug in let me see what happens.

(Refer Slide Time: 20:15)

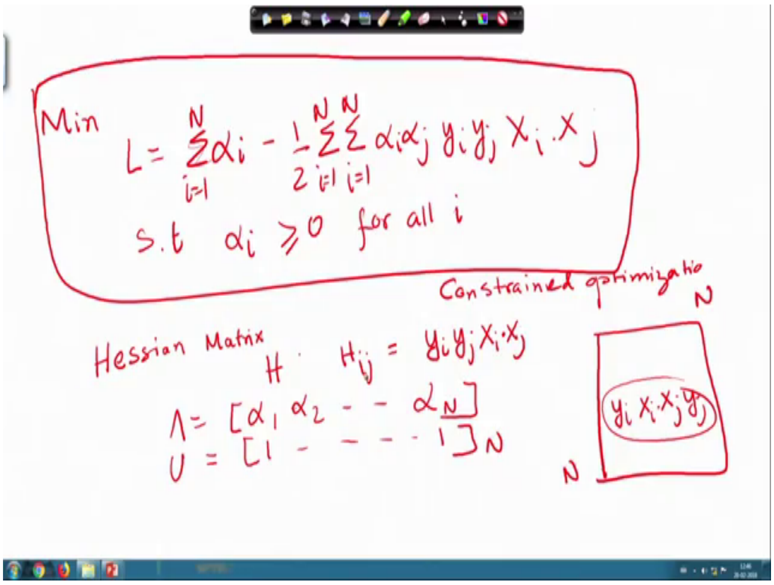


These 2 e height at the minima, so if I plug in this value of W here and here, I get I am just doing some algebra this value of W L, now substitute this one now the third term ok.

So, actually if you just plug in the value of W and just see that you when 2 summations will come and you will get a equation of this form. Now this quantity is 0 and this quantity is there, note one more thing is that see y i y j is 1 and minus 1 or 1 or minus 1 excise a vector excise a vector, but here I am taking the dot product of 2 vectors dot product W transpose W dot product; dot product of 2 vectors is a scalar this is a scalar alpha is a scalar. So, this entire quantity becomes a single scalar value right.

So, let me tidy up a little bit so I can write L as see even though this seems very cumbersome equations the actual algebra is very simple school level algebra, you just if you just go through it with open heart, will easily understand it. This is my L that I want to minimize subject to for all the way should I properly right.

(Refer Slide Time: 24:52)

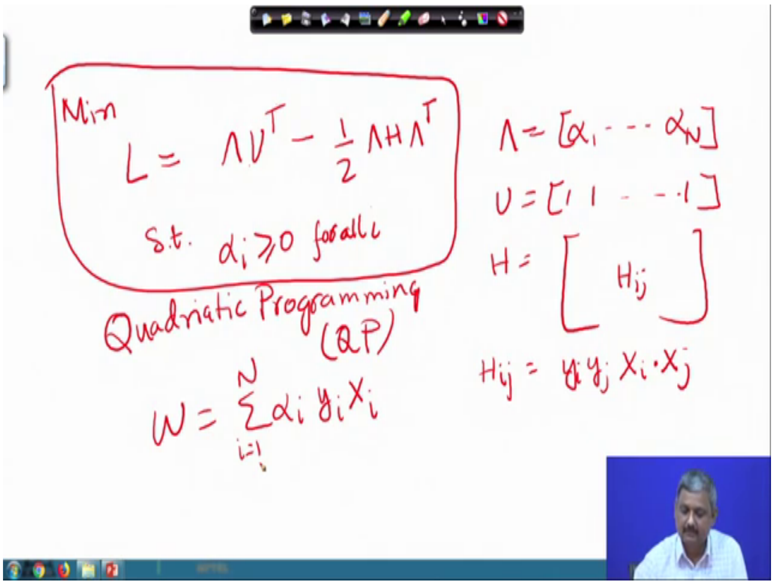


Yes let me properly right it once believe me I am copying correctly, i this is my constrained optimization problem. Note that now I have only 1 variable W and b are gone by this substitution of the minima; I have only the Lagrange multipliers as free variables ok.

So, even though this equation looks very cumbersome let me write a small matrix form of this which will be more intuitive. So, let me define some matrix so what I do I form a n by n matrix n is the number of training points, I pick up every pair of training examples xi and xj I take that dot product and I multiply by them their class levels. So, this way I get every entry of this n mind in matrix I call it a hessian matrix and then let me define this to be this vector capital alpha to be this all LaGrange multiplier vectors and u to be a vector of nN 1 with this notation; note this matrix is very critical, I take every pair of xi xj multiply them multiplied by their class levels I form a matrix.

So, if there n points I get n by n matrix. So, I can rewrite this equation as like this you.

(Refer Slide Time: 27:57)



If you remember the previous notations, I have made a small mistake. So, this only free variable is this alpha is so this is quadratic in alpha you have this square term and can be solved by a numerical method called quadratic programming QP and I get my value of alpha n, I plug them in this equation to get my w; how to get b, I will explain in my next lecture. So, with this so form this h matrix solve this optimization problem using QP get values of alpha i plug in alpha i here to get W up to this you have them. So, slope up the line you could find out next will discuss how to find b and some geometrical significance of these Lagrange multipliers. So, I stop here today I will continue my discussion in my next lecture.

Thank you.