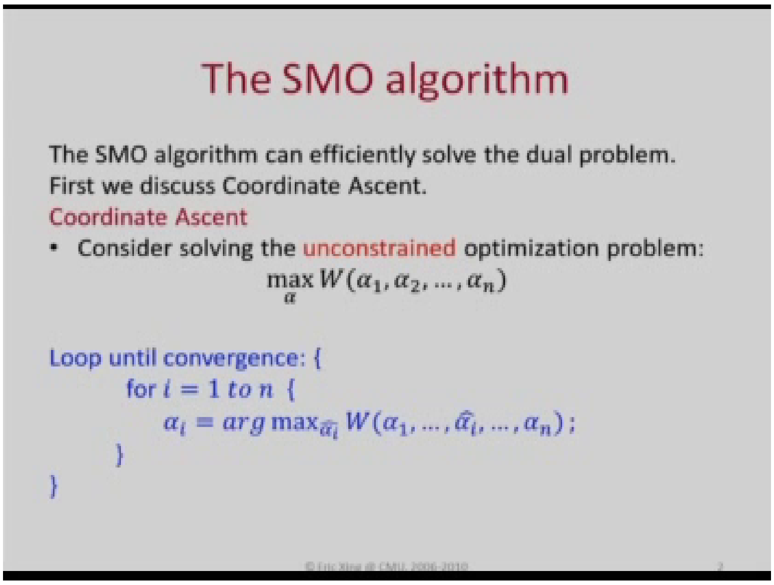
Introduction to Machine Learning
Prof. Sudeshna Sarkar
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur

Module – 5
Lecture – 25
SVM: Solution to the Dual Problem

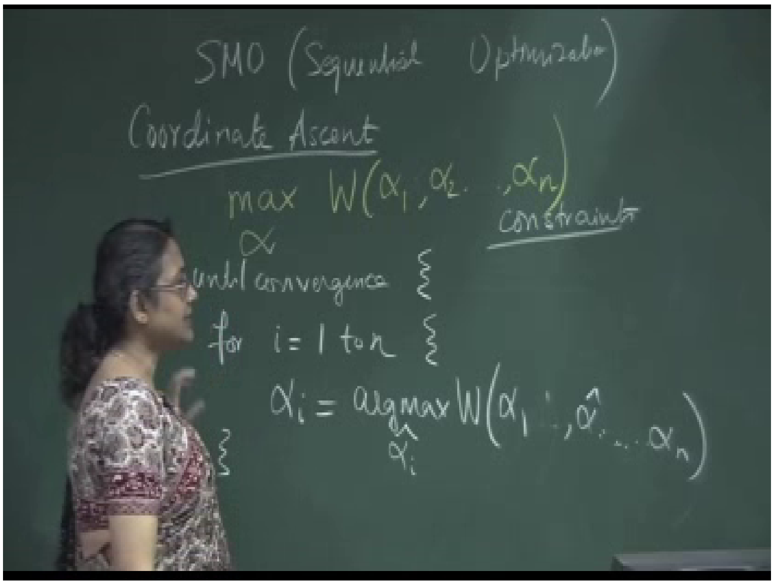
Good morning, this is the last lecture corresponding to SVM. We will very briefly discuss how we get a solution to the SVM problem. We have seen how we formulate the SVM as a quadratic optimisation problem and we came up with the dual formulation which has certain features which makes it easy to deal with kernel function, and we can solve this quadratic optimisation problem with the standard QP solver. But in fact, these with particular dual formulation it can be solved very efficiently and I will talk about the solution strategy in brief, an efficient solution strategy for dual SVM in brief, just to introduce you, but we will not get into very lot of details about this.

(Refer Slide Time: 01:13)



Now, the algorithm which can be used rather an algorithm which can be used for solving the dual problem is the SMO algorithm which stands for sequential optimisation.

(Refer Slide Time: 01:22)



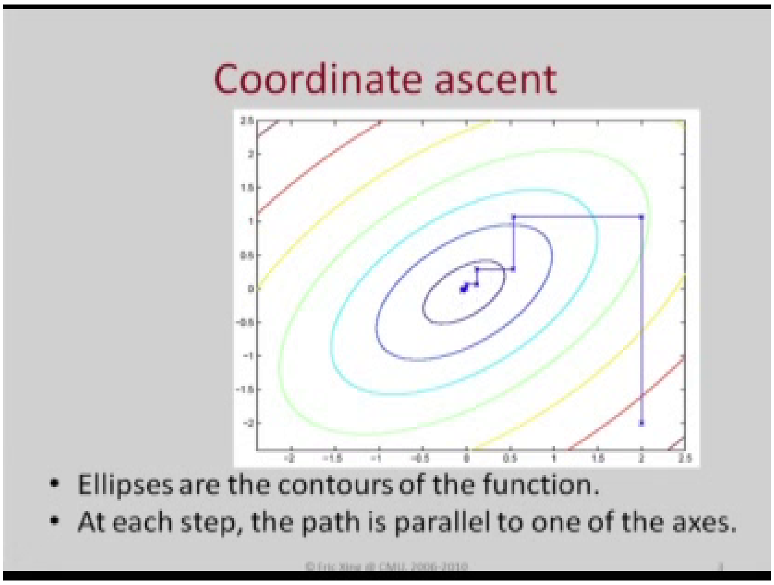
In sequential optimisation what we do is that, we have a number of variables and we keep the rest of the variables fixed and we only optimise one or two of the variables with respect to the objective function keeping the rest of the variables fixed. Before we talk about the sequential SMO algorithm, we will look at coordinate ascent. In coordinate ascent what we do is that this is the general form of solving an optimisation problem. Suppose, the optimisation problem is given by general optimisation problem, you have to maximise the parameters alpha of a function and suppose these are your variables; alpha 1, alpha 2, alpha n and you have to find the maximum value of this function with respect to alpha.

Now, in coordinate ascent what we do is that we loop until convergence for i equal to 1 to n. What we do is that we take alpha, we keep alpha 1 to alpha 1 minus 1 alpha i plus 1 to alpha n that is we keep rest of the parameters fixed and find out that value of alpha i for which this value is the highest. So, we can write alpha i equal to that value of estimate of alpha i where this w you know. So, where we keep the rest of the alpha i is fixed and we only change alpha i and then we end.

So, we can see, we can decide sequence of the alpha. Let us say, we will do first alpha 1, first optimise alpha 1 keeping alpha 2 alpha n fixed, we have find an initial value of alpha 1, alpha 2, alpha n which satisfies the constraint then we will keep the remaining alphas fixed and only change one alpha. So, that the value of this expression is optimised while satisfying all the constraints right we have some constraints also. So, these constraints must be satisfied and on satisfying this constraint, we find that value of alpha i for which this is the highest then i will keep this alpha i fixed and the rest of them fixed and take some alpha i plus 1 and we will change this.

So, sequentially we will do the optimisation with respect to this parameters alpha. So, this is the coordinate ascent algorithm of course, the solution that we get may depend on the sequence in which we choose the alphas.

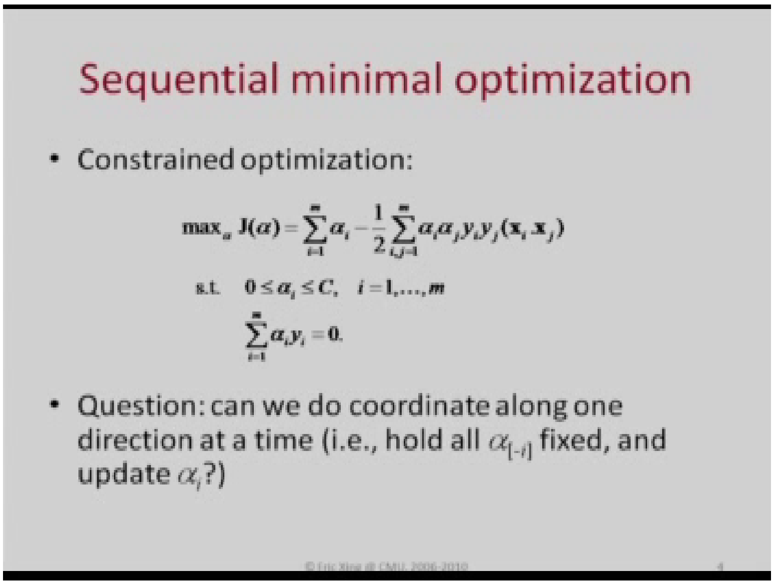
(Refer Slide Time: 05:08)



Now, this coordinate ascent algorithm can be visualised by in this slide. Suppose, here these different contours is ellipses are the different contours of the function right. So, initially suppose, I choose the value of alphas, so that I start from here, then I only change on one of the particular parameters right and we get here then we change in another parameter, we come here change in another parameter, we come here change in another parameter, we come here change in another parameter.

So, these contours correspond to the value of the function and the value of the function is minimum here after starting from here by optimising on one of the attributes at a time. We make this steps and these steps are we can see are on paths parallel to one of the axis and until we get to one of the minimum or wherever we get in. So, this is the idea of coordinate ascent.

(Refer Slide Time: 06:14)

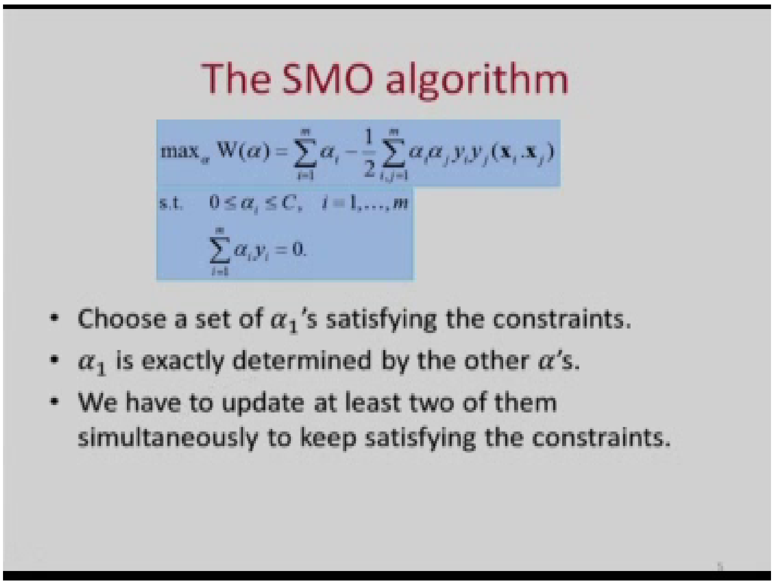


Now, this sequential minimal optimisation or SMO can be applied, we can see whether it can be applied to the dual formulation of SVM. So, if we look back at the dual formulation we are trying to maximise, find the values of alpha. So, the j alpha is maximised and j alpha is sigma alpha i minus half sigma alpha i alpha j y i j x i x j and these are the constraints alpha i is between 0 and c sigma i y i equal to 0. Now, the question is can we apply coordinate ascent directly here? Now, one of the things you notice is this constraints sigma i sigma alpha i y i equal to 0.

Now, in order to solve this, what we have to do is first we have to come up with some values of alpha 1, alpha 2, alpha m. So, that these constraints are satisfied alpha i is between 0 and c and sigma alpha i y i equal to 0. Now, you notice from this is that if you have m minus 1 of the alpha i is if you fixed and you want to change another alpha i, you cannot because given m minus 1 values the other value is completely determined. So, you cannot independently change only one of the alpha i's.

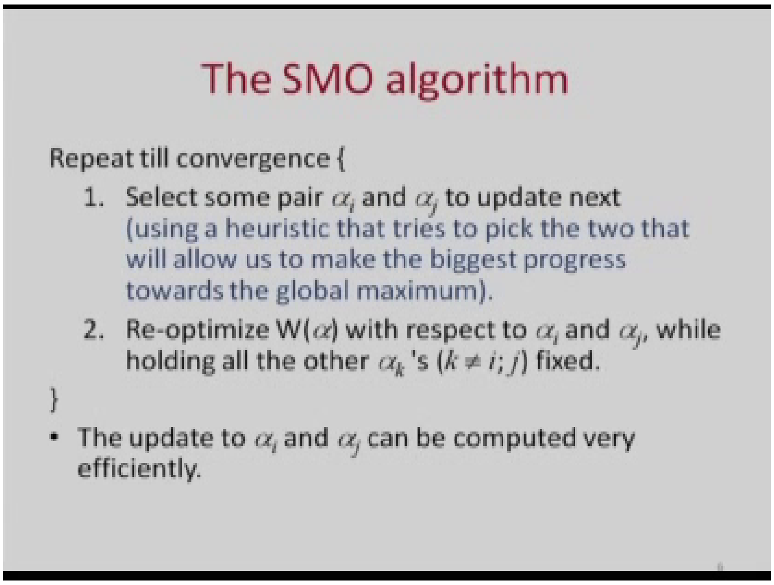
So, this algorithm where we sequentially optimised one attribute at a time cannot be applied to the solution of this particular formulation, but what we can do is that we can optimise with respect to two attributes at a time. So, we cannot update on only one attribute.

(Refer Slide Time: 08:11)



So, in the SMO algorithm, what we do is that we take two attributes at time to optimise them. So, this is the formulation that we initially choose a set of alpha i's, which satisfy all the constraint. Now, alpha 1 is exactly determined by the other alphas. So, what we have to do is that we update two of the attributes at a time.

(Refer Slide Time: 08:44)



So, this brings us to the SMO algorithm which can be used to solve the dual formulation of SVM. We repeat till convergence we select a pair alpha i and alpha j to update and we can use a heuristic to decide which alpha i alpha j pair to take, we can pick the two that allow us to make the biggest progress towards the global maximum to choose the alpha i alpha j to address next can be done based on a heuristic function.

Now, after choosing this pair alpha i and alpha j i reoptimise w alpha with respect alpha i and alpha j, while holding the other alpha case fixed. Now, it can be shown that the updates to alpha i and alpha j can be computed very efficiently. We will not talk about those details in this class, but I just want to tell you is that if the m minus to alpha z fixed, we get some constraints on these two alphas and this particular optimisation problem of optimising alpha i and alpha j keeping the rest of them fixed can be done quite efficiently and based on that, we have an efficient sequential minimal of optimisation problem to find a solution to the dual formulation of the SVM

With this we come to the end of our lecture on support vector machine. We have seen that support vector machines come up with linear decision surfaces and you change their formulation, so that you can accommodate noise and it is possible to convert a transform the original feature space to a new features space where the decisions surface are linear with respect to the new feature space. So, while having a linear decision surface you can also in affect have a non-linear function. However, the function that we have is linear in terms of either the original features space or the new features space.

In the next week, we will study neural networks and we will see in neural networks we can have highly non-linear functions and for that we wait till next week.

Thank you very much.