Transcriber Name: Saji Paul
Statistical Learning for Reliability Analysis
Prof. Monalisa Sarma
Subir Chowdhury School of Quality and Reliability
Indian Institute of Technology, Kharagpur

Lecture - 60
Tutorial on SVM

(refer time: 00:35)

Hi guys. So, the last tutorial of this class last tutorial as well as the last lecture. So, now as usual in all my tutorials the way I have covered. The first will be solving few objective type equations and then we will go on to solve few problems. Like here but one thing is that for solving problems for SVM classifier it is very difficult to get some toy problems for that. Because here solving some problems means we will be; needing some small problem.

That is way we can find out the optimization function and what to say from the optimization function we need to compute the MMH and if we need to do transformation also so it is very difficult to do using pen and paper and assets form, formulates such a toy example it is really difficult. So, the problems what I have kept here is just one simple toy example I have kept for classifications.

But the other problems just how to use the Lagrangian multiplier from how to use, how to solve the optimization problem using the language that is the convex optimization following. Using Lagrangian multiply I have just given some problems on that because that experience will also help.
(refer time: 01:44)

And now first coming to the objective type questions. So, these are also very easy ones. So, first is support vector machine is based on what? It is based unsupervised learning algorithm; it is based on the unsupervised learning algorithm, regression analysis, logistic analysis. So, support vector machine is not based on the regression or logistic and logistic regression and it is a classification supervised SVM we have used for classification.

So, when we have used first classification, we have used our training data has the class level. So, definitely when our training data is a class level so answer to this question is the supervised

learning algorithm.
(refer time: 02:26)

Then the vectors which are lying on the decision boundary that basically gives the complexity of the SVM also. What are those vectors called? Those vectors are called training vectors, those vectors are called test vectors. What are training factors? Training factors are which those vector which you use to train our SVM that is called training vectors. What is a test vector? Test vector test data which you want to test that is called the test data test back to this and the support vector training vector due to noise out of question.

So, this is the lying on the distant boundary this we call it a support vector and support vector also gives the complexity of the SVM as well.
(refer time: 03:08)

So, support vector machine can be used for this question support vector machine this we have seen that we have used support vector machine for classification type. But however, support vector machine can also be used for regression and also be used for ranking. So, we have not seen those two, this is a bit advanced topic. If you are interested you can just learn it by yourself. Once you have the idea the thing is that basically in SVM also the amount of lecture I have covered in SVM.

That is actually if you consider the SVM literature such a huge literature that is for just covering SVM literature. If I try to cover go into the depth of SVM go into every integrity of SVM it will be taking at least say 15, 20 lectures which is really not feasible. Here the objective behind is that once you know what is once you can understand the concept once you know what it is then you yourself can do some self-study and you yourself can learn those things it becomes very easier.

Once you know the direction know the concept then you can go ahead. So, that is a; be that logic actually have what to say discuss the SVM. I did not go into depth of the SVM so now similarly this SVM machine which we have used for only classification. But however, it is it can be also used for regression analysis and ranking tasks. So, the answer is this.
(refer time: 04:29)

The goal of SVM is to find the optimal separating hyperplane such that so what is the goal of SVM? The goal of SVM is to find out the maximum margin hyperplane, is not it? That is our goal. So, what it is given? It maximizes the margin of the hyperplane, it minimizes the margin of the hyperplane that is definitely not, it minimizes training error maximizes test error. When we are from trying to from our classification and always in any tasks our goal is to minimize training

error minimize test error both the time.

We do not need more test error and less training error or the vice versa, we need optimal of both. So, this is also definitely out of position so it maximizes the margin of the hyperplane, this is our goal.
(refer time: 05:17)

It is not here consider the following data set. We are performing a binary classification problem using linear SVM classify. The points are circled red that are representing support vectors. So, these are the support vector this is one support vector, this is one support vector, this is one support vector. If I remove any one red points from this data if I remove any one red point if I remove this or if I remove this or if I remove this.

If I remove any one of these data does the decision boundary will change? Yes. The equation here the answer to this is yes, because the support vectors decide the decision boundary. It is the support vectors which decide the distance boundary. So, if the support vector if I play around with a support vector definitely my decision boundary will also change. So, does the decision boundary will change if you remove any one point? Answer is yes.
(refer time: 06:12)

Next question is the same question. If you remove any non-red circle points if I remove any of this point, I do not reform the support vector. It may remove any of these points does the decision boundary will change? No, the decision boundary will not change because the decision boundary is very much dependent on the support vectors.
(refer time: 06:28)

What is true about the kernel in SVM? Kernel function map low dimensional data to higher dimensional space this is obviously does that, is not it? Kernel function map is that it implies the similarity in the higher in the transform space taking the data from the input space, is not it? So, it is a similarity function so both are correct. So, only one is true, both one and two are true this is the answer.
(refer time: 07:02)

Which is not used as a kernel function in support vector machine? Sigmoid kernel is used we have seen that, radial bias RBF kernel also is we have seen that, the table in the last lecture I have given some lists of some kernel functions where this is the polynomial kernel also, we have seen where we have the speed to the power p, p is a user defined. We can use any non-linear polynomial definitely it has to satisfy certain constant and only we can say is a kernel

function. So, this is the answer.
(refer time: 07:35)

What is the purpose of the kernel trick? What is the purpose? Now you know what is the purpose of the kernel treatment. Let us see what are the options given. To transform the problem from regression to classification is it the purpose definitely not. To transform the data from non-linearly separable to linearly separable, yes of course. To transform this problem from supervised to unsupervised, definitely not none of the above definitely not B is the answer.
(refer time: 08:03)

Which one of the following is a correct equation for a linear separator that is a hyperplane all symbols bear the usual meaning. Correct equation of a linear separator, which one? Definitely this one is not a linear separator less than equals to zero, in the hyperplane should be equals to zero. So, this is the one this. If I take decide and it will be equals to zero so in a hyper plane it will be when it is when it lies on the hyperplane and the point will lie on the hyper plane.

The value will be equal to zero. But here on the above the hyper plane will get a positive value below the hyper plane will get a negative value. So, this is the equation.
(refer time: 08:46)

Now some problems, as I told you these problems are not classification problem see.
(refer time: 08:55)

Suppose a company produced two goods x and y. As per company's policy total 42 goods need to be produced each month. The total production cost of the company follows a cost function so this is the cost function of the production cost. So, it produces two types of put x and y and this is the cost function. The owner wants to minimize the production cost while maintaining the constraint that the total number product in each month is $x + y = 42$.

Find a solution for x y using the Lagrangian multiply method. Even if you have mentioned Lagrangian we could we know we can very well use this Lagrange multiplier method because this is the optimization problem where we need to minimize this subject to this constraint. So, which one will be will it is a convex automation problem or with equality constraint? This is the equality constraint we have $x + y = 42$.

So, equality constraint means we will have will find out the Lagrangian, once we find out the Lagrangian and Lagrangian will do the differentiation of Lagrangian within each input attributes.

So, in our support vector we have seen we have used the Lagrangian we have differentiated a Lagrangian multiplier with the different attributes. The different attributes we have differentiated with different x i.

If there are n dimensional so δ (L) / δ (1 x 1), δ (L)/ δ (x 2), δ (L) / δ (x 3) likewise δ (L) / δ(x n) then also we have differentiated with respect to B δ (L) / δ (B), is not it? That is how we have found out the different equations. And then once we found out a different expression then we found out the values of the W and b. And of course, we have also differentiated and δ (L) with respect to the different dummy parameters this λ also, δ (L / λ).

How many dummy parameters? If there are two dummy parameters δ (L) / δ (λ 1) δ (L) / δ (λ 2). The number of dummy parameters a day will have to differentiate with respect to those also. So, the Lagrangian will differentiate the lagrangian and with respect to all of the input parameters as well as the all of the dummy parameters and then we have a couple of equations.

We will solve these equations to find the value. We are keeping in the mind constraint that we need to satisfy. So, that was equality constraint problem.
(refer time: 11:21)

So, here also it is a equality constraint problem. So, like we have to minimize this subject to this. So, how we can form the Lagrangian? This is how we can form the Lagrangian. So, this is f x + λ of the constraint so my constraint is this x + y. So, this I can write it in this form, is not it? 42 - x - y = 0 I want to equalize it to 0 so this is my what to say constraint. Constraint = 0 equality constraint so this is here.

So, I will do the differentiation of L with respect to x, differentiation of L with respect to y, differentiation of L with respect to λ.
(refer time: 12:05)

So, with respect to x I got this, with respect to y I got this with respect to λ I got this. Then I solved this tree expression and I found this is the value for x. This is this very simple equation so I could easily solve it but for a bit complex equation I will have to use some numerical techniques method. So, I got this x value this y value this and λ value this. So, the company must produce 25 of the first group and you see this satisfies the constraint also x + y = 42.

It is satisfying the constraint as well. So, the company must produce 25 of the first good and 17 of the second good to minimize the cost maintaining the constraint.

(refer time: 12:46)

Again, another one problem of optimization, solve the following optimization problem. Maximize this subject to this. Remember our convex optimization problem, we use for minimizing a particular function subject to inequality constraints where are inequality constrained is less than equals to. If you can remember our inequality constraint was less than equals to and we are of minimizing the optimization function.

So, if it is greater than equal to then what we will do? We will be in our Lagrangian will be subtracting it. So, this maximization problem we can make it minimization easily we can do that. So, our Lagrangian is this.
(refer time: 13:29)

I have changed it to minimization because this was maximization and this since it was greater than so I have used minus. Instead of plus you put it less than I would have used plus of λ 1 into these plus of λ 2 into this. So, this is my Lagrangian and in the Lagrangian I have to have some KKT constraint remember for what to say convex optimization problem you for inequality constraint so I have to define the Lagrangian and then it should satisfy certain KKT constraints.

So, what are the KKT constraint? KKT constraint is that δ (L) / δ x 1 δ (L) / δ (x 2) whatever I get that those two is equals to 0 and I do not differentiate with λ 1 and λ 2. I do not differentiate L with the dummy parameters for the inequality constant rather I have some other con constraint KKT constraint if you can remember KKT constraint. So, what are the KKT constraint?
(refer time: 14:26)

So, first I have been δ (L) / δ (x 1), δ (L) / δ x 2 these are the two KKT constraint. Now distinct this I have just kept it to just to remain remind you that this is not required for inequality constraint. For equality constraint only we differentiate with the dummy parameter. So, for inequality we do not differentiate this, this is not required. Then another what are the other KKT constraint?

Other KKT constraints are the constraints inequality constraints what is already there. So, these are the inequality constraint λ 1, λ all my λ 1, λ 2 should be greater equals to zero λ i should be greater equals to zero, this is one KKT constraint and then whatever the constraint specified. Then another one is λ i x i = 0 this is one KKT constraint. Λ i x i = 0 and then x i = 0 and then this is λ i greater equals to zero.

These are the different KKT constraint, is not it? So, we have this is one KKT constraint x 1 + x 2 less than equals to four and my λ 1 greater equals to zero I can write it this way, this is equals to 0, this into this. And similarly, it is x 1 + 3 x 2 - 9 this is less than equals to 0 and this λ to greater equals to 0. I can write as this this is one KKT constraint this is one KKT, this is one KKT constraint and these are the two KKT constraint.

We have the different KKT constraint. Now using this KKT constraint we will be solving the value finding out the value for x 1 and x 2 taking different combination of λ 1 and λ 2 value. We have two dummy parameters that is λ 1 and λ 2 and it specified that λ i should not be greater it should be greater or equal to zero, it cannot be less than zero.
(refer time: 16:25)

So, I will take different combination, I will take λ 1 λ to 0, λ 1 greater than 0, λ 2 = 0, λ 1 = 0, λ 2 greater than zero, λ 1 greater than zero, λ 2 greater than zero for all these combinations there will be four such combination. For so four such combination I will be using these four combinations and will try to find out the values.
(refer time: 16:49)

So, basically these are my obtained constraints whatever I have in the last slide I have written it here. Now with different case this already we have solved it, you can just go through different steps I do not need to explain this. We will take for each case we will take separately. First case is λ 1 λ 2 = 0. If λ 1 and λ 2 = 0 what does this expression becomes? This expression become - 2 of x 1 - 4 = 0, this expression become - 2 of x^2 - 4 = 0.

And λ 1 = 0 λ 2 = 0 then what happen? If λ 1 = 0 λ 2 = 0 then we will be using this x 1 + x 2 and this. Suppose λ 1 is greater than 0, if λ 1 is greater than 0 then I can write x 1 + x 2 - 4 = 0 then I can use this also as the equation. Again, if λ 2 is greater than 0 then this will be equals to 0, x 1 + 3 x 2 - 9 = 0. So, that way I will be having all the equation then I can find out all the solutions and once finding out for there are total four cases as I told you.

For the four cases you will be finding different solution. Now from among this solution you will have to find out which is the feasible solution. if there is one feasible solution then that is the only solution. If there is more than one for feasible solution then among the feasible solution you will have to find out which is the optimized solution and accordingly that is the solution.
(refer time: 18:18)

So, you can just go through the steps in fact please do not go through these steps you do it yourself. And once you get the answer then you try to match it with this.

(refer time: 18:29)

(refer time: 18:31)

So, this is how decided I am not going to all the steps you can do it on its own. So, we got the finally we got the solution x 1 and x 2 = 2 and it satisfies the constraint also. What is the constraint x + x 2 greater equals to 4 that is it is satisfying the constraint as well.
(refer time: 18:47)

Now here one is a small example of a classification problem. As I told you it is a toy answer this is not at all a realistic because for classification for developing an SVM we need more and more observation. We need more and more training data with a class level then only we can come up with a proper maximum margin hyperplane. So, it is just a toy example just to see just to utilize the concept. So, suppose there are three points, so what are my points?

This is one point, this is another point, this is my third point that means it is that each point is a two dimensional it is in a two dimensional space. Here P 1 and P 2 are the negative class this two belongs to negative class and P 3 of positive class two class negative and one positive. Compute the decision boundary as learned by the SVM classifier.
(refer time: 19:36)

So, first what will I do? So, this I will consider this is my decision boundary W x + b = 0. So, this is in two dimensional so I will be having two parameters w 1 and w 2. So, w 1 x 1 + w 2 x 2 + b = 0. So, we need to compute the values w 1, w 2 and b. Find this SVM learning problem is given by minimize? We will need to minimize this subject to this for all i. So, first we will define the Lagrangian that so my Lagrangian is this, minimize this.

So, it is greater then I will since my convex optimization problem it is less than equals to zero. Since; I have >= 0 so I will be using minus. So, this is how I have defined my Lagrangian. So, I have defined my Lagrangian now I will have to find out a different KKT constraint.
(refer time: 20:33)

So, these are the data points I have given here in the figure x 1, x 2, x 3 these are a different data point and other this is a class level which is already given class level and these are the λ value λ 1, λ 2, λ 3 we will have to find out this these three values now. So, now the KKT constraint this is one KKT constraint δ L / δ w. So, then this already we know λ i and y i should be equals to zero.

So, in our case n = 3 so we will first we have to find out what is w, w it says we know this is the w this is the formula for w. so putting this y i - and y i I will take value at a minus or plus then similarly we found out the values for w. And this is another one expression λ i w i = 0 means λ greater equals to zero. So, λ greater equals to zero minus means this is our λ this is equals to zero.

So, we will get this so these are the two equations. Simple putting it and putting it in the equation and trying to solve it that is all.
(refer time: 21:41)

Now basically if I dual problem is maximizing this, is not it? Subject to this greater equals to zero. Now we will put all this in this dual problem then we find L is equals to this just simply putting this here. Λ y i λ j y j x i x j and whatever data we have just put in the data, simple dot product. Excuse me if you do it you can simply do the dot product and see you will get this and simplifications and using this equation 1, we have this λ 1 + λ 2 = λ 3.

Using this we will get this is my Lagrangian. So, now I founded the constraint λ L by what to say differentiation of L / δ λ 1 / δ of λ 2 we found these different values. So, solving we got the λ value so λ 1 we got 0, λ 2 1 / 8 and λ 2 = λ 3 both is 1 / 8. So, I got a value λ 1, λ 2 and λ 3. If any values if the λ value is 0 that means that is not a support vector. So, my first point is not a support vector, the second and third point is my support vector.
(refer time: 23:04)

So, here this is not a support vector because λ 1 value is zero for value is zero means it does not lie in the parallel line, is not it? So, since it is zero so it is not a support vector these two are the support vector. But actually, this is a bit trivial see here, this is let us come to that later. So, now here we got this support vector λ 1, λ 2, λ 3 value we got.
(refer time: 23:30)

Once we know this then from here, we can find out the W values as well. We know λ 1, λ 2, λ 3 values we can put it here and we can find out the w value so we got w 1, w 2. So, we need to find out w 1, w 2 two values we will put it here and we found out the values. Once we found out the W values, we can find out the b values also. Once we find out the λ values w value, b value so we can find out the equation of the hyper plane.
(refer time: 23:54)

Now my equation of the hyper plane turns out to be x 2 = 0, very trivial example. Just this example is to show just to practice whatever you have just learned very simple example. Now here this x 2 = 0 means my hyperplane is this is my hyperplane and this are this point and this point I got support vector. See if this is my hyper plane this point, I got this is at a distance of two I got a this is my support vector this is a distance of 2 I got a support vector.

Since this is the x axis which is my hyper plane and that means distance of two my these are the two parallel lines. So, these are my distant boundaries, is not it? Now since again has surprisingly this I got as a support vector and this point if this point also lies in this side did not get as a support vector, why? So, this is a bit contrary what we have learned, is not it? Because if this is a support vector; definitely this also should be a support vector.

But actually, considering the case that this is just we are walking with three data's in considering three data's this. If considering this as a support vector a sufficient what to say it is actually it makes it sufficient for this data. If I just consider this as a support data this is no longer it needs to be considered as a support vector. In fact, when we see that some such a case whenever the line is parallel to the x axis the hyperplane is parallel hyperplane is on the x-axis or hyperplane is on the y axis.

So, then what happens? The decision boundary is just parallel to this two lines. So, in that case among there will be there may be many data which falls on the line parallel to that but all those data is not necessary to be considered as a support vectors. Just considering few of those as the support vectors what will help us in solving the problem? Because the main reason behind is that so it is very what to say first of all the reason behind is that since we are working in a very small example.

Very small and working in a small data we will definitely not get a proper SVM. So, to move the data we train to move the better and SVM we get. So, once we get a better SVM stand like which will be in a support vector which is not in the support vector then it is easier to find out. Now it is very toy example so it makes no sense just to find see how it works basically that is why we have kept this example.
(refer time: 26:36)

So, now another one, consider following two dimensional two dimensional data point for a binary classification problem. A non-linear transformation is used to transform the input vector. We have we are using this transform to infer to a transform space, this is the transform space. So, what is my transform function? Φ 1 of x 1 x 2 is this function and φ 2 of x 1 x 2 is this function. This is how we are transforming.

This one this whole thing we are transforming to φ 1 of x 1 x 2 to a different dimension and this whole thing we are transforming using φ 2 x 2 x 2 to a different function. Transform the data to the transform says is there any specific observation.
(refer time: 27:22)

So, this is my original data x 1, x 2. Suppose this original data is given and the class level are also given, the original data is given and the class level is written. Now from this original data I tried finding out my Z values z 1 and z 2. What is my z 1? This is z 1 and this is z 2, I tried finding my z 1 and z 2 value. Now I will try to these are also two attributes these are also two this is also two dimensional this is also two dimension.
(refer time: 27:51)

I tried plotting this data. See when I plotted this data so these are initially these are my data's. When I tried plotting this data if you see if you know you will get this sort of there is from this figure. We can see from this figure edition they are not linearly separable, is not it? Now once I have transformed it when I transform it so these are my values. Now in my transform space you see it is very much linearly separable. These are the data usually this set us triangle.

Now here in the transform space all into comes to one side, all the square comes to the other side is very much it has become very much linearly separable. So, we can see that the data points in our original input says on the left is not linearly separable. But when the data points are projected from input space to the transform space, they become linearly separable.
(refer time: 28:53)

So, with this I end this lecture I hope you have enjoyed this course. So, with that I had like to tell thank you to all of you and then conclude this course, thank you.