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 - 55
Support Vector Machine (Part - Ⅱ)

(refer time: 00:34)

Hi guys, in continuation of our lecture on SVM so today is the second lecture on that. So, and we have in our last lecture we have seen that for a linear SVM our objective is to find out the what to say margin basically for the maximum margin. So, as like what we can separate the data into the both sides of it, is not it. So, nowadays minute of a margin our objective is to find out the maximum margin this is also called maximum margin hyperplane.

And then until now we were just discussing about for binary classification on (01:07) classification of two classes. So, in this context now in this class will be covering how to compute the margin we have seen that maximum margin what is maximum margin hyperplane that we have seen. Now in this lecture we will see how to compute the maximum module hyperplane MMH.

And then this computing this maximum margin hyperplane disk can basically be formulated as an optimization problem. So, we will see how to formulate this as an optimization problem and this to solve this optimization problem we use Lagrangian multiplier method which will be discussing in this lecture. So, now first is the computing margin of maximum margin hyperplane. (refer time: 01:47)

So, in the last lecture we have already learned that a equation of the hyperplane is in the matrix form that is W X + b = 0 this already we have seen is not it. So, what is W is? W is equals to different w 1 w 2 depend on how many parameters access how many features how many parameters. So, accordingly that we have the different coefficient w 1 to w n if access m features then X = x 1 to x n and b is a real constant.

We have seen that is not it and we have also seen that we have level plus as one plus and mind this is the plus class this is the minus class and we can predict the class level Y for any test data given any test data X we can predict a class level Y this is the way Y is equals to plus X if W X + b is greater than 0 what happens. If this is the hyperplane anything which falls in this range above the hyperplane it is above the hyper plane.

It will be W X + b it will be greater than zero and below the hyper plane this side it is W X + b is less than 0. So, we will get a X class class if W X + b is greater than 0 we will get a minus class if W X + b is less than zero, this already we have seen.

(refer time: 03:07)

Now in the equation W X + b = 0, what does W represents, W represents the orientation business is not it and b represents the intercept. Now this any straight line if Y = m x + C what happens if we just scale up or down by dividing a non-zero constant what happens we get the same hyperplane. If we scale it if we multiply it or if we divide it by a non-zero constant, we get the same hyper plane. So, there can be infinite number of solution using various scaling factors.

All of them geometrically representing the same hyperplane, if you just scale up by multiplying it by some numbers or by dividing by some numbers both W and b will essentially get the same hyperplane.
(refer time: 03:54)

So, this there can be infinite number of such solutions, now so this leads to a confusion, so avoid this confusion what we do we can make W and b unique by adding a constraint. So, what we do we want to make this unique so that if we add a constant so this unnecessary confusion will not be there. So, what did you do basically the in other words this is the hyper plane and we draw two parallel lines which we already have seen in our last lecture is not it.

So, and this helps us also to avoid this confusion of infinite number of hyper planes. So, what we do is that I am just writing it here so basically W X + b is equals to if I write k plus minus k plus minus k because k plus is above the hyperplane minus is below the hyper plane, so plus minus K if I divide it by k both sides so this, I can write it in this form more generalized form. So, W X dot X + b dash is equals to plus minus 1.

So, now it may be noted that this W dot W X dot X + b dash plus minus 1 it represents two hyperplanes parallel to each other which you have already seen. So, now this W X b dash is unnecessary it is it will create confusion just for clarity instead of W X this is nothing this the parameters only is not it, so instead of that we can just write W X + b = plus minus 1, just two parallel lines on the both the side of the hyper plane, now we got this.

Now having this understanding now, we are in a position to calculate the margin of the hyperplane. Our objective is to find out the maximum hyperplane maximum the hyperplane which will have a maximum distance. So, now we are in a position to calculate the margin. So, how we will calculate.
(refer time: 05:45)

So, here we have some we will see two different method first method say suppose X 1 and X 2 are two points on the distant boundaries. These are the two distant boundaries say this is b 1 say this is b 2 or whatever it is ordinary person. So, X 1 and X 2 are the two points on a decent boundaries b 1 and b 2 respectively does what happens I can write W X 1 + b is equals to if it is in this boundary.

So, I can write this and if it is in this boundary then I can write W X 2 + b = - 1 or if I take both together I can I will get this is not it w dot X 1 - X 2 = 2. In fact, if you see here in this figure see this is one vector this X 2 X 1 what are these are vectors is not it this X 2 is one vector this is one vector. So, now then what is this is nothing but X 1 - X 2 vector is not it. So, if I want to find out the magnitude of this vectors what how will I write.

Say d is nothing but this vector X 1 minus here what I have written d is nothing but this vector. So, what is the magnitude of this vector? Basically, why I want to find out this is basically the distance between two vectors is not it this is because this is one vector this is one vector and so this is the distance let me rub it and show it again. So, this is one vector this is another one vector so this will basically give the distance between these two vectors.

So, now distance between these two vectors if I write it d, I will get this I can write it in this way because this is nothing but a dot product W dot X 1 minus it is nothing but a dot product. When we write the since this is a dot product if I take the magnitude of this vectors can I not write it in this way 2 by what is this and what is these equals to this whenever when I want to find out the magnitude of W it is nothing but square root of w m square + w square + up to w m square.

It is in a m dimensional space since we are considering X as m attribute so w will be also in the m dimensional space. So, the magnitude of w will be this so what we got does distance between two vector is 2 by magnitude of w, this is how this is one method of calculating the margin this.
(refer time: 08:13)

Let us see the second method of calculating the margin. Second method of calculating the margin so now see consider the two parallel hyper plane say this and this it is the same hyperplane what we have usually considered. Initially what we have considered let me write it here w 1 x 1 + w 2 x 2 + b = 1 and another one is w 1 x 1 + w 2 x 2 + b = - 1 is not it, it is a plus minus 1. So, if I take b decide what will happen so it will be w 1 x 1 + w 2 x 2 = 1 - b.

And similarly, this I will get a - 1 - b, so 1 - b if I write b 1 and - 1 - b, if I write b 2, if I this I write b 1 and this I write b 2 then I can this two equation I can write it in this format. So, do not get confused why suddenly I got b 1 and b 2 and this is nothing b 1 is 1 – b, b 2 is - 1 - b. So, these are the two hyper planes now these are the two hyper planes because we have already seen. Now if this is so how distance between, we want to find out the maximum distance between these two hyperplane.

So, max distance between these two hyper plane is this distance is not is this d, so how do I find d. So, this d is nothing but this AC this AC. So, if I was if somehow, I can find out the value of AC that AC is nothing but d, so how we will use some simple simplification to find out the value of AC. Now from this two line H 1 and H 2 whichever line whichever hyper plane you consider.

So, what will be the slope, slope is nothing but the - w 1 by w 2 from this line you can find out the slope. So, what is the slope? Slope is nothing but tan theta and what is that this is equal to minus w 1 by w 2.
(refer time: 10:11)

So, now here if you see what is this angle is theta, then what will be this angle, this angle will be 180 - theta. So, now if I find out the; sign of sine of this angle and now this is what this is the hypotenuse this is the perpendicular this is the base is not it. So, in triangle ABC AB is the hypotenuse AC is the perpendicular and the distance between this H 1 and H 2 I this AC is the distance between about to say H 1 and H 2.

So, now if I try to find out the sign of this angle what is this 180 - theta, 180 - theta is nothing but this AC by AB perpendicular hypotenuse sin theta. So, what we got this is one expression we got, so here AC = AB dot sin theta so if somehow if I can find out this value of sin theta and value of AB then I am done I got the value of AC. So, now here how do I find out the value of AB?

AB is easy to find out what is the distance from here this whole distance minus this distance is not it. This is AB distance from 0 to B. If this distance minus this distance will give me AB distance, so this is how I got AB and I know tan theta is also – w 1 by w 2. If I know tan theta, I can easily find out what is sin theta, sin theta I found out this simple calculation. So, I found AB, I found sine theta, so I found AC.

So, this is my AC, AC is equals to magnitude of b 2 - b 1 by root about w 1 square + w 2 square so what is b 2 - b 1 what is my b 2, b 2 is - 1 – b 1 is 1 - b. If I see so this is basically b 2 - b 1 is nothing but this is nothing but actually 2 by is not it this is not nothing but that we got the same thing here also we got. The first method also we got the distance between the two back to hyperplane here also we have got this and using the next method also.

This is what I got in SVM literature this margin is famously written as mu of W b, this margin this is called mu of W b, so this is my margin now.
(refer time: 12:43)

So, now I know this is the margin now what is that actually basically we will have to find out that finding the maximum margin hyperplane so that training phase of SVM involves estimating the parameters W and b for a hyperplane from agreement training data. So, what is the hyperplane our hyper plane is W x + b = 1 is not it so plus minus 1. So, now basically our hyper plane is W x + b = 0 the W x + b + - 1 is the parallel lines which is parallel to the hyper plane in equal distance.

So, now our in what we need to find out essentially in all or all the analysis what we have seen regression analysis and classification and whatever we just need to find out the parameters, here what is the parameters, the W and b is the parameters. So, now we have to choose the parameters in such a way that the following two inequalities are satisfied these are the two inequalities, one is W dot x i + b is greater equals to 1 and W dot x i + b less equals to - 1.

If this is for the plus class and this is for a minus class mentioning we are doing for binary classification so one is below the hyper plane. So, below the hyperplane it should be - 1 above the hyper plane it should be plus. So, we have to choose the, what to say parameter in such a way so that these inequalities are satisfied. So, this condition impose the requirement that all the training tuples from class Y is equals to plus must be located on or above the hyperplane W x + b = 1.

All the tuples belonging to this class where it will be located on or above it will be located on this line or above. Similarly, any tuples belonging to two this class this class will be located in this hyperplane or below.
(refer time: 14:37)

So, now so these are the two inequalities which are always seen if this two in equal this if we summarize together. So, these two increases we can write it in this form by combining these two together is not it can we not write it yes say this is if W y = 1 for some simplification we are writing is this plus and minus basically plus class and minus class. So, I am writing W i = 1 another one is W i = - 1 there is nothing but the one is above the hyper plane class and below the hyperplane class.

So, if this is the case if I put these two together I can this is my inequality now joining these two together. So, W i this is my inequality now. Note that any tuples that lie on the hyperplane H 1 and H 2 are called as support vectors. That tuples that lie on the hyperplane those are those we call it support vectors this is very important like here we see this is one support vector. This is one support vector which lie on the hyper plane that is called the support vector.

Essentially the support vectors are the most difficult that to classify because why it is most as something which is away from the hyperplane which is much above the hyper plane definitely it will be in a plus class, I know something which is a below much below the hyperplane it will wind a minus class I know. But something which is on the support vector means on the what to say on this plane this is most difficult to classify is not it that is another thing.
So, because the distance is very less the support vectors are the most difficult apples to classify and I give the most important information regarding the classification most important reform information regarding the classification means we will see later. The support vector actually defines the complexity of the SVM and so that is why we call it; it gives the most important information regarding classification.

And it has some in support vectors also gives up will visually see that. So, in the following we just discussed the approach of finding this minimum maximum margin hyperplane and the support vectors, how do we find the support vectors and what are the disks maximum margin hyperplane.
(refer time: 16:53)

So, now we have seen the; what is the hyperplane how we have to estimate we have seen this are the two techniques. One is the first method this is the first method this is our margin is not it, this is our margin. The second method also we got the same thing this is our margin this is nothing but the same thing two by magnitude of W the same thing. So, this is the margin so now and but at the same time it should satisfy the constant we have also seen that it should satisfy this constraint is not it should satisfy this constant.

So, now what we need? We want that our margin should be this margin what we have seen a margin is nothing by 2 by magnitude of W is not it that is the margin and this is what is them this is the constraint this should satisfy. So, now if cannot we; not put it in as an optimization problem. So, what is the optimization problem? Basically, first we need to maximize the margin this maximum margin hyperplane means our margin should be as high as possible we need to maximize this margin.

So, now we need to maximize this if I write it in the opposite way numerator becomes denominator if I put it in numerator then this is equivalent to minimizing the following. This is maximum this is maximizing then I can write as mu dash W of b is equals to this is if minimizing this then instead of if I write it in this pattern then I need to maximize this value but I am doing it opposite.

So, I am telling I need to minimize this and under the constraint subject to constraint this. So, in nutshell the learning test of SVM can be formulated as the following constant optimization problem what is this minimize this subject to this. So, this is how I formulate this as an optimization problem I need to minimize this, minimize what W by 2 and subject to W i y i into W dot x i + b and it should be greater equals to 1 subject to this.
(refer time: 19:02)

So, now so this is our optimization problem. So, the above constraint the above stated constant observation problem is popularly known as convex optimization problem. This optimization this type of optimization problem is also known as convex optimization problem. Where the object why what is the specialty about this objective function is quadratic in constraint here the objective function is quadratic is not it and the constraints are linear in the parameters W and b.

So, the specialty is the objective function is quadratic and the constraints are linear. So, this type of things we call it a convex optimization problem and for convex optimization problem the well-known technique is the Lagrangian multiplier method is used to find out the solution for such optimization problem. So, now we will learn this Lagrangian multiply method, how once we know this Lagrangian multiplier method then we will use this to find out our objective here is to find out the SVM.

SVM means our objective here is to find out the maximum margin hyperplane. Finding out the maximum margin hyperplane, essentially what we need to find out, essentially, we need to find out W we need to find out b is not it; we need to find out the support vectors those are the things we need to find out. So, for that we have formulated it as an optimization problem, what is the optimization minimize this subject to this.
(refer time: 20:32)

So, Lagrangian multiplier method follows two different space depending two different steps depending on the type of the constraint. So, does what is Lagrangian multiply method? It has two different technique one is we minimize this function for us also we have a minimization function subject to it has a equality condition subject to g i x = 0 means it has some constraint which is linear constant and this is equals to zero that is my constraint here so, i = 1 to p means there can be p constraint any number of constraints.

So, another is inequality constraint minimize this function subject to let us constrain h i x less than equals to 0 and similarly there may be p number upon strain. So, this is equality constraint optimization problem this is inequality constraint optimization problem. See inequality constraint my constraint is equals to 0 and inequality constraint my constant is less than equals to zero and what in both the case we have to minimize the function.

So, our case is also our case is what we have to minimize a function but our case is subject to h i x if I consider is h i x is greater equal to 0 is not it greater equals to 1. It is an inequality constraint.
(refer time: 21:58)

So, now how we do the how we define the Lagrangian multiply in this way. So, this is does we introduce a new parameter which you call it is a dummy parameter that is lambda so Lagrangian of this how we write it as f X that is the function plus all this constraint summation of lambda i g i summation of all the constraint multiplied by Lambda. For each constant we will use one lambda 1 dummy variable.

If there is one constraint, so lambda one g i if there is two constant lambda 1 g i + lambda to g 2, 3 constant lambda 1 g 1 + lambda 2 g 2 + lambda 3 g 3 likewise. So, it is summation of all the constraint multiplied by the dummy variable and plus f X this is my Lagrangian this is how I write it. Now what I will do is that I will do take the first order derivative of this with respect to each parameter of x each attribute of X first order derivative of this equation with respect to each attribute of X.

So, if x has d attributes so I will be del L by del x i del L by del x 2 del L by del x 3 del L by del x 4 up to up to del L by del x d. So, this is the first thing and since we are trying to find out optimum, so definitely does whatever will do the differentiation first order differentiation will equalize it to zero so if there are d attributes so from this del L by del x i total, we will be getting how many equation total will be getting d equations.

Next again this will also differentiate it with respect to lambda i so if there are p lambdas so we will differentiate all lambda del L with respect to all the lambdas lambda 1 lambda 2 lambda 3. And similarly, we will be getting total p equations. So, here how many equations will get we will get d + p equations. And so, from this d + p equations we will be able to find out the value and equalizing all to 0 and from this will be able to find out the values of this all x 1 to x d and all the lambda i's lambda i lambda 1 lambda 2 up to lambda p.

From this g + p equations will be able to get this value. This is the first one equality constraint

optimization problem how to get the values of the different X how to get a what to say we have to minimize the function. So, what is the value of the minimize function we can know the value of the minimize function for that we will have to know the value different parameters of the X for what parameters of the X we get the minimum value is not it. So, this gives us the different values of the attributes X which will give us the minimum value.
(refer time: 24:48)

So, here is an example also suppose we need to minimize this x + 2 y subject to this constraint x square + y square - 4 = 0. So, Lagrangian how we write so see it was a function of x y and we have used, another one dummy so x y lambda. So, what is this is our function is not it plus lambda into the constraint since there is only one constant so this is just one lambda see is not it, the same thing how we have written f X how we have written f X.

This is how we have written is not it f X plus summation of all the; constant multiplied by lambda I so here there is only one constraint so just one lambda so this is what my L. Now how what will I do I will differentiate this with respect to x, I will just differentiate this with respect to y because there are two attribute x and y. So, first I will differentiate with respect to x, then I will differentiate with respect to y that is one.

Next, I will again differentiate this equation with respect to lambda just raise 1 lambda so 1. So, difference in this with respect to s I got this with respect to y I got this, then with respect to lambda if I differentiate, I got this. So, I got three equations three equation how many unknown three unknown x y and lambda. So, solving this I got this value these are the different x values y values and lambda values.
(refer time: 26:14)

Now what will I do so I got different x value and y value, now my con thing is that I need to find out for which x value and y value there are different x so it is like simple optimization problem you know optimization problem you know genetic algorithm how do we do we take out the different values of x and y so, similarly we got the different values of x and y now which values of x and y actually satisfies our function we need a value which is minimize that x + 2 y should be minimized and at the subject to a particular constraint so from this.

We got different values we for x and y we for x y lambda we got two, two values so from this which value will actually we will give us the minimum value that we have to find out. We will have to solve each and find out so total there are how many choice the total there are 2 to the power 3 choices each value has three each for each variable we got two, two values and there are three parameters so total there are two to the battery choice.

For Lambda is equals to this we take x we take different x and y values definitely lambda will not be used in the equation when we will try to solve because lambda was a dummy parameter which we have just introduced. So, we will take different x and y values we try to substitute it and what we got if we use first this index function we got - 10 by root 5. Similarly, if I take lambda is equals to minus and this x and y this values, I got this lambda is not necessary you

just take the different values of x and y.
Different combination of x and y with a different combination of x and y which gives the minimum at the same time this also said to be satisfied x y + y square - 4 = 0 this should be sanctified and this should be the minimum value. Here we have seen in this two you can calculate all the values and see we have seen has its minimum value at this, this value. You calculate it and see we got it for x = - 2 by root 5 and y = - 4 by root 5 we got the minimum value.

So, now what is the value of x and y x value of x and y is this is the value of x and y. Essentially will try to do will try to solve it you taking different combination of x and y. Take one x another one y again keeping that x change the y, now again the change keep the x constant and change both the x likewise you will be getting all the combinations put it in the function see whichever gives us the lowest minimum value.

But at the same time whether a constraint is satisfied or not it is in the minimum and the constant is satisfied that is my solution that is for equal constraint.
(refer time: 28:52)

Now for inequality constraint inequality constraint is a bit complexity is a bit more than the equality how it is not difficult. So, the method of solving this problem is quite similar to Lagrange multiplier method described previously here it starts with the same with the Lagrangian same the earlier how we have used Lagrangian here also we are using the Lagrangian same thing nothing till now nothing is different.

But in addition, we use some additional constraint which we call KKT constraint. So, this is in Karush- Kuhn Tucker constraint in short it is called KKT constraint. So, for inequality constraint in addition to days this Lagrangian this is we call it Lagrangian when we introduce the dummy variable lambda, we call it Lagrangian and in addition to this Lagrangian we introduce some other constrain.
(refer time: 29:46)

So, what are these constraints say one is first of all the one constraint is that this Lagrangian will have to derive it develop to the first order derivation for all each for each attribute of x i which also we have done for equality constraint. Same wise we will have to do it for each parameters of x i = 0 and for the equally constraint we have also done first order derivation with respect to lambda here we will not do that.

Instead, we have this tree constant what is one is lambda I should be greater or equal to zero that is one constraint my lambda I value should be greater or equal to zero it should not be less than zero my x i value will be less or equal to zero, excited a constraint my constraint value will be x less or equals to 0 and my lambda i x i should be equals to 0. See these are the three constant. Lambda I greater equals to zero please remember when we will solve it.
We will be needing this lambda I beta = 0 h i that is the constraint less equals to 0 and lambda i h i will be equal to zero. Lambda is h i = 0 means either lambda i = 0 or h i = 0. Solving the

above equation, we can find the optimum value of f x now if we solve this, we can find out the optimum value of f x.
(refer time: 31:12)

So, let us see an example suppose we need to minimize this is subject to this constraint. Subjects to this constraint we need to measure here in the inequality constraint our constraint is in less than sign so less than equals to zero. So, accordingly we will just do a bit of manipulation and will bring so this will become x + y - 2 less than equal to 0 is not it and this will become x - y from this we can write x - y less than equals to 0.

If you bring x this side of sine gets changed so x - y less than equals to 0 so these are by two constraints so there are two constraints so I will be using lambda 1 and lambda. So, first is this my function what is my Lagrange? Lagrangian is my function this is my function plus I have two constraint so summation of both the constraint multiplied after multiplication with the dummy parameter so this is one this is the second.

So, what is my KKT constraint, KKT constraint is first del L by del x here there are two parameters x and y so del L by del x whatever I get that s = 0 then del L by del y = 0. So, del L by del x i got this is equals to zero first order derivative simple del L by del y i got this is equals to 0. So, remember what was the what was my KKT constraint one is del L by all the attributes of x they will x i so here there are two attributes x and y that we have done.
(refer time: 32:46)

And what is the second constraint second constraint is my this all by what to say if this was my second constraint, I have written it in the second constraint is all my del lambda 1 lambda 2 is greater equals to 0, is not it. Lambda 1 lambda 2 is greater equals to 0 and what is my the third constraint third constraint is my all my h i h i is less equals to zero is not it whatever my constant h i less equals to 0 so this is the diff I can write it into x less equals to 0 format also.

That means x + y - 2 less equals to 0 then x - y less equals to 0. So, lambda 1 lambda 2 greater than zero dash less equals to 0 then what is the next constant lambda into h i = 0 so this is lambda into h i = 0 lambda into h i = 0 this is one constraint this is one this is one. So, these are my KKT constant. So, this many equations I got. Now with these equations I got all this equation so with this equation I will be able to find out the value of x y and a lambda 1 and lambda 2.

So, now I will have to find out the minimum value of f x y subject to this constraint subject to this KKT constant all my KKT constraints which has been satisfied. So, based on that we will have to find out the value.
(refer time: 34:05)

Now to solve KKT constant we have to check the following test. So, first if I take lambda 1 = 0 lambda 2 = 0 my what is my lambda 1 what is my constraint, my constraint is this is the constant lambda 1 lambda 2 greater equal to zero it cannot be less than zero. So, if I take lambda 1 and lambda 2 = 0 if I take lambda 1 and lambda 2 = 0 in this equation what happens lambda 1 = 0 lambda 2 = 0.

Then what happens, I got this is 0 this is 0 so I get 2 into x - 1 = 0 I get this expression similarly from this expression I get this 2 into y - 3 = 0. From this two equation I get x = 1 y = 3 if I solve it simple equation simple equation you solve it so but however when I get x = 1 y = 3 what is my constraint my constraint is x + y 1 of the constant is x + y is less than equals to - 2. This is one of my constraint another constraint is y is greater equals to x.

The first constraint only plus y less than equals to less than equals to two it is not satisfied with this value x 1 = y x = 1 y = 3 this is not satisfied so this is not a feasible solution. So, if I took lambda 1 is also zero, lambda 2 = 0 the value of x and I what I got that is not a feasible solution now let me take another one combination. So, another one combination let me take lambda 1=0 lambda 2 not equals to 0.

If my lambda 2 is not equals to 0 first look here if lambda 2 is not equals to zero, you see this expression equation if lambda 2 is not equals to 0 then definitely x - y = 0 again is not it. So, if x - y = 0 so from this what will I get x = y I go to another one if lambda 1 = 0 lambda 2 not equals to 0 putting these values here in this two expression. I got this is one equation this is another equation putting value lambda 1 = 0 and lambda 2 is not equals to lambda 1= 0 it is gone.

So, these are the three equation I got from this I found out this value x = 2 y = 2, lambda 2 = - 2 but what is given my lambda 1 lambda 2 should be greater equals to 0. So, here I got lambda 2 = - 2 so that means this is it violates my assumption so this is also not a feasible solution. So, I took lambda 1 0 lambda 2 zero so if I see here lambda 1 lambda 2 different combination I can take this zero both zero this I got this is not a feasible solution.

Then I take this is zero and this is not zero this also I found not a feasible solution then I take this is not zero and this is zero, I will find out the solution another is both not zero both lambda 1 lambda 2 not 0. These are a different solution I can get.
(refer time: 37:13)

So, next I took lambda 1 not zero lambda 2 = 0. Accordingly, I found these are the equations and this value I got this is lambda 1 = 2 this is also feasible because it is given lambda 1 lambda to beta equals to zero and I got x = 0 y = 2 and it satisfies the constraint also, what is my constant my constant is x + y should be less equals to 2 x + y less is equals to 2. So, it is satisfying the constrain also.

So, this is a feasible solution and a final solution what do you find lambda 1 is not equals to zero both lambda 1 and lambda to not equals to zero. This what I got this also if I solve it, I got x = 1 y = 1 but lambda 2 = - 2 my lambda so lambda should be greater equals to that is also one of the constant KKT constant. So, this is also not a feasible solution so just this is the feasible solution. So, what is the feasible solution of this optimization problem x = 0 y = 2.

That is my feasible solution that is how we find it Lagrange in multiplier using the Lagrangian

multiply method we how we find out the values of the values of the f x and y values of the function which will give us the minimum value of the function. So, that we have seen in for some example.
(refer time: 38:29)

Now basically we will be using this method definitely our is inequality constrained and we will be using the second method that is the Lagrangian multiplier for inequality constant. We will be using that method to solve our hyperplane problem that was our problem, for that problem we have used we have formulated as the optimization problem. Now we will be solving this optimization problem using the Lagragian and multiplier which we have just now learned to solve that.

So, now coming to the conclusion in this lecture we learn about the computation of the margin of the hyperplane separating the class in the class of linear SVM. We have seen how to compute the margin we have seen two different two different methods so what is the margin is 2 by w so, how we have computed we have seen then we formulated the optimization problem to obtain the optimal hyperplane with a maximum margin.

To solve the optimization problem Lagrangian multipliers were used we got a basic introduction of how the Lagrangian multiplies are equalized we have seen that. So, in this lecture we will attempt to solve the optimization problem for linear SVM.
(refer time: 39:33)

So, with that I end this class thank you guys.