NPTEL

NPTEL ONLINE CERTIFICATION COURSE

Introduction to Machine Learning

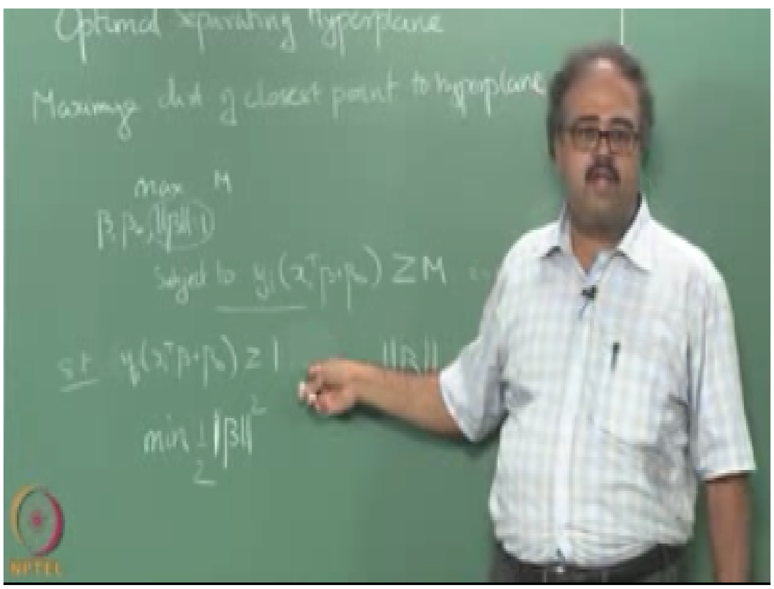
Lecture 27

Prof. Balaraman Ravindran
Computer Science and Engineering
Indian Institute of Technology Madras

Support vector machine 1
Formulation

Also we are looking at linear classification and I will quickly remind you of the properties of hyper planes that we wrote down in the last class.

(Refer Slide Time: 00:27)



Ok will denote by f of X right so the hyper plane is essentially given by solving the effects equal to zero okay and what I really want is, So I do not care about the other properties what I really want is for you to recall that the sign distance to the hyper plane is given by f of X okay. so then we looked at the perception learning algorithm so does a problem with the perception learning algorithm convergence yeah.

So if it is linearly separable it will convert but it might converge slowly if it is not linearly separable it' will cycle okay but if it is linearly separable what can you say about the convergence apart from the fact it slow it depends on the starting point right it is not definite as to where it will actually converge right there is no particular solution to which it will converge right so now what were going to try and do is try to characterize a specific optimal solution.

Right we will first start by considering the case of linearly separable data so just like whatever I have drawn here so he data point is given to you is actually perfectly separable by ahyperplane okay right so we will start with this case now I am going to try in characterize what i mean by nor I mean by an optimal separating hyper plane give me some options for what could be optimal think this distance is some of what are you some of distances is maximum between the points and separating ahead of me themed low some of the distance some of the distance of all the data points to separating epically maybe this point is power exactly right so that makes a lot of sense so that is exactly what we are going to use right so we are going to maximize.

Distance of closest point to the hyper plane right so essentially so if you think of this data right so that is close but this is closer right so if I want to maximize the distance of the closest point what should I do move it like that or like that whatever some summer moved further away from this right so what how much further away can I move it until other side also i say the closest point from both sides should be at the same distance.

Right so on both classes right the closest point should be at the same distance from the hyper plane right and I have to choose an appropriate orientation for the hyper plane so that this distance is maximized so that is essentially what we are going to try and do so since of erasing the hyper plane rowing it again I just move the data point so that this is closed up now rights now I'm essentially going to have as slab in some sense around the separating hyper plane which will have no points.

Right so on the thickness of the slab will be the same on either side of the hyper plane. So that is essentially what am looking for so this is called the margin whatever you cleaned up around the hyper plane is called the margin so that is why these kinds of optimal hyper plane classifiers or sometimes known as math’s margin classifiers right because they are trying to maximize the margin is the distance of the closest point to the hyper plane is the margin.

So in fact so the margin would be that so we know what X I transpose π + π naught this lets the distance signed distance from the hyper plane so x y I so that I always get it positive right so essentially what I am saying is I look at the quantity this is essentially the distance a data point is away from the hyper plane right I am saying that every data point has to be at least me away from the hyper plane that is my constraint.

Right so go through all my training data points 1 to N and I am saying that every data point should be at least a distance e m away from the hyper plane and under that condition maximize him right so I cannot make me arbitrarily large because I might not be able to find a π which will satisfy for every data point okay so this is how I will write down the optimization function what we wanted was maximize the distance of the closest point to the hyper plane.

So now what I am going to say every point should be at least MMA now maximize I am this will automatically maximize the distance of the closest point right so what do you think will be the distance of the closest point whatever me you end up with right so whatever is optimal me for this will be the distance of the closest point and that will be the margin so that is essentially what you are doing is you are directly maximizing the margin okay.

So we have this constraint that name β should be one that because we do not want the solution to blow up arbitrarily so instead of that right will get rid of that by that if I did not have if you didn't have the constraint of non π equal to one I can arbitrarily make π large here right and make things larger than m2.so now I am normalizing by π right so that I do not have to worry about that I can remove this constraint here that instead of that I put the name Bata herein the denominator in the constraint it make sense right but then I can do something more interesting now into that right now let us step back and think about it for a minute so if something satisfies this constraint right.

I can arbitrarily scale it right so it will still satisfy the constraint you think of it because I have norm π on this side so if a π already satisfy this constraint okay I can just scale the π ok and then that will satisfy the constraint as well correct so I can arbitrarily set right I still have to find out the direction of orientation of the π I'm just saying that whatever orientation of the π you pick you normalize it so that the norm of π is 1 by M.

So now what happens if I set that so people with me so far so kind of I started with this constraint started with that optimization problem made the assumption that well norm π is 1 by M right and I came up with this problem so then nothing just little man geometric I mean algebraic manipulation in fact it is geometric as well I am just not drawing the geometry here right but then the objective function also now can change it is a maximizing e m again minimize norm π because now π is 1 by M right.

So I can do that subject to the constraints here right I am going to do all of that so that it makes it easy for me to take derivatives minimize and things like that right so I just made it a squared function so that can manipulate it more easily does not matter minimizing norm π is the same as minimizing meter square right so norms or anyway positive or non-negative yeah is it clear so far right so essentially.

Now I can go and say that this margin is actually going to be what right so another way of thinking about it is what I am trying to find is a minimum norm solution such that all the data points are what are correctly classified right so what does this mean when y I x I transpose π + π naught is greater than or equal to 1that essentially means that X I is in the right side of the hyper claim right remember that so y is + 1 then this is also positive right .

Then I mean I at least + 1 right so not only it should this be positive there has one and then you will get a 1 here and similarly if Y is -1 which is the other side of the hyper plane so this product this term has to be at least -1 right so that you will get a + one it will be greater than or equal to + one so you know that all the data points are correctly classified correct and minimizing π essentially it is the smallest possible π I so finding the smallest both π set data points are correctly classified and not only correctly classified they are at least certain distance away from the habit.

IIT Madras Production

Funded by
Department of Higher Education
Ministry of Human Resource Development
Government of India

www.nptel.ac.in

Copyrights Reserved