NPTEL

NPTEL ONLINE CERTIFICATION COURSE

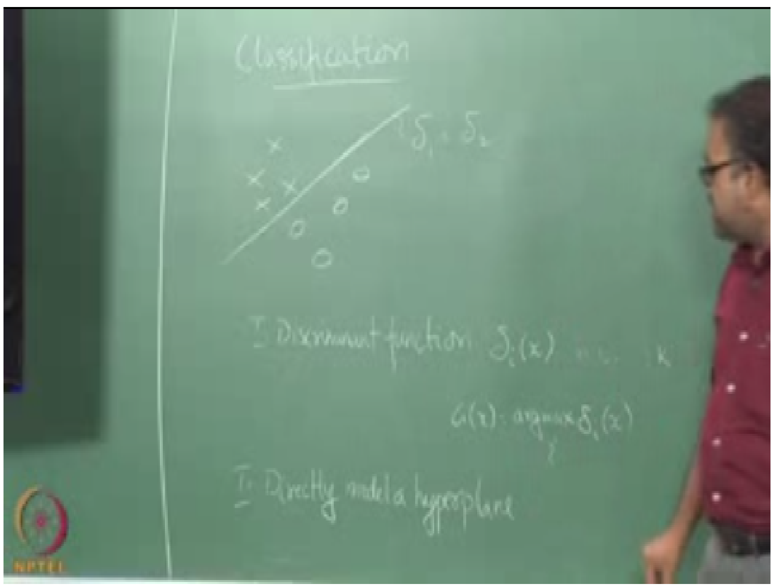
Introduction to Machine Learning

Lecture 19

Prof. Balaraman Ravibdran
Computer Science and Engineering
Indian institute of technology

Linear Classification

(Refer Slide Time: 00:16)



So we move on from linear regression linear methods for regression to linear methods for classification and right, so far we have been looking at linear methods for regression but I did tell you that you could do quote-unquote nonlinear regression also by doing appropriate basis transformations. So what do I mean by linear methods for classification linear regression you can understand right, so the response is going to be a linear function of the inputs right, so what do I mean by linear classification? So when I am going to separate the positive classes or when I am going to separate two classes the boundary of separation between the two classes will be linear.

So that is what I mean by linear classification, so this boundary that I draw between two classes will be linear, so you can think of when we did look at an example in the first class where we had drawn quadratics and phases and things like that right, so but instead of that we will assume that the classification surface or the separating hyper plane will be here well or the separating surface will be a hyper plane. So there are two classes of approaches that will look at for classification for linear classification and the first one is essentially on modeling a discriminate function.

So one rough way of thinking about it is to say that I am going to have a function for each class right and if the function for class I output for class I is higher than for all the other classes I will take classify the data point as belonging to class A right, so I am going to have some function so for each class I will have a function right and so depending on which is the highest I will output it to that, so this is essentially the idea behind discriminate functions all right so I am going to have to figure out a way to learn this left eyes okay so suppose let us just keep it simple think of a two class problem okay.

Here a question okay they think of a two class problem and I have δ 1 and Δ 2 rights so where will be by separating hyper plane? Wherever Δ 1 = Δ 2 right, so when Δ 1 is greater than Δ 2it is class 1 when Δ 2 is greater than Δ 1 it is class 2 like wherever they are equal it will be this a boundary right, so this will essentially be okay, so if I need this to be a linear surface right, so what conditions should Δ 1 Δ 2 satisfy should they be linear not necessarily but okay this is efficient condition if you are linear yeah the surface will be linear.

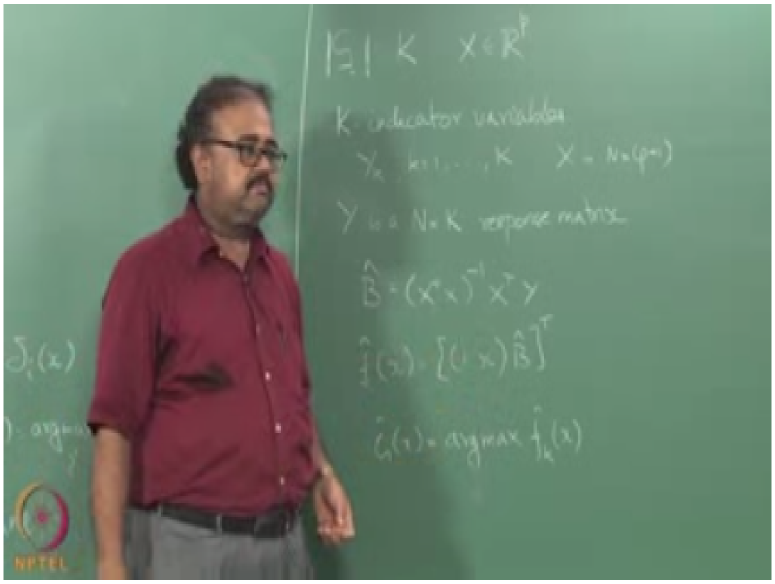
So what else can they be they can be non linear as long as I have some kind of a monotone transformation of them which will become linear okay, so we will see examples of this will actually look at discriminate functions okay or will yeah, so we look at the assumptions which will appear to be, we are doing something nonlinear heavily nonlinear but at the end of the day you will find that the surface will be linear okay the separating surface will be linear, so we look at that as we go along right.

So the few approaches that we look in this class or essentially linear regression you could do a linear regression and try to treat that as in India as your discriminate function it for each class you could do we talked about this in the very first class right or the 2nd class yeah, so where it could do a linear regression on an indicator variable, so that will give you a discriminate function or you could do logistic regression or it could do linear discriminate analysis which is like participant component regression but taking into account the class labels you will think of deriving directions and which will be doing the classification.

You look at the 3 of those the II class of methods which will come to this directly moderate hyper plane so it is related to this in some sense right, so if I give you the discriminate functions, so I can always recover the hyper plane but here instead of trying to do a class wise discriminate function will directly try to model a hyper plane okay, so this is II class of problems, so we look at one classic approach for doing that which is the perceptron right and we will also talk about some more recent well founded ways of doing that.

Which is essentially looking at the question of what an optimal hyper plane is right and trying to solve for it directly, so these are the two approaches we look at right, so this basically just setting things up, so people remember the basic set up for classification right.

(Refer Slide Time: 07:05)



So I am going to assume that I have some space G which has K classes, so I will first convenience an indexed M 1 to K next is going to come from R^p as before right and the output is going to come from this space G okay, so that is our setup and so if there are K classes I am going to have when have K indicator variables.

The way we talked about one of K encoding so one of these K indicator variables will be one for any input right depending on what class that data point belongs to like, when this is assuming that my x has so assuming I have augmented once so my β hat is equal to X^T X^{-1} X^T that is linear regression for you right so I can do just do linear regression on my response matrix, so β is capitalized here because it is also a matrix right, so one so this is capital β if you have one column for each of the classes right.

So each class I have a set of β so I can produce a vector of outputs here right given an input X by essentially taking the product with the β right that gives me a vector of outputs f and finally class label that is sent to the data point this arc max of they have f right, so I am going to get a vector of hips one for each class right and the one, that I assign finally is the one that gives me the maximum output okay, so I am not that any complex math here at all, so only bit of math here we already saw in the very first linear regression fitting okay yeah, because I wanted to add that I want to make it a P plus one thing right.

So x is the input the input data point I add a 1 to the front of it to for the bias okay, so what does it mean what does this fk f of x mean, so what do these if f of x mean know that is fine every sudden you need you need semantics you can associate with the f_k yeah, so if you think about it so whenever the input take as of some classes let us pick a particular class let us call it J right let us have a class or even make it more confident class 3 ok the input belongs to class 3 whenever, the input belongs to class 3 okay white 3 will be one where is the training data if you look at it, whenever the input belongs to class 3 right y_3 will be one.

So if you think about it if you look at the expected output that you should get for a particular x the expected output you should get for particularly x is okay how many what is the average number of times it is going to be one so I am going to see the x again and again and again right whenever that the x belongs to class 3 the output will be 1 and the x does not belong to class 3 the output will be 0 right I see many times I see x okay, so what is the output I expect it is the average of the outputs right, the prediction should be the average of the outputs does it make sense, so I have many x,x,x there are different x they are the same x ok many times I am getting x again and again so sometimes it is class 3 sometimes it is not sometimes it is not sometimes it is clustering okay.

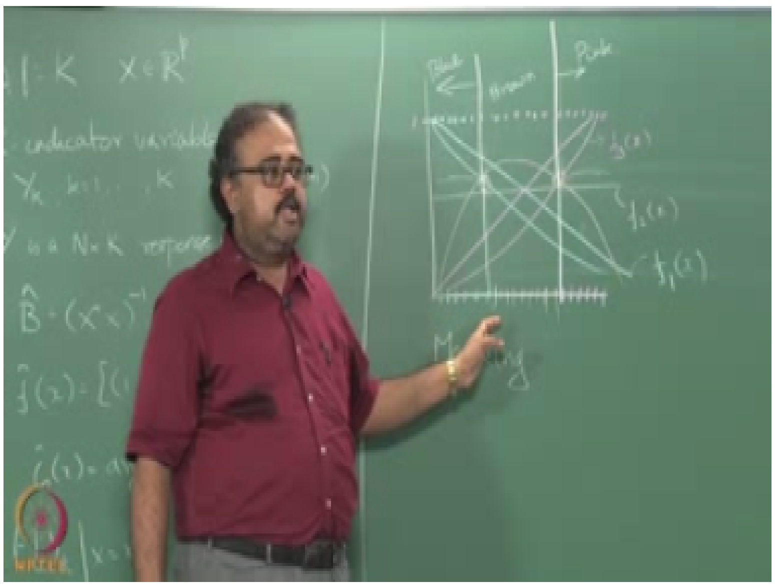
So if you take the average of all of this outputs what am I getting dealer probability that x is class 3 right if you take the average of those outputs I am getting the probability that x is class 3 right and we know that when I am trying to do the linear regression what I am trying to predict is the expected value right, ideally I should be trying to predict the expected value of this but since its linear you will not be able to get there but we are trying to do is probability of ah that is all problem of using linear regression and that is what I am coming to it right.

So you cannot really interpret these as probabilities because linear regression is not constrained right, so we will come to that in a minute, how to fix that in a minute, so but this is one I am working upto that it is telling you the interpretation of what you want to do is that it is a probability. So I really would like to interpret this right, so the expected value of y_k given x you would ideally like it to be probability that the output is K given the input is x right, so this is ideally this is what you want and the linear regression gives you hope of getting there right and people sometimes still use linear regression because it is easy to use.

You do and other things we would have to think of other ways of getting to it right I will come to that in a minute, so before that I just want to point out one other pitfalls of using linear regression for classification right but is it clear any questions this is same this is remember the indicator variable thing, so it is either 0 or 1okay what do you mean behold the linear regression will work right, so I can do linear regression, so I mean as just as a method of using it right you can see how it was going to work I am going to do linear regression give me a minute huh.

What it means is we would ideally like it to mean this it is not going to mean that okay, I will just give just hold or I will do this example and then you can come back and ask me this question okay.

(Refer Slide Time: 16:50)



So I am going to assume there is a single input direction, so let us say that okay so let us say that there are data points, here that belong to one class okay data points here that belong to another class right. So if you think about it let us say this is this is encoded by pink right so the training data right will look like this right and 0 elsewhere for pink training data for blue will look like this and 0 elsewhere.

So now if I try to fit a straight line to this so what do you think will happen so I will get a line that goes like that right I will probably get a line that goes like that right, so this is essentially what your outputs will look like, so directly trying to interpret this as probabilities is not a good idea obviously right but you can see that wherever this is greater than this okay that should probably belong to class blue, where ever it is pink is greater than blue it should belong to class pink right. So at least this much you can conclude from the output of the linear regression.

So that is essentially how you would interpret the output? So whenever one output is greater than the other or greater than all the others you will assume that it is the correct class directly interpreting that as probability it is a problem, so this is what you would like to do that is what I said right but you do not want to do this okay having said this let us see how visible this color is, suppose I have a 3 class okay they are sitting in the middle like this okay, so the outputs were this will be somewhere there right. Now if I try to fit a straight line for this what is going happen?

Hey remember the rest of the points are all sitting here right they are a bunch of 0 here a bunch of 0 here and a bunch of once there, so I try to do linear regression on this so I am going to get that line like that I know what is the problem with that blue and pink completely dominated, there is no apart part in the input space, where Brown no part of the input space where Brown actually dominates right, the output of Brown never dominates anywhere yeah. So this will be right, so this is essentially what your f_1 f_2 f_3 will be so it turns out that for class you will never output any input point.

As class okay, so this problem is called this problem is called masking okay so this is one thing which you have to be aware of well you are doing linear regression for making your predictions okay is there any way to get to over Basking anything close, so instead of looking at paths you just look, at higher order basis transformations right instead of regressing on x right that is what we did here right. So in some regressing on x if I regress on x squared I am going to get different outputs okay today okay good return but next time actual English not just coffee right.

So if I am going to do that essentially I am going to get curves that look like that, with interesting curve is this guy how is this brown curve and you will click okay, so these are the crossover points it is anything, that this side will be blue anything to that side will be pink and anything in between will be brown okay, you can see okay. So but remember the input space is just on this line okay, so here this is the output whatever is going up is the output, so the input is only on this line okay just a single dimensional input. So it is no region but say it is only a line segment here so in this part of the input space it will be blue this part it will be brought in this part it will be pink thank the almost ideal except there is a small here.

That is the just drawing it so you can you can choose appropriate data points is that you can actually with the with the quadratic transformation if you regulars on x squared you can recover the actual boundaries okay so the rule of thumb is if you have K classes in your input data you need at least K - 1 basis transformation so in fact with a lot of work you can show that even with x squared regression you will have masking if you have four classes so in four classes you have to regress on the cubic transformation okay so that you can still get away with it.

IIT Madras Production

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

www.nptel.ac.in