NPTEL

NPTEL ONLINE CERTIFICATION COURSE

Introduction to Machine Learning

Lecture 51

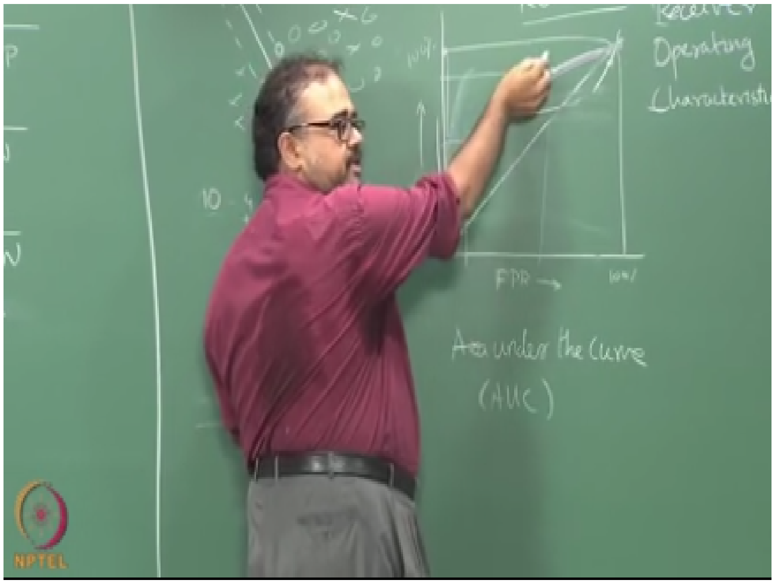
Prof. Balaraman Ravindran
Computer Science and Engineering
Indian Institute of Technology Madras

The ROC Curve

So when I build the classifier typically right, so what is going to happen if I'm whether I am building this hyper plane business or whether I am using a discriminant function right, there is an additional thing which I can use to tune how am I going to finally assign the labels right, so let us say I have a two class problem so then you can say that I learn a discriminant Δ and typically what we do if Δ is less than 0 we assign it to one class, Δ is greater than 0 we assign it to another class, correct.

Yes ,so what if I tell you that no, no if Δ is less than some θ you assign it to one class if it's greater than θ assign it to another class, right so what will this let us do let me move things around a little bit right, so essentially what will happen is I can by moving by changing this θ I can take some points which are originally classified as positive and make them negative right, so one way or the other rights i just keep sliding θ around, right, so essentially.

(Refer Slide Time: 01:38)



Right, so that is a line that marks where they are equal to 0 right, when I say that is greater than some θ so that essentially means I could either have a line this side right or I could have a line that side right, and as you keep moving this what happens when you move here so this will become 0 that will become class 4 right this is already class 4 now this will additionally become class 4 at likewise when I move it that side now this will become class x right, so I can actually change where I pitch my line right and I can get different performance right.

So what is important here is okay, I have figured out that right but then I can move this that way or that way is little bit I can change me what is it whatever I want to do right, so I can increase the precision of my classifier by just saying okay I have learnt the classifier right I have learnt whatever it is i have learnt the hyper plane but instead of looking at 0 right, I will just move it a little bit that same right when looking at the point where the probability is 0 I will move it a little bit that side or that this side so that I can change my classification that I give right.

So given that I can do something like this right, so how do I know i have got a good classification right, how to put it I put it in another way right, so I am asking you to give me a classifier right, but I am not going to tell you right what is the precision I need or what is a recall I need from this classifier right, so you give me a classifier and then i am going to figure out what is the θ I need to set so that i get a good precision or a good recall or a good the classification error whatever it is.

Next I want to be able to tune this and figure out where I am going to settle down right, so one way of summarizing all of this is something called the ROC curve right, so the x-axis is a true positive rate the y-axis is the false positive rate so what I mean by this so at any point I am going to look at how many true positives could I possibly get right, now that is my denominator how many of them have actually obtained that is my numerator, okay let us take a simple example.

Suppose I have, I say I have 10 data points okay I am going to make it very simple let us say I have 10 data points right and 4 are positive 6 are negative okay, 4 are positive 6 are negative right and I have a classifier okay, so of these 4 points it gives me 3 here and 1 here right, yeah, so I got it right okay, fine so it manages to tell me that 4 of the negative points are actually negative right, and 3 of the positive points are actually positive right and one of the positive points it classifies as negative and 2 of the negative points it classifies as positive right.

So the true positive rate for this is essentially 3/4 right, the false positive rate for this is false positive 2/6 there are totally 6 things which I can say as false positives right of which only 2 I get as false positive so I am not 2 bad right so that is the thing, so typically what you what you would want is maybe I flip this thing around right, is that TPR on this FPR and always get confused with this right, so when I make when I say something right I need to go up yeah, so that is TPR and this is FPR right, yeah so I ideally want a curve that goes like this okay.

So I should get a 100% true positive rate before I start saying anything is false positive right, does it make sense so here is 100% and here is 100% so at this point I am classifying everything as positive at this point I am classing classifying everything is positive right, so i will have a 100% true positive rate because i got everything is positive and I 100% false positive rate so everything that I could say is false, falsely positive I am saying is falsely positive.

So at this point I am classifying everything as positive right, but then what I would really like to see is as I when should my false positive rate start going up after I have achieved 100% in the true positive axis right, after I have classified every positive point as positive then if I still ask you to move your classifier more that side then I should start getting the negative points as positive right, so you should really go up all the way here before you start moving this way right.

And what about that guy this is essentially random behavior right, I mean so for every true positive guy I get a equal fraction of false guys as positive so essentially I am this flipping coins and telling you whether it is positive or negative right, so I am just flipping coins and telling you that this essentially gives me that line right, so the further up you are if the probability of you saying positive is higher.

I am tossing a coin like you give me a data set it will give me a data point I will toss a coin and will tell you whether it is positive or negative okay, if the probability of it coming a positive is slow then it will be somewhere here the probability of it coming a positive is high then I will be somewhere here right, makes sense right so that is this line this is bad right, you want to be above that line you never want to be below that line right this is essentially random that you want to be above that line.

So typical curves you find will be something like this okay, so obviously you will not get that right ideally would like to get some curves like this right, make sense questions so far so the steeper this rise is the better it is for you right, so sometimes what happens you get curves that rise very steeply and then do that then a good or a bad ROC curve. Yeah, nobody said depends yeah, of course we all know the right answer is depends.

Depends on how yeah, what type of performance I am looking at right, so when will this be good right, so when I want to achieve a middling true positive rate right, so this is about middling rates about 50% true positive see that means of all the people with dingo I want to capture at least 50% of them okay, without putting too many people on quarantine right, so this is a good classifier.

But then if I see if you really want to get 90% right then this becomes unacceptably high false positive rates, well this might not be that bad right want to get 90% so this might not be that bad a false positive rate right, but then at this point this classifier is better so if i want for 50% false positive rate this classifier is better for 90% false positive rate this classifier is better, so there is no sure fire way of saying this is better that is better without knowing what you want from it just because I drew a curve that went below the random line really does not mean that this classifier is uniformly bad, right.

So if you want to show true dominance you have to show that one curve is above the other throughout right, this white line truly dominates the pink line that then now in such cases I can say white is better than pink right, but not in these cases in this case this curve is actually better for some points some operating points it is not better for some other operating points, so nobody asked me what ROC stood for.

Yeah, receiver operating characteristics I think I wrote curve next to it ROC curve okay, so this essentially was used in olden days when people are talking about radars and things like that so false detection true detection versus false detection right, and then you choose your operating point right based on how much, how sensitive you wanted it to be do you want it to capture everything that came your way in which case it shows a different operating point for your detector right, so that is why these curves came about and you can use it for the same purpose in your machine learning evaluation right, you can use it to figure out which point in the space of parameters that you want your classified operatives okay, makes sense everyone get what ROC curves are all about okay.

So the thing with the ROC curves is unfortunately and people do not really use it this way when they have when they run experiments right, so what do they end up doing they do not want to look at the curve right they do not want to look at this curve and try to sit down and do an analysis and write papers because they want to run 100 of experiments they want to generate 100 H curves and they wanted an automatic way of comparing the curves,

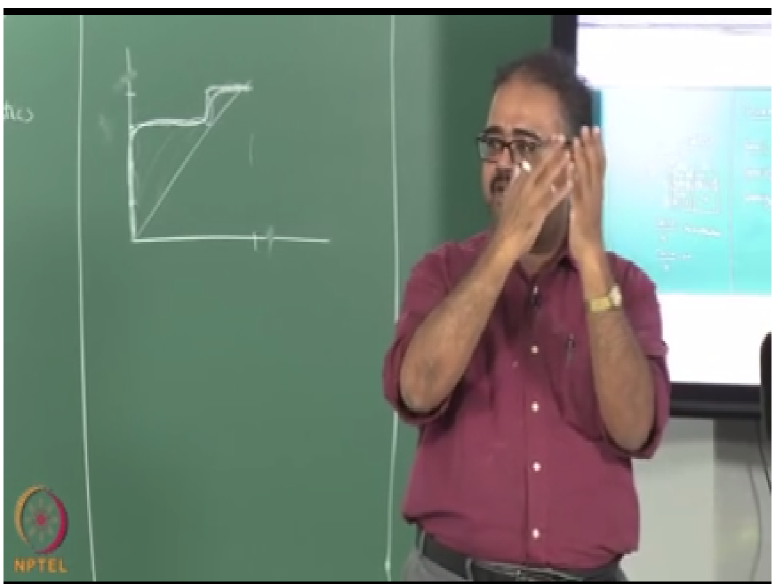
So what they did was they use this measure call, so area under the curve the uses measure called area under the curve or a AUC right, when they typically mean the area under the curve they do not mention it but they mean the ROC curve right, so the area under the curve is essentially this, so the assumption is the larger the area under the curve the better your classifier is because the ideal classifier is the one that goes like that right, so it gets all the positives before it gets a negative so the area under the curve is one for ideal classifier right.

And the area under the curve is 0.5 for random so if you are somewhere between 1 and 0.5 you are better than random so higher the value the better it is okay, all of this is fine provided all your curves are of similar shapes right, but you could get funky curves like this so what do you think went wrong here so the data is something like this right, so there are lots of 0 here and soon so forth and then are a few more x is that and then a few more 0 there right, so if I want to get those x as correct I have to get lot of the 0 as x’s before I get those x is correct so that is essentially what is happening here.

I am getting more and more of the 0 both as x is here and then suddenly i get those last two pieces last two x's and then I get them right, so that is essentially what this means so that your negative class okay, is lying between two bunches of positive data points so you are getting all the initial set of positive data points and then before you get the remaining positive data points you are getting a very, very large fraction of the negative data points and then you are getting the positive data points right.

So this could be an indication that your encoding feature encoding is wrong right or you need to go to a different dimension so you need to do an expansion of the dimension so that you can get all the positives before you get the negative ones, but people unfortunately do not actually look at the ROC in fact there is all this code bases for generating AUC directly you just done your experiments feed the data feed the classifier to your AUC generating package and then out comes the string of numbers nobody even plots the ROC and looks at it anymore right, so you can actually get insights about what is happening by looking at ROC okay.

(Refer Slide Time: 17:11)



So how would you actually go about plotting the ROC okay, so here is a very simple way of thinking about it right ,so I am going to take all these data points right and arrange them in descending order of their likelihood of being positive it could be anyway I choose to do the right so if I am going to slide this thing around okay, the farther I am from the hyper plane right, so the less likely I will be further I am on that side from the hyper plane the less likely I'm going to be positive right so I will start off from here and then I will slowly increases or if I am doing a neural network I can look at the probability of the classification right.

I can look at what is the probability this data point is going to be positive right, so like that so whatever is it I am going to arrange it in descending order let us say that I do not need to do the particular class so let us say that I am re arranging it like this right, so this is the true classes I have arranged it in this order okay, now I choose a threshold above which I will classify this as positive right, so let us say choose a threshold here right, so what do I get no, 1/4th right, so I will just go up by 1/4th here right.

Next what do i do next I move it down one more data point then what we get up I move up by another one for if I see another plus I move up with another 1/4th likewise next down I will move up to another 1/4th okay, I am assuming all of this is 1/4th so that means that will be 1 okay, now what do I see say negative point what do I do, I do not go down I go right I go right by 16, and I go right then another negative point I go right another negative point I go right, right and then Bob then right, right ,yeah.

So my ROC curve is actually like that because I have only ten points it does not look like a curve anymore right this is all the estimates I can do but this is my curve right, that make sense right and this is my random in fact I should cross the random curve at some point because 3/4th what is that 2/3 okay that is fine I will still be about random okay that is fine right, so that is my, that is how I draw the ROC curve right, fairly simple right.

So you arrange it in descending order of it being positive right, so whenever you see a positive data point as you go down that list you keep going up whenever you see a negative data point you keep going right the step you go up is one by the number of positive data points step you go right this one by the number of negative data points. Once you do this now you can compute this in the end of this curve fairly easily. So the probability with which my classifier thinks this point is positive right, or whatever measure weight whatever measure I am using so it could be the distance from the hyper plane whatever is the measure the closer it is to the hyper plane the less likely it is to be positive right.

I mean the further it is from a hyper plane the more likely it is positive, so whatever is the thing I arrange it according to the criteria that I am using right or if using a discriminant function the larger the value for the positive things then the higher up the ranking it will be right, so like that and I just do this in descending order of what is the probability i think it will be positive or what is the likelihood i think it will be positive right.

So this is roc curve, any questions on ROC yeah that is why i said you start of from the leftmost end and then you keep ranking them according to whatever you start of from the one end of it and then you keep ranking them according to that when there are other ways of actually once you train a classifier when you do the hyper plane thingy there other ways are figuring out what the probability should be right, so and from that you can you do not have to actually shift the hyper plane around so you figure out what the probability of the classification will be and then arrange them in the descending order of that.

Because you can always say that okay, get my SVM to run give me the output give me the distance from the hyper plane for all the data points I will convert that to a probability and then I will use the probability for making the prediction if it is at least point three probable that it should be positive and classified as a positive class and in decision trees again it is easy to do probabilities so how will you do probability in decision trees, look at the leaves the data point lies in and look at the fraction of data points belonging to one class right, then you can do probability in decision trees.

So likewise SVM is the only thing that is tricky but you can there is a way of converting the distance from hyper planes into probabilities, anything else I want to say here so I am not sure if I am going to get to this later so let me actually make the note now there is a another of supervised learning problem right we talked about regression, we talked about classification that is a another problem called ranking problem learning to rank right, so it is inspired by information retrieval right but it is used in a variety of other settings right so I am not interested only knowing it is positive or negative okay, that is not the problem I want to know a ranking okay, so where is this appropriate.

So suppose I am yeah, so this is something which people do when they are trying to match protein structures so I will have one structure for a protein right, so and I want to figure out all proteins that look similar to this right which will probably have the same functions that as the original protein that I'm looking at so that I am not interested in you telling me if there is similar or not similar I actually want you to rank it right, or there is another question.

So I want you to build a recommender system for me right, I want you to predict whether I like a movie or not like a movie right, but I do not want you to end up giving me like or not like I want you to give me a ranked list of movies so okay, this guy is coming in okay these are all his history last movies that he has seen okay, and he is an old guy so these are the set of movies I am going to recommend to him right, but in an order so this is what learning to rank means I am not just interested in yes or no answers but within the s i am interested in ranking order right.

It turns out that one of the ways of improving the performance in the ranking right, is not to look at this precision and recall and Felicity and sincerely in thing it is trying to directly optimize the AUC, trying to improve the things essentially what it means is the more relevant items you are trying to push up in the ranking right, so improving the AUC means what you are essentially making the curve go steeper and steeper right, improving the AUC means you have to make this curve rise steeper and steeper that is why this one will get a higher area under the curve right.

So that they essentially would correspond to making more and more of the positive points come higher up the ordering right, because this is a rank order right so you are trying to push these higher up their ordering so that essentially gives you the ranking effect.

IIT Madras Production

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

www.nptel.ac.in