NPTEL

NPTEL ONLINE CERTIFICATION COURSE

Introduction to Machine Learning

Lecture 50

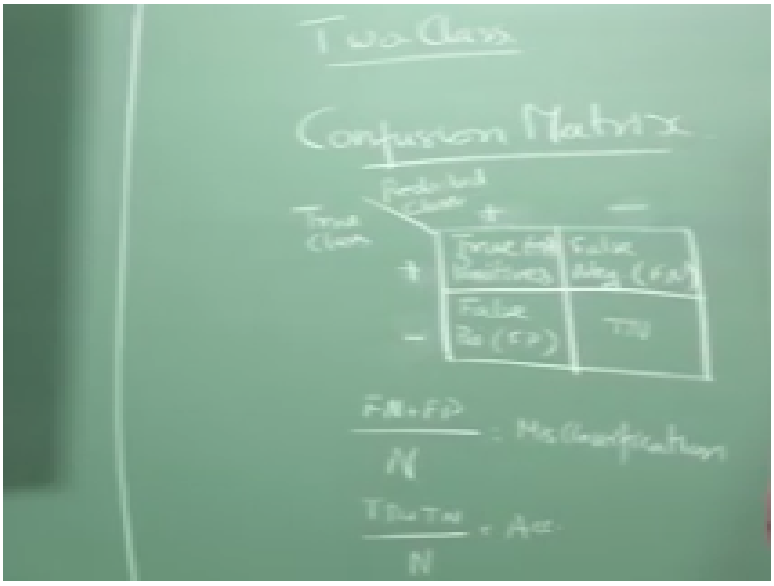
Prof. Balaraman Ravindran
Computer Science and Engineering
Indian Institute of Technology Madras

2 Class Evaluation Measures

Okay, good. So today we will look at some measures that are typically used in classification. And so, I will primarily focus on two class problems okay. The main reason is that more often than not the classification problems will encounter is for two classes, you know want to identify it as belonging to a class or not belonging to a class right so multiclass classification is little rarer to combine right.

And frankly two class classification has a lot more richer set of measures that people have proposed and usually the ones that we use for multiclass classification or extensions of these measures to multiclass okay.

(Refer Slide Time: 01:02)



So the first thing I want to introduce you to is something called the confusion matrix okay. So is nothing to do with the understanding of the course material or anything so far okay, it is something completely different right. So I say it is a true class problem, let us say the classes are 0 and 1 right. So I am going to say that, so I am going to form this matrix so the true classes are on the rows, the predicted classes are on the columns right.

I mean it does not matter if you do the transpose of this as long as you remember the meaning of the numbers that going here okay. Let us, I am going to change this slightly, so I am going to call class instead of calling them 0 and 1 as we have done so far okay. So I am going to call them positive and negative. So it makes it little easier for me, so what is the positive class, typically the class of interest to us right.

So what we will denote as positive class is the class of interest to us, what we denote as negative class is the class that we do not want right. So when I say that in some problem that the positive class is that person is suffering from dengue, it does not mean that dengue is something positive okay. It just means it is a class of interest to us, so just remember this okay. So when I say positive class it is a class of interest okay.

And more often than not your positive class in the population will be small right, hopefully right, I mean not too many of you have dengue right. So the positive class will be small and the negative class will be large right. So and we have to worry about getting the, what does it mean, of course yes. I mean that is my interest, so I am a doctor, I need patients to pay me right. So obviously the people were sick or more interesting to me then the people who are healthy right.

So that is the class of interest and the negative class is the other class, I am not so much interested in right. So things that go in here right, so with the true class is positive right and I am also predicting it to be positive okay, these things are called true positive okay, otherwise known as we will denote them as PP right, there are true positives okay. And what about these guys, they are true negatives right.

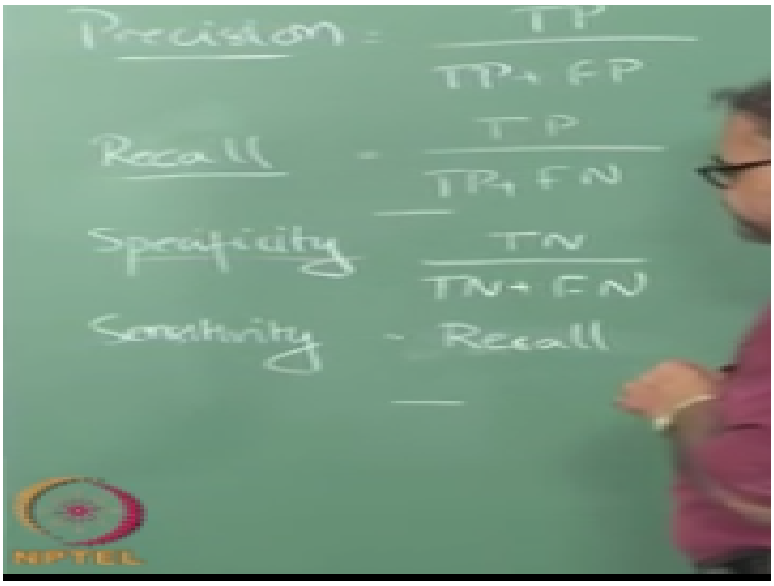
They are actually true class is negative and the predicted class is negative there are true negatives right. So that is why I said, you just had to remember which the true positive, which is the true negative and other things here. So whether you write it this way or whether you write the transpose it does not matter right, as long as you flip these three things, these two things around so what about these guys, false right.

So what about this, some of you know about true positive false for everything right, everyone has been telling me all of this, so where have we encountered this before, nowhere so glaringly obvious okay good, great.

So now what is the most common classification thing that we know about this classification error right so accuracy right so -1 the miss classification error is known as accuracy right what will be the miss classification error here where n is the total number of radar points, right this is what miss classification so what about accuracy.

Yeah so n is the total number of radar points okay it is not the total number of negative points in case somebody is worried about that but I need a symbol for total number of negative points also I will introduce something later right this is known as accuracy right as you can see is 1- which classification error right and what are the other things that we know of that are popular right so you can take a lot of different ratios of these numbers right and come up with different evaluation measures right.

(Refer Slide Time: 06:39)



So I will talk about a few popular once so anyone know what prison is true positive so what does it mean so the classifier is going to tell you so many data points are of the positive class right the classifier is telling you so many data points are of the positive class how many of them are really positive right so the denominator is also those points which the classifier is telling are positive, right so true positive + the false positive are all the points that the classifier tells you are positive and the true positive are the once that are truly positive right this ratio give me the what is called the precision right.

So precision can be defined per class if you will if you want to rights so here I am doing this only for a two class thing that is why we are talking about positives but suppose I had k class problem I can treat it like a 2 class problem and I can give you the precision for any class, so suppose class I am interested in some class k okay I just keep that as the positive class and everything else as a negative class right now I can create a true positive true negatives, so I can actually talk about precision for class k precision class I and so and so forth.

It can for a multi class problem I can talk about precision right but typically this defined for 2 class problem, so there is a completely measure for precision know as recall right so what is recall all is negative, so what is recall essentially there are so many positive points in the data right there are positive points in the data of these positive points what fraction is the classifier telling you are positive.

Right we get that all the positive points in the data which is true positive + false negative right how what fraction of it is the classifier telling you are positive so that is recall right so one way of thinking about it is so precision recall actually originate from information retrieval they are not originally for the classification domain they are originally proposed as a measures for evaluating information retrieval, so what do you mean by information retrieval so there is some repository of documents right.

Then you type in a search query right and then I give you back results corresponding to that search query right suppose there are 10 results that I give to you right of this 1 how many truly relevant to the query, what is that precisions suppose there are 50 documents in repository that are true relevant to the query how many of them appear in that 10 that is recall right so that is why it is called recall so how many of them to by actually recall from the repository so I have this huge repository of documents, how many of the truly relevant documents to by recall from the repository so that is why is call recall and procession you can see this is actually miss normal for people who are used to measurements right, what is precession in measurements sorry, how there is no not enough closeness in precession right.

No, elect guys come on how many elect guys are still in the class, 2,3,5,6,7,8 okay yeah, so what is precession sorry, how we elects valid precession if we go to measurements right, precession is essentially how many digits you are going to actually measure it to okay, that is precession you did not be accurate okay, the accuracy tells you how correct you are right, we can be less precise and very accurate and it can be very precise and very inaccurate I mean I can give you absolutely random number to 10 decimal places so I will be very precised, right.

But I cannot make any guarantees about the accuracy, but precession here is nothing to do with that okay, so precession here is essentially all to do with correctness, right so if I giving you 10 answers how many of them are correct. So why is this is a good measure, when this is a good measure already I told you an example, information retrieval right, so what characterizes information retrieval if you think about it, why cannot I use accuracy in information retrieval as a measure, why do I need something new, right.

Why cannot I use the following measure okay, I type in a query I give you back 10 results okay that means of the remaining documents okay, so I have rejected all of them right, so these 10 documents I have classified as being relevant to your query the remaining documents I have

classified as not relevant to your query, so among those I have classified as a relevant how many are truly relevant, among those I classified as not relevant, how many are truly not relevant and I can do an accuracy, right.

I can do miss classification error, so is that not a good measure for information retrieval. Yeah, so there will be like I said 50 documents at extra relevant to your query but you might have a 10 million document copy though if you Google you have several 100 billion documents as your copyrights right, and of which 10 or 15 might be relevant to your query right. So if I just use accuracy as a evaluation measure right, I need to be really, really precise in the measurement sense to make out any difference between two algorithms, because they mostly they will be correct.

And because of large fraction of the data I am going to say is irrelevant and I will be correct right, suppose I have two million documents and I am returning 10 to you right, and there are only 50 things that are relevant so basically I am right mostly right, a few things I miss here and there but I am correctly most of the time, because I have said large fraction of the irrelevant documents are truly irrelevant, right so that way I am good right, so that is not a good measure.

So when there is extreme class imbalance right, accuracy is not a good measure right, only 1% of your data is of positive class then if I say everything is negative class I am 99% accurate, but my precession will be what, you can define it to be 0 but in mathematically it will be undefined because I said everything is negative class I have no true positive, no false positive okay, so but you can define it to be 0 if you want and recall will also be 0 in that case, okay.

But quite often so what you will find is that if you try to increase you precession right, your recall will fall, when you try to increase your recall your precession will fall right, why is that so if you want to pull in more of the positive class right, if you try to pull in more of the positive if you want to predict suppose there are 40 documents that are relevant and I want to predict all 40 of them as a being relevant so the easiest way to ensure that is predict everything as being relevant, right so may recall will be very high.

Right which will be 100% right but the position will be too low depends on what is my universe of documents position will be too low so and if you want to have very high precision what we

have to do predict only sure documents select no documents that go back and define zero base zero as one instead of defining it has zero right.

So we will define zero and zero minute ago and define it has one right because if undefined you can do whatever you want with it right so select no documents is saying as 100% precise obviously because you cannot point out the mistake I have made in giving documents back to you so we will recall and we will suffer because recall will be zero right.

So there is always precision and recall right so we have to figure out where you want to pitch your algorithm right so typically people draw what are known has PR curves right precision and recall curves and how do you think this PR curves look like, like this, like this yeah like that here it really want to be here right.

You do not have high precision and high recall but then you can compare algorithm so you compares again this PR curves right for example again let us go on I will tell you little bit more as we get long right so there is another measure which is especially popular in medical literature call specificity so what is specificity something different from what we are seeing so far it is what does it mean yeah so what does it mean what does its schematic of this.

So schematics of this if I say that something is if I have a high sensitivity sorry if I have high specificity it means that if I say something is negative then it is for truly it is really negative right so why it is good thing to have exactly right so this is very useful in medical test so I run the test and I said that you do not have malaria right or well must be topical you do not have dengue right.

Then you really should not have dengue it should not be say okay he said he does not have dengue but then really has as a 50% of chance he actually has dengue even though the test says that you do not have dengue okay that is the bad thing to have right so in such cases specificity is very important so if you are building a classifier right that predicts whether a person is suffering from a particular disease or not.

Then it should have a high specificity okay the flips side of specificity is right is a terminology that comes from medical domain right so sensitivity is true positive by true positive plus false positive and specificity is too negative by true negative by false negative okay it looks like at the other things that you leave out here okay this is specificity this is sensitivity.

The two measures that you just like a precision and recall and information retrieval in medical literature you have sensitivity and specificity okay sensitivity is just like precision well specificity is the opposite of that right so sensitivity says okay how likely you have to sensitivity is recall I think sensitivity is recall sorry yeah sensitivity is recall so it just says how likely are you to diagnose the disease right.

If there are so many patients with this particular disease how likely I am to find a patient with the disease right so what fraction of the patient is fully discover if there are the disease so that is the sensitivity and specificity is essentially if at all you do not have disease how likely is that you do not have disease okay right.

So for regression we already looked at a classify measure so more or less is the same thing right so basically you look at squared error right so all the interesting things are with the classification right for regression we look at a squared error or you can do an absolute error also if you want to evaluate the how good your regression fit is you can do absolute error we can do squared error whatever if we can use so one more thing which I want to talk to you about.

IIT Madras Production

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

www.nptel.ac.in