**Learning Analytics Tool**

**Professor. Ramkumar Rajendran**

**Department of Educational Technology**

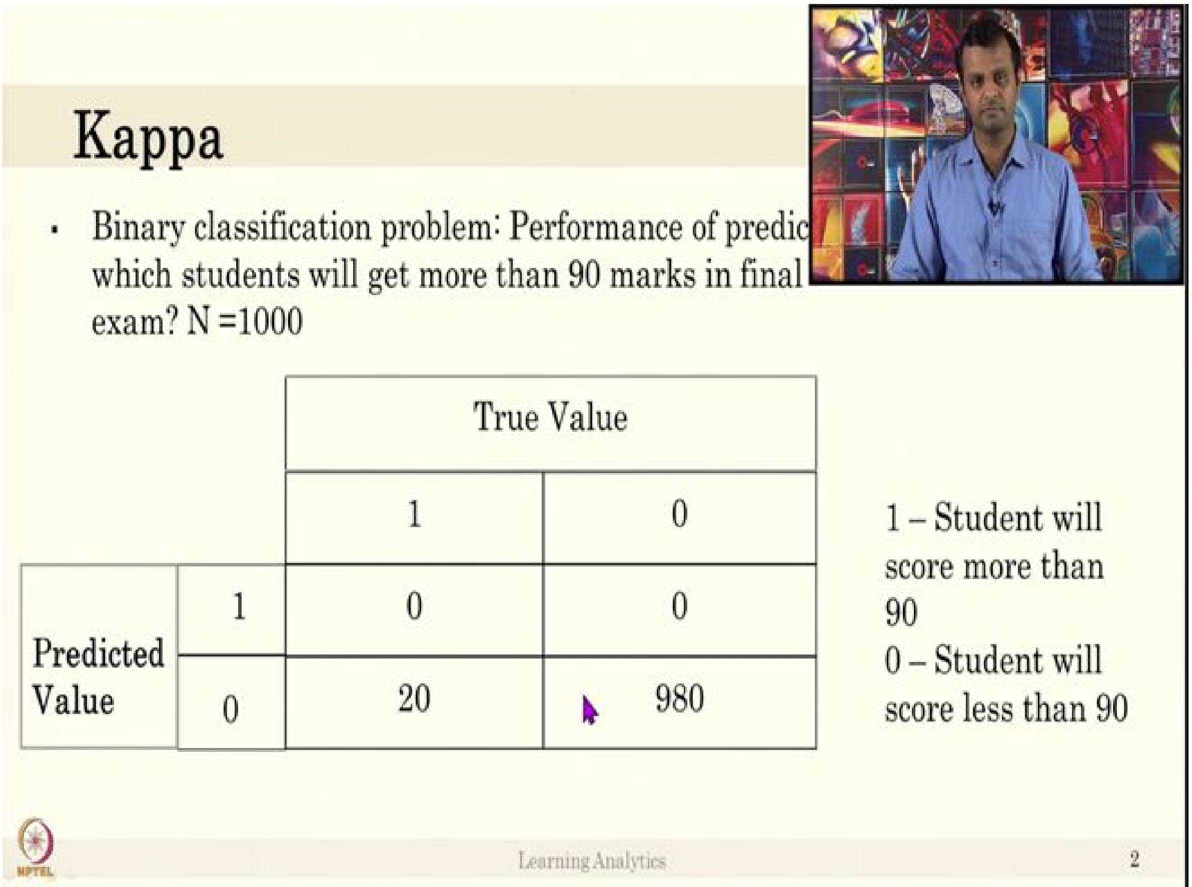
**Indian Institute of Technology, Bombay**

**Lecture No. 3.6**

# Performance Metrics - III

Welcome back to Learning Analytics Tool Course. Let us continue performance metrics in machine learning classifiers.

(Refer Slide Time: 00:28)



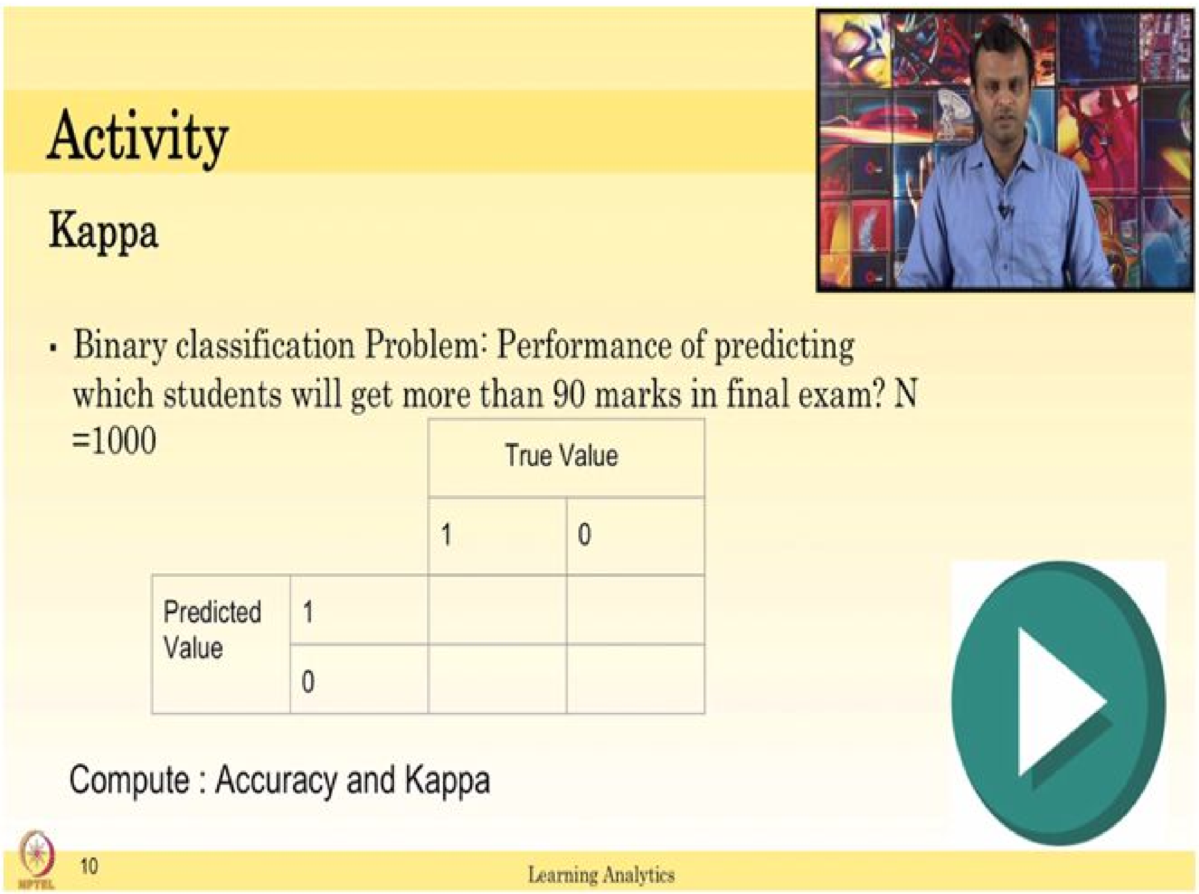

So, we saw that Kappa is used for evaluating performance i.e. to pick which classifier is better if you have imbalanced data set and. There is a binary classification problem we saw that, out of 1000 students, we want to predict how many students will get more than 90 marks in the final exam. Consider the table looks like this,

| | | True Value | |
|---|---|---|---|
| | | 1 | 0 |
| Predicted Value | 1 | 0 | 0 |
| | 0 | 20 | 980 |

that is 980 students got less than 90 marks and the machine/classifier predicts it. Also, the other 20 students who actually got more than 90 marks the classifier predicts them as students who got 90 marks or more than 90 marks.

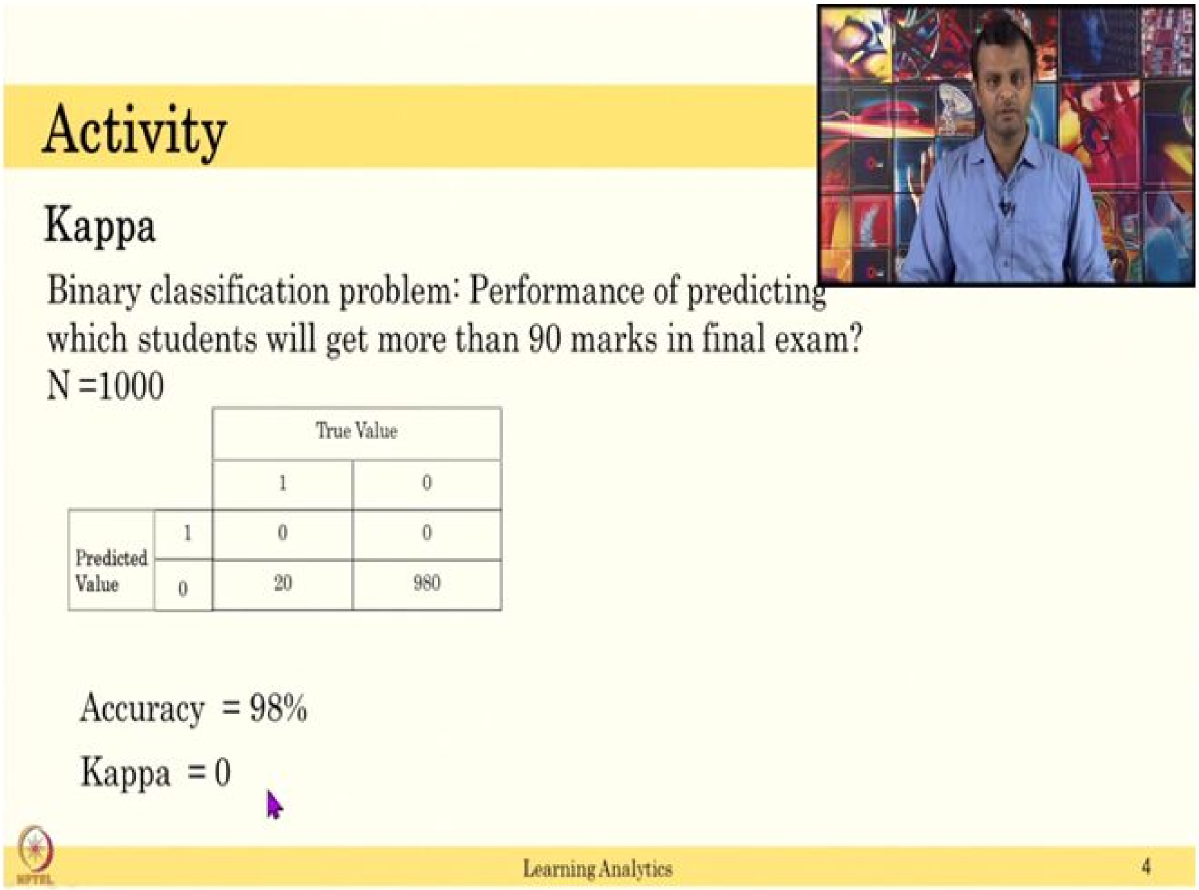
Consider this table what will be the Kappa score? And what will be the accuracy? The classifier did nothing i.e. the classifier did not create any rule, you might have given him 10 features does not matter, the classifier created a 0 rule classification problem. It simply classified everything into one class or it can be one simple rule. It classifies all the data into majority classes (0 here). So, the classifier did simple classification, classifying everything into one majority class.

(Refer Slide Time: 01:47)



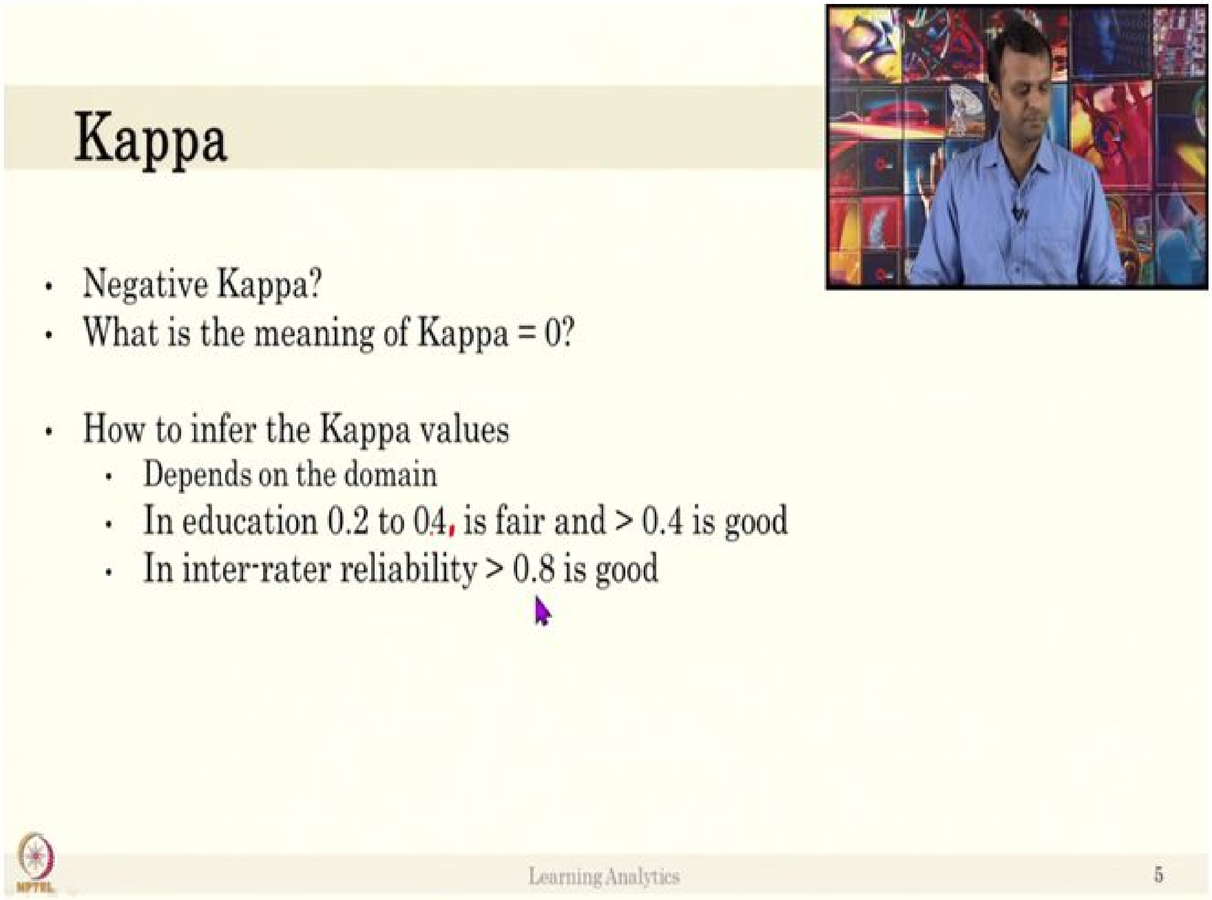

So in this problem, compute accuracy and Kappa and infer why the Accuracy values and Kappa Score like that? After writing down your answer, please resume continuing.

(Refer Slide Time: 02:08)




So binary classification problem, Let us look at accuracy and Kappa. Accuracy is 98 percentage because 980 students out of 1000. So 98 percentage, there is no doubt and Kappa is 0.

(Refer Slide Time: 02:22)



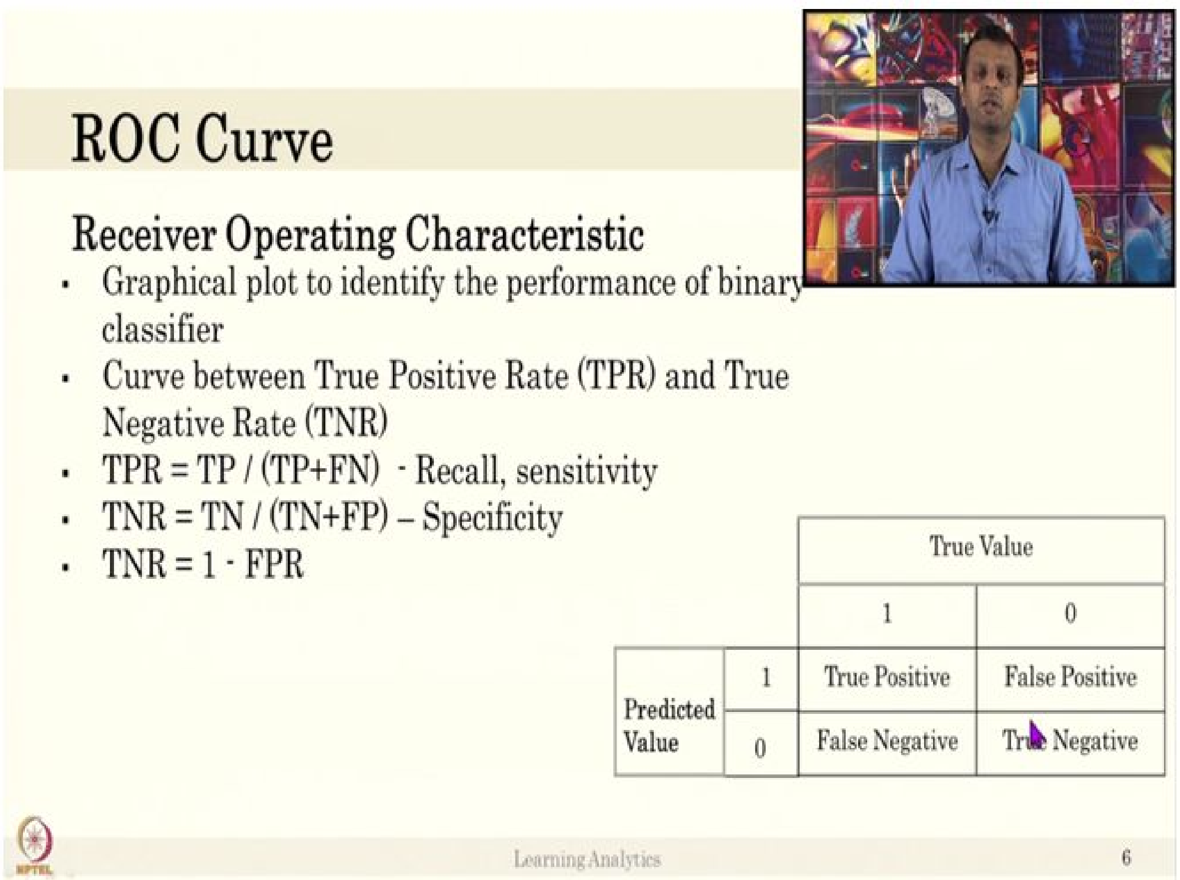

So let us see, can you have a negative Kappa? So, think of a problem and create your own confusion table and try to get a negative Kappa in it. If negative Kappa, what is the meaning of it? If it is negative, which means the classifier is doing very poorly compared to even by chance prediction. So, what is the meaning of Kappa as 0? It means the classifier is simply performing prediction equal to the chance level.

So, how to infer the Kappa values. So in the last slide, we motioned that Kappa equal to 0.4 means what? Whether 0.4 is good or bad? There is no definite answer for that I said that depends on the domain. For example, in the education domain, 0.2 to 0.4 is fair if you get a Kappa score, above 0.4 is considered to be good. And if you have multi-class problems, say you have to classify the students into multi-class like the students will get passed, the student will get more than 50 marks to 60 marks or 60 to 70 marks.

If a multi-class problem, the Kappa score of 0.4 is also considered to be good. The values 0.4 here it is not this. But for an Inter-rater reliability problem, like what we saw in our last class. We should have an agreement between two raters at least 0.8 are greater than 0.8. If you have an agreement say I want to 0.6, it is not good to use that inter-rater tool for the observation.

So, make sure that you have inter-rater reliability greater than 0.8. And if they are not achieving it. Train the researchers again discuss with the co-researchers and find out where is the mistake? Why mistakes have happened? Again, do the measuring (student's frustration or something like that). Then after measuring the student's affective states, again, compute the Kappa and make sure that you get better than 0.8. If you are not `having 0.8, you may not be able to publish your research in well-reputed journals.

(Refer Slide Time: 04:40)



So, let us move on to the other metric. In this class, we will see the two more metrics to pick the better ML classifier that is the one as Receiver Operating Characteristic Curve. This particular ROC Curve comes from signals or audio signals transmission receiving electronics and communication. So what is the Receiver Operating Characteristic Curve? It is actually a graphical plot to identify the performance of the binary classifier.

So, it is a graph between True positive rate and True negative rate. So, given this table True positive, False positives, remember this table two classes ago. So, True positive rate is

$$\frac{TP}{TP+FN}$$

It is also called Sensitivity. The true negative rate is

$$\frac{TN}{TN+FP}$$

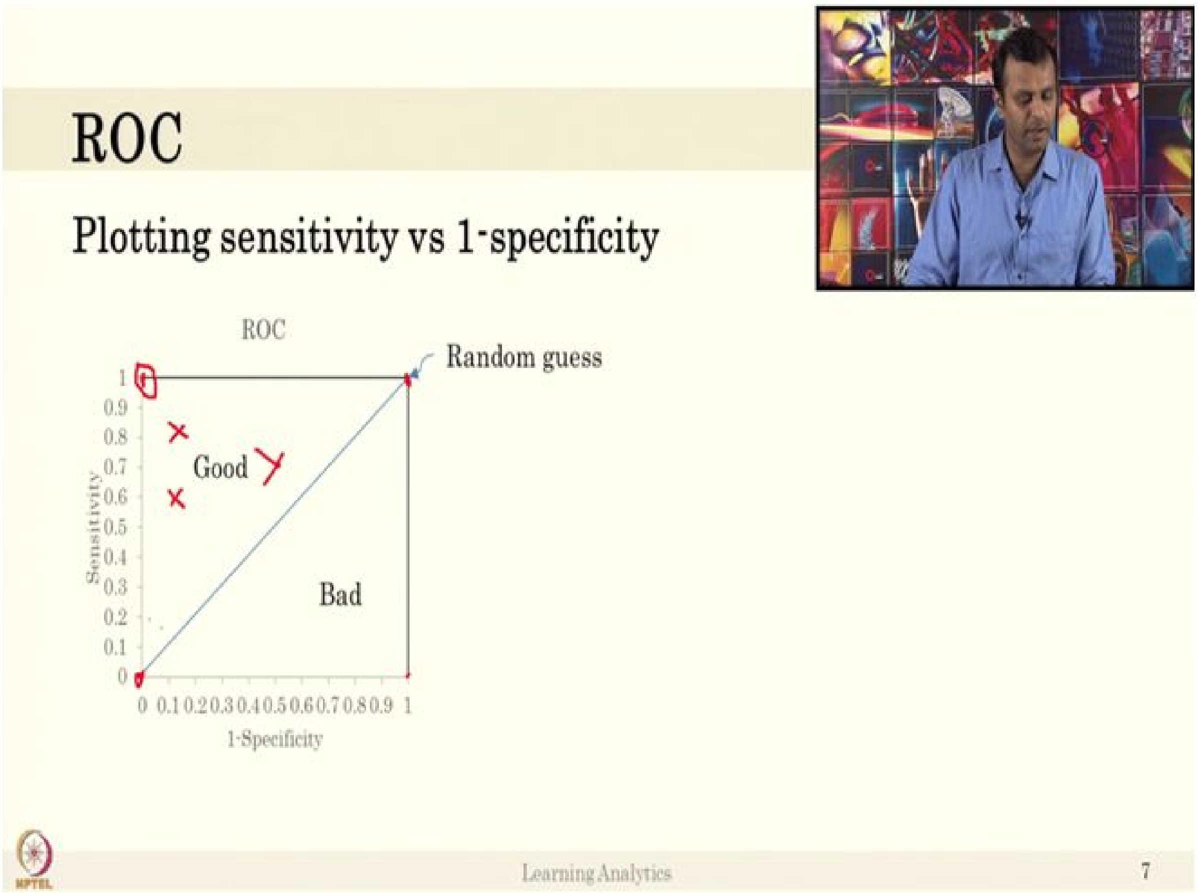
it is called specificity. And True negative rate is

$$= 1 - FPR$$

False-positive rate(FPR) is

$$= \frac{FP}{FP+TN}$$

(Refer Slide Time: 05:51)



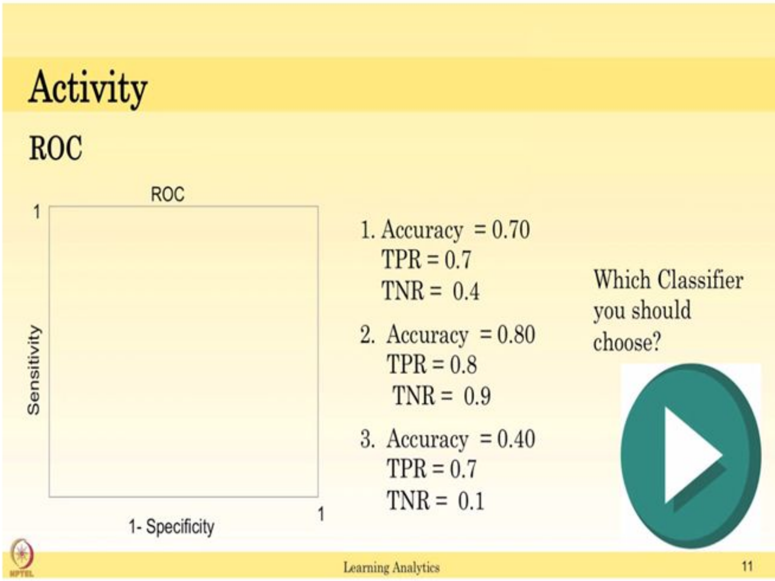

So, the receiver operating curve is plotting sensitivity versus (1-specificity). Let us consider this is the plot. And you have values 0 to 1 here and 0 to 1 here. So, let us see, if you have a classifier which gives very good, very good specificity, the classifier will be here and sensitivity that is recall rate is very low then this value will be here.

So, the values lie below this line which indicates the classifier is performing bad.

The values which lie above this considers the classifiers are performing good. For example, if you have a perfect classifier, which has high recall and say high specificity. So the value will be here, is the best classifier. This is the best, these two are just a random guess or average. Do not even consider pick the classifier in this line. Classifiers perform below this is the worst, do not even consider picking the classifier.

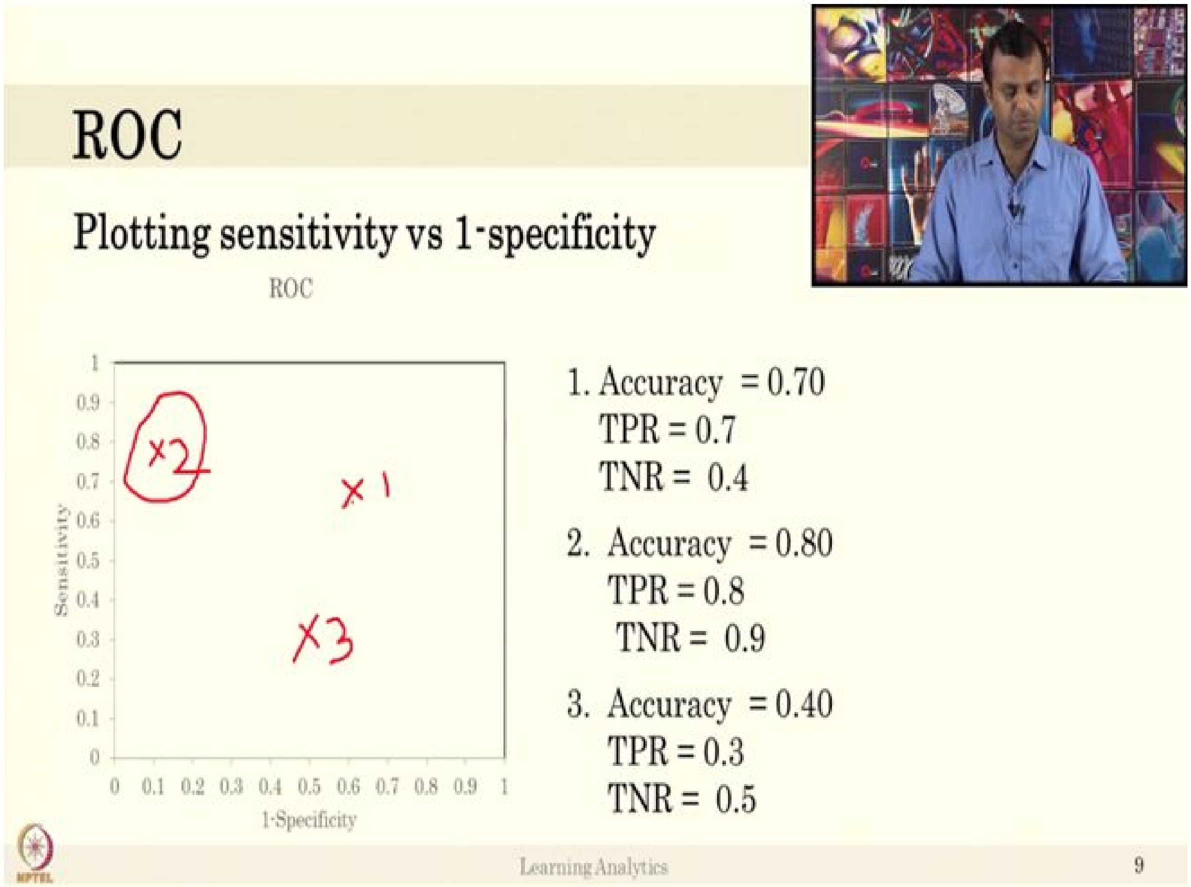
Let us do a small activity, then we will come back to this our ROC Curve in detail.

(Refer Slide Time: 08:14)




So, let us start a small activity to understand the ROC Curve better. So, you have a values Accuracy, TPR and TNR. You know that TNR is the specificity. So, we have 3 classifiers and the performance is given to you. And which classifier you have chosen? Use the ROC Curve to plot and pick the right classifier. After you pick the right classifier, you can resume the video to continue.

(Refer Slide Time: 08:48)




Let us see which classifier is doing good. You might have done it, I will just repeat it again. So we do not need accuracy for ROC at all. There's no point in giving that here, So TPR True positive rate that is sensitivity is 0.7, for classifier 1 TNR is 0.4 which means 1 -TNR is 0.6, So, it will be like here, So, X will be at (0.7,0.6), this is classifier 1.

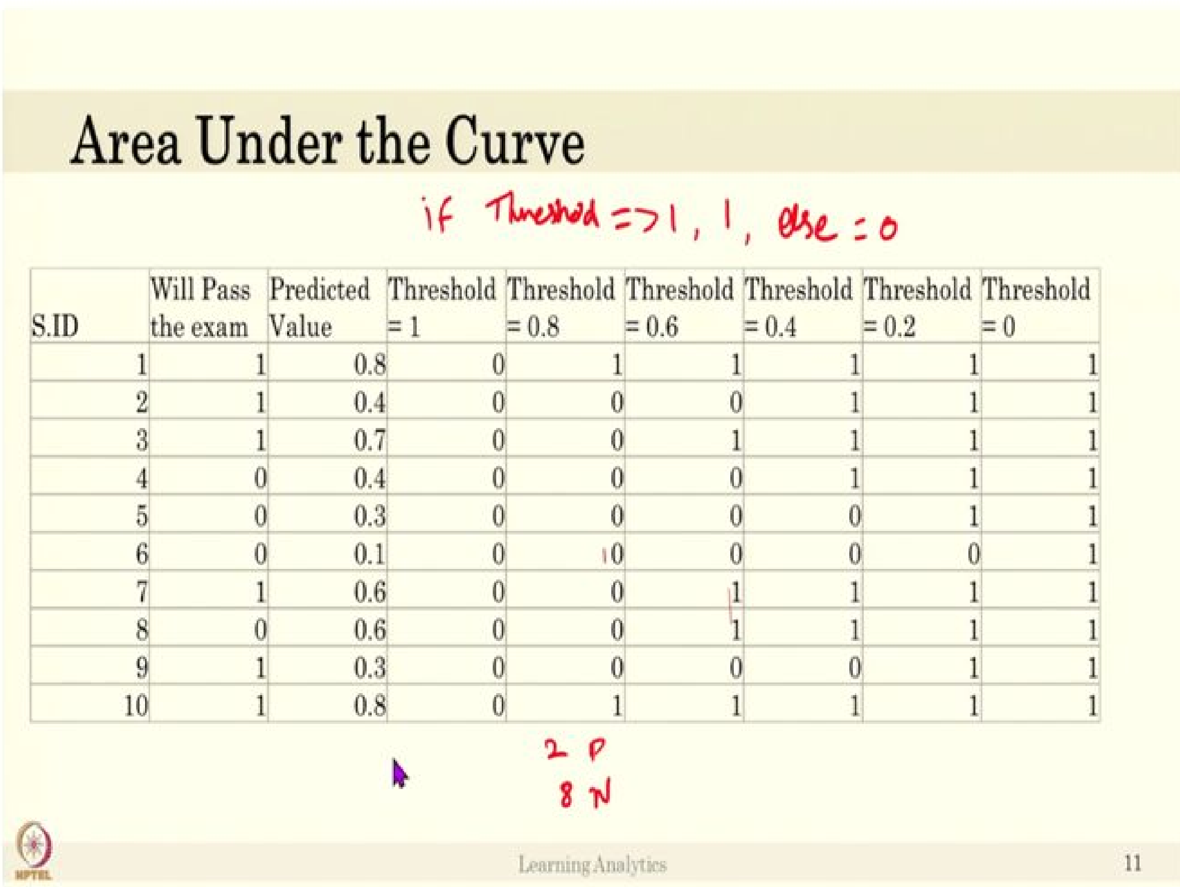
So, this is classifier 2 and TPR is 0.8 and this is 0.9, TPR equal to 0.8 and TNR is 0.9 and it will be like here at (0.8,0.1), so here this is classifier 2. And TPR is 0.3 and TNR is 0.5, if TNR is 0.5 then 1-specificity will be 1-0.5. So, we can see that the best classifier is perfectly the one which is here the classifier 2.

If you computed like that, it is good. So, you understood what is ROC Curve is. If not, please look at the slides and also check Wikipedia. What is ROC Curve means? It is simply the curve between sensitivity and 1-specificity to measure which classifier is doing good, which assumes both precision and recall also True positive rate and True negative rate.

(Refer Slide Time: 10:51)

So, we saw what is ROC. The other important metric in machine learning to pick the right classifier with the right threshold is called Area under the curve. Assume a binary classifier developed to classify whether your student will pass the exam or not. But the classifier response is not simple 0 or 1 instead it is skewed the probability value of being 1 or 0 is in the range of 0 to 1.

(Refer Slide Time: 11:20)



# Area Under the Curve

if Threshold =>1, 1, else =0

| S.ID | Will Pass the exam | Predicted Value | Threshold = 1 | Threshold = 0.8 | Threshold = 0.6 | Threshold = 0.4 | Threshold = 0.2 | Threshold = 0 |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 0.8 | 0 | 1 | 1 | 1 | 1 | 1 |
| 2 | 1 | 0.4 | 0 | 0 | 0 | 1 | 1 | 1 |
| 3 | 1 | 0.7 | 0 | 0 | 1 | 1 | 1 | 1 |
| 4 | 0 | 0.4 | 0 | 0 | 0 | 1 | 1 | 1 |
| 5 | 0 | 0.3 | 0 | 0 | 0 | 0 | 1 | 1 |
| 6 | 0 | 0.1 | 0 | 0 | 0 | 0 | 0 | 1 |
| 7 | 1 | 0.6 | 0 | 0 | 1 | 1 | 1 | 1 |
| 8 | 0 | 0.6 | 0 | 0 | 1 | 1 | 1 | 1 |
| 9 | 1 | 0.3 | 0 | 0 | 0 | 0 | 1 | 1 |
| 10 | 1 | 0.8 | 0 | 1 | 1 | 1 | 1 | 1 |

2 P
8 N



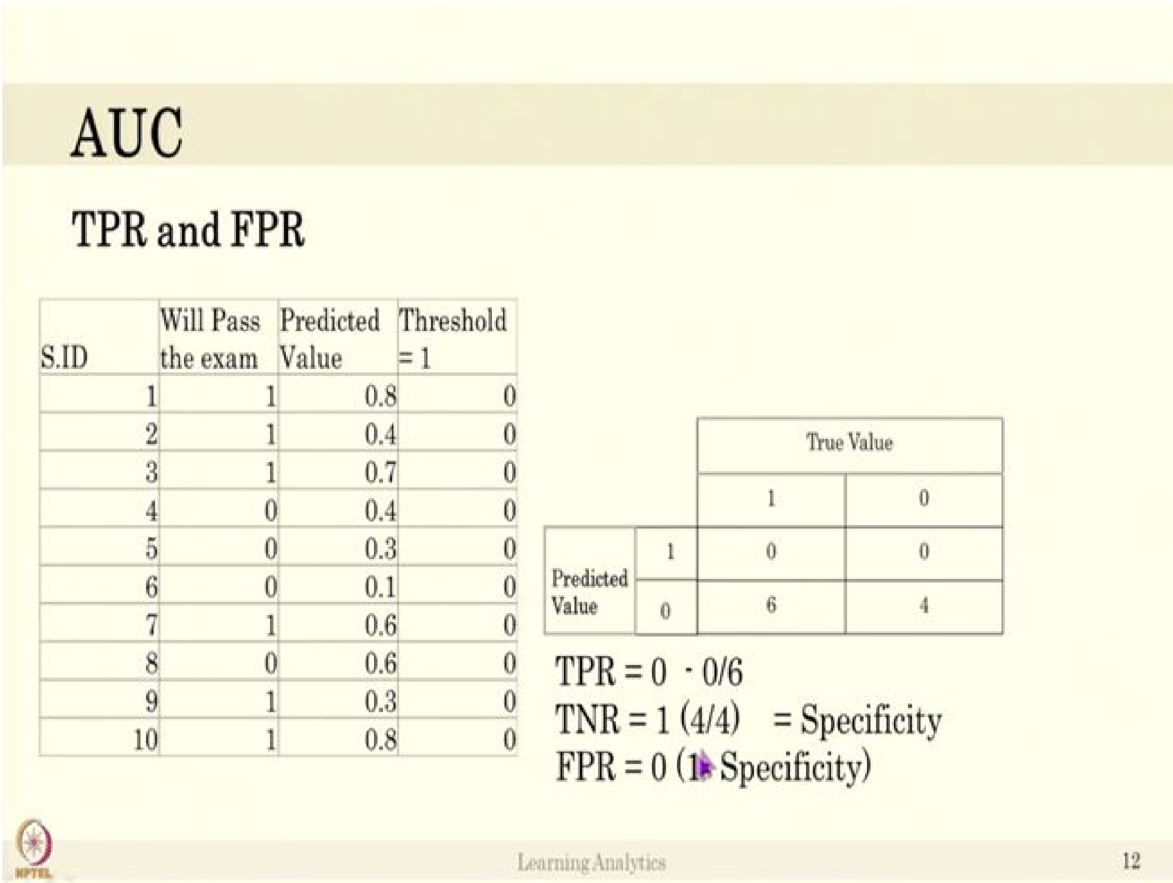
Consider this is the table. This is the True value. There are like 10 students, there are 3 students (1,2,3) will pass the exam. Next 3 will not pass, like that. This is a predicted value instead of the predicted value being 1, or 0, we saw that in a previous slide of previous classes. Here it will give that some probability say, 0.8, 0.4, 0.7 some values like this. So, you can apply a threshold on this particular, this particular column then you can create your own classes you can classify these students into class 1 or 0.

This is a testing dataset. Consider that. So, if I apply threshold is equal to 1 what happens none of them is equal to 1. So, we classify everything as 0. If the threshold equal to or greater than 1, you assume the value is equal to classes 1 else you assume the class equal to 0, if you have applied this threshold and you see this 0.8 is not equal to a greater than 1 which means, all of them are classified as a 0.

If we apply that threshold is 0.8, this 0.8, so, this classify student 1 and 10 as 1and everything else as 0. There are 2 positive classes and 8 negative classes here. 2 will pass, 8 will not pass. If you applied the threshold is 0.6, this classifies SID - 1,3,7,8,10 as pass and other students as "not pass".

Similarly we can do for threshold equal to 0.4, 0.2. If threshold equal to 0 everyone will be considered to be pass. So, let us see, you have the option to choose which threshold to select for your classifier. Let us assume you have the option to choose a threshold for your classifier based on the performance. So in this in this slide, we are thinking about the classifiers which give a probability value instead of the either 1 or 0 output. That will be most of the time like that, we will be dealing with in our classifiers.

(Refer Slide Time: 13:56)




So, let us assume only threshold equal to 1 and let us compute the precision and recall. So, as I discussed in the previous slide, we have 10 students and their value assume threshold equal to one, we will classify all the students as "will not pass". For example, as we assume everybody is not passed but 4 students are actually classified correctly (SID 4,5,6 and 8).

So, the True positive rate will be True positive rate is

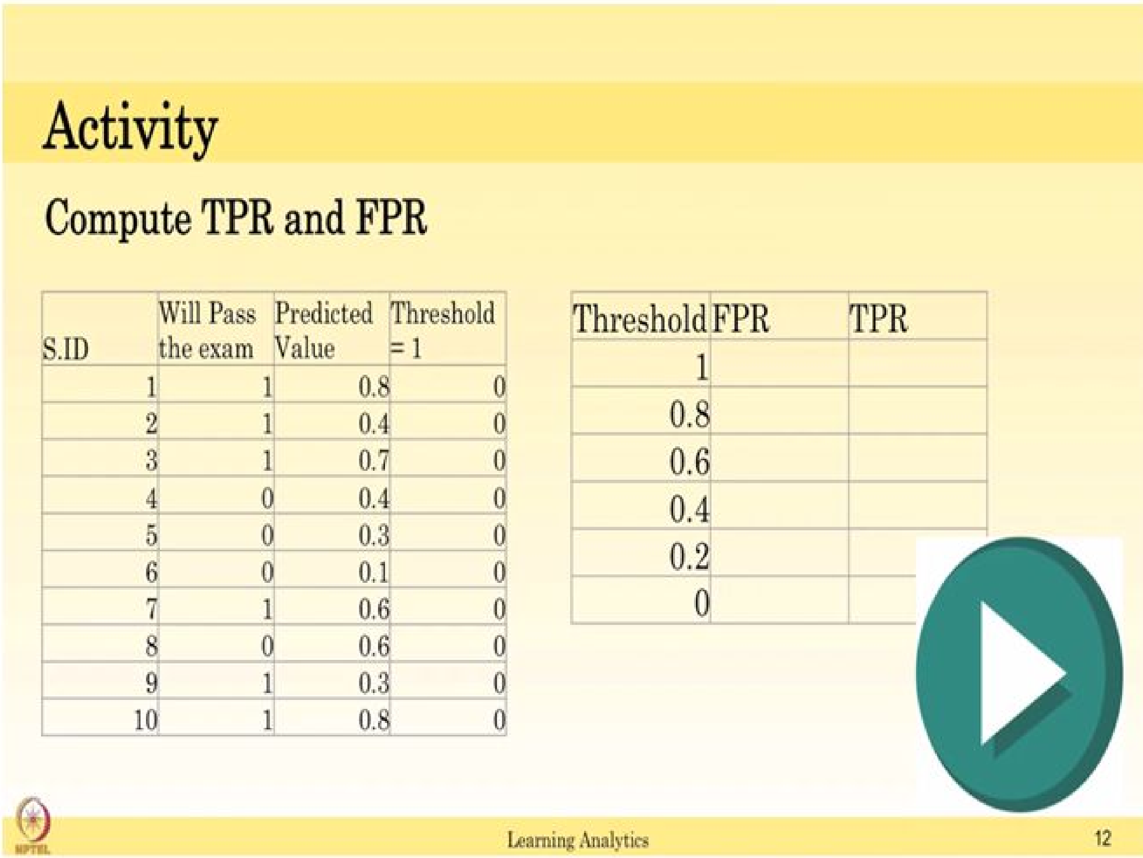
$= \frac{0}{6} = 0$

So, True negative rate is

$= \frac{4}{4} = 1$

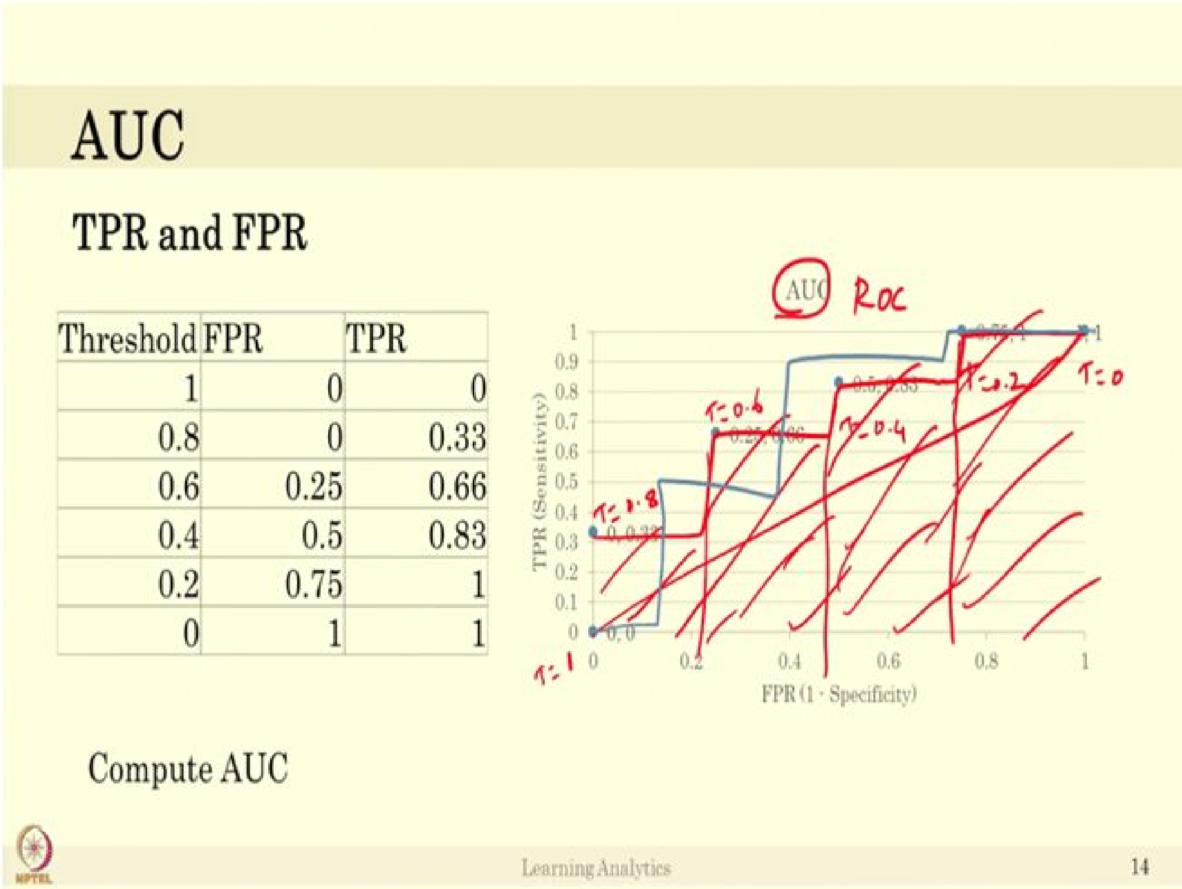
and that is called specificity and the False positive rate is 1-specificity that is what you want to plot in the graph.

(Refer Slide Time: 14:49)




So, let us do a small activity. I want you to stop the video and really go and compute False positive rate and True positive rate for all threshold values. Use the table given in a previous slide, go and compute threshold for each and every threshold False positive, True positive. Please compute so that if you have any mistakes which can be corrected and if I made any mistakes you can inform me in the forums. After writing down all the False positive and True positive please resume to continue.

(Refer Slide Time: 15:24)



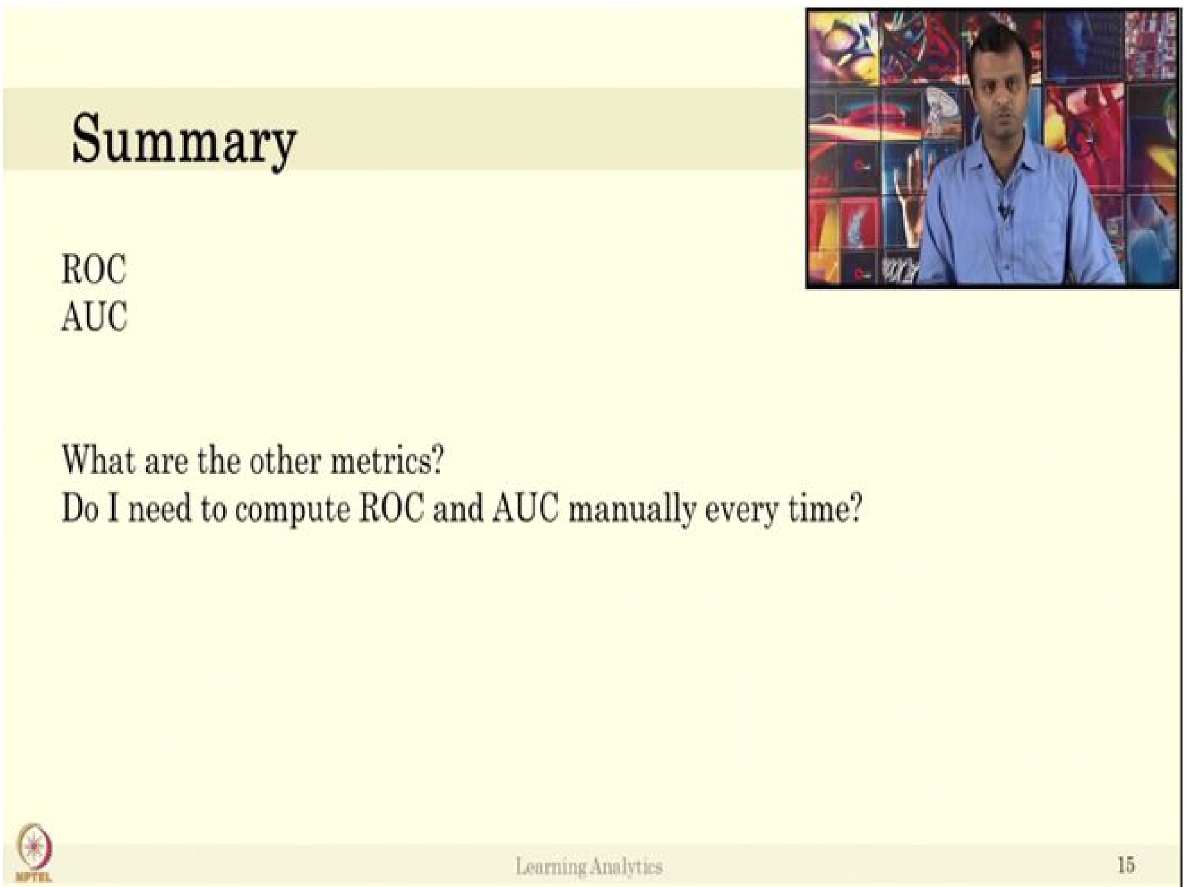

So, here is the False positive rate and True positive rates for all the threshold values and we also plotted them in the curve, in the AUC Curve. So, In Area Under Characteristic curve that is similar to ROC. So, Receiver Operator Characteristic curve we are going to compute AUC in the ROC it is not AUC.

So, in our ROC curve, we have plotted FPR versus TPR, FPR is 1-specificity. So, for a threshold equal to 1,the value is (0,0). For threshold equal to 0.8 the False positive rate is 0 but True positive rate is 0.33. So the point will be at (0,0.33). For threshold equal to 0.6 False positive rate is 0.25 and 0.66.

By you now know that if the value lies in this line that is, this line, it is kind of not good random goes something like that. And anything above is good. So, this classifier is doing good. Let us draw the curve. The curve is not exactly the curve instead is a step curve. The area under curve is you have to compute the area under all of this, that is all this area should be computed. So, I am not going to compute the areas here, but you have to compute all the areas if you want.

So this is one classifier, and based on the different threshold, we can pick which threshold is good. There might be another classifier, say, another classifier which might have different values that might give a different area under curve. So based on the area under curve, we can pick the better classifier. More importantly, it is not about picking the better classifier it is important to pick which threshold value you need for your classifier.

(Refer Slide Time: 17:55)




So in this video, we saw what is the Receiver Operating Characteristic Curve and also how to compute Area Under Curve and to pick right classifier also to pick the right threshold. So, is there any other metrics in machine learning to pick the right classifier? Yes, a lot of other metrics are like A’. So, but we will stop here. We know these 3-4 metrics. First to pick the right classifier for the binary classification problems.

So, do you need to compute ROC, AUC manually every time? I said no, simply use tools or the libraries available in each scripting language to compute it. So, the idea here is that you have to understand what is ROC and what is Area Under Curve. How it is computed so that when someone shows you the ROC curve or Area Under Curve, you should know, this means this, I should pick the right classifier.

The issue operating curve shows me there is more area it, it covered. So, I might pick this one. So, the best classifier always, will look like, completes the Area Under Curve will be one like it completes all the Area but it is highly unlikely that we see that kind of classifiers. Thank you.