

Confusion Metrics : Used to describe the performance of a classification model on a set of test data.

TP : The value of actual class is yes and the value of predicted class is also yes.

FP: When actual class is no and predicted class is yes

TN: The value of actual class is no and value of predicted class is also no

FN: When actual class is yes but predicted class is no

Accuracy:

(correctly predicted observation) / (total observation) = (TP + TN)/(TP + TN + FP + FN)

Precision :

(correctly predicted Positive) / (total predicted Positive) = TP / TP + FP

Recall :

(correctly predicted Positive) / (total correct Positive observation) = TP / TP + FN

F1 Score :

Harmonic mean of recall and precision

$= 2 * (\text{recall} * \text{precision}) / (\text{recall} + \text{precision})$

What is precision and recall ?

Recall is the % of correct items retrieved.

Precision is the % of retrieved items that are correct.

- Recall: correctly predicted 1's / all 1's = $TP / (TP + FN)$
- Precision: correctly predicted 1's / all predicted 1's = $TP / (TP + FP)$

Suppose in a village of 100 people, 30 people got infested with a rare virus, which starts showing visible signs in a month. You are a clinician and you decided to use a detector to identify infested people before it's too late . Your device detected that 40 out of 100 people were infested with virus (Detected Positive). Your findings now mapped with actual cases as follows:

	Detected NEGATIVE Cases	Detected POSITIVE Cases	
Actual NEGATIVE Cases	50	20	
Actual POSITIVE Cases	10	20	= 30
		= 40	

Precision and Recall of your detector can be understood by answering following questions.

Recall: What percentage of people actually infested with virus were detected correctly using your device? = $20/30$ or 66.67%

Precision: What percentage of people detected positive using your device were actually infested with virus? = $20/40$ or 50.00%

In Machine Learning, performance measurement is an essential task. So when it comes to a classification problem, we can count on an AUC - ROC Curve. When we need to check or visualize the performance of the multi-class classification problem, we use the AUC (**Area Under The Curve**) ROC (**Receiver Operating Characteristics**) curve. It is

one of the most important evaluation metrics for checking any classification model's performance. It is also written as AUROC (**Area Under the Receiver Operating Characteristics**)

ROC (*Receiver Operating Characteristic*) Curve tells us about how good the model can distinguish between two things (*e.g If a patient has a disease or no*).

The AUC is the area under the ROC curve. This score gives us a good idea of how well the model performs. If AUC score of 0.5, means that the model is performing poorly and its predictions are almost random.

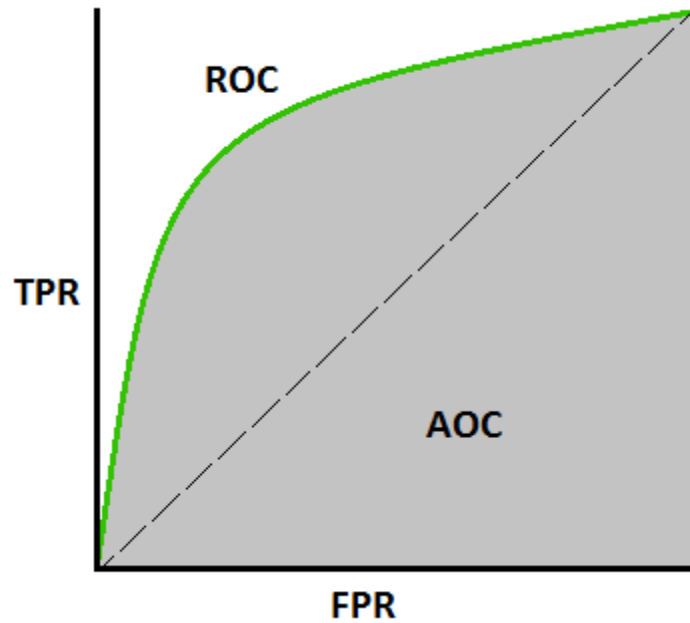
ROC (*Receiver Operating Characteristic*) Curve tells us about how good the model can distinguish between two things (*e.g If a patient has a disease or no*).

The AUC is the area under the ROC curve. This score gives us a good idea of how well the model performs. If AUC score of 0.5, means that the model is performing poorly and its predictions are almost random.

What is the AUC - ROC Curve?

AUC - ROC curve is a performance measurement for the classification problems at various threshold settings. ROC is a probability curve and AUC represents the degree or measure of separability. It tells how much the model is capable of distinguishing between classes. Higher the AUC, the better the model is at predicting 0 classes as 0 and 1 classes as 1. By analogy, the Higher the AUC, the better the model is at distinguishing between patients with the disease and no disease.

The ROC curve is plotted with TPR against the FPR where TPR is on the y-axis and FPR is on the x-axis.



AUC - ROC Curve [Image 2] (Image courtesy: My Photoshopped Collection)

Defining terms used in AUC and ROC Curve.

TPR (True Positive Rate) / Recall /Sensitivity

$$\text{TPR / Recall / Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

Image 3

Specificity

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

Image 4

FPR

$$\text{FPR} = 1 - \text{Specificity}$$

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

Image 5

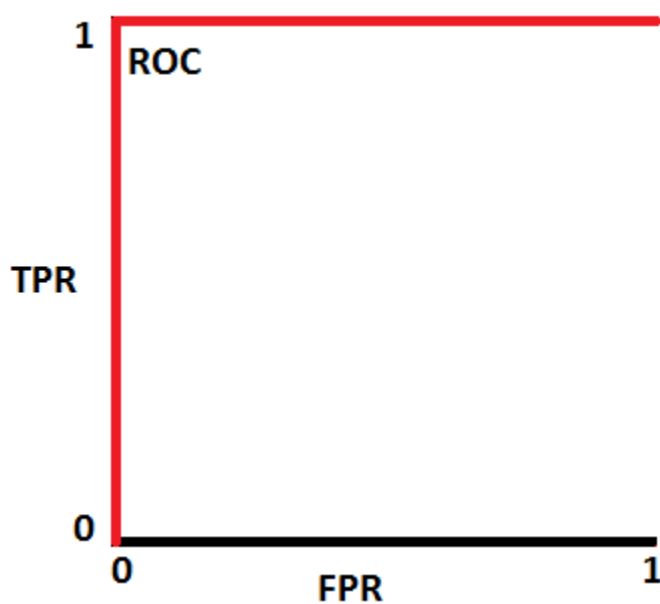
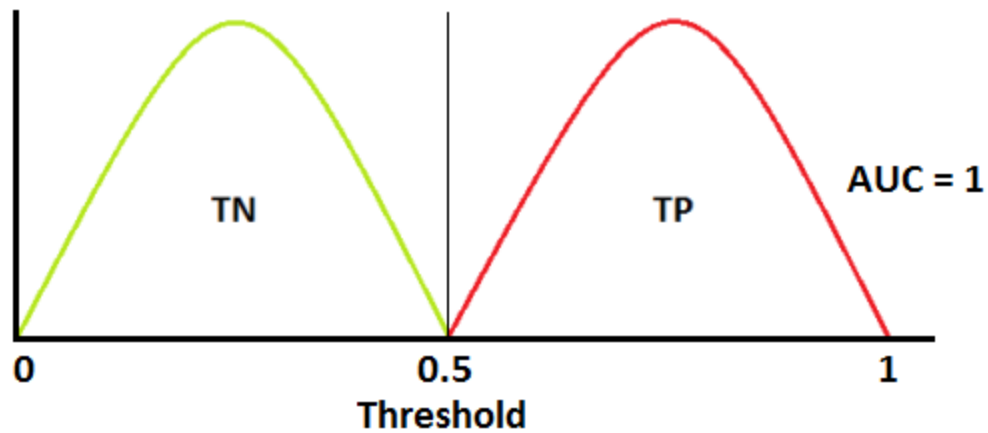
How to speculate about the performance of the model?

An excellent model has AUC near to the 1 which means it has a good measure of separability. A poor model has an AUC near 0 which means it has the worst measure of separability. In fact, it means it is reciprocating the result. It is predicting 0s as 1s and 1s as 0s. And when AUC is 0.5, it means the model has no class separation capacity whatsoever.

Let's interpret the above statements.

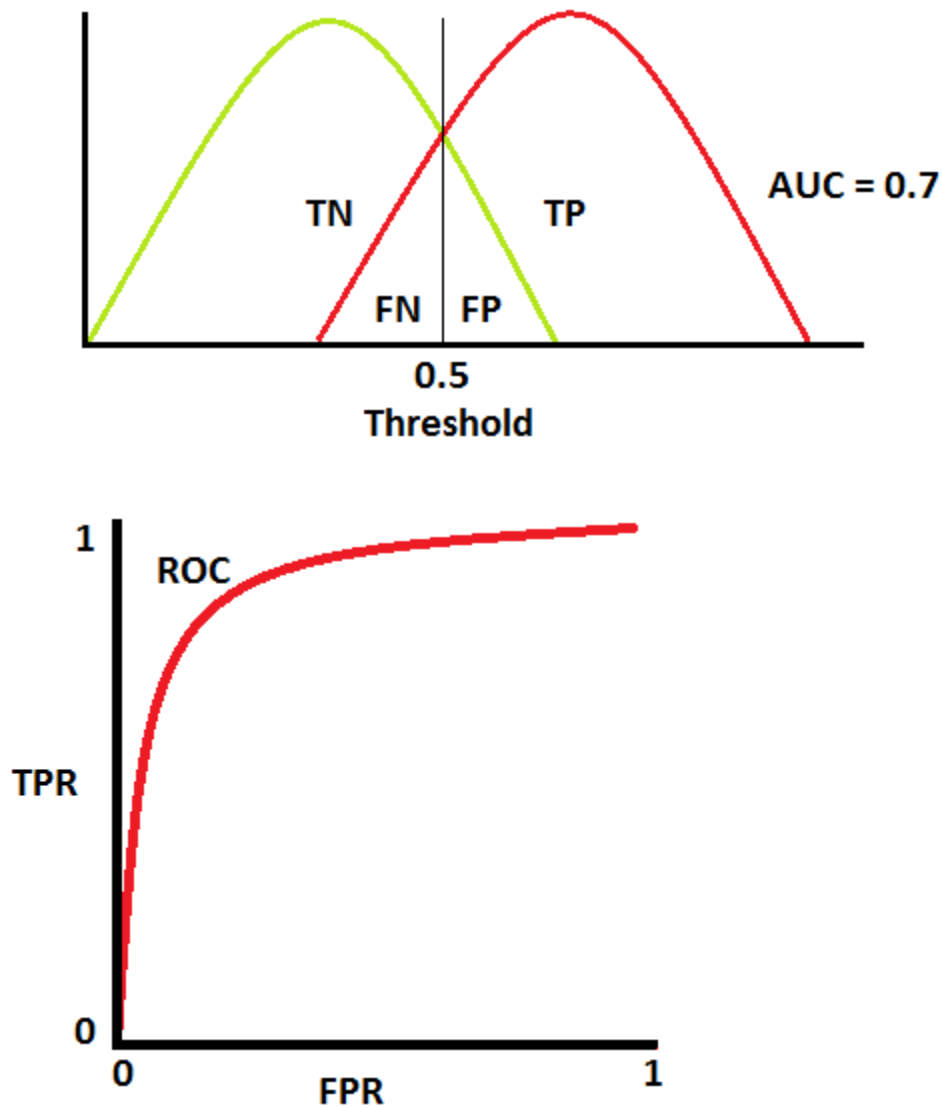
As we know, ROC is a curve of probability. So let's plot the distributions of those probabilities:

Note: Red distribution curve is of the positive class (patients with disease) and the green distribution curve is of the negative class (patients with no disease).



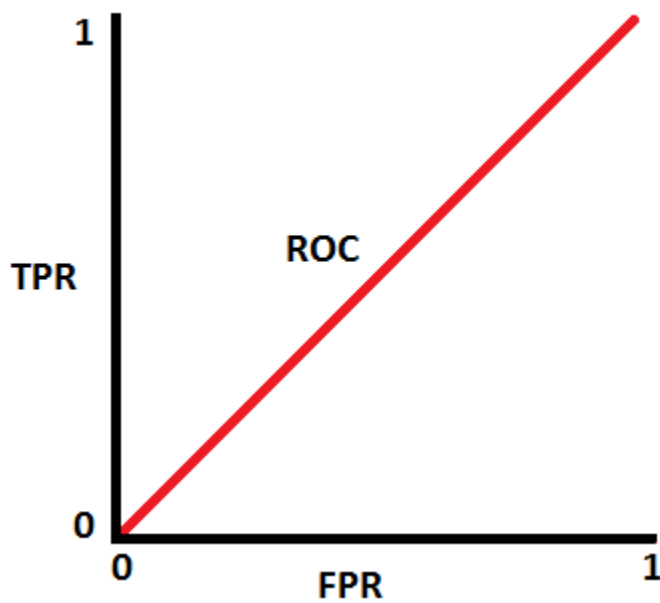
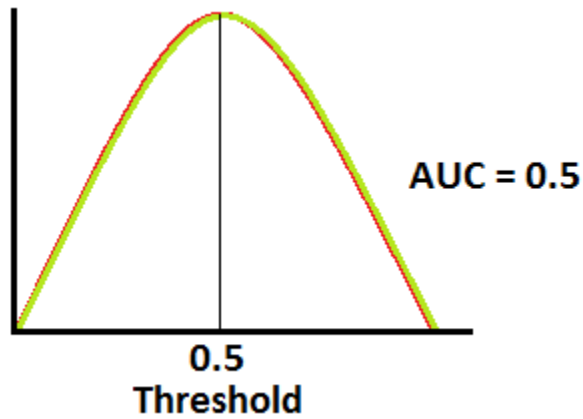
[Image 6 and 7] (Image courtesy: My Photoshopped Collection)

This is an ideal situation. When two curves don't overlap at all means model has an ideal measure of separability. It is perfectly able to distinguish between positive class and negative class.



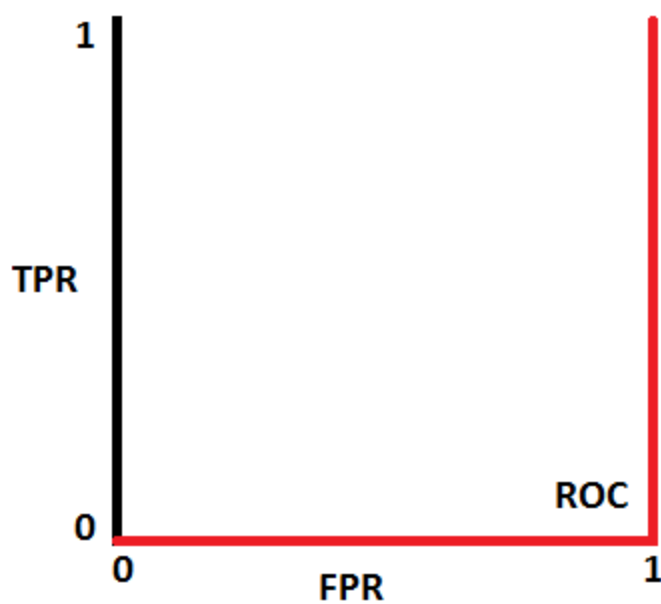
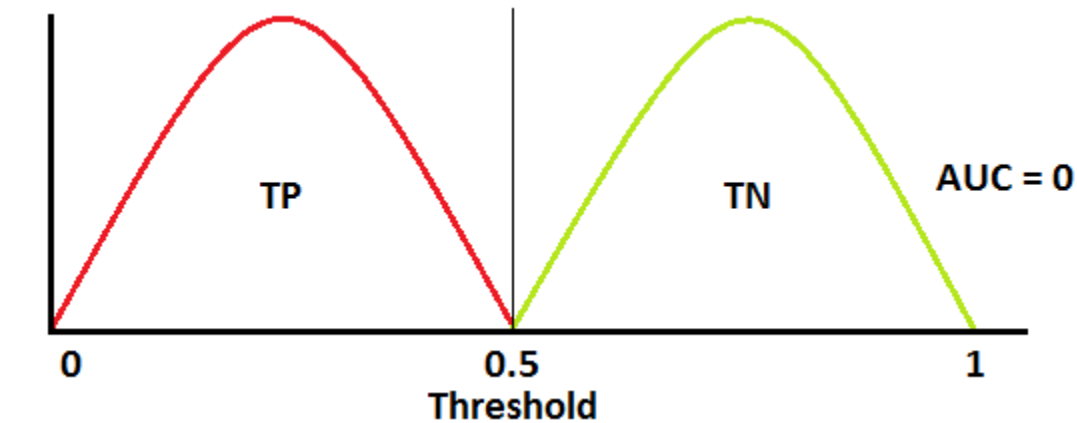
[Image 8 and 9] (Image courtesy: My Photoshopped Collection)

When two distributions overlap, we introduce type 1 and type 2 errors. Depending upon the threshold, we can minimize or maximize them. When AUC is 0.7, it means there is a 70% chance that the model will be able to distinguish between positive class and negative class.



[Image 10 and 11] (Image courtesy: My Photoshopped Collection)

This is the worst situation. When AUC is approximately 0.5, the model has no discrimination capacity to distinguish between positive class and negative class.



[Image 12 and 13] (Image courtesy: My Photoshopped Collection)

When AUC is approximately 0, the model is actually reciprocating the classes. It means the model is predicting a negative class as a positive class and vice versa.

The relation between Sensitivity, Specificity, FPR, and Threshold.

Sensitivity and Specificity are inversely proportional to each other. So when we increase Sensitivity, Specificity decreases, and vice versa.

Sensitivity \uparrow , Specificity \downarrow and Sensitivity \downarrow , Specificity \uparrow

When we decrease the threshold, we get more positive values thus it increases the sensitivity and decreasing the specificity.

Similarly, when we increase the threshold, we get more negative values thus we get higher specificity and lower sensitivity.

As we know FPR is $1 - \text{specificity}$. So when we increase TPR, FPR also increases and vice versa.

TPR \uparrow , FPR \uparrow and TPR \downarrow , FPR \downarrow

How to use the AUC ROC curve for the multi-class model?

In a multi-class model, we can plot the N number of AUC ROC Curves for N number classes using the One vs ALL methodology. So for example, If you have **three** classes named **X**, **Y**, and **Z**, you will have one ROC for X classified against Y and Z, another ROC for Y classified against X and Z, and the third one of Z classified against Y and X.