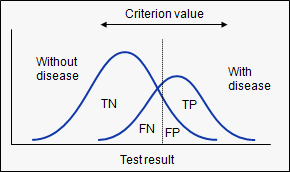
**ROC Curve:**

The ROC curve is a fundamental tool for diagnostic test evaluation. In a ROC curve the true positive rate (Sensitivity) is plotted in function of the false positive rate (100-Specificity) for different cut-off points of a parameter. Each point on the ROC curve represents a sensitivity/specificity pair corresponding to a particular decision threshold. The area under the ROC curve (AUC) is a measure of how well a parameter can distinguish between two diagnostic groups (diseased/normal).

The diagnostic performance of a test, or the accuracy of a test to discriminate diseased cases from normal cases is evaluated using Receiver Operating Characteristic (ROC) curve analysis. ROC curves can also be used to compare the diagnostic performance of two or more laboratory or diagnostic tests.

When you consider the results of a particular test in two populations, one population with a disease, the other population without the disease, you will rarely observe a perfect separation between the two groups. Indeed, the distribution of the test results will overlap, as shown in the following figure.



For every possible cut-off point or criterion value you select to discriminate between the two populations, there will be some cases with the disease correctly classified as positive (TP = True Positive fraction), but some cases with the disease will be classified negative (FN = False Negative fraction). On the other hand, some cases without the disease will be correctly classified as negative (TN = True Negative fraction), but some cases without the disease will be classified as positive (FP = False Positive fraction).

**Schematic outcomes of a test**

The different fractions (TP, FP, TN, FN) are represented in the following table.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Disease** |  |  |  |  |  |  |
| **Test** | **Present** | **n** |  | **Absent** | **n** |  | **Total** |
| **Positive** | True Positive (TP) | *a* |  | False Positive (FP) | *c* |  | *a + c* |
| **Negative** | False Negative (FN) | *b* |  | True Negative (TN) | *d* |  | *b + d* |
| **Total** |  | *a + b* |  |  | *c + d* |  |  |

The following statistics can be defined:

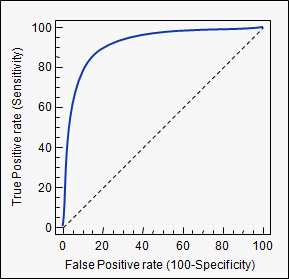
|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Sensitivity** | |  | | --- | | a | | a + b | |  | **Specificity** | |  | | --- | | d | | c + d | |
| **Positive Likelihood Ratio** | |  | | --- | | Sensitivity | | 1 - Specificity | |  | **Negative Likelihood Ratio** | |  | | --- | | 1 – Sensitivity | | Specificity | |
| **Positive Predictive Value** | |  | | --- | | a | | a + c | |  | **Negative Predictive Value** | |  | | --- | | d | | b + d | |

* **Sensitivity:** probability that a test result will be positive when the disease is present (true positive rate, expressed as a percentage).   
   i.e. a / (a + b)
* **Specificity:** probability that a test result will be negative when the disease is not present (true negative rate, expressed as a percentage).   
   i.e. d / (c + d)
* **Positive likelihood ratio:** ratio between the probability of a positive test result given the presence of the disease and the probability of a positive test result given the absence of the disease,

i.e. True positive rate / False positive rate = Sensitivity / (1-Specificity)

* **Negative likelihood ratio:** ratio between the probability of a negative test result given the presence of the disease and the probability of a negative test result given the absence of the disease, i.e.   
  i.e. False negative rate / True negative rate = (1-Sensitivity) / Specificity
* **Positive predictive value:** probability that the disease is present when the test is positive (expressed as a percentage).   
   i.e. a / (a + c)
* **Negative predictive value:** probability that the disease is not present when the test is negative (expressed as a percentage).   
   i.e. d  / (b + d)

In a Receiver Operating Characteristic (ROC) curve the true positive rate (Sensitivity) is plotted in function of the false positive rate (100-Specificity) for different cut-off points. Each point on the ROC curve represents a sensitivity/specificity pair corresponding to a particular decision threshold. A test with perfect discrimination (no overlap in the two distributions) has a ROC curve that passes through the upper left corner (100% sensitivity, 100% specificity). Therefore the closer the ROC curve is to the upper left corner, the higher the overall accuracy of the test.



### Area under the curve**:**

When using normalized units, the area under the curve (often referred to as simply the AUC) is equal to the probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one (assuming 'positive' ranks higher than 'negative').

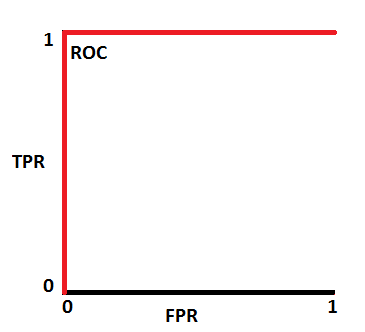
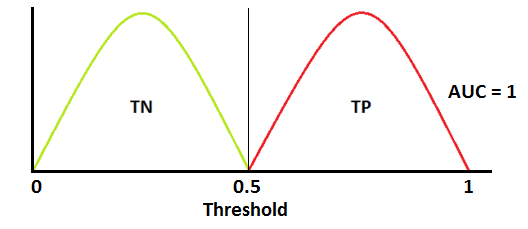
**How to speculate the performance of the model?**

An excellent model has AUC near to the 1 which means it has good measure of separability. A poor model has AUC near to the 0 which means it has 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 model has no class separation capacity whatsoever.

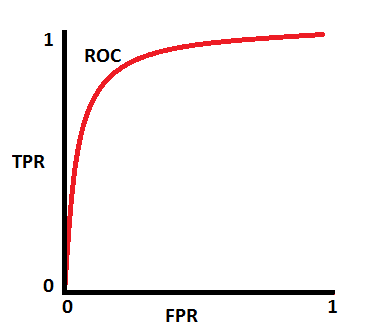
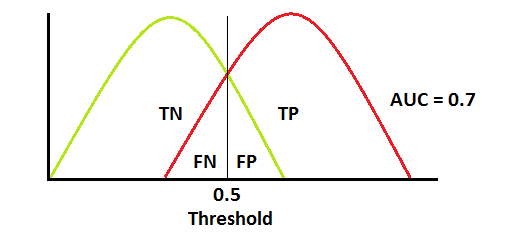
Let’s interpret above statements.

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

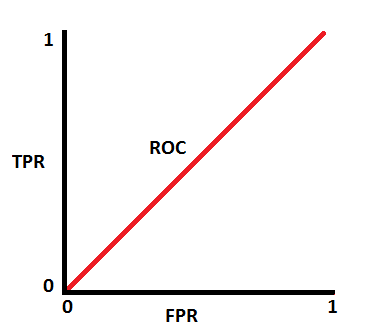
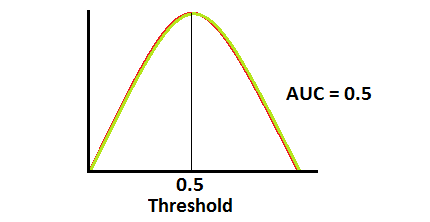
Note: Red distribution curve is of the positive class (patients with disease) and green distribution curve is of negative class(patients with no disease).



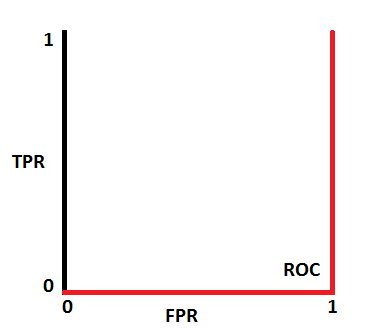
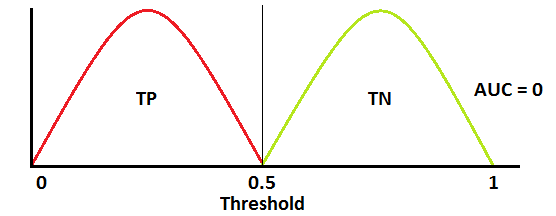
This is an ideal situation. When two curves don’t overlap at all means model has an ideal measure of separability. It is perfectly able to distinguish between positive class and negative class.



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



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



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

**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⬆️, Specificity⬇️*

*and*

*Sensitivity⬇️, Specificity⬆️*

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 - specificity. So when we increase TPR, FPR also increases and vice versa.

*TPR⬆️, FPR⬆️ and TPR⬇️, FPR⬇️*

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

In multi-class model, we can plot N number of AUC ROC Curves for N number classes using 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 a third one of Z classified against Y and X.

Following are a few thumb rules for **ROC curve**:

* .90-1 = excellent (A)
* .80-.90 = good (B)
* .70-.80 = fair (C)
* .60-.70 = poor (D)
* .50-.60 = fail (F)

We see that we fall under the excellent band for the current model. But this might simply be over-fitting. In such cases it becomes very important to to in-time and out-of-time validations.

**Points to Remember:**

1. For a model which gives class as output, will be represented as a single point in ROC plot.

2. Such models cannot be compared with each other as the judgement needs to be taken on a single metric and not using multiple metrics. For instance, model with parameters (0.2,0.8) and model with parameter (0.8,0.2) can be coming out of the same model, hence these metrics should not be directly compared.

3. In case of probabilistic model, we were fortunate enough to get a single number which was AUC-ROC. But still, we need to look at the entire curve to make conclusive decisions. It is also possible that one model performs better in some region and other performs better in other.

**Advantages of using ROC**

ROC curve on the other hand is almost independent of the response rate. This is because it has the two axis coming out from columnar calculations of confusion matrix. The numerator and denominator of both x and y axis will change on similar scale in case of response rate shift.

Ref: https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5