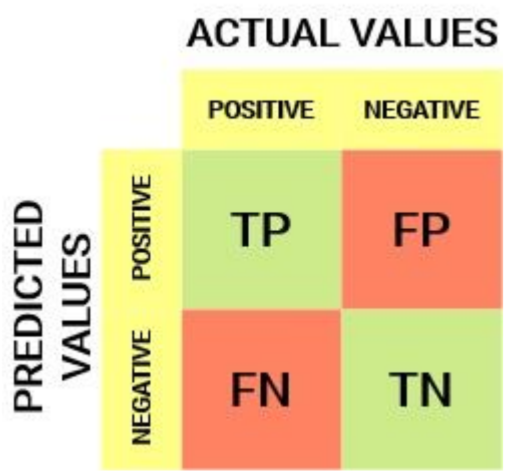
AUC ROC Curve In Machine Learning

Introduction

You've built your machine learning model – so what's next? You need to evaluate and validate how good (or bad) it is, so you can decide whether to implement it. That's where the AUC-ROC curve comes in.

The name might be a mouthful, but it is just saying that we are calculating the "Area Under the Curve" (AUC) of the "Receiver Operating Characteristic" (ROC).

What are Sensitivity and Specificity?



Sensitivity / True Positive Rate / Recall

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

Sensitivity tells us what proportion of the positive class got correctly classified.

A simple example would be determining what proportion of the actual sick people were correctly detected by the model.

Specificity / True Negative Rate

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

Specificity tells us what proportion of the negative class got correctly classified.

Taking the same example as in Sensitivity, Specificity would mean determining the proportion of healthy people who were correctly identified by the model.

Probability of Predictions

A machine learning classification model can be used to directly predict the data point's actual class or predict its probability of belonging to different classes. The latter gives us more control over the result. We can determine our own threshold to interpret the result of the classifier.

Probability of Predictions

A machine learning classification model can be used to directly predict the data point's actual class or predict its probability of belonging to different classes. The latter gives us more control over the result. We can determine our own threshold to interpret the result of the classifier.

An ROC curve, or receiver operating characteristic curve, is like a graph that shows how well a classification model performs. It helps us see how the model makes decisions at different levels of certainty. The curve has two lines: one for how often the model correctly identifies positive cases (true positives) and another for how often it mistakenly identifies negative cases as positive (false positives). By looking at this graph, we can understand how good the model is and choose the threshold that gives us the right balance between correct and incorrect predictions.

The **Receiver Operator Characteristic (ROC)** curve is an evaluation metric for binary classification problems. It is a probability curve that plots the **TPR** against **FPR** at various threshold values and essentially **separates the 'signal' from the 'noise.'** In other words, it shows the performance of a classification model at all classification thresholds. The **Area Under the Curve (AUC)** is the measure of the ability of a binary classifier to distinguish between classes and is used as a summary of the ROC curve.

The higher the AUC, the better the model's performance at distinguishing between the positive and negative classes.