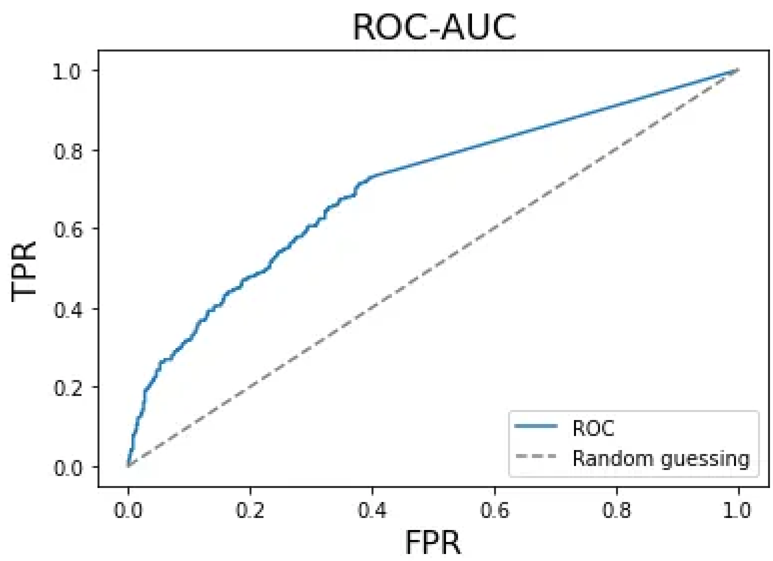
Mathematics behind ROC-AUC interpretation

A mathematical explanation of one of the most used interpretation of ROC-AUC

Nathan · Follow

Published in Analytics Vidhya

4 min read · Aug 11, 2020



An illustration of the ROC curve

Motivation

As I learned about Data Science and Machine Learning online, I often found myself wondering what were the maths behind some of the techniques commonly used. While I was able to find it online most of the time (thanks to Medium and StackExchange for instance), I could not find a proper mathematical explanation of the maths behind the ROC-AUC interpretation.

This article aims at mathematically demonstrating why the Area Under the Receiver Operating Characteristics, commonly referred to as ROC-AUC, can be interpreted as

$$P(X_1 > X_0) = P(X_1 - X_0 > 0)$$

where :

- X_1 is a continuous random variable giving the “score” output by our binary classifier for a randomly chosen **positive** sample
- X_0 is a continuous random variable giving the “score” output by our binary classifier for a randomly chosen **negative** sample

This interpretation is very useful because it basically tells us that ROC-AUC gives us the probability that our classifier gives a higher score to a positive sample than to a negative one! So, it is really a measure of the quality of our classifier.

But where does it come from exactly?

Definitions and preliminary results

First, some definitions :

- Let X_1 and X_0 be defined as above
- Let f_1 and f_0 be, respectively, the density function of X_1 and X_0
- Let F_1 and F_0 be, respectively, the repartition function of X_1 and X_0
- **True Positive Rate (TPR)** and **False Positive Rate (FPR)** have their usual meaning, i.e. :

$$TPR = \frac{TP}{P} \quad FPR = \frac{FP}{N}$$

TP stands for True Positive, P for Positive, FP for False Positive, N for Negative

We can already observe that, for a ***classifier threshold T***, a randomly chosen positive sample would be correctly classified (true positive) if $X_1 > T$. So, for a randomly chosen positive sample, the probability of correctly classifying it is $P(X_1 > T)$. By definition of the ***TPR***, it corresponds to the probability of correctly classifying a randomly chosen positive sample, so $TPR(T) = P(X_1 > T) = 1 - P(X_1 \leqslant T) = 1 - F_1(T)$. (1)

This also means, by definition of the density function, that :

$$TPR(T) = \int_{T}^{+\infty} f_1(x)\,\mathrm{d}x$$

By expressing $P(X_1>T)$ with the density function

Similarly, we can show that ***$FPR(T) = 1 - F_0(T)$*** (2)

Demonstration

Now let's dig into the calculus!

By definition of the ROC, we have that :

$$ROC - AUC = \int_{0}^{1} TPR(FPR)\,\mathrm{d}FPR$$

$$= \int_{0}^{1} TPR(FPR^{-1}(x))\,\mathrm{d}x$$

A very basic expression of ROC-AUC

By using this change in variable :

$$T = FPR^{-1}(x) \iff x = FPR(T)$$

the integral becomes :

$$\int_{+\infty}^{-\infty} TPR(T) \times FPR'(T)\,\mathrm{d}T$$

Now, thanks to (2) we know that we can express this integral as :

$$\int_{+\infty}^{-\infty} TPR(T) \times (-f_0(T))\,\mathrm{d}T = \int_{-\infty}^{+\infty} TPR(T) \times f_0(T)\,\mathrm{d}T$$

It stems from the fact that FPR(T) = 1- F_0(t) so FPR'(T)=-f_0(t)

Thanks to (1) we know that this can be expressed as :

$$\int_{-\infty}^{+\infty} \int_{T}^{+\infty} f_1(x)\,\mathrm{d}x \times f_0(T)\,\mathrm{d}T$$

By using this change in variable for the inner integral :

$$v = x - T$$

the integral becomes :

$$\int_{-\infty}^{+\infty} \int_{0}^{+\infty} f_1(v+T)\,\mathrm{d}v \times f_0(T)\,\mathrm{d}T$$

$$= \int_{0}^{+\infty} \int_{-\infty}^{+\infty} f_1(v+T)\ \times f_0(T)\,\mathrm{d}T\,\mathrm{d}v$$

By swapping the integrals

and by using this change in variable for the inner integral :

$$u = v + T$$

it becomes :

$$\int_{0}^{+\infty}\int_{-\infty}^{+\infty} f_1(u) \times f_0(u-v)\, du\, dv$$

Do you get where we're going? Yes, right to the **convolution theorem**!
First, let's point out that since $f_0(t)$ is a density function of X_0, $f_0(-t)$ is a density function of $(-X_0)$.
Then, according to the convolution theorem and assuming the convergence, a density of $X_1 - X_0 = X_1 + (-X_0)$ is :

$$\int_{-\infty}^{+\infty} f_1(u) \times f_0(u-v)\, du$$

This means that :

$$P(X_1 > X_0) = P(X_1 - X_0 > 0) = \int_{0}^{+\infty}\int_{-\infty}^{+\infty} f_1(u) \times f_0(u-v)\, du\, dv$$

And eventually we have that :

$$P(X_1 > X_0) = ROC - AUC$$

Thank you for reading this far!
I hope you enjoyed this article and it helped you better understand what are the mathematics behind the most common and most powerful interpretation of the ROC-AUC.
Feel free to drop a comment should you find any miscalculation and/or error.

Machine Learning | Data Science | Mathematics