

Comprehension of the AUC-ROC curve

A deeper look into AUC-ROC...



Kurtis Pykes [Follow](#)

Feb 18 · 7 min read ★

Predictive modelling is the problem of developing a model using historical data to make a prediction on new data where we do not have the answer. — Jason Brownlee



Figure 1: Restriction sign at Area 51 — source: <https://www.atlasobscura.com/places/area-51>

When we do supervised learning, we often try to optimize our algorithm for some

...
...
...
...
...
...

could possibly be a metric that we try to optimize for depending on the problem we are trying to solve.

Unfortunately, Area 51 has nothing to do with AUC (or does it? hmm) and I have no insights there so we can ignore *figure 1*— but it is a cool photo, however in this post, I aim to break down the AUC-ROC metric. By the end of this post, you will know:

1. What the ROC curve is.
2. How to create the ROC curve.
3. The formulas to derive the metrics used when plotting the ROC curve.
4. How to interpret the ROC curve.
5. What the Area under the curve is.

If you have not come across what a Confusion Matrix is before, I suggest opening another window and acquainting yourself with my last post on that subject matter. When you are comfortable with the confusion matrix, understanding the ROC curve would be a breeze.

Confusion Matrix “Un-confused”

Breaking down the confusion matrix

[towardsdatascience.com](https://towardsdatascience.com/breaking-down-the-confusion-matrix-3a3e1f3a23)

What is ROC?

The ROC curve is a graphical plot. Its purpose is to illustrate our classification models

We will start by recapping the confusion matrix briefly. The confusion matrix segments our predictions into 2 types of errors and correct predictions as shown in *figure 2*. Type I error is known as the false alarm since we say positive when it is actually negative, and Type II error is a when our classification model predicts negative when the true label is positive — Therefore leaving us with only True positive and True negative labels.

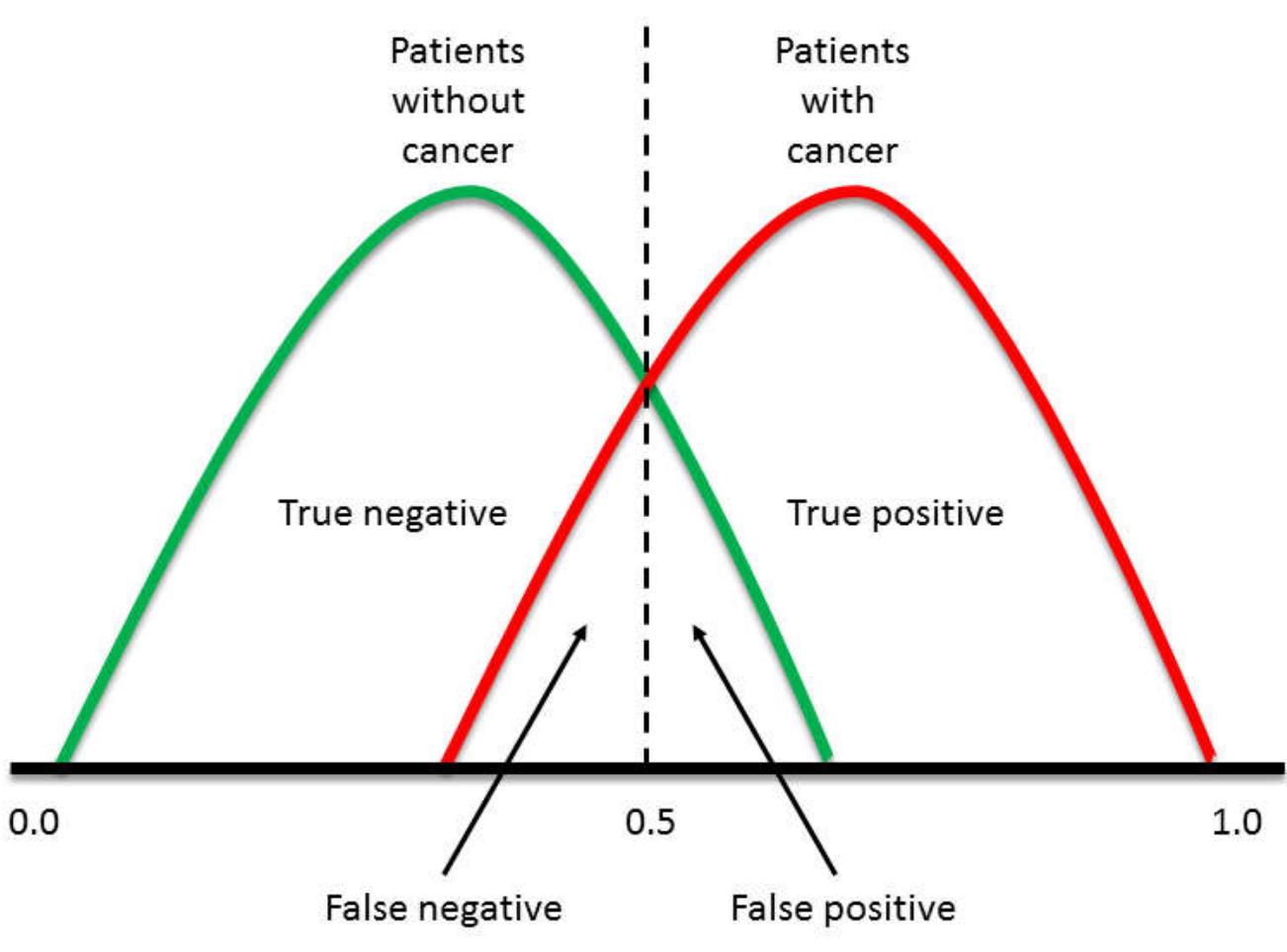


Figure 2: Plot of a models ability to discriminate between classes for different thresholds ranging from 0–1.

Source: <https://stackoverflow.com/questions/57280577/sensitivity-specificity-plot-python>

The dashed vertical line on *figure 2* indicates the current model's threshold. We can alter the threshold of the model by moving that line right or left along that axis, however, this causes a trade-off in sensitivity or specificity depending on which way we move it. Don't worry if you do not understand those two terms yet, see the different trade-offs the model is making if we move it to the left or right.

How to create an ROC curve

We can now infer the model's ability to discriminate the classes at various thresholds by plotting the true positive rate (TPR) against the false positive rate (FPR) at the various threshold settings between 0–1, which is exactly what the ROC curve is. I know what you are thinking... How do we derive these metrics? Great question.

The True Positive Rate (also referred to as sensitivity or recall) is a measure of the number of actual positives that are correctly identified as such from our model's predictions. This is important because when a classifier makes predictions, we as the data scientist should want to know the proportion of correct predictions made out of all of the actual correct labels present in our data (as well as the ones that had errors and why) since in an ideal world we would want all of our predictions to be correct.

However, since we do not live in an ideal world (or share you can share why you think we do with me on twitter — @KurtisPykes), and even us humans, possibly the most intelligent life on earth, can struggle to distinguish between classes, for instance, whether a cell is malignant or not. The amount of false alarms that we classify is known as the False Positive Rate.

• • •

Formulas to derive the metrics

To derive the True Positive rate, we must identify all of the True Positive predictions from our model — correctly predicted positive predictions — and divide it by the True positives and the positive labels that our model classified as negative (False negatives — type II errors).

True Positive

TDD —

Figure 3: Formula for Sensitivity (True positive rate).

On the other hand, to derive the False Positive rate from our model's predictions, we must identify the False positive predictions from our model — the false alarms — and divide it by the False positives and the negative labels that our model classified as negative (True Negatives — correct rejections).

$$\text{FPR} = \frac{\text{False Positive}}{\text{False Positive} + \text{True Negative}}$$

Figure 4: Formula for False positive rate (1- Specificity).

To review, so far we have learnt that an ROC curve is a graphical plot that illustrates our classification models ability to distinguish between classes at various thresholds. Additionally, we now know that this plot is made by plotting the True Positive rate as a function of the False Positive rate and we know how to derive those metrics. Fantastic! It's good enough we know this information, but what does it mean and what can we infer from this to drive business value?

• • •

How to interpret the ROC curve

We now know that this plot is made by plotting the True Positive rate as a function of the False Positive rate.

If we plot the False Positive Rate along the x-axis with values ranging from 0–1 to represent the different probability thresholds, and plot the True Positive Rate along the y-axis from 0–1 representing the probability thresholds. when $x = v$ (when the False

When our model makes more positive predictions it increases our True Positive rate and our false positive rate, in turn decreasing our True negative rate. Simply put, we are trading off predicting the negative class for more positive predictions.

Therefore, when $x=y$ our model is as good as a random guess since the model's true positive rate is increasing with the more positive predictions we make — in Lehman's terms, it is struggling to distinguish between the classes. In some textbooks, you may see the False positive rate being referred to as 1- Specificity since by increasing the False positive rate we offset the True Negative rate (correct rejections) because we make more positive predictions as we move along the x-axis.

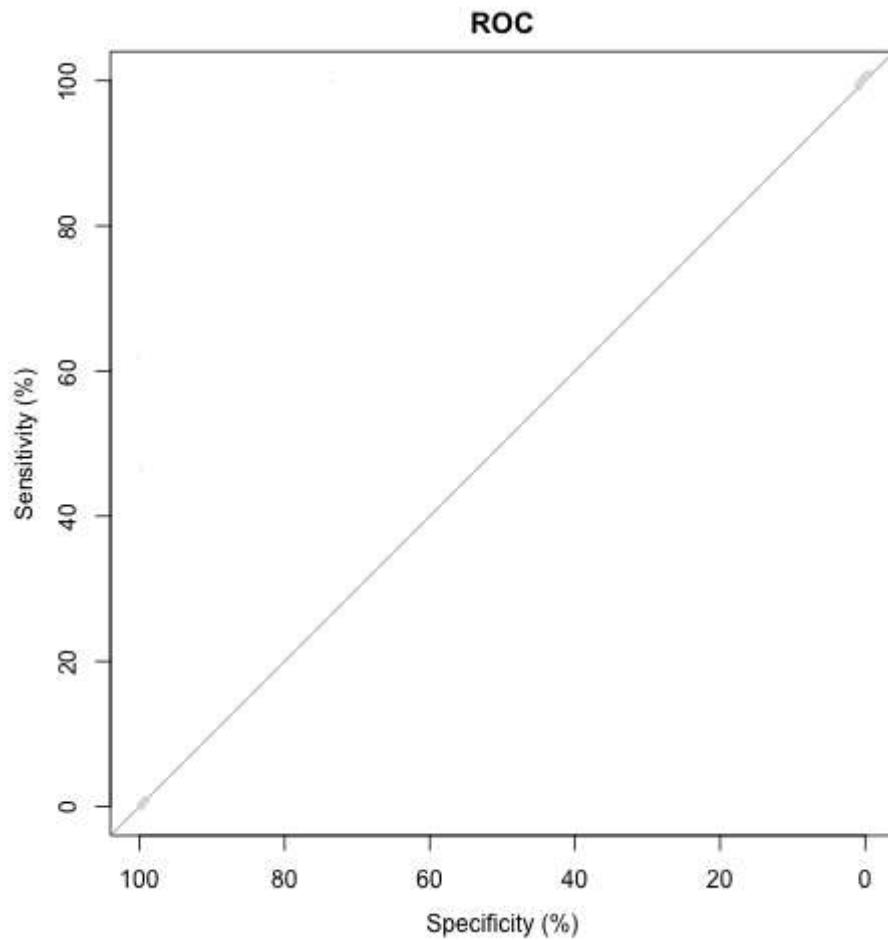


Figure 5: ROC curve showing model that is randomly guessing

Note: Specificity (a.k.a selectivity and True negative rate) is a measure of the amount of

$$\text{TNR} = \frac{\text{True Negative}}{\text{True Negative} + \text{False Positive}}$$

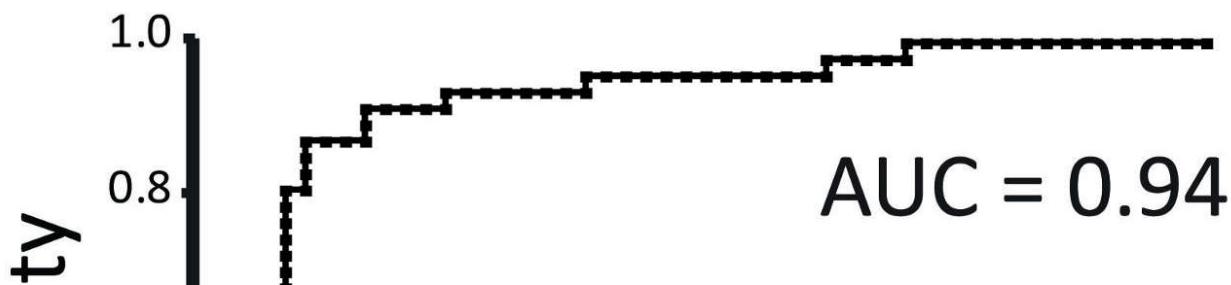
Figure 6: Formula for Specificity (True Negative rate).

You may be beginning to get the gist of the ROC curve by now, if not refer to *Figure 5* and try to digest what is happening in the graph. However, if you are gaining a deeper intuition of the ROC curve, you may have realized that a model that can distinguish between the classes well is one that has a low False Positive rate and a high True Positive rate — visually, this means that the plots will be hugging the top left corner of the chart (*see figure 7*). This brings us on to the final section of this post.

As data scientists, it is our task to be able to understand the type of problem that we are trying to solve. This helps us know what to optimize for in different settings for example, in a case where we are trying to predict malignant tumors, we may want to set our threshold to optimize for sensitivity, so that we are cautious to not say that a patient does not have a malignant tumor when they actually do, whereas in a spam detection setting we don't want to classify important mail as spam so we optimize for specificity.

What is the AUC?

If we want to quantify the ability of our model to separate the classes (the ROC) we use the AUC, which stands for Area under the Curve. Area under is the percentage of the plot that is under the ROC curve.



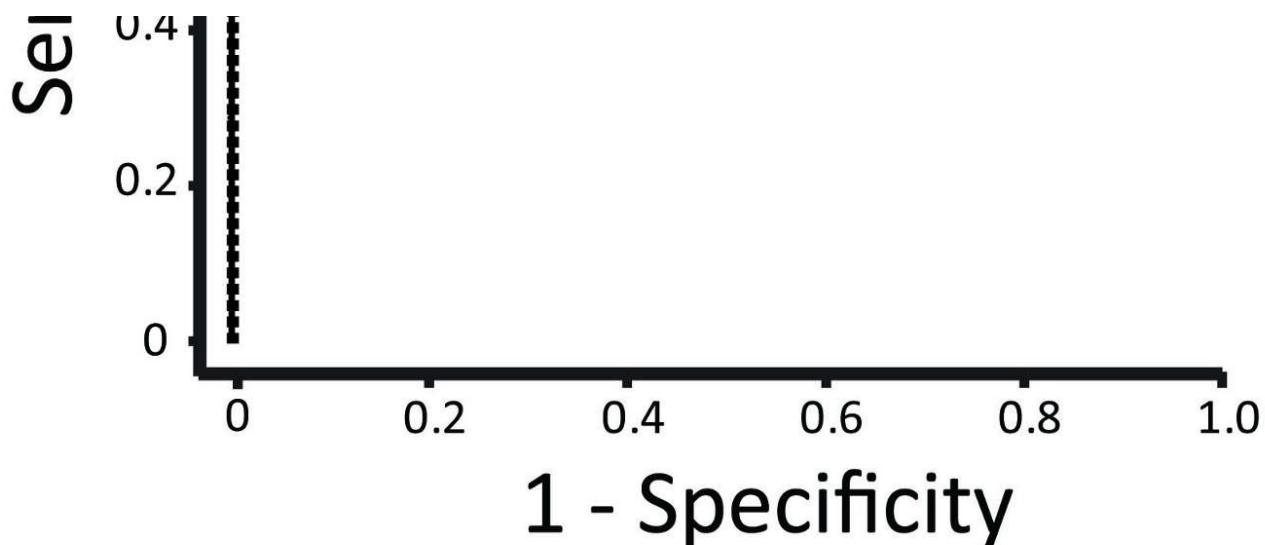


Figure 7: ROC curve that shows a model that is discriminating between classes well with high AUC. Source: <https://stats.stackexchange.com/questions/260845/what-is-more-important-in-a-test-set-sensitivity-specificity-or-roc-auc>

You now know what AUC-ROC is, how to create the plot, how to derive the TPR and FPR, how to interpret the ROC curve, and what the AUC curve is. That concludes this post, though you may want to extend this foundation of knowledge you have received to see how AUC-ROC can be extended for multi-class classification task of which our classifier aims to classify our instances into one of three or more classes, or you may also want to know how AUC-ROC works with imbalanced data. Below, I have referenced some good sources for you to extend your knowledge!

• • •

References:

Data School. (Nov 20, 2014). *ROC Curves and Area Under the Curve (AUC) Explained*. <https://www.youtube.com/watch?v=OAl6eAyP-yo>

USMLE Biostatistics. (Apr 29, 2016). *Sensitivity vs. Specificity: Trade-Off*. <https://www.youtube.com/watch?v=yax-n3ROboE>

Wikipedia. *Multiclass Classification*.

https://en.wikipedia.org/wiki/Multiclass_classification

Data Science Machine Learning Metrics

About Help Legal