Understanding Confusion Matrix

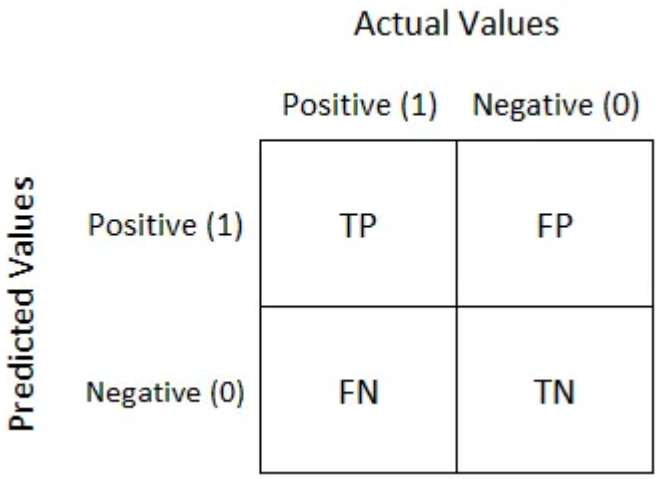
When we get the data, after data cleaning, pre-processing, and wrangling, the first step we do is to feed it to an outstanding model and of course, get output in probabilities. But hold on! How in the hell can we measure the effectiveness of our model. Better the effectiveness, better the performance, and that is exactly what we want. And it is where the Confusion matrix comes into the limelight. Confusion Matrix is a performance measurement for machine learning classification.

This aims to answer the following questions:

1. What the confusion matrix is and why you need it?
2. How to calculate Confusion Matrix for a 2-class classification problem?

What is Confusion Matrix and why you need it?

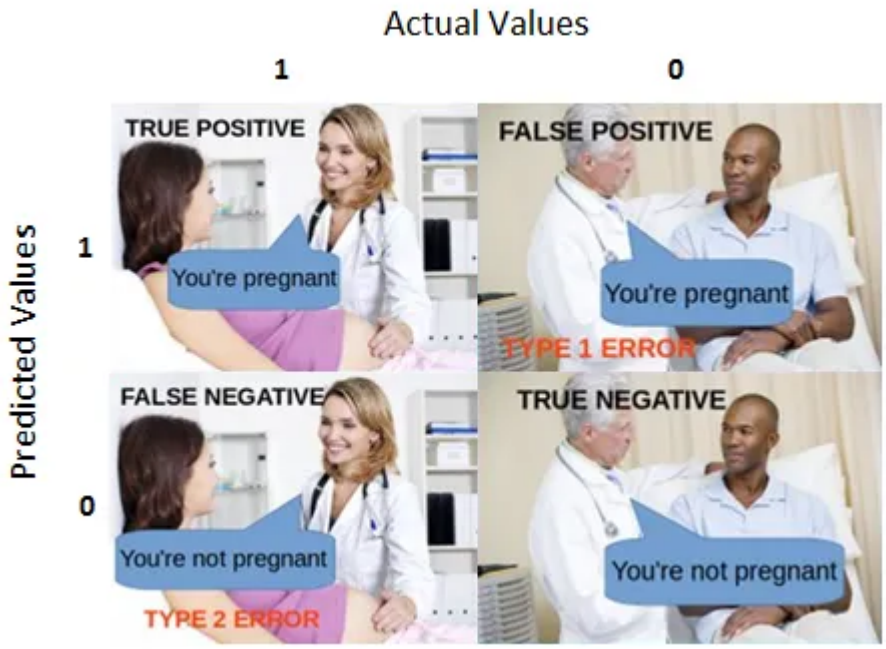
Well, it is a performance measurement for machine learning classification problem where output can be two or more classes. It is a table with 4 different combinations of predicted and actual values.



Confusion Matrix

It is extremely useful for measuring Recall, Precision, Specificity, Accuracy, and most importantly AUC-ROC curves.

Let's understand TP, FP, FN, TN in terms of pregnancy analogy.



Confusion Matrix

True Positive:

Interpretation: You predicted positive and it's true.

You predicted that a woman is pregnant and she actually is.

True Negative:

Interpretation: You predicted negative and it's true.

You predicted that a man is not pregnant and he actually is not.

False Positive: (Type 1 Error)

Interpretation: You predicted positive and it's false.

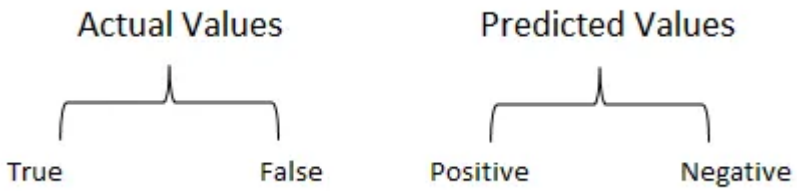
You predicted that a man is pregnant but he actually is not.

False Negative: (Type 2 Error)

Interpretation: You predicted negative and it's false.

You predicted that a woman is not pregnant but she actually is.

Just Remember, We describe predicted values as Positive and Negative and actual values as True and False.

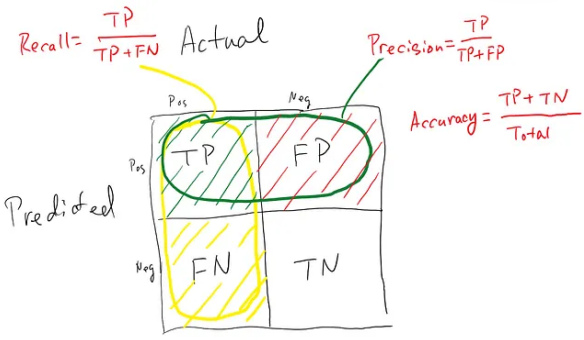


Actual vs Predicted values

How to Calculate Confusion Matrix for a 2-class classification problem?

Let's understand the confusion matrix through math.

y	y pred	output for threshold 0.6	Recall	Precision	Accuracy
0	0.5	0			
1	0.9	1			
0	0.7	1			
1	0.7	1	1/2	2/3	4/7
1	0.3	0			
0	0.4	0			
1	0.5	0			



Confusion Matrix [Image 5 and 6] (Image 5 courtesy: My Photoshopped Collection) (Image 6 courtesy: I can not find the source. If you know please comment. I will provide appropriate citations. :D)

Recall

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

The above equation can be explained by saying, from all the positive classes, how many we predicted correctly.

Recall should be high as possible.

Precision

$$Precision = \frac{TP}{TP + FP}$$

The above equation can be explained by saying, from all the classes we have predicted as positive, how many are actually positive.

Precision should be high as possible.

and

Accuracy

From all the classes (positive and negative), how many of them we have predicted correctly. In this case, it will be 4/7.

Accuracy should be high as possible.

F-measure

$$F\text{ - measure} = \frac{2*\text{Recall}*\text{Precision}}{\text{Recall} + \text{Precision}}$$

F1 Score [Image 9] (Image courtesy: My Photoshopped Collection)

It is difficult to compare two models with low precision and high recall or vice versa. So to make them comparable, we use F-Score. F-score helps to measure Recall and Precision at the same time. It uses Harmonic Mean in place of Arithmetic Mean by punishing the extreme values more.