Confusion Matrix

A confusion matrix is a summary of prediction results on a classification problem. The confusion matrix shows the ways in which your classification model is confused when it makes predictions.

Classification accuracy alone can be misleading if you have an unequal number of observations in each class of if you have more than two classes in your dataset.

Classification accuracy is the ration of correct predictions to total predictions made

Classification accuracy = correct predictions / total predictions

Classification accuracy can also be easily be turned into a misclassification rate or error rate by inverting the value such as:

Error rate = 1 – (correct predictions / total predictions)

Classification accuracy can hide the detail you need to diagnose the performance of your model. Hence we use the confusion matrix

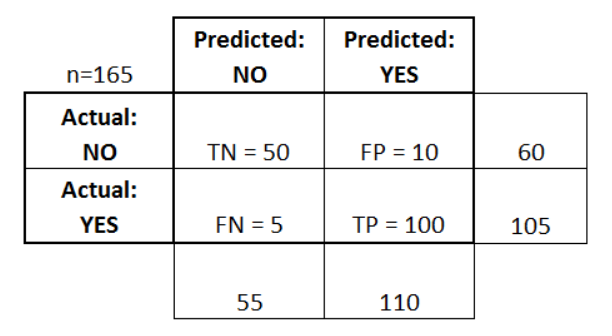


Figure 1 Example Confusion Matrix

The above confusion matrix is from a disease prediction project. The classifier made a total of 165 predictions. The different terms in the above confusion matrix are defined below:

* True Positive (TP): These are cases in which we predicted yes, and they do have the disease
* True Negatives (TN): We predicted no, and they don’t have the disease
* False Positives (FP): We predicted yes, but they don’t actually have the disease
* False Negatives (FN): We predicted NO, but they actually do have the disease

Following is a list of rates that are often computed from a confusion matrix for a binary classifier:

* Accuracy: (TP + TN) / Total
* Misclassification Rate: (FP + FN) / Total
* TP Rate: TP / Actual Yes (Also known as Sensitivity or Recall)
* FP Rate: FP / Actual No
* TN Rate: TN / Actual No (Also known as Specificity)
* Precision: TP / Predicted Yes
* Prevalence: actual Yes / total

Other terms are also worth mentioning:

* Null error rate: This is how often you would be wrong if you always predicted majority class. For the above example, the NER would be 60/165 = 0.36 because if you always predicted yes, you would only be wrong for the 60 “NO” cases. This can be a useful baseline metric to compare your classifier agent. However, the best classifier for a particular application will sometimes have a higher error rate than the null error rate, as demonstrated by the Accuracy Paradox [https://en.wikipedia.org/wiki/Accuracy\_paradox]
* Cohen’s Kappa: This is essentially a measure of how well the classifier performed as compared to how well it would have performed simply by chance. In other words, a model will have a high Kappa score if there is a big difference between the accuracy and the null error rate.
* F score: Weighted average of the true positive rate and precision
* ROC Curve: Commonly used graph that summarizes the performance of a classifier over all possible thresholds. It is generated by plotting the TP (y-axis) against the FP (x-axis) as you vary the threshold for assigning observations to a given class.

Cumulative Accuracy Profile [https://en.wikipedia.org/wiki/Cumulative\_accuracy\_profile]