Performance Metric

B ANIRUDH SRINIVSAN
COE17B019

1. Introduction

To understand the efficiency of a classification model, we tend to use different metric to find its efficiency. Some commonly used metric are derived from confusion matrix are,

- Accuracy
- Sensitivity and Specificity
- Precision
- F1 score

2. Some terms before the metric analysis

- **True Positive (TP)**

  A true positive is an outcome where the model correctly predicts the positive class.

- **True Negative (TN)**

  A true negative is an outcome where the model correctly predicts the negative class.

- **False Positive (FP)**

  A false positive is an outcome where the model incorrectly predicts the positive class. Also known as Type I error.

- **False Negative (FN)**

  A false negative is an outcome where the model incorrectly predicts the negative class.Also known as Type II error.

3. Confusion Matrix

A confusion matrix is a table that is often used to describe the performance of a classification model (or "classifier") on a set of test data for which the true values are known.

Lets take an example.

n=165	Predicted: NO	Predicted: YES
Actual: NO	50	10
Actual: YES	5	100

Figure 1: Example dataset

Let us consider a binary classification case of yes and no

By definition, true positives are those where the model predicted yes and the actual value is also yes, and true negatives are those where the model predicted no and but the actual value is also no.

Similarly, false positives are those where the model predicts yes and the actual value is no, and false negatives are those where the model predicted no and but the actual value is yes.

Therefore, the confusion matrix for the above dataset would be,

n=165	Predicted: NO	Predicted: YES	
Actual: NO	TN = 50	FP = 10	60
Actual: YES	FN = 5	TP = 100	105
	55	110	

Figure 2: Confusion Matrix for the above example

The following metric are calculated using confusion matrix.

4. Performance Metric

4.1 Accuracy

This is the measure of overall, how often is the classifier correct.

$$Accuracy = \frac{(TP + TN)}{totalnumberofsample} \tag{1}$$

In our case,

$Accuracy = \frac{(100+50)}{165} = 0.91$

Accuracy may not be a good measure if the dataset is not balanced.

For imbalanced data, accuracy is defined as,

$$Accuracy = \frac{(Sensitivity + Specificity)}{2} \tag{2}$$

4.2 Sensitivity and Specificity

Sensitivity is the measures the proportion of actual positives that are correctly identified to the total number of positive cases.

Specificity is the measures the proportion of actual negatives that are correctly identified to the total number of negative cases.

$$Sensitivity = \frac{TP}{(TP + FN)} \tag{3}$$

$$Specificity = \frac{TN}{(TN + FP)} \tag{4}$$

In our case,

$Sensitivity = \frac{100}{(100 + 5)} = 0.95$

and, $Specificity = \frac{50}{(50 + 10)} = 0.83$

Sensitivity is also called Recall.

4.3 Precision

Precision is the measures the proportion of actual positives that are correctly identified to the total number of cases classified as positive.

$$Precision = \frac{TP}{(TP + FP)} \tag{5}$$

In our case,

$Precision = \frac{100}{(100 + 10)} = 0.91$

4.4 F1-Score

So ideally in a good classifier, we want both precision and recall to be one which also means FP and FN are zero. Therefore we need a metric that takes into account both precision and recall. F1-score is a metric which takes into account both precision and recall and is defined as follows:

$$F1\_score = 2 * \frac{Precision * Recall}{(Precision + Recall)} \tag{6}$$

F1 Score becomes 1 only when precision and recall are both 1. F1 score becomes high only when both precision and recall are high. F1 score is the harmonic mean of precision and recall and is a better measure than accuracy.

In our case,

$F1\_score = 2 * \frac{(0.91 * 0.83)}{(0.91 + 0.83)} = 0.87$