



Precision and Recall

Data Boot Camp
Lesson 21.2



Precision & Recall

Confusion Matrix

A confusion matrix is a table used to describe the performance of a classifier by comparing the predicted and actual values. Consider the following matrix where the classes are “Cancer” or “No Cancer.”

Actual Class	Predicted Class		
		Cancer	No Cancer
	Cancer	True Positive	True Negative
	No Cancer	False Positive	False Negative



True Positive (TP): The predicted class and the actual class are the same. Both predicted Cancer.



True Negative (TN): The predicted class and the actual class are the same. Both predicted No Cancer.



False Negative (FN): The actual class was Cancer, but the prediction was No Cancer.



False Positive (FP): The actual class was No Cancer, but the prediction was Cancer.

Accuracy

Accuracy is the ratio of correctly predicted observations to the total number of observations.

$$\text{Accuracy} = (\text{TP} + \text{TN}) / n$$

Actual Class	Predicted Class		
		Cancer	No Cancer
	Cancer	True Positive	True Negative
	No Cancer	False Positive	False Negative

Accuracy

Accuracy is the ratio of correctly predicted observations to the total number of observations.

$$\text{Accuracy} = 85/100 = 0.85$$

Actual Class	Predicted Class		
		Cancer	No Cancer
	Cancer	25	10
	No Cancer	5	60

Precision

Precision is the ratio of correctly predicted positive observations to the total predicted positive observations (i.e., of all the samples we classified as Cancer, how many are actually Cancer?).

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$$

Actual Class	Predicted Class		
		Cancer	No Cancer
	Cancer	True Positive	True Negative
	No Cancer	False Positive	False Negative

Precision

Precision is the ratio of correctly predicted positive observations to the total predicted positive observations (i.e., of all the samples we classified as Cancer, how many are actually Cancer?).

$$\text{Precision} = 25/30 = .8333$$

Actual Class	Predicted Class		
		Cancer	No Cancer
	Cancer	25	10
	No Cancer	5	60

Recall

Recall is the ratio of correctly predicted positive observations to the total predicted positive observations (i.e., of all the actual Cancer samples, how many did we classify as Cancer?).

$$\text{Recall} = \text{TP} / (\text{TP} + \text{TN})$$

Actual Class	Predicted Class		
		Cancer	No Cancer
	Cancer	True Positive	True Negative
	No Cancer	False Positive	False Negative

Recall

Recall is the ratio of correctly predicted positive observations to the total predicted positive observations (i.e., of all the actual Cancer samples, how many did we classify as Cancer?).

$$\text{Recall} = 25/35 = .714$$

Actual Class	Predicted Class		
		Cancer	No Cancer
	Cancer	25	10
	No Cancer	5	60

F1 Score

The F1 score is the harmonic average of the precision and recall, where an F1 score reaches its best value at 1 (perfect precision and recall) and worst at 0.

$$F1 = 2 * (\text{precision} * \text{recall}) / (\text{precision} + \text{recall})$$

Actual Class	Predicted Class		
		Cancer	No Cancer
	Cancer	True Positive	True Negative
	No Cancer	False Positive	False Negative

F1 Score

The F1 score is the harmonic average of the precision and recall, where an F1 score reaches its best value at 1 (perfect precision and recall) and worst at 0.

$$F1 = 2 * ((.8333 * .714) / (.8333 + .714)) = 0.77$$

Actual Class	Predicted Class		
		Cancer	No Cancer
	Cancer	25	10
	No Cancer	5	60



Questions?