# 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.”

| | Predicted Class | | |
|---|---|---|---|
| Actual Class | | Cancer | No Cancer |
| | Cancer | True Positive | False Negative |
| | No Cancer | False Positive | True 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. **<u>** Commonly used KPI</u>**

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

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

# Accuracy

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

$$\text{Accuracy} = (25+60)/(25+10+5+60) = 85/100 = 0.85$$

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

# False Positive Rate/( 1-[SPecificity or True Negative Rate] )

False Positive Rate (FPR) is the ratio of incorrectly predicted negative observations to the total negative observations (i.e., of all the non-Cancer samples, how many did we classify as Cancer?). **<u>** Important/informative KPI in many industries</u>**

** Waste (pregnancy test), **"Type I" error** (pregnant man).

| | Predicted Class | | |
|---|---|---|---|
| **Actual Class** | | Cancer | No Cancer |
| | Cancer | True Positive | False Negative |
| | No Cancer | False Positive | True Negative |

$$FPR = FP / (FP + TN)$$

# False Positive Rate/( 1-[SPecificity or True Negative Rate] )

False Positive Rate (FPR) is the ratio of incorrectly predicted negative observations to the total negative observations (i.e., of all the non-Cancer samples, how many did we classify as Cancer?).

$$FPR = 5 / (5+60) = 5/65 = 0.0769$$

| | Predicted Class | | |
|---|---|---|---|
| Actual 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?).

** How often am I correct when I identify positive cases?

** Reduces “Type I” error, reduces waste.

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

**Precision = TP / (TP + FP)**

# 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?).

**Precision = 25/(25+5) = 25/30 = .8333**

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

# Recall/Sensitivity/Probability Of Detection/True Positive Rate

Recall (POD) 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?). **<u>** Crucial KPI in most industries</u>**

** The ability of identify positive cases.

** Misdiagnosis, 1 - **"Type II" error** (pregnant woman misdiagnosed).

| | **Predicted Class** | | |
|---|---|---|---|
| **Actual Class** | | Cancer | No Cancer |
| | Cancer | True Positive | False Negative |
| | No Cancer | False Positive | True Negative |

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

# 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?).

**Recall = 25/(25+10) = 25/35 = .714**

| | Predicted Class | | |
|---|---|---|---|
| Actual 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.

** Confidence in the ability to identify positive cases correctly.

** Reduces both error Types

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

**F1 = 2 * ( (precision * recall) / (precision + recall) )**

# 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.

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

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

# Questions?