Metrics

CS230 Fall 2018
Section 10

Overview

- Shortcomings of Accuracy
- Recall
- Precision
- F1

Ground truth

Normal

Pneumonia

Model1

Model2

Model3

Ground truth

Model1 Accuracy = ?

Model2 Accuracy = ?

Model3 Accuracy = ?

Ground truth

Model1 Accuracy = 9/10

Model2 Accuracy = 8/10

Model3 Accuracy = 5/10

Is Model1 the best model?

Model1 -

Recall

Recall

- Out of all the patients with pneumonia, how many did the model predict as having pneumonia?

- (No. of patients that have pneumonia and are predicted as having pneumonia by the model) / (No. of patients that have pneumonia)

Recall: (No. of patients that have pneumonia and are predicted as having pneumonia by the model) / (No. of patients that have pneumonia)

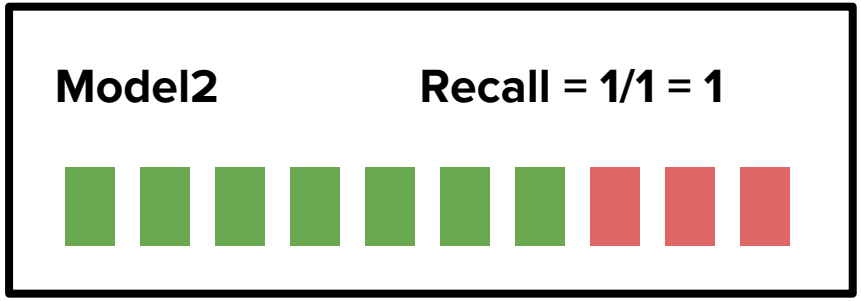
Ground truth

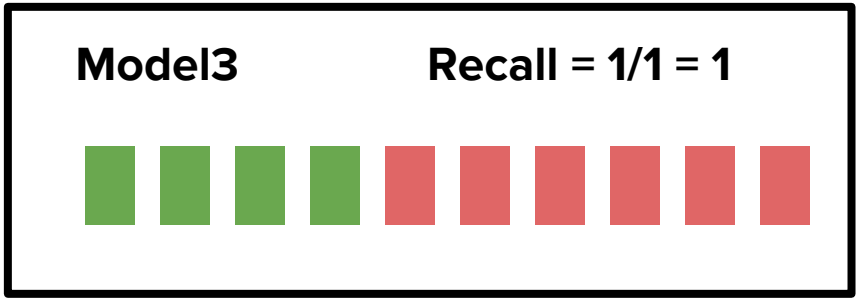
Model1 **Recall = ?**

Model2 **Recall = ?**

Model3 **Recall = ?**

Recall: (No. of patients that have pneumonia and are predicted as having pneumonia by the model) / (No. of patients that have pneumonia)





Are Model2 and Model3 equally good?

Model2

Model3

Precision

Precision

- Out of all the patients that are predicted to have pneumonia, how many actually have pneumonia?

- (No. of patients that have pneumonia and are predicted as having pneumonia by the model) / (No. of patients that are predicted to have pneumonia)

Precision: (No. of patients that have pneumonia and are predicted as having pneumonia by the model) / (No. of patients that are predicted to have pneumonia)

Ground truth

Model1 **Precision = ?**

Model2 **Precision = ?**

Model3 **Precision = ?**

Precision: (No. of patients that have pneumonia and are predicted as having pneumonia by the model) / (No. of patients that are predicted to have pneumonia)

Ground truth

Model1 **Precision = 0/0**

Model2 **Precision = ⅓ = 0.3**

Model3 **Precision = ⅛ = 0.16**

F1 score

Which one is better?

Classifier	Precision	Recall
A	95%	90%
B	98%	85%

Image credits - Machine Learning Yearning, Andrew Ng, Pg 21

Hard to compare!

Classifier	Precision	Recall
A	95%	90%
B	98%	85%

Image credits - Machine Learning Yearning, Andrew Ng, Pg 21

Other metrics

- PR AUC
- ROC AUC
- Sensitivity (used mostly in the medical field)
- Specificity (used mostly in the medical field)

Look at papers to find the metric used in the space that you’re working in!

F1 score

$$F_1 = 2 \cdot \frac{\mathrm{precision} \cdot \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}}$$

F1 score

Classifier	Precision	Recall	F1 score
A	95%	90%	**92.4%**
B	98%	85%	**91.0%**

Image credits - Machine Learning Yearning, Andrew Ng, Pg 22

Back to our original example

Ground truth

Model1 F1 = ?

Model2 F1 = 0.5

Model3 F1 = 0.29