DATA 301: Precision/Recall, F1 Score, Confusion Matrix

Topics

Classification metrics
Accuracy, and why it's not as useful
Precision/Recall
F1 Score
Confusion Matrix

Classification

These metrics apply to classification not regression problems.

Classification: predict a class
Regression: Predict a number

Accuracy, and why it's not as useful

It's a fine measure if you have a balanced dataset. Or roughly the same number of every class in the dataset.

Accuracy fails if you have an unbalanced dataset, for instance

- Cancer diagosis dataset
- Or credit card fraud dataset

There are likely to be very few positive cancer diagosis, or frauds in the above datasets.

For each of these datasets your model can always predict the majority class and score high accuracy. (The more unbalanced, the better the accuracy)

A better way- Precision/Recall/F1

Precision: Out of all found how accurate were predictions?

$$P = \frac{T_p}{T_p + F_p}$$

Recall: How accurate if you consider all?

$$R = \frac{T_p}{T_p + F_n}$$

F1 score: The harmonic mean of precision and recall

$$F1 = 2\frac{P \times R}{P + R}$$

Tp true positives
Fp false positives
Tn true negatives
Fn false negatives

Precision/Recall/F1

$$P = \frac{T_p}{T_p+F_p} \qquad R = \frac{T_p}{T_p+F_n} \qquad F1 = 2\frac{P \times R}{P+R}$$

Example: If a database has 100 items, 60 of which are relevant.

If your algorithm finds 50 items, 40 of which are relevant then

P= 40/(40+10) =0.8
R= 40/60 =.66
F1= 2*(.8*.66)/(.8+.66) = .723

You strive for high precision and recall

Precision/Recall/F1 for Multiple classes

When you have more than 2 classes:

Precision: Sum of all the true positives over the sum of all the true positives plus the sum of all the true negatives

Recall: Sum of all the true positives over the sum of all the true positives plus the sum of all the false negatives

F1 score: Still the harmonic mean of precision and recall

Precision/Recall/F1 for Multiple classes Example

A multiclass classification problem with 3 classes.
A,B,C
With a 1:1:100 class ratio (100 times as many C's as A's)

If a dataset has 10,000 C's it will have 100 A's and B's

Precision/Recall/F1 for Multiple classes Example

A multiclass classification problem with 3 classes.
A,B,C
With a 1:1:100 class ratio (100 times as many C's as A's)

If a dataset has 10,000 C's it will have 100 A's and B's

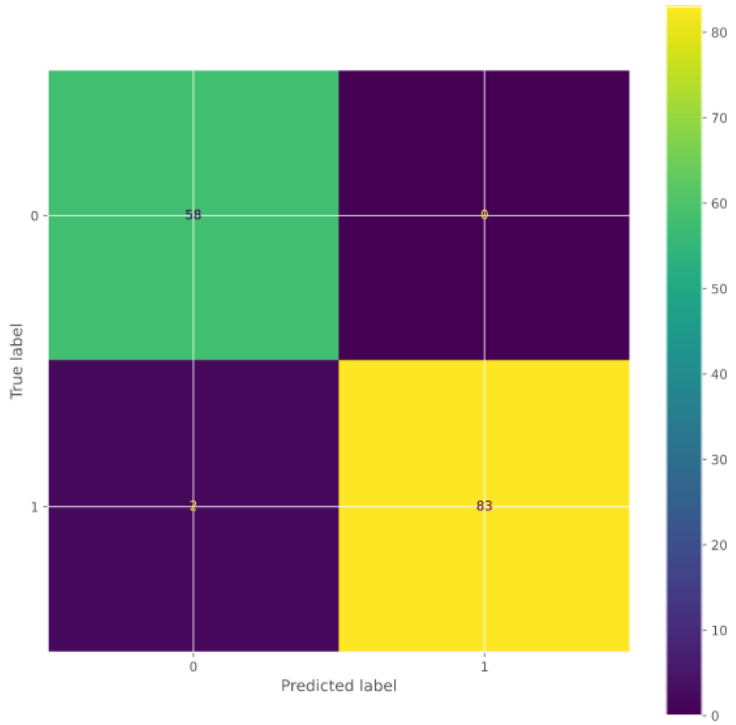
$$P = \frac{T_p}{T_p + F_p}$$

A model predicts 70 A's, 50 are correct and 20 wrong
Predicts 150 B's, 99 are correct and 51 wrong

Precision=(50 +99)/((50+99) + (20 + 51))
=.677

Confusion Matrix

Very simple display:
shows the number right and wrong for every class



Summary

Accuracy is misleading, especially if your dataset is imbalanced
Precision/Recall
F1 Score
Confusion Matrix- just a visual way to check your models predictions against correct values for all classes