Evaluation Metrics

By Trần Minh Dương - Learning Support

Overview

How do you know if your model is truly performing well? Evaluation metrics provide the answer. They act as a scorecard, measuring your model's ability to make accurate and meaningful predictions. From Accuracy to F-score, these metrics highlight strengths, reveal weaknesses, and guide improvements.

1. Confusion Matrix, TP, FP, TN, and FN

The **confusion matrix** is a table that summarizes how successful the classification model is at predicting examples belonging to various classes. The key terms are:

- **True Positives (TP):** Correctly predicted positive cases.
- **False Positives (FP):** Incorrectly predicted as positive (false alarms).
- **True Negatives (TN):** Correctly predicted negative cases.
- **False Negatives (FN):** Incorrectly predicted as negative (missed positives).

	Predicted: 0	Predicted: 1
Actual: 0	True Negative (TN)	False Positive (FP)
Actual: 1	False Negative (FN)	True Positive (TP)

2. Evaluation Metrics

Accuracy:

The proportion of correct predictions out of the total predictions.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}}$$

Precision:

How many of the predicted positive cases are truly positive.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

Recall:

How many of the actual positive cases are correctly predicted.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

F-score:

The harmonic mean of Precision and Recall, balancing the two.

$$\text{F-score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2 \cdot \text{TP}}{2 \cdot \text{TP} + \text{FP} + \text{FN}}$$

3. Illustrative Example:

Imagine during the COVID-19 epidemy, you develop a testing kit that can determine if a person has COVID or not.

- TP: People tested positive and actually have covid

- TN: People tested negative and really do not have covid

- FP: People tested positive but do not have covid

- FN: People tested negative but actually have covid (Extremely dangerous)

- Accuracy: Of all the people tested, how many are correctly diagnosed (TP + TF / n)

- **Precision**: Of all positively-tested people, how many truly have covid ? (TP / TP + FP)

- **Recall**: Of all the people having covid, how many were spotted ? (TP / TP + FN).
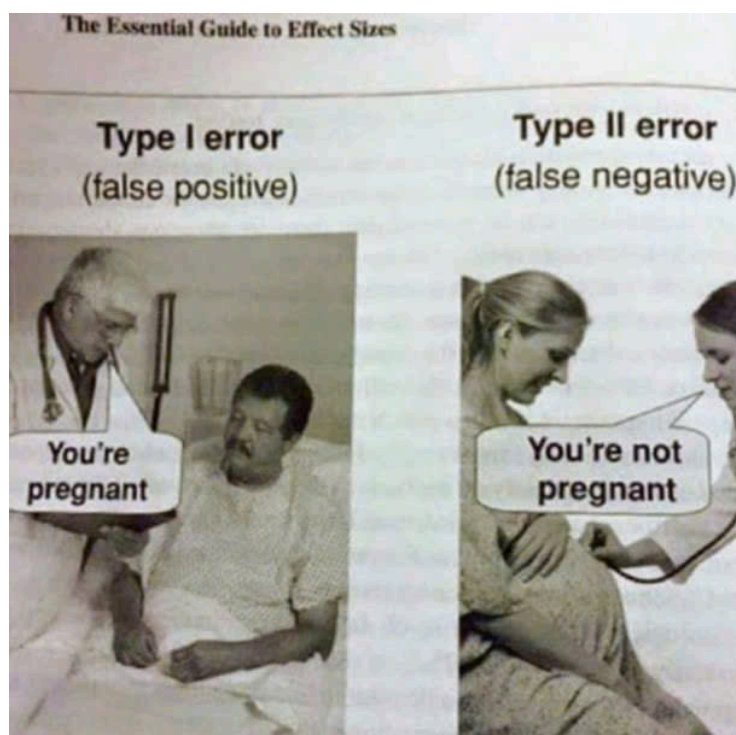
Exercise

Given the following confusion matrix:

n = 192	Predicted: 0	Predicted: 1
Actual: 0	118	12
Actual: 1	47	15

Compute the Accuracy, Precision, Recall, and F-score for this model.

```
In [1]:  TP = 15
         FP = 12
         TN = 118
         FN = 47

         Accuracy = (TP + TN) / (TP + FP + TN + FN)
         Precision = TP / (TP + FP)
         Recall = TP / (TP + FN)
         F_score = 2 * (Precision*Recall) / (Precision+Recall)

         print(f"Accuracy: {Accuracy:.4f}")
         print(f"Precision: {Precision:.4f}")
         print(f"Recall: {Recall:.4f}")
         print(f"F_score: {F_score:.4f}")
```

```
Accuracy: 0.6927
Precision: 0.5556
Recall: 0.2419
F_score: 0.3371
```

This document was created in Jupyter Notebook by Trần Minh Dương (tmd).

If you have any questions or notice any errors, feel free to reach out via Discord at @tmdhoctiengphap or @ICT-Supporters on the USTH Learning Support server.

Check out my GitHub repository for more projects: GalaxyAnnihilator/MachineLearning .