Classification Metrics

The issue with Accuracy?

Ans: The accuracy metric fails when data is imbalanced.
- Consider there are 90 data points of class 1 and 10 data points of class 0.
- If the model predicts every data point as class 1.

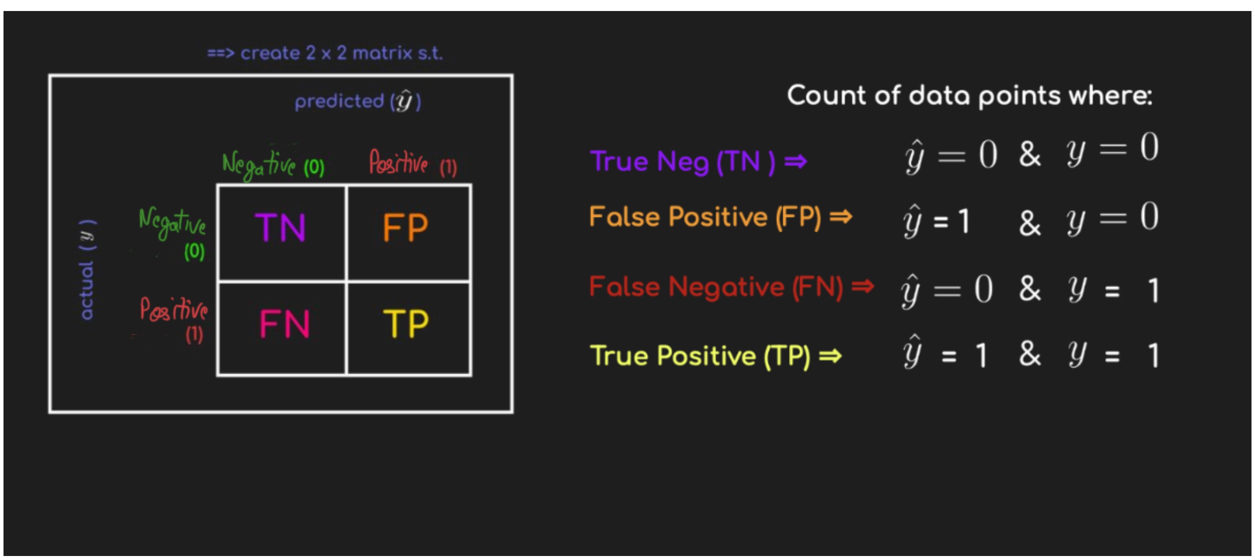
⇒ The accuracy of the model is 90%, which is completely wrong.

What other metric to use?

Ans: Confusion matrix

When a model predicts, there can be 4 scenarios:

A. **True Positive (TP):** Model predicts True, actually is True
B. **False Positive (FP):** Model predicts True, actually is False
   - Aka Type 1 Error
C. **True Negative (TN):** Model predicts False, actually is False
D. **False Negative (FN):** Model predicts False, actually is True
   - Aka Type 2 Error



Note:
- For a dumb model that predicts everything as negative, FP = TP = 0
- For an ideal model that has no incorrect classification, FP = FN = 0

How to use CM to determine if a model with high accuracy is good?

Ans. High accuracy can be deceiving. It is only a good model if:

- Both TP and TN should be high
- Both FP and FN should be low

Given a CM, how can we calculate actual positives?

Ans. TP + FN = P (total actual positives)

⇒ Similarly, FP + TN = N (total actual negatives)

How to calculate accuracy from the confusion matrix?

Ans: Accuracy: $Accuracy = \frac{Correct\ Predictions}{Total\ Number\ of\ Predictions} = \frac{TP+TN}{TP+TN+FN+FP}$

Which metric to use, when we cannot afford to have any false positives?

Ans: Precision: It tells us out of all points predicted to be positive, how many are actually positive.

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

For ex:

- Misclassification of a spam email as not spam is somewhat acceptable i.e FN
- However, classifying an important mail as spam can lead to major loss i.e. FP

⇒ i.e. reducing FP is more critical.

What is the range of precision values?

Ans. Between 0 to 1.

Which metric to use, when we cannot afford to have any false negatives?

Recall / Sensitivity / Hit Rate: It tells us out of all the actually positive points, how many of them are predicted to be positive

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

For ex:

- Classifying a healthy person as cancerous and carrying out further testing is somewhat acceptable
- However, classifying a person with cancer as healthy can be a life-death situation.

⇒ Here, reducing FN is more critical.

Note:

- **True Positive Rate (TPR):** $TPR = \frac{TP}{TP+FN}$**;** is the same as Recall

What is the range of recall values?

Ans. Between 0 to 1.

What other metrics to look for?

- True Negative Rate (TNR): $TNR = \frac{TN}{FP+TN}$

- TNR tells out of all the actual negative points, how many have been predicted as False.

- Also called as **Specificity / Selectivity**

- False Positive Rate (FPR):$FPR = \frac{FP}{FP+TN}$

- Intuitively, it tells, out of all data points that are actually negative, how many are misclassified as positive

- False Negative Rate (FNR):$FNR = \frac{FN}{FN+TP}$

- Intuitively, it tells, out of all data points that are actually positive, how many are misclassified as negative

What is the F1 score? When is it used?

When both Precision and Recall are equally important measures for the model evaluation, then we use F1-Score:

$$F1\ Score\ =\ \frac{2\times(Precision\ \times Recall)}{Precision+Recall}$$

For example, in a fintech company, giving out loans,

- Loss to business if they give loan to people who are unable to repay it (FP)
- Also, a loss if they miss out on good people who will be able to repay (FN)

⇒ Here, we need to focus on both FP and FN.

Note:

- F1-score is just the Harmonic mean of Precision and Recall.
- The F1 score Is also Useful when data is imbalanced.
- Range - [0, 1]

Why do we take harmonic mean in the F1 score instead of arithmetic mean?

Harmonic Mean penalizes the reduction in Precision and Recall more than Arithmetic Mean.

Can we adjust the F1 score to give more preference to precision or recall?

A beta parameter can be added to make the F1 Score have more attention on either the Precision score or Recall score.

$$Fbeta\ =\ \frac{(1+beta^2)\ \times Precision\times Recall}{beta^2\times Precision+Recall}$$

- Beta = 2 if Recall is more important than Precision.
- Beta = 0.5 when Precision is more important.