Assessing Classification Model Performance Using the Confusion Matrix

Total Population
N

ACTUAL CLASS

PREDICTED CLASS	Actual Positive $AP = TP + FN$	Actual Negative $AN = FP + TN$
Predicted Positive $PP = TP + FP$	**True Positive** TP	**False Positive** Type I error FP
Predicted Negative $PN = FN + TN$	**False Negative** Type II error FN	**True Negative** TN

Favor **Precision** when false positives are costly.

Precision
also Positive Predictive Value (PPV)
Proportion of predicted positives that were right
$$PPV = \frac{TP}{PP}$$

$PPV = 1 - FDR$ ↑↓ $FDR = 1 - PPV$

Incorrectly reject the null hypothesis H_0

False Discovery Rate
How often model incorrectly discovered alternate hypothesis
$$FDR = \frac{FP}{PP}$$

Negative Predictive Value
Proportion of predicted negatives that were right
$$NPV = \frac{TN}{PN}$$

$NPV = 1 - FOR$ ↑↓ $FOR = 1 - NPV$

Incorrectly fail to reject the null hypothesis H_0

False Omission Rate
How often model incorrectly omitted an alternate hypothesis
$$FOR = \frac{FN}{PN}$$

Favor **Recall** when false negatives are costly.

Recall, Sensitivity
True Positive Rate (TPR)
Coverage of actual positive sample
$$TPR = \frac{TP}{AP}$$

$TPR = 1 - FNR$ ↑↓ $FNR = 1 - TPR$

Miss Rate
False Negative Rate (FNR)
Type II Error Rate (β)
$$FNR = \frac{FN}{AP}$$

Higher **Specificity** indicates fewer false positives.

Fallout
False Positive Rate (FPR)
Type I Error Rate (α)
$$FPR = \frac{FP}{AN}$$

$FPR = 1 - TNR$ ↑↓ $TNR = 1 - FPR$

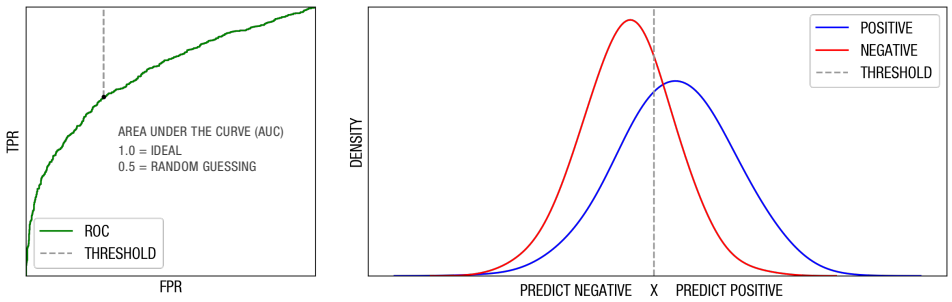
Specificity
True Negative Rate (TNR)
Coverage of actual negative sample
$$TNR = \frac{TN}{AN}$$

Accuracy
How much did the model get right overall?
$$ACC = \frac{TP + TN}{N}$$

$ACC = 1 - ER$ ↑↓ $ER = 1 - ACC$

Error Rate
How much did the model get wrong overall?
$$ER = \frac{FP + FN}{N}$$

Receiver Operating Characteristic (ROC) Curve
Illustrate the tradeoff between model specificity and recall



F_1 Score
Harmonic mean of precision and recall
$$F_1 = 2\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

F_β Score
A generalization of F_1 score such that recall is β times more important than precision
$$F_\beta = (1 + \beta^2)\frac{\text{Precision} \times \text{Recall}}{\beta^2 \times \text{Precision} + \text{Recall}}$$

smaller β → emphasize precision, accept more FNs
larger β → emphasize recall, accept more FPs

Michael Burkhardt (mihobu.lol) Updated 2024-11-19