Binary Classification Metrics Cheat Sheet

ROSCODE TECH

roscoekerby

September 14, 2025

This sheet covers useful metrics for binary classification, with consistent definitions, meanings, and worked examples. It also includes threshold-dependent curves for quick intuition.

Contents

1 Confusion Matrix and Notation

	Predicted Positive	**Predicted Negative**
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

Total samples: $N = TP + TN + FP + FN$.

Running example. Unless stated otherwise, use: $TP = 50$, $FN = 10$, $FP = 5$, $TN = 35$, so $N = 100$.

2 Core Metrics

Accuracy

$$Accuracy = \frac{TP + TN}{N}$$

Meaning: Fraction of all predictions that are correct.

Example: $\frac{50+35}{100} = 0.85$.

Sensitivity, Recall, True Positive Rate (TPR)

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

Meaning: Ability to find actual positives.

Example: $\frac{50}{50+10} = 0.83$.

Specificity, True Negative Rate (TNR)

$$Specificity = \frac{TN}{TN + FP}$$

Meaning: Ability to correctly reject actual negatives.

Example: $\frac{35}{35+5} = 0.875$.

Precision, Positive Predictive Value (PPV)

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

Meaning: Reliability of positive predictions.

Example: $\frac{50}{50+5} = 0.91$.

Negative Predictive Value (NPV)

$$NPV = \frac{TN}{TN + FN}$$

Meaning: Reliability of negative predictions.

Example: $\frac{35}{35+10} = 0.78$.

F1 Score

$$F1 = \frac{2\,(\text{Precision} \cdot \text{Recall})}{\text{Precision} + \text{Recall}}$$

Meaning: Harmonic mean of precision and recall. Encourages balance.

Example: With Precision = 0.91 and Recall = 0.83: $F1 = \frac{2 \cdot 0.91 \cdot 0.83}{0.91+0.83} \approx 0.87$.

Balanced Accuracy

$$\text{Balanced Accuracy} = \tfrac{1}{2}\,(\text{Recall} + \text{Specificity})$$

Meaning: Equal weight for positive and negative classes. Useful with imbalance.

Example: $\frac{1}{2}(0.83 + 0.875) = 0.8525$.

Error Rates (FPR, FNR)

$$FPR = \frac{FP}{FP + TN}, \qquad FNR = \frac{FN}{TP + FN}$$

Meaning: FPR is the false alarm rate on negatives. FNR is the miss rate on positives.

Example: $FPR = \frac{5}{40} = 0.125$, $FNR = \frac{10}{60} \approx 0.167$.

Matthews Correlation Coefficient (MCC)

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

Meaning: Correlation between prediction and truth. Robust under class imbalance.

Example: $\frac{50 \cdot 35 - 5 \cdot 10}{\sqrt{55 \cdot 60 \cdot 40 \cdot 45}} \approx 0.74$.

Cohen's Kappa

$$\kappa = \frac{p_o - p_e}{1 - p_e}, \quad p_o = \frac{TP + TN}{N}, \quad p_e = \frac{(TP{+}FN)(TP{+}FP) + (TN{+}FP)(TN{+}FN)}{N^2}$$

Meaning: Agreement beyond chance.

Example: $p_o = 0.85$, $p_e = \frac{3300+1800}{10000} = 0.51$, so $\kappa = \frac{0.85-0.51}{1-0.51} = 0.69$.

Youden's J

$$J = \text{Sensitivity} + \text{Specificity} - 1$$

Meaning: Single summary of informedness.

Example: $0.83 + 0.875 - 1 = 0.705$.

Geometric Mean (G-Mean)

$$G\text{-}Mean = \sqrt{\text{Sensitivity} \cdot \text{Specificity}}$$

Meaning: Balance between catching positives and rejecting negatives.

Example: $\sqrt{0.83 \cdot 0.875} \approx 0.852$.

---

3 Calibration Metrics

Log Loss (Cross Entropy)

$$\text{LogLoss} = -\frac{1}{N} \sum_{i=1}^{N} \Big[y_i \log(p_i) + (1 - y_i) \log(1 - p_i) \Big]$$

Meaning: How well predicted probabilities match outcomes. Lower is better.

Example: If a positive is predicted with $p = 0.9$, contribution is $-\log(0.9) = 0.105$. If it was actually negative, contribution would be $-\log(1 - 0.9) = 2.303$.

Brier Score

$$\text{Brier} = \frac{1}{N} \sum_{i=1}^{N} (p_i - y_i)^2$$

Meaning: Mean squared error of predicted probabilities. Lower is better.

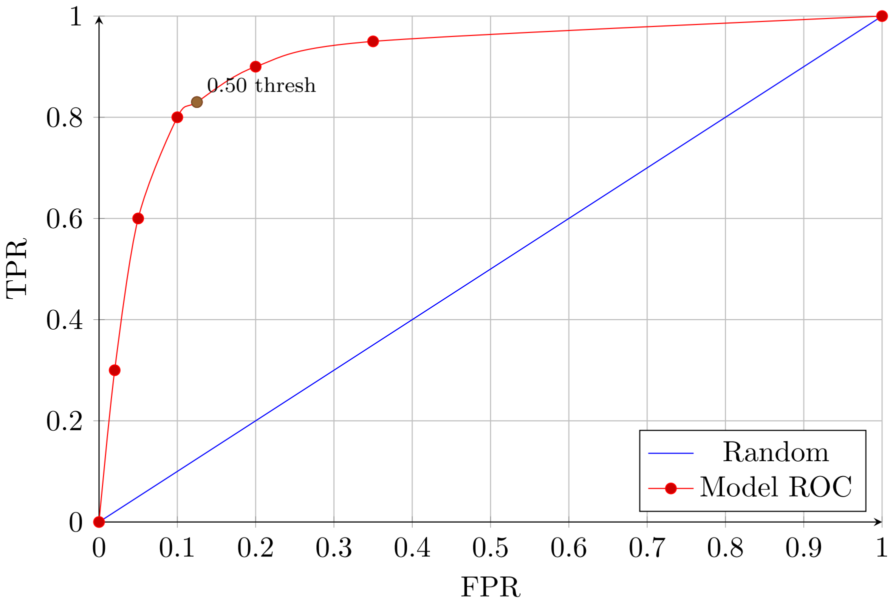
Example: For three cases with $(p, y) = (0.8, 1), (0.4, 0), (0.3, 1)$: score $= \frac{(0.2)^2 + (0.4)^2 + (0.7)^2}{3} \approx 0.203$.

---

4 Threshold Curves

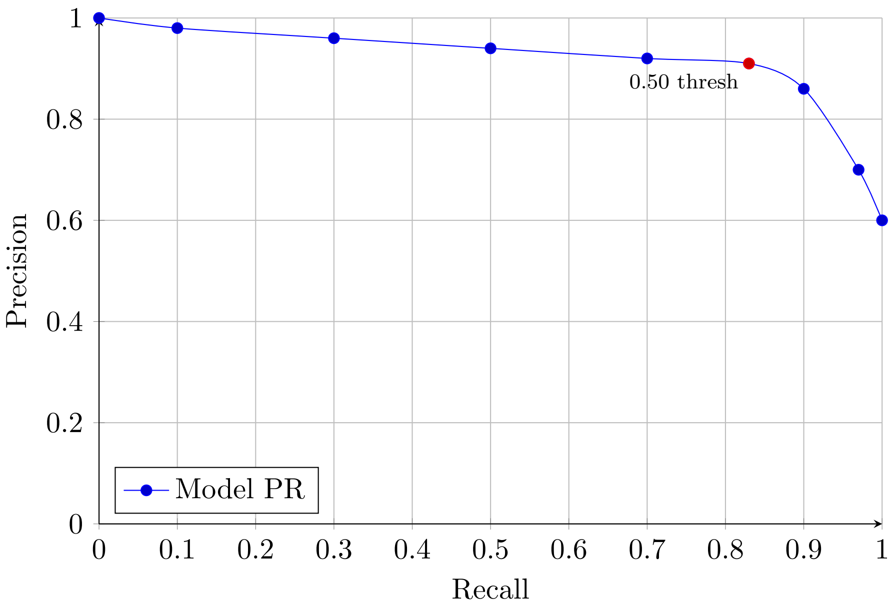
ROC Curve and AUC

Formula. Plot $TPR = \frac{TP}{TP+FN}$ against $FPR = \frac{FP}{FP+TN}$ while sweeping the decision threshold from 1 to 0. **Meaning.** Trade-off between sensitivity and false alarms. **Example.** At threshold 0.50: $TPR = 0.83$, $FPR = 0.125$.



Precision–Recall Curve

Formula. Plot Precision $= \frac{TP}{TP+FP}$ versus Recall $= \frac{TP}{TP+FN}$ across thresholds. **Meaning.** Often more informative than ROC when positives are rare. **Example.** At threshold 0.50: Precision = 0.91, Recall = 0.83.



---

5 Quick Tips

- Use Balanced Accuracy, MCC, F1, and PR curves when classes are imbalanced.
- Move the threshold to trade precision for recall, or the reverse. Choose based on the application cost.
- F1 ignores true negatives. If true negatives matter, consider MCC or Balanced Accuracy.

- AUC is threshold independent, yet PR curves can be more revealing when positives are rare.
- Check calibration with Log Loss or Brier before using probabilities downstream.