Introduction to Machine Learning: Evaluation and Training Math Review

Evan Misshula

June 11, 2025

Definition

- A confusion matrix is a table that summarizes the performance of a classification model.
- It compares the predicted labels with the actual labels.
- Especially useful for binary or multiclass classification.

Binary Confusion Matrix

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

Notation

Let:

- $y_i \in \{0, 1\}$ be the true label
- $\hat{y}_i \in \{0, 1\}$ be the predicted label

Then:

- TP: $\sum \mathbb{1}_{\{y_i=1 \wedge \hat{y}_i=1\}}$
- TN: $\sum \mathbb{1}_{\{y_i=0 \wedge \hat{y}_i=0\}}$
- FP: $\sum \mathbb{1}_{\{y_i=0 \wedge \hat{y}_i=1\}}$
- FN: $\sum \mathbb{1}_{\{y_i=1 \wedge \hat{y}_i=0\}}$

Accuracy

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

- Proportion of total predictions that were correct.
- Best used when classes are balanced.

Precision

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

- Among predicted positives, how many were actually positive?

Recall (Sensitivity, TPR)

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

- Among actual positives, how many did we correctly predict?

F1 Score

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

- Harmonic mean of precision and recall.
- Balances false positives and false negatives.

Specificity

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

- Among actual negatives, how many did we correctly predict?

Generalization

- For K classes, the confusion matrix is $K \times K$
- Entry (i, j) is the number of times class i was predicted as class j

Use Cases

- Visualizing classifier performance
- Identifying types of errors
- Computing per-class precision and recall

Key Takeaways for the Confusion Matrix

- Confusion matrix is foundational for classifier evaluation.
- Metrics derived from it (precision, recall, F1) offer deeper insight than accuracy alone.
- Always inspect confusion matrices especially on imbalanced datasets.

Motivation

- ROC AUC is a standard metric for evaluating binary classifiers.
- Focuses on ranking predictions rather than absolute accuracy.
- Especially useful with imbalanced data or when decision thresholds vary.

What Is the ROC Curve?

What Is the ROC Curve?

- Receiver Operating Characteristic (ROC) curve:
  - A graphical plot that shows the trade-off between True Positive Rate (TPR) and False Positive Rate (FPR).
- The curve is constructed by sweeping a decision threshold over the predicted probabilities output by the model.

Understanding the Threshold

- Most classifiers (like logistic regression) output a probability score $\hat{p} \in [0, 1]$.
- We need to decide: at what probability value do we say "yes, this is a positive"?
- This cut-off value is called the threshold.

Example:

- If threshold $= 0.5$:
  - $\hat{p} \geq 0.5 \Rightarrow$ predict positive
  - $\hat{p} < 0.5 \Rightarrow$ predict negative
- Lowering the threshold means more predictions are labeled positive, increasing TPR but also increasing FPR.
- Raising the threshold means fewer predictions are labeled positive, which may reduce FPR but also lower TPR.

Each point on the ROC curve corresponds to:

- A different threshold
- A pair (FPR, TPR) computed using that threshold
- Sweeping the threshold from 0 to 1 traces out the entire ROC curve