

ROC Curve and Its Usage in Machine Learning

Rewan Khaled

August 29, 2025

1 Introduction

The Receiver Operating Characteristic (ROC) curve is a graphical tool used to evaluate the performance of a binary classifier. It plots the True Positive Rate (TPR) against the False Positive Rate (FPR) at various threshold settings.

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad \text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}}$$

where TP, FP, TN, and FN are the counts of true positives, false positives, true negatives, and false negatives, respectively.

2 Constructing the ROC Curve

To construct the ROC curve:

1. Obtain classifier scores or probabilities for all instances.
2. Choose a threshold to classify instances as positive or negative.
3. Calculate TPR and FPR at that threshold.
4. Vary the threshold from 0 to 1 and plot TPR vs FPR.

3 Interpretation

- The closer the ROC curve is to the top-left corner, the better the classifier.
- A diagonal line (from bottom-left to top-right) represents random guessing.

4 Area Under the Curve (AUC)

The AUC measures the overall performance:

$$\text{AUC} = \int_0^1 \text{TPR}(\text{FPR}) d(\text{FPR})$$

- AUC = 1.0: perfect classifier
- AUC = 0.5: random classifier
- Higher AUC indicates better performance

5 Practical Usage

ROC curves are widely used in:

- Medical diagnostics (disease detection)
- Spam detection
- Fraud detection

Advantages:

- Evaluates classifier performance across all thresholds
- Works well with imbalanced datasets

Limitations:

- Only applicable for binary classification
- May be misleading with highly imbalanced classes

6 Conclusion

ROC curves provide an intuitive way to visualize classifier performance and compare models. By analyzing TPR vs FPR across thresholds, one can select optimal operating points and assess model reliability.