



JOHNS HOPKINS

WHITING SCHOOL
of ENGINEERING

685.621 Algorithms for Data Science

Supervised Learning: Model Evaluation

Why Do We Need Model Evaluation

- **Purpose of Evaluation:**

- Determines how well the model generalizes to unseen data.
- Helps compare different models objectively.
- Identifies overfitting and underfitting.

- **Two types of evaluation:**

- **Training Performance:** How well the model fits the training data.
- **Generalization Performance:** How well the model performs on unseen data.

Breaking Down Predictions

- **True Positives (TP):**
Correctly classified positive cases.
- **False Positives (FP):**
Incorrectly classified as positive.
- **True Negatives (TN):**
Correctly classified negative cases.
- **False Negatives (FN):**
Incorrectly classified as negative.

		PREDICTED LABEL	
		Positive	Negative
TRUE LABEL	Positive	TP	FN
	Negative	FP	TN

Accuracy: Is It Always Reliable?

- **Pros:**

- Simple to understand and compute
- Works well when class distribution is balanced

- **Cons:**

- Misleading in imbalanced datasets.
- Doesn't distinguish between different types of errors

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

Precision vs. Recall

- **Precision (Positive Predictive Value)**

- Measures how many predicted positives were correct.
- High precision means fewer false positives.

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

- **Recall (Sensitivity/True Positive Rate)**

- Measures how many actual positives were detected
- High recall means fewer false negatives

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

- **Use cases:**

- **High Precision Needed?** Spam detection (false positives matter)
- **High Recall Needed?** Disease diagnosis (false negatives matter)

The Harmonic Mean of Precision and Recall

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

- **Why use F1-Score?**

- Balances Precision and Recall
- Useful for imbalanced datasets.
- Helps when neither metric alone tells the full story.

Measuring Model Discrimination

- **ROC Curve**

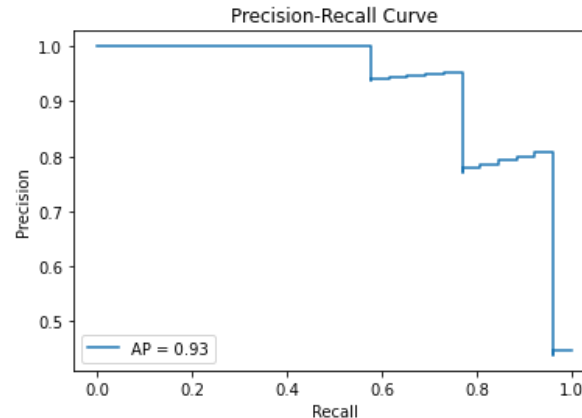
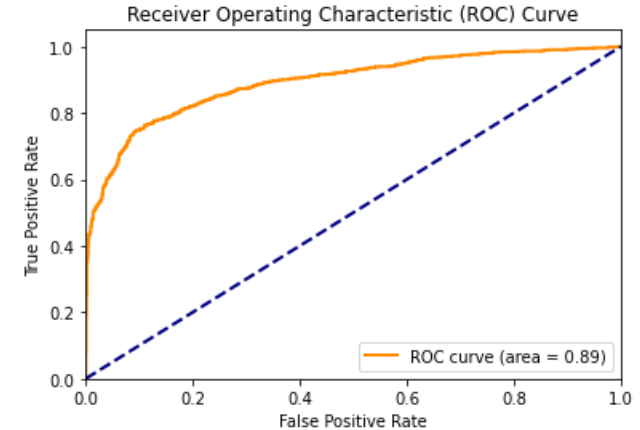
- Plots **True Positive** Rate vs. **False Positive** Rate

- **AUC (Area Under Curve)**

- **AUC** = 1.0 → Perfect classifier
 - **AUC** = 0.5 → Random Guessing

- **Precision-Recall**

- More **informative** than ROC when dealing with **imbalanced** data.
 - Highlights performance on the **positive class**.





JOHNS HOPKINS

WHITING SCHOOL
of ENGINEERING

© The Johns Hopkins University 2024, All Rights Reserved.