

## Lesson 27 Summary: Classification Metrics

Data Science with Python – Key Concepts

Data Science Program

## Classification Metrics

### Confusion Matrix

TP, FP, TN, FN  
Base for all metrics  
Row=actual, Col=pred

### Precision & Recall

Precision:  $TP/(TP+FP)$   
Recall:  $TP/(TP+FN)$   
Tradeoff exists

### F1 Score

Harmonic mean  
 $2 * P * R / (P + R)$   
Balanced metric

### ROC Curve & AUC

TPR vs FPR at thresholds  
AUC: 0.5=random, 1.0=perfect

### Threshold Tuning

Default 0.5 may not be optimal  
Adjust based on cost of errors

*precision\_score | recall\_score | f1\_score | roc\_auc\_score*

Choose metrics based on what errors cost your business

## Foundation of classification metrics:

- **TP:** True Positives (correctly predicted positive)
- **FP:** False Positives (Type I error)
- **TN:** True Negatives (correctly predicted negative)
- **FN:** False Negatives (Type II error)

---

All classification metrics derive from TP, FP, TN, FN

## When accuracy fails:

- **Imbalanced data:** 99% negative = 99% accuracy by guessing
- **Cost asymmetry:** Missing fraud is worse than false alarm
- **Alternative:** Balanced accuracy = mean of per-class recalls

---

Accuracy is misleading for imbalanced datasets

## Complementary metrics:

- **Precision:**  $TP / (TP + FP)$  – of predicted positive, how many correct?
- **Recall:**  $TP / (TP + FN)$  – of actual positive, how many found?
- **Tradeoff:** Increasing one often decreases the other

---

High precision = few false alarms; high recall = few missed

## Balancing precision and recall:

- **Formula:**  $2 \times \frac{P \times R}{P + R}$
- **Property:** Harmonic mean (penalizes imbalance)
- **Range:** 0 to 1 (higher is better)

---

F1 is useful when you care equally about precision and recall

## Visualizing threshold tradeoffs:

- **X-axis:** False Positive Rate ( $FPR = FP / (FP + TN)$ )
- **Y-axis:** True Positive Rate ( $TPR = Recall$ )
- **Curve:** Shows TPR vs FPR at all thresholds

---

The diagonal line represents random guessing

# AUC (Area Under Curve)

## Single number summary:

- **AUC = 0.5:** Random classifier
- **AUC = 1.0:** Perfect classifier
- **Interpretation:** Probability of ranking positive higher

---

**AUC is threshold-independent and robust**

## Choosing the right cutoff:

- **Default 0.5:** May not be optimal
- **Cost-based:** Minimize expected cost
- **F1-optimal:** Threshold that maximizes F1

---

**Business requirements should drive threshold choice**

## sklearn's summary function:

- **Per-class:** Precision, recall, F1 for each class
- **Averages:** Macro, weighted, micro
- **Support:** Number of samples per class

---

Use `classification_report()` for comprehensive overview

### Essential Commands:

Metric	Code
Confusion matrix	<code>confusion_matrix(y, pred)</code>
Precision	<code>precision_score(y, pred)</code>
Recall	<code>recall_score(y, pred)</code>
F1	<code>f1_score(y, pred)</code>
ROC AUC	<code>roc_auc_score(y, prob)</code>
Report	<code>classification_report(y, pred)</code>

---

Match your metric choice to your business objective