Precision, Recall, and Accuracy

Introduction

Precision, recall, and accuracy are key metrics used to assess a model's performance. Understanding these metrics helps in selecting the right model and fine-tuning it for better results.

1. Accuracy

Definition: Accuracy is the ratio of correctly predicted observations to the total observations. It answers the question: "What fraction of the total predictions were correct?"

Formula: $\text{Accuracy} = \text{TP+TN} / \text{TP+TN+FP+FN}$

- **TP** (True Positives): Correctly predicted positive observations.
- **TN** (True Negatives): Correctly predicted negative observations.
- **FP** (False Positives): Incorrectly predicted positive observations.
- **FN** (False Negatives): Incorrectly predicted negative observations.

Example: If you have 100 predictions, and 90 are correct, the accuracy is 90%.

2. Precision

Definition: It is the ratio of correctly predicted positive observations to the total predicted positive observations.

It answers the question: "How many of the predicted positives are actually positive?"

Formula: $\text{Precision} = \text{TP} / \text{TP+FP}$

Example: If a model predicts 10 positive cases, and 7 are actually positive, the precision is 70%.

3. Recall/ Sensitivity/ True Positive Rate

Definition: It is the ratio of correctly predicted positive observations to all observations in the actual class. It answers the question: "How many of the actual positives were captured by the model?"

Formula: $\text{Recall} = \text{TP} / \text{TP+FN}$

Example: If there are 20 actual positive cases, and the model correctly identifies 15, the recall is 75%.

Code Example:

```python
import numpy as np
from sklearn.metrics import accuracy_score, precision_score, recall_score

y_true = np.array([1, 0, 1, 1, 0, 1, 0, 0, 1, 1]) # Actual labels
y_pred = np.array([1, 0, 1, 0, 0, 1, 0, 0, 1, 0]) # Predicted labels

accuracy = accuracy_score(y_true, y_pred)

precision = precision_score(y_true, y_pred)

recall = recall_score(y_true, y_pred)
```