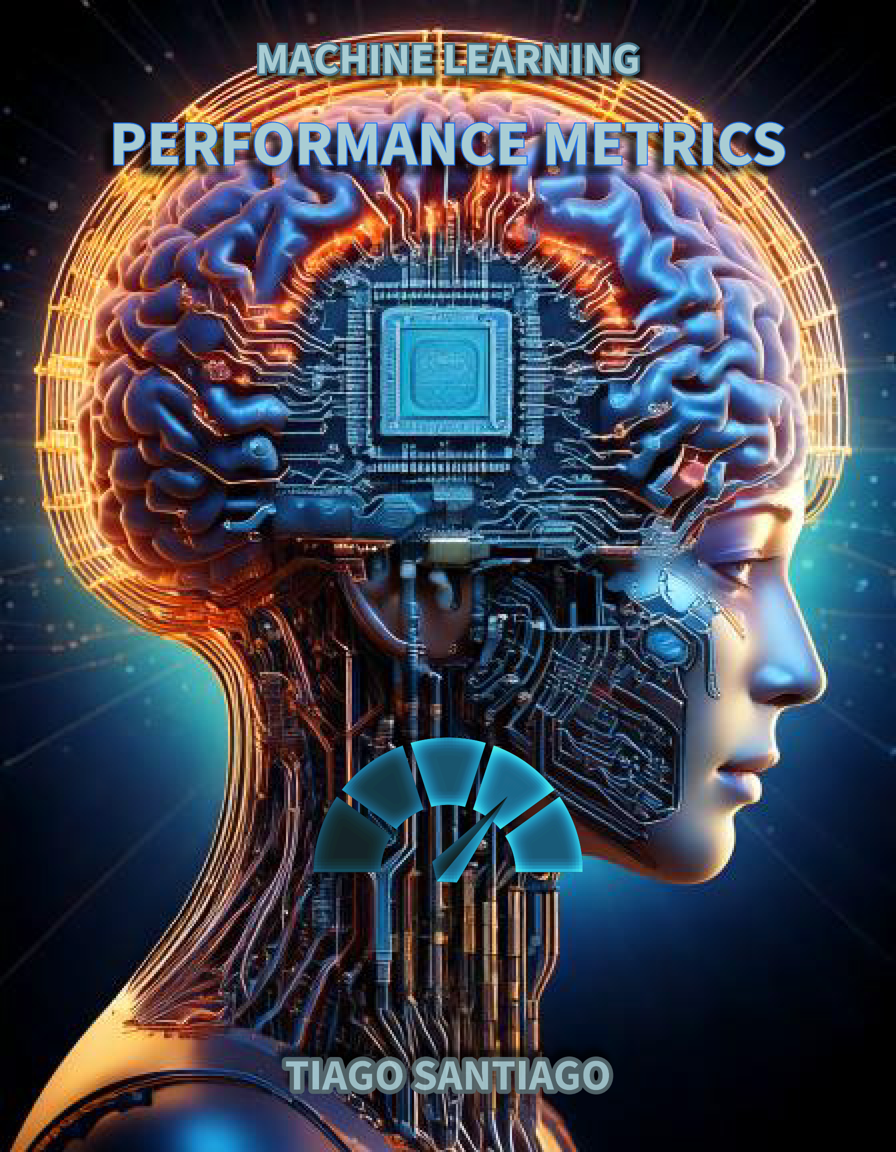
MACHINE LEARNING
PERFORMANCE METRICS



TIAGO SANTIAGO

Table of Contents

Introduction to Machine Learning Performance Metrics 1
1. Accuracy 2
2. Precision 3
3. Recall (Sensitivity) 4
4. F1-Score 5
5. Mean Absolute Error (MAE) 6
6. Mean Squared Error (MSE) 7
7. R-squared (Coefficient of Determination) 8
8. Logarithmic Loss (Log Loss) 9
9. Area Under the Curve (AUC) 11
10. Confusion Matrix 12
Conclusion 14

Introduction to Machine Learning Performance Metrics

Machine learning models rely on performance metrics to evaluate their effectiveness. This e-book provides an introduction to some of the most commonly used metrics, offering definitions, use cases, and examples for better understanding.

1. Accuracy

Accuracy measures the proportion of correct predictions out of the total number of predictions made.

Formula:

```
Accuracy = (True Positives + True
Negatives) / Total Predictions
```

Example:

A classifier predicts whether an email is spam. Out of 100 emails, it correctly identifies 90 as either spam or non-spam. The accuracy is:

```
Accuracy = (90) / (100) = 0.9 or 90%
```

2. Precision

Precision measures the proportion of true positive predictions out of all positive predictions made by the model. It is important when the cost of false positives is high.

Formula:

```
Precision = True Positives / (True Positives + False Positives)
```

Example:

A medical test predicts whether a patient has a disease. Out of 50 positive predictions, 40 are correct. The precision is:

```
Precision = (40) / (40 + 10) = 0.8 or 80%
```

3. Recall (Sensitivity)

Recall measures the proportion of true positive predictions out of all actual positive cases. It is crucial when the cost of false negatives is high.

Formula:

```
Recall = True Positives / (True
Positives + False Negatives)
```

Example:

In the same medical test, there are 60 actual positive cases, and the test correctly identifies 40 of them. The recall is:

```
Recall = (40) / (40 + 20) = 0.67 or 67%
```

4. F1-Score

The F1-score is the harmonic mean of precision and recall, providing a balance between the two metrics.

Formula:

```
F1-Score = 2 * (Precision * Recall) /
(Precision + Recall)
```

Example:

With a precision of 80% and a recall of 67%, the F1-score is:

```
F1-Score = 2 * (0.8 * 0.67) / (0.8 +
0.67) ≈ 0.73 or 73%
```

5. Mean Absolute Error (MAE)

MAE measures the average absolute difference between predicted and actual values in regression tasks.

Formula:

```
MAE = (Σ|Predicted - Actual|) / Total
Predictions
```

Example:

For predicted house prices `[250k, 300k, 400k]` and actual prices `[260k, 310k, 390k]`:

```
MAE = (|250-260| + |300-310| + |400-
390|) / 3 = 10k
```

6. Mean Squared Error (MSE)

MSE measures the average squared difference between predicted and actual values, penalizing larger errors.

Formula:

```
MSE = (Σ(Predicted - Actual)^2) / Total Predictions
```

Example:

Using the same house prices:

```
MSE = ((250-260)^2 + (300-310)^2 + (400-390)^2) / 3 = 100k^2
```

7. R-squared (Coefficient of Determination)

R-squared evaluates how well the predictions explain the variance in the actual data.

Formula:

```
R^2 = 1 - (Σ(Predicted - Actual)^2 /
Σ(Actual - Mean)^2)
```

Example:

For house prices, if the variance of residuals is 50k^2 and the total variance is 200k^2:

```
R^2 = 1 - (50 / 200) = 0.75 or 75%
```

8. Logarithmic Loss (Log Loss)

Log Loss measures the performance of a classification model by comparing the predicted probabilities with the actual class labels.

Formula:

```
Log Loss = - (1/N) Σ [y * log(p) + (1-y) * log(1-p)]
```

Example:

For a binary classification problem with true labels `[1, 0, 1]` and predicted probabilities `[0.9, 0.1, 0.8]`:

```
Log Loss = -(1/3) * [(1*log(0.9)) +
(0*log(0.1)) + (1*log(0.8))] ≈ 0.164
```

9. Area Under the Curve (AUC)

AUC evaluates the performance of a classification model by analyzing the tradeoff between true positive rate and false positive rate across different thresholds.

Formula:

AUC is the area under the Receiver Operating Characteristic (ROC) curve.

Example:

For a model with an ROC curve, the AUC is calculated numerically. If the AUC is 0.85, the model distinguishes between classes 85% of the time.

10. Confusion Matrix

A confusion matrix provides a summary of prediction results, showing the counts of true positives, true negatives, false positives, and false negatives.

Structure:

```
              Pred. Pos.    Pred. Neg.
Actual Pos.     TP              FN
Actual Neg.     FP              TN
```

Example:

For a binary classification: - True Positives (TP): 50 - True Negatives (TN): 40 - False Positives (FP): 10 - False Negatives (FN): 5

The confusion matrix is:

```
              Pred. Pos.    Pred. Neg.
Actual Pos.      50             5
Actual Neg.      10            40
```

Conclusion

Understanding these metrics is crucial for evaluating and improving machine learning models. Each metric serves a specific purpose, and choosing the right one depends on the problem at hand.