

# AI and Machine Learning Algorithms and Techniques

David<sup>1</sup>

<sup>1</sup> Santiago.

## 1. Evaluation Metrics for Supervised Learning Models

### 1.1. Introduction

Evaluation metrics play a critical role in assessing the performance of supervised learning models. These metrics help us understand how well a model predicts outcomes and whether it can generalize to unseen data. Different tasks, such as classification and regression, require different evaluation metrics.

By the end of this reading, you'll be able to:

- **Identify key evaluation metrics:** Understand and describe the most commonly used evaluation metrics for classification and regression models.
- **Apply metrics to model performance:** Evaluate the performance of ML models using appropriate metrics such as accuracy, precision, recall, mean squared error (MSE), and R-squared.
- **Choose the right metric for the task:** Select the most suitable evaluation metric based on the specific problem and dataset characteristics, ensuring accurate model assessment.

### 1.2. Evaluation Metrics for Classification Models

Classification models predict discrete outcomes, such as whether an email is spam or not, or

whether a customer will churn or remain. Below are some key evaluation metrics used to assess the performance of classification models:

#### 1.2.1. Accuracy

Accuracy measures the percentage of correct predictions out of all predictions made.

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}}$$

**Example:** If a model correctly predicts 90 out of 100 instances, its accuracy is 90%. However, accuracy may not always be the best metric for imbalanced datasets.

#### 1.2.2. Precision

Precision measures the percentage of true positive predictions out of all positive predictions that the model makes. It is important in cases where false positives are costly.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

**Example:** In spam detection, precision is the proportion of emails predicted as spam that are actually spam.

#### 1.2.3. Recall (Sensitivity or True Positive Rate)

Recall measures the percentage of true positive predictions out of all actual positives. It is important when the cost of missing positive instances is

high.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

**Example:** In a cancer detection model, recall is the proportion of actual cancer cases that the model correctly identifies.

#### 1.2.4. F1 Score

The F1 score is the harmonic mean of precision and recall. It provides a balanced metric when both precision and recall are important, especially for imbalanced datasets.

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

**Example:** A model with high precision but low recall or vice versa will have a lower F1 score.

#### 1.2.5. Confusion Matrix

A confusion matrix is a table used to summarize the performance of a classification model. It shows the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN).

	Predicted Positive	Predicted Negative
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

From this matrix, you can calculate accuracy, precision, recall, and other metrics.

#### 1.2.6. ROC Curve and AUC

The receiver operating characteristic (ROC) curve plots the true positive rate (recall) against the false positive rate at different threshold levels. The area under the curve (AUC) measures the overall performance of the classifier. AUC ranges from 0 to 1, where a value closer to 1 indicates a better-performing model.

### 1.3. Evaluation Metrics for Regression Models

Regression models predict continuous values, such as house prices or temperatures. The following metrics are commonly used to evaluate regression models:

#### 1.3.1. Mean Squared Error (MSE)

MSE measures the average squared difference between the predicted and actual values.

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

**Example:** If the predicted price is \$200,000 and the actual price is \$250,000, the MSE for that prediction is  $2,5 \times 10^9$ .

#### 1.3.2. Root Mean Squared Error (RMSE)

RMSE is the square root of the mean squared error, which brings the error metric back to the same units as the target variable.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

**Example:** If the MSE is \$625,000,000, then the RMSE will be \$25,000.

#### 1.3.3. Mean Absolute Error (MAE)

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

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

**Example:** In a weather forecasting model, MAE tells you the average difference between the predicted and actual temperatures.

### 1.3.4. R-squared (Coefficient of Determination)

R-squared explains the proportion of variance in the dependent variable that is predictable from the independent variable(s).

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

**Example:** An  $R^2$  value of 0.9 means that 90 % of the variance in house prices is explained by the model's input features.

### 1.3.5. Adjusted R-squared

Adjusted R-squared adjusts the  $R^2$  value based on the number of features in the model.

$$\text{Adjusted } R^2 = 1 - \left( \frac{(1 - R^2)(n - 1)}{n - p - 1} \right)$$

Where  $n$  is the number of data points and  $p$  is the number of predictors.

**Example:** If adding more features to a model decreases the adjusted  $R^2$ , it suggests that the additional features are not improving the model.

## 1.4. Choosing the Right Evaluation Metric

Choosing the right evaluation metric depends on the problem and the nature of the data. For ins-

tance:

- For imbalanced classification problems, use precision, recall, F1 score, or ROC AUC instead of accuracy.
- For regression models, if large errors are particularly undesirable, consider using RMSE or MSE. If you want a metric that is less sensitive to outliers, use MAE.
- For complex models, look at  $R^2$  and adjusted  $R^2$  to assess how well the model explains the variance in the target variable.

## 1.5. Conclusion

Evaluation metrics are essential for understanding and improving ML models. By using the right metrics, you can accurately assess model performance, make necessary adjustments, and ensure that your model is well suited for the task at hand. Whether you're working with classification or regression problems, these metrics will provide you with the insight needed to create reliable and effective machine learning solutions.

## 2. Model evaluation metrics

### 2.1. aa