What is the definition and example of Misclassification Rate in Machine Learning?

Authored by

stats writer

June 29, 2024

Misclassification rate, also known as error rate, is a metric used in machine learning to evaluate the accuracy of a classification model. It measures the percentage of incorrectly classified data points from the total number of data points in a dataset. This metric is important in determining the effectiveness and performance of a classification model. For example, if a model has a misclassification rate of 10%, it means that 10 out of 100 data points were classified incorrectly. A lower misclassification rate indicates a more accurate model.

Misclassification Rate in Machine Learning: Definition & Example

In machine learning, misclassification rate is a metric that tells us the percentage of observations that were incorrectly predicted by some .

It is calculated as:

Misclassification Rate = # incorrect predictions / # total predictions

The value for misclassification rate can range from 0 to 1 where:

0 represents a model that had zero incorrect predictions.1 represents a model that had completely incorrect predictions.

The lower the value for the misclassification rate, the better a classification model is able to predict the

outcomes of the .

The following example show how to calculate misclassification rate for a in practice.

Example: Calculating Misclassification Rate for a Logistic Regression Model

Suppose we use a logistic regression model to predict whether or not 400 different college basketball players get drafted into the NBA.

The following confusion matrix summarizes the predictions made by the model:

		Predicted	
		Drafted = Yes	**Drafted = No**
Actual	**Drafted = Yes**	**120** (True Positive)	**40** (False Negative)
	Drafted = No	**70** (False positive)	**170** (True Negative)

Here is how to calculate the misclassification rate for the model:

Misclassification Rate = # incorrect predictions / # total predictionsMisclassification Rate = (false positive + false negative) / (total predictions)Misclassification Rate = (70 + 40) / (400)Misclassification Rate = 0.275

The misclassification rate for this model is 0.275 or 27.5%.

This means the model incorrectly predicted the outcome for 27.5% of the players.

The opposite of misclassification rate would be accuracy, which is calculated as:

Accuracy = 1 - Misclassification rateAccuracy = 1 - 0.275Accuracy = 0.725

Pros & Cons of Misclassification Rate

Misclassification rate offers the following pros:

It's easy to interpret. A misclassification rate of 10% means a model made an incorrect prediction for 10% of the total observations.It's easy to calculate. A misclassification rate is calculated as the number of total incorrect predictions divided by the total number of predictions.

However, misclassification rate has the following con:

It doesn't take into account how the data is distributed. For example, suppose 90% of all players do not get

drafted into the NBA. If we have a model that simply predicts every player to not get drafted, the model would have a misclassification rate of just 10%. This seems low, but, but the model is actually unable to correctly predict any player who gets drafted.

In practice, we often calculate the misclassification rate of a model along with other metrics like:

Sensitivity: The "true positive rate" - the percentage of positive outcomes the model is able to detect.Specificity: The "true negative rate" - the percentage of negative outcomes the model is able to detect.F1 Score: A that tells us the accuracy of a model, relative to how the data is distributed.

By calculating the value for each of these metrics, we can gain a full understanding of how well the model is able to make predictions.

Additional Resources

The following tutorials provide additional information about common machine learning concepts: