

Introduction to evaluating binary classifiers: Takeaways



by Dataquest Labs, Inc. - All rights reserved © 2021

Concepts

- Prediction accuracy is the simplest way to determine the effectiveness of a classification model. Prediction accuracy can be calculated by the number of labels correctly predicted divided the total number of observations:

$$Accuracy = \frac{\# \text{ of Correctly Predicted}}{\# \text{ of Observations}}$$

- A discrimination threshold is used to determine what labels are assigned based on their probability. Scikit-learn sets the discrimination threshold to 0.5 by default when predicting labels.
 - For example, if the predicted probability is greater than 0.5, the label for that observation is 1. If it is less than 0.5, the label for that observation is 0.

- There are four different outcomes of a binary classification model:

- **True Positive:** The model correctly predicted the label as positive.
- **True Negative:** The model correctly predicted the label as negative.
- **False Positive:** The model falsely predicted the label as positive.
- **False Negative:** The model falsely predicted the label as negative.

- Sensitivity or True Positive Rate, is the proportion of labels that were correctly predicted as positive. Mathematically, this is written as:

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

- Sensitivity helps answer "How effective is this model at identifying positive outcomes?"

- Specificity or True Negative Rate, is the proportion of labels that were correctly predicted as negative. Mathematically, this is written as:

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

- Specificity helps answer "How effective is the model at identifying negative outcomes?"

Resources

- [Sensitivity and Specificity](#)
- [Discrimination threshold](#)