

Classification Metrics - Introduction



(<https://github.com/learn-co-curriculum/dsc-classification-metrics-intro/issues/new/choose>)

Introduction

Classification models can be more complex to evaluate than regression models. There are more trade-offs involved as well as different metrics that can be used.

Evaluating Classifiers

We'll look at the practicalities of evaluating logistic regression models based on precision, recall, and accuracy to evaluate other classifiers.

We also take a little time to look at how to plot a confusion matrix for a logistic regression classifier and introduce a couple of key concepts for determining the optimal precision-recall trade-off for a given classifier - Receiver Operating Characteristic (ROC) curves and AUC (the Area Under the Curve).

Class Imbalance Problems

We then introduce the concept of class imbalance. Imagine a classifier for cancer where only 1 screened individual in 1000 is sick. You could obtain over 99 percent accuracy by just saying everyone is fine, but that wouldn't be a very useful approach. We look at the ideas of class weights and over/undersampling and how they can be used to work with highly imbalanced classes.

Summary

Many of the concepts around model evaluation will be useful whenever you're trying to solve a classification problem.

How do you feel about this lesson?



Have specific feedback?

Tell us here!  (<https://github.com/learn-co-curriculum/dsc-classification-metrics-intro/issues/new/choose>)

