Introduction to the Problem

Bookmark this page

To emphasize the problem of relying on a single metric such as accuracy for prediction, we'll compare scenarios from different countries' COVID-19 response.

Problem Setting

At the peak of the COVID-19 pandemic, hospital authorities had to make a call about who to admit and who to send home given the limited available resources. Our problem is to have a classifier that suggests whether a patient should be immediately admitted to the hospital or sent home.

The Data

The data use the following primary predictors:

- age
- sex
- cough, fever, chills, sore throat, headache, fatigue

The outcome is a classification prediction to indicate the **urgency of admission**.

- Positive: indicates a patient that was admitted within 1 day from the onset of symptoms.
- Negative: indicates everyone else.

> **Issues**
>
> While this case study tries to mimic a real-life problem, it is important to note that this analysis is for **educational purposes only.**
>
> - The data is sourced through online forms and thus is of questionable accuracy.
> - A large portion of the original dataset collected had missing values. This was ignored for a simpler analysis.
> - The entire premise of predicting urgency of admission is false because some people had to wait longer to be admitted because of lack of hospital beds and resources.

For this problem setting, we examine two different models: Logistic Regression and kNN Classification. The goal is to train both these models and report the accuracy.

Discussion Board (External resource)

Haga clic en Aceptar para que su nombre de usuario y dirección de correo electrónico se envíen a una aplicación de terceros.

Aceptar

‹ Previous Next ›

Calculator