Classification

Bookmark this page

Classification is the problem of predicting a discrete label- for example, predicting whether the patient has the value "Yes" or "No" in the AHD (heart disease) column.

Skip to after table

Age	RestBP	Thal	...	AHD
63	145	Fixed	...	No
67	108	Normal	...	Yes
...	...	...	...	...
41	130	Reversable	...	No

A key example of classification is medical diagnosis. We can use one (or more) of the features in the data to try and predict whether AHD will be true or false. This can help doctors find patients most at risk of developing heart disease. Similar to regression, we can use any number of features as the inputs to our model.

Our model's response/target variables in a classification problem are categorical. Some examples of categorical targets are:

- The **binary** target of whether a patient has heart disease
- The **3-dimensional** target of a self-driving car scanning a traffic light to determine if the light is green, yellow, or red.
- The **5-dimensional** target of a weather-prediction model that is trying to determine if tomorrow's description will be Rain, Snow, Cloudy, Sunny, or Partly Cloudy
- Grouping customers into similar "types"

In the following section we will focus on two algorithms for classification:

- k-Nearest Neighbors classification (similar to the kNN algorithm for regression)
- Logistic regression

Discussion Board (External resource)

Click OK to have your username and e-mail address sent to a 3rd party application.

Legal

Connect

Calculator