Predictive Modeling

Brady Lamson

2022-03-03

Notes taken from Modern Data Science for R textbook. These notes will shamelessly rip everything from this chapter, none of this is my own work.

Vocab

- **Null Model:** Otherwise known as the *mean* or *no predictor* model. The null model only contains one intercept where the intercept is the mean of Y. The simplest prediction in a regression that you can imagine is:
  - Using the mean of the target ($\hat{Y}$) as prediction
  - ie choosing a slope of 0

Packages (tidymodels):

- **parsnip:** Workhorse package for model fitting
- **yardstick:** Provides model evaluation metrics

General Notes

First separate the data set into two pieces by sampling the rows at random. A sample of **80%** of the rows will become the **training data set**, with the other **20%** functioning as the **testing** or "hold-out" data set.

- initial_split() divides the data
- training() and testing() recover the two smaller data sets.