Assignment 3

October 19, 2020

All files referred to in this homework can be found on CourseWorks.

Your hand-in should be made on Gradescope. Your submission is going to be a jupyter notebook that contains answers to all the questions. Please use headings to separate your notebook into answers to each question.

To see examples of how to write latex code in jupyter, see this link.

You must submit *two* files:

1. An `.ipynb` file (i.e. jupyter notebook file) with your data analysis for answering questions
2. A pdf file with your jupyter notebook output. Please use the workflow described here for generating this pdf.

1 Bootstrap Description

Suppose that we have a dataset $(x_i, y_i)_{i=1}^n$, and we fix a value $\bar{x}$. Further suppose that we are going to build a predictor for the response $\bar{y}$ associated to $\bar{x}$ using some statistical learning method. Describe how we might estimate the standard deviation of our prediction. You must explicitly define every variable and equation.

2 Bootstrap for Estimating Standard Errors of Logistic Regression Coefficients

Solve exercise 6 from Chapter 5 of ISLR, in python.

You can use either the `statsmodels` or `sklearn` package for this. I recommend using `statsmodels`, since it has better support for this kind of statistical analysis (note also that `sklearn` regularizes by default, so you must turn that off if you use `sklearn`).

For `statsmodels`, after building a model `m`, you can use `m.summary()` to get the standard errors of the coefficients.

For 6.b, write a function `boot_fn` that works as described in ISLR. Instead of the R library function `boot`, you must write your own: write a function `boot(data, fn, R)` where `data` is a pandas dataframe, `fn` is a function that computes a statistic, and R is the number of replicates. You can use resample from `sklearn` to generate individual bootstrap samples.

3 Cross-Validation on Simulated Data

Solve exercise 8 from Chapter 5 of ISLR, in python.

4 Ridge Regression Effect of λ

Solve exercise 4 in Chapter 6 of ISLR

5 Comparing Lasso, Ridge, and Least Squares

Solve exercise 9 from Chapter 6 of ISLR. You only need to complete questions (a),(b),(c),(d), and (g). For question (g), you only need to compare the three approaches from (b), (c), and (d).

You can use `scikitlearn` LinearRegression, Ridge, Lasso.

If you prefer `statsmodels`, then you can use regularized for lasso and ridge.