BA222 Problem Set 5

Jetson Leder-Luis

Due: Tuesday Dec 12 2023 at 5PM
No Late Psets will be Accepted

This is a simple problem set that should take you no more than 1-2 hours.

Implement Lasso

- Load the CruiseData file. This contains information on cruises, with information about the company, the boat, and the crew size. We will be trying to predict the crew size.
- Store the crew as the Y variable.
- Drop the crew size and ship name, and store all of the other variables as your X variable.
- Run `pd.get_dummies` to turn the categorical variables into dummies, and add a constant to this data frame.
- Run Lasso on your big set of possible X with the dummies and constant. Use $\alpha = 0.7$. Look at your model results.
- Run a regular OLS regression on the variables that the lasso model picks.
- For comparison, run an OLS regression on the big data frame with all of the dummies and variables.

Please answer the following questions in a short writeup. Turn this writeup in as a PDF. You do not need to submit your code.

1. How many variables does Lasso pick, and which ones?
2. What is the fitted equation of the OLS model with those variables?
3. What is the R^2 of the OLS model with those variables? How does that compare to the OLS model with all of the variables and dummies?
4. Include a screenshot of your OLS model using the variables picked by Lasso.