

# Problem Set 4

Econ 4676: Big Data and Machine Learning for Applied Economics

**Due Date:** December 3 at 11:00 am.

## 1 Theory Exercises: Choose your own adventure.

1. For this part, you formulate your own theoretical question based on the topics learned so far. It can be anything you want as long as it relates to the course. I will especially reward questions related to the topics in this problem set. Your grade will depend on the “ingenuity” of the question and the “correctness” of the answer

## 2 Empirical Problems

Please note that I’ll base part of your grade on your following Jacob’s rubric posted in the course communication channel. The rubric details how to structure the repository and the document.

The repo link to create your submission is <https://classroom.github.com/g/M7b6bOyL>.

### 2.1 *“Look Momma: I’m a machine learnist”*

1. You made it to the end. Congratulations!!! There is one more step. In the last three problem sets, you learned various new tools. I want you to go back to one and re-estimate your model using at least two new tools.
2. I expect you to include a nice table with descriptive statistics and a brief write up about the table. Be sure to mention if you made any adjustments/transformations or dropped observations.
3. Describe the models you used in the previous problem set and in this one. What are the advantages and disadvantages of the methods?
4. Present and compare (can be in a table or a graph) your prediction errors from your previous methods and the new methods. Was there a significant improvement?