Machine Learning - Problem Set 1

PPHA 30545 - Professor Clapp
Winter 2023

This assignment must be handed in via Gradescope on Canvas by **11:45pm Central Time on Wednesday, January 18th.** You are welcome (and encouraged!) to form study groups (of no more than 2 students) to work on the problem sets and mini-projects together. But you must write your own code and your own solutions. Please be sure to include the names of those in your group on your submission.

You should submit your answers as a single PDF containing BOTH a write-up of your solutions that directly integrates any relevant supporting output from your code (e.g., estimates, tables, figures) AND your code appended to the end of your write up. You may type your answers or write them out by hand and scan them (as long as they are legible). Your original code should be a Python (*.py) file converted to PDF format. The use of Jupyter Notebook (*.ipynb) files is discouraged.

Regardless of how you submit your answers, be sure to make it clear what question you are answering by labeling the sections of your write up well. Also, be sure that it is immediately obvious what supporting output from your code (e.g., estimates, tables, figures) you are referring to in your answers. In addition, your answers should be direct and concise. Points will be taken off for including extraneous information, even if part of your answer is correct. You may use bullet points if they are beneficial. Finally, for your code, please also be sure to practice the good coding practices you learned in PPHA 30535/6 and comment your code, cite any sources you consult, etc.

You are allowed to consult the textbook authors' websites, Python documentation, and websites like StackOverflow for general coding questions. You are not allowed to consult material from other classes (e.g., old problem sets, exams, answer keys) or websites that post solutions to the textbook questions.

1. Do the following questions from Chapter 2 of the *Introduction to Statistical Learning* textbook:
   (a) Question 3
      - Note: in part (a), you can sketch the curves by hand. You don't need to plot the curves in Python (although you're welcome to if you'd like to).
   (b) Question 5
   (c) Question 10
      - Note: each observation in the "Boston" housing data set is a different census tract.
2. Do the following questions from Chapter 3 of the *Introduction to Statistical Learning* textbook:
   (a) Question 3
   (b) Question 15