Machine Learning - Problem Set 3

PPHA 30545 - Professor Clapp
Winter 2021

This assignment must be handed in via Gradescope on Canvas by **11:45pm Central Time on Tuesday, February 23rd.** You are welcome (and encouraged!) to form study groups (of no more than 3 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 code as a single Python (*.py) file and the write up of your solutions as a single PDF. For the former, 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. For the latter, you may type your answers or write them out by hand and scan them (as long as they are legible).

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 5 of the *Introduction to Statistical Learning* textbook:
   (a) Question 6
   (b) Question 8
2. Do the following questions from Chapter 6 of the *Introduction to Statistical Learning* textbook:
   (a) Question 11
      i. In part (a), use best subset, forward stepwise, & backwards stepwise selection
      ii. In part (b), compare the results of using the mathematical-adjustment approaches (C_p, AIC, BIC, & adjusted R^2) to using 5-Fold Cross-Validation (5FCV)