Homework 5: High dimensional problems

This homework covers material covered in Chapter 6 of *Introduction to Statistical Learning with Applications in R (ISLR)*. Some questions come directly from the book exercises, others are new questions.

Complete this homework as a jupyter notebook titled "LASTNAME_Homework6.ipynb" posted on your GitHub account. Turn in your assignment by emailing a link to the notebook to timothyv@andrew.cmu.edu by no later than 3pm on Tuesday April 26th, 2018.

Conceptual: Short answer questions and proofs. Be concise. Show stepwise solutions to proofs where appropriate

1. What is the "curse of dimensionality" in a regression context? How does it impact your model fit, and how does it affect the interpretability of your results? Be specific and define what a "high dimensional" statistical problem is.

2. Describe conceptually how subset selection, shrinkage models (e.g., ridge regression, lasso), and dimensionality reduction approaches (e.g, principal component regression) approaches deal with the curse of dimensionality. Contrast how each approach differs from the others.

3. For parts (a) through (c), indicate which of i. through iv. is correct. Justify your answer.
        (a) The lasso, relative to least squares, is:
                i. More flexible and hence will give improved prediction accuracy when its increase in bias is less than its decrease in variance.
                ii. More flexible and hence will give improved prediction accuracy when its increase in variance is less than its decrease in bias.
                iii. Less flexible and hence will give improved prediction accuracy when its increase in bias is less than its decrease in variance.
                iv. Less flexible and hence will give improved prediction accuracy when its increase in variance is less than its decrease in bias.
        (b) Repeat (a) for ridge regression relative to least squares.
        (c) Repeat (a) for principal component regression relative to least squares.

Applied: Submit all R code, along with plots and written responses.

Do applied exercises 6.8 & 6.9.