Problem set 3
Stat 154, SP 14; N. El Karoui
Tentative due date: 2/28/14, 3:45pm

In Hastie-Friedman-Tibshirani (henceforth HTF), please do the following problems:

1. 2.7
2. 2.8
3. 3.2
4. 3.12

Problem 5 ["Simple" linear regression"] Analyze the data given in the file brains.csv. The data is about average brain weight vs average body weight for roughly 60 species of mammals.

Problem 6 [Sparse methods in practice: Tracking an index] In the HW1 folder, under Data, there are two datasets, one with the daily value of the SP500 index for the last few years. One with the daily stock price of around 500 major stocks for the last few years. (They are the components of the SP500 index as of 9/17/13).

One concern for individual investors is the high transaction costs associated with buying stocks. Hence, one might be interested in portfolios with contain few different stocks.

Please do the following:

1. Using methods seen in class, construct a sparse/parsimonious portfolio that tracks/replicates approximately the SP500 index. Explain your methods and how you judge the quality of your replicating portfolio.
2. Suppose you are allowed to change your portfolio every 60 days. Find sparse portfolios that track the SP500 index well over 60 days. How stable are your portfolios? Explain what penalty you could use to try to get portfolios that change little over time.
3. Suppose now that you are not allowed to owe a negative amount of stock. (In practice, having a negative amount of stock is allowed, this is called "shorting" a stock. Though for various reasons, you might not want to be short a stock.) Explain how this change the optimization problems you are trying to solve.
4. Suppose now that you are only interested in replicating the returns of the SP500. (The daily return of a stock at time t is $R_t = S_t/S_{t-1} - 1$, where S_t is the price of the stock.) How does this change your answers to questions 1 and 2?

NOTE: the data was obtained from CRSP through Wharton's WRDS. This is intended only for scholarly research and use in this class. Using it for other purposes would be a violation of their end-user agreement.