Stat 610 Homework 7

For this assignment:

- Due Friday, November 13, 11:59pm
- You may work in groups of up to three.

Background

In this assignment, you will look into the time it takes for two different optimization methods to run as the number of predictor variables gets large.

Assignment

1. Make five simulated data sets $X^{(i)}, i = 1, \dots, 5$, where $p = (10, 50, 100, 200, 500)$ and $n = 2000$ such that

$$X^{(i)} \in \mathbb{R}^{n \times p_i}, \quad X_{jk}^{(i)} \sim N(0, 1)$$

2. Set $\beta^{(i)} \in \mathbb{R}^{p_i}$ with $\beta^{(i)} = (1, \dots, 1)$ and $\epsilon \in \mathbb{R}^n$ with $\epsilon_j \sim N(0, 1)$. Finally, generate $y^{(i)} = X^{(i)}\beta^{(i)} + \epsilon$

3. Write down the log likelihood, the gradient of the log likelihood, and the Hessian of the log likelihood for the linear model

$$y \sim N(X\beta, I_n)$$

4. Write out what a Newton step would be to maximize the log likelihood in the model above.

5. Write out what a gradient descent step would be to minimize the negative log likelihood in the model above.

6. Write a function that uses Newton's method to maximize the likelihood in the model above. Remember that we know that Newton's method will converge in one step, so you don't need to bother writing a loop here.

7. Write a function that uses gradient descent with backtracking line search for the example above.

8. Using `microbenchmark`, measure the amount of time it takes to estimate β using using Newton's method with the five data sets $(X^{(i)}, y^{(i)})$, $i = 1, \dots, 5$ that you created in the first two parts.

9. Using `microbenchmark`, measure the amount of time it takes to estimate β using Newton's method with the five data sets $(X^{(i)}, y^{(i)})$, $i = 1, \dots, 5$ that you created in the first two parts.

10. Plot the amount of time it takes to estimate β using Newton's method vs. p. Also plot the cube root of the time it takes to estimate β using Newton's method. The second plot should look linear.
11. Plot the amount of time it takes to estimate β using gradient descent vs. p. For what value of a does the plot of gradient-descent-time^a vs. p look approximately linear?

Submission parameters

Submit two files:

- A pdf writeup containing the answers to the questions.
- A file containing the code you used.