Penn State STAT 540
Homework #3, due Thursday, April 5, 2018 at midnight

What you have to submit in a Canvas submission folder: (i) Your `R` code, (ii) a pdf file that contains a clear writeup for the questions below. Note that your code should be readily usable, without any modifications. For example, to read in the data make sure the code accesses the website rather than a directory on your computer.

1. Find the MLE for a logistic regression model. For $i = 1, \dots, n$

$$\begin{aligned} Y_i &\sim \text{Ber}(p_i(\theta)), \quad \text{where } \theta = (\beta_1, \beta_2) \\ p_i(\theta) &= \exp(\beta_1 X_{1i} + \beta_2 X_{2i})/(1 + \exp(\beta_1 X_{1i} + \beta_2 X_{2i})), \end{aligned} \tag{1}$$

   where each Y_i is a binary response corresponding to predictors X_{1i}, X_{2i}. Write a Newton-Raphon algorithm to find the MLE $(\hat{\beta}_1, \hat{\beta}_2)$. The data for this problem are available here: `http://personal.psu.edu/muh10/540/data/logReg.dat`

   (a) Provide details about how you set up the algorithm, including any analytical quantities you derived. You should also include the stopping criteria, how you obtained initial values, how you tuned s_n (step size/learning rate) and other relevant information, e.g. did you run the algorithm using multiple initial values?

   (b) Provide your pseudocode for the main algorithm. You can use notation defined in class (and specified above) in order to make the pseudocode succinct and easy to read, e.g. $d\ell_{12}(\beta_1^{(n)}, \beta_2^{(n)})$.

   (c) Provide the MLE as well as estimate of the standard error, along with asymptotic 95% confidence intervals.

   (d) Report the total computing time taken by your algorithm (e.g. you could use `system.time`). If you have 2-3 different versions of the algorithm, report timings for each version.

   (e) Provide the computational cost of the algorithm (flops) per iteration as a function of n. Also provide the number of iterations it took before the stopping criteria was met. Of course, the number of iterations will vary by initial conditions so you should provide at least 2-3 different counts depending on where you started.

   (f) Summarize in 1-2 sentences the sensitivity of the algorithm to the initial value, that is, does the algorithm work well regardless of where you start or does it need to be in a neighborhood of a certain value? (This is not required, but because this is a simple 2-D optimization problem, you could also provide the 2-D log-likelihood surface to obtain insights.)

2. Implement a gradient descent algorithm (work with the negative log likelihood). Answer (a)-(f) for this algorithm.

3. Now implement a stochastic gradient descent algorithm and repeat (a)-(f).

4. Compare the three algorithms above: Newton-Raphson, gradient descent, and stochastic gradient descent. Provide a 1-2 sentence summary of which algorithm you would recommend and why. Then provide a more detailed comparison, for example how stable the algorithms are with respect to initial values, how sensitive they are to choice of s_n, a comparison of the total computational cost, and the computational cost per iteration.

5. E-M algorithm: Return to a lightbulb lifetimes problem similar to the problem in a previous homework. Assume lightbulbs made by a particular company are independent and gamma distributed with parameters α, β (parameterization is such that expected value is $\alpha\beta$). Suppose in an experiment m bulbs are switched on at the same time, but are only completely observed up to time τ. Let the lifetimes of these bulbs be $A_1, A_2, \ldots, A_m$. However, since the bulbs are only observed till time τ, not all these lifetimes will be observed. Now suppose that at time τ, the experimenter observes the number of lightbulbs, W, still working at time τ, and the lifetimes of all lightbulbs that stopped working by τ. For convenience, denote these bulbs that stopped working by time τ as $A_1, A_2, \ldots, A_{m-W}$. Hence, the missing information consists of the lifetimes of the lightbulbs still working at time τ, $A_{m-W+1}, \ldots, A_m$. For a particular experiment, let τ be 200 days and $m = 300$. The data on the lightbulb lifetimes for the bulbs that stopped working by τ are here: `http://personal.psu.edu/muh10/540/data/bulbsHW3.dat` Assume that the remaining bulbs were still working at time τ. Find the MLE for (α, β) using the E-M algorithm. Repeat parts (a)-(f) from above for this algorithm as well.