Homework 5, CS 4220, Spring 2021
Instructor: Austin R. Benson
Due Friday, April 30, 2021 at 3:44pm ET on CMS (before lecture)

Policies

Submission. Submit your write-up as a single PDF on CMS: `https://cmsx.cs.cornell.edu`.
Coding questions. You can use any programming language for the coding parts of the assignment. Include your code in your write-up.
Typesetting. Your write-up must be typeset with LaTeX.
Collaboration. Please discuss and collaborate on the homework, but you have to write your own solutions and code.
Resources and attribution. Feel free to use any resources that might be helpful, and provide attribution for any key ideas. We only ask that you work on the problems in earnest. Please do not hunt for solutions with a search engine.

Problems

1. *Learning preferences.*

   Discrete choice is a setting where an individual chooses some item from a discrete set of alternatives, such as choosing a TV show to watch on Netflix, selecting a car to buy after test driving a few, or voting for a political candidate.

   One simple mathematical model of discrete choice behavior is the following. Suppose that there is a universe $Y = \{1, 2, \ldots, n\}$ of items, and each item i has some inherent real-valued score $\alpha_i > 0$, $i \in Y$. Let $C \subset Y$ be the discrete set of alternatives available to the decision maker; we will call C the *choice set*. Then the *logit model* posits the following random model for how someone chooses an item from the choice set C:

   $$\Pr(\text{choose item } i \text{ from choice set } C \subset Y) = \frac{\alpha_i}{\sum_{k \in C} \alpha_k}.$$

   For this problem, suppose that we have data on m choices $(x_1, C_1), \ldots, (x_m, C_m)$, where $C_j \subset Y$ is the jth choice set and $x_j \in C_j$ is the chosen item (note that the choice sets might be different). We would like to estimate the scores $\alpha_1, \ldots, \alpha_n$. Similar to the last homework, we will use maximum likelihood estimation for this task.

   (a) One way to manage the positivity constraint $\alpha_i > 0$ is to let $\alpha_i = \exp(u_i)$ for some $u_i \in \mathbb{R}$, so that

   $$\Pr(\text{choose item } i \text{ from choice set } C \subset Y) = \frac{\exp(u_i)}{\sum_{k \in C} \exp(u_k)}.$$

   Given some candidate set of values $u \in \mathbb{R}^n$, the likelihood function is

   $$L(u; \{(x_j, C_j)\}_{j=1}^m) = \prod_{j=1}^m \frac{\exp(u_{x_j})}{\sum_{k \in C_j} \exp(u_k)}.$$

   Similar to the last homework, first find an expression for the negative log-likelihood $N(u; \{(x_j, C_j)\}_{j=1}^m)$. Then, for $\beta > 0$, let

   $$f(u) = \frac{1}{m} N(u; \{(x_j, C_j)\}_{j=1}^m) + \beta \|u\|_2^2,$$

and find an expression for $\nabla f(u)$. (Here, you can just write $\frac{\partial}{\partial u_j} f(u)$ for arbitrary j.)

(b) The "log-sum-exp" function lse: $\mathbb{R}^d \to \mathbb{R}$ is defined as

$$\text{lse}(z) = \log\left(\sum_{k=1}^{d} \exp(z_k)\right)$$

Show that the Hessian of lse evaluated at $z_1, \ldots, z_d$ is

$$\text{diag}(p_z) - p_z p_z^T,$$

where $p_z \in \mathbb{R}^d$ has ith entry equal to $\frac{\exp(z_i)}{\sum_{k=1}^{d} \exp(z_k)}$ and $\text{diag}(p_z)$ means the diagonal matrix with diagonal entries given by p_z.

Next, show that the Hessian is positive semidefinite. *Hint: show that the Hessian is (weakly) diagonally dominant and apply the Gershgorin circle theorem.*

(c) Using part (b), explain why the Hessian H of f is symmetric positive definite.

(d) We could represent a matrix-vector product with H implicitly. Using part (b), explain how to compute a matrix-vector product Hv in $O(n + \sum_{j=1}^{m} |C_j|)$ time.

(e) We could also represent H as a sparse matrix. Characterize which entries of H are non-zero.

(f) Implement the following two ways of finding search directions from a current guess u_k:

  (i) the negative gradient $-\nabla f(u_k)$, and

  (ii) an approximate Newton direction from approximately solving $H(u_k)p_k = -\nabla f(u_k)$ with the conjugate gradient method.

In the latter case, you should compute matrix-vector products using the ideas in either part (d) or part (e). Please use a built-in library implementation of CG (e.g., `IterativeSolvers.jl` with `LinearMaps.jl` for Julia).

For this part, you only need to turn in code.

(g) Download the `hotel-choice.txt` file of choice data at https://github.com/arbenson/cs4220_2021sp/tree/main/hw5data. Run each of the two methods from part (f) for 50 steps. Make a plot that shows $f(u_k)$ as a function of k (the number of steps).