Biostat 561: Homework 7

Instructor: Amy Willis, Biostatistics, UW

15 May, 2019

Homework 7 due May 22, 3:15pm

Link to Homework 7 submission: https://classroom.github.com/a/Y_RepAkO

In this homework, you are expected to do some math, research. and self-teaching. These expectations are consistent with the graduate-level designation and enrolment criteria for this course.

Question 1: Maximum likelihood estimation

Consider the problem of estimating the parameter $\beta \in \mathbb{R}^p$ of a Poisson generalized linear model

$$Pr(Y_i = y_i | X_i) = \frac{\lambda_i^{y_i} e^{-\lambda_i}}{y_i!}$$

where $\lambda_i = exp\left(X_i^T \beta\right)$.

(a) Implement `poisson_glm_gradient_descent()`, a function that takes in a vector $Y \in \mathbb{R}^n$, a matrix $X \in \mathbb{R}^{n \times p}$, and step size multiplier γ and returns the vector $\hat{\beta} \in \mathbb{R}^p$ that is the maximum likelihood estimate of β obtained using gradient descent.

(b) Implement `poisson_glm_newton()`, a function that takes in a vector $Y \in \mathbb{R}^n$, a matrix $X \in \mathbb{R}^{n \times p}$ and returns the vector $\hat{\beta} \in \mathbb{R}^p$ that is the maximum likelihood estimate of β obtained using using the Newton-Raphson algorithm.

Illustrate that both of your methods work and produce similar estimates on the dataset `eg_data.RDS`. A reasonable sanity check is to confirm your result using R's `glm()` function. Note that going through the code of `glm()` will probably be more confusing than helpful.

In your response, explicitly derive the objective function, and its gradient and Hessian.

Question 2: Stan with R

Repeat Question 1, but this time write a Stan program to obtain the posterior distribution of β_j for $j = 1, \ldots, p$. Investigate various choices of prior distribution and justify your final choice.

You should rely heavily on the RStan documentation to learn how to use Stan with `R`. I suggest confirming that you can run one of the RStan examples before modifying it for the Poisson GLM.

Notice how you needed to do *no math* to use Stan! However, it is difficult to know exactly how Stan works and how well it is doing without advanced training in Bayesian statistics. Use Stan with caution!

Question 3: Teach yourself about an algorithm

Choose an algorithm that you have heard about or read about in the statistical literature. Research the algorithm, and summarise your findings in a report of approximately one page. Describe the purpose of the algorithm, outline the algorithm, and give intuition or results that explain why the algorithm works.

Consider including an example of your own to show how it works. You will learn nothing if you copy someone else's example!

This question is deliberately designed to be open-ended to give you the opportunity to teach yourself about something that you have come across or are interested in. I hope it helps you with your thesis research!