HOMEWORK 2

1. 证明题

(1) Suppose for the ith subject we observe x_i and y_i. Let $p(x_i; \beta) = P(Y = 1|X = x_i)$. Maximum likelihood estimation:

$$\ell(\beta) = \sum_{i=1}^{N} \left\{ y_i \log p(x_i; \beta) + (1 - y_i) \log(1 - p(x_i; \beta)) \right\}$$
$$= \sum_{i} \left\{ y_i x_i^\top \beta - \log\left(1 + \exp(x_i^\top \beta)\right) \right\}$$

Please derive the blue part.

(2) Write Newton-Raphson algorithm to estimate logistic regression.

Reminder: you need to derive the equation

$$\frac{\partial^2 \ell(\beta)}{\partial \beta \partial \beta^\top} = -\sum_{i} x_i x_i^\top p(x_i; \beta)\{1 - p(x_i; \beta)\}.$$

Generate $X = (1, X_1, X_2)$, where $X_j \sim N(0, I_N)$.

Set true parameter $\beta = (0.5, 1.2, -1)^\top$.

Set $N = 200, 500, 800, 1000$.

Estimate β using NR algorithm for $R = 200$ rounds of simulation. For each round of simulation, terminate the iteration when $\max_j |\widehat{\beta}_j^{old} - \widehat{\beta}_j^{new}| < 10^{-5}$. Denote $\widehat{\beta}_j^{(r)}$ as the estimation of β_j in the rth round of simulation. Then please: for each j, draw $(\widehat{\beta}_j^{(r)} - \beta_j)$ in boxplot for $N = 200, 500, 800, 1000$.

2. H1B 签证申请影响因素分析。

编程语言可以使用 R/python，可对应使用 Rmarkdown 或 Jupyter notebook 输出报告。具体任务见 word 文档。

最后以 HTML 的形式提交案例分析报告、以 PDF 或 HTML 形式提交证明题。报告

中需包括题目内容中涉及的代码和相关文字解释、结果分析。

提交时间：3 月 30 日 20:00 之前。请预留一定的时间，迟交作业扣 3 分，作业抄袭 0 分。