

HOMEWORK 3

一、利用 Newton-Raphson 算法实现逻辑回归：

1. 设置真实的参数值 $\beta = (-0.5, 0.5, -1, 1)^\top$
2. 生成 $\mathbf{X} = (1, X_1, X_2, X_3) \in \mathbb{R}^{N \times 4}$ ，其中 $X_j \in \mathbb{R}^N \sim N(\mathbf{0}, \mathbf{I}_N)$; $\forall j = 1, 2, 3$
3. 设置样本量为 $N = 200, 500, 750, 1000$ ，在每个样本量下，重复生成不同的数据（第 r 次实验的随机种子数设置为 r ）并利用 Newton-Raphson 算法估计参数 β ，重复次数设置为 $R = 200$
4. 如果记第 r 次重复估计得到的第 j 维参数为 $\hat{\beta}_j^{(r)}$ ，在样本量为 $N = 200, 500, 750, 1000$ 下，对每一个维度 ($j = 1, 2, 3, 4$) 作出 $\hat{\beta}_j^{(r)} - \beta_j$ 的箱线图，并通过对比不同样本量 N 下的图象得出你的结论

注意：Newton-Raphson 算法需自己实现，不能调包，算法迭代终止条件设置为

$$\max_j |\hat{\beta}_j^{old} - \hat{\beta}_j^{new}| < 10^{-5}$$

同时提示：

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

二、征信系列-用户行为数据分析。

编程语言可以使用 R/python 等一切能满足题目要求的语言。具体任务见“金融岗位招聘分析任务”文档。

最后以 HTML/PDF 的形式提交报告。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。（提示：R 语言可用 Rmarkdown 输出分析报告；Python 可用 Jupyter 输出 HTML 报告；也自己进行格式调整后输出 PDF 格式的报告）

提交时间：3 月 28 日，晚 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。