Homework assignments Week 12 (Students should submit their homework before 10 a.m. on December 15, 2021.)

3. Derivation of the Lipschitz constant for the logistic loss for regression estimation (2%)

Assume that we have observed data $\{y_i, \mathbf{x}_i\}_{i=1}^n$, where $y_i \in \{0, 1\}$ is the label of observation i and $\mathbf{x}_i$ is a $(p+1)$-dimensional vector of the corresponding covariates. Our aim is to build a regression model using $\mathbf{x}_i$ to predict y_i. A common way is to model the log odds of the probability that $y_i = 1$ as a linear combination of $\mathbf{x}_i = (1, x_{i1}, x_{i2}, \cdots, x_{ip})$:

$$\log\left[\frac{\mathbb{P}(y_i = 1)}{1 - \mathbb{P}(y_i = 1)}\right] = \mathbf{x}_i^T \boldsymbol{\beta} = f_i.$$

The corresponding *minus* log likelihood function of $\boldsymbol{\beta}$ is

$$L(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\left\{-y_i f_i + \log\left[1 + \exp(f_i)\right]\right\}. \tag{1}$$

where $f_i = \mathbf{x}_i^T\boldsymbol{\beta}$. With (1) we use the following gradient algorithm to obtain an estimate of $\boldsymbol{\beta}$:

$$\boldsymbol{\beta}^{r+1} = \boldsymbol{\beta}^r - c\nabla L(\boldsymbol{\beta}^r). \tag{2}$$

To run (2), we need to know the value of the stepsize c. Which of the following statements are *true*? To ensure the **descent property** will be guaranteed for the sequence $\{\boldsymbol{\beta}^r\}_r$ for evaluating the loss $L(\boldsymbol{\beta})$,

a. We set

$$c = 4.$$

b. We set

$$c = \frac{1}{\lambda_{p+1}},$$

where λ_{p+1} is the smallest eigenvalue of the Gram matrix $n^{-1}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^T$.

c. We set

$$c = \frac{4}{\lambda_1},$$

where λ_1 is the largest eigenvalue of the Gram matrix $n^{-1}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i^T$.

Programming work

4. The gradient algorithm for logistic regression estimation (6%)

In this programming work we will build a gradient algorithm to find an estimate of regression coefficients in logistic regression model. To run such an algorithm, you need to:

1. Construct an iterative scheme based on the gradient algorithm.

2. Specify a **stepsize** for the iterative scheme. Here you are allowed to use whatever way you like to specify the **stepsize**.

3. Specify (a) a **stopping criterion**, (b) a **tolerance** for the error and (c) the **maximum number of iterations** for stopping the iterative scheme. Here you are allowed to use whatever way you like to specify the **stopping criterion** like the following one:

Some measure on error $\leq$ tol OR The number of iterations $>$ max_iter.

However the tolerance for the error and the maximum number of iterations should be

$$\begin{aligned} \text{tol} &= 5 \times 10^{-6}, \\ \text{max\_iter} &= 10,000. \end{aligned}$$

Data generation: We let the number of observations $n = 200$ and the number of covariates $p = 10$. We use the following model to generate the data:

$$
\begin{aligned}
\boldsymbol{\beta}^{\text{true}} &= (-1, 1, -1, 1, -1, 1, -1, 1, -1, 1), \\
\mathbf{x}_i &= (x_{i1}, x_{i2}, \cdots, x_{i10}), \\
x_{i1} &= 1 \text{ for } i = 1, 2, \cdots, 200, \\
x_{ij} &\sim \text{Normal}(0, 1) \text{ for } i = 1, 2, \cdots, 200 \text{ and } j = 2, 3, \cdots, 10, \\
y_i &\sim \text{Bernoulli}\left(\frac{\exp(\mathbf{x}_i^T \boldsymbol{\beta}^{\text{true}})}{1 + \exp(\mathbf{x}_i^T \boldsymbol{\beta}^{\text{true}})}\right) \text{ for } i = 1, 2, \cdots, 200.
\end{aligned}
$$

Tasks: Report line plots of the following two settings:

1. The x-axis is the number of iterations r and the y-axis is $||\nabla l(\boldsymbol{\beta}^r)||_2$, the Euclidean norm of the gradient of the loss function you use in your iterative scheme;

2. The x-axis is the number of iterations r and the y-axis is $l(\boldsymbol{\beta}^r) - l(\boldsymbol{\beta}^*)$, the difference between the loss functions evaluated at the current update $\boldsymbol{\beta}^r$ and the *optimizer* $\boldsymbol{\beta}^*$. The optimizer $\boldsymbol{\beta}^*$ can be obtained from functions or software for carrying out logistic regression estimation available in your programming environment.

Remark: You may specify y-axis at the logarithm (with base 10) scale in the above two plots to make the plots more readable.