Supplementary Document to "A Third-Order Majorization Algorithm for Logistic Regression with Convergence Rate Guarantees"

Licheng Zhao, Rui Zhou, Wenqiang Pu, Tianyu Qiu, and Qingjiang Shi

In this supplementary document, we provide a detailed derivation of the second- and third-order derivatives of the cross-entropy loss function $f(x) = -y\log(\sigma(x)) - (1-y)\log(1-\sigma(x))$ where $y = \{0,1\}$ and $\sigma(x) = 1/(1+\exp(-x))$ is the Sigmoid function.

I. First-order Derivative

We notice that

$$\sigma'(x) = \sigma(x)(1-\sigma(x)). \tag{1}$$

Therefore,

$$\begin{aligned} f'(x) =& -y\frac{1}{\sigma(x)}\sigma(x)(1-\sigma(x)) - (1-y)\frac{1}{1-\sigma(x)}\cdot \\ & -(\sigma(x)(1-\sigma(x))) \\ =& -y(1-\sigma(x)) + (1-y)\sigma(x) \\ =& -y+\sigma(x). \end{aligned} \tag{2}$$

II. Second- and Third-order Derivatives

With the first-order detivative, the derivation of $f''(x)$ is straightforward:

$$f''(x) = \sigma(x)(1-\sigma(x)). \tag{3}$$

The third-order derivative $f'''(x)$ can be further obtained with the chain rule:

$$\begin{aligned} f'''(x) =& \sigma'(x)(1-\sigma(x)) + \sigma(x)(-\sigma'(x)) \\ =& \sigma(x)(1-\sigma(x))^2 - \sigma^2(x)(1-\sigma(x)) \\ =& \sigma(x)(1-\sigma(x))(1-\sigma(x)-\sigma(x)) \\ =& \sigma(x)(1-\sigma(x))(1-2\sigma(x)). \end{aligned} \tag{4}$$

The derivation process is thus complete.

This work was supported in part by the National Nature Science Foundation of China (NSFC) under Grant 62206182, 62201362, and 62101350.

Licheng Zhao, Rui Zhou, Wenqiang Pu, and Tianyu Qiu are with Shenzhen Research Institute of Big Data, Shenzhen 518172, China (email: {zhaolicheng, rui.zhou, wpu, qiutianyu}@sribd.cn). Qingjiang Shi is with the School of Software Engineering, Tongji University, Shanghai 201804, China, and also with Shenzhen Research Institute of Big Data, Shenzhen 518172, China (email: shiqj@tongji.edu.cn).