## **Functions of Logistic Regression**

Xueyuan Gong
Department of Computer and Information Science
Faculty of Science and Technology
University of Macau
Macau, China
{yb47453}@umac.mo

## 1. LOGISTIC REGRESSION

Given a data set  $X \in \mathbb{R}^{n \times (m-1)}$  and a label set  $Y \in \mathbb{R}^{n \times 1}$ . Let two indices i and j be subject to  $1 \le i \le n$  and  $2 \le j \le m$  respectively. Thus, a data value is denoted as  $X_j^i$  and a label value is denoted as  $Y^i$ . For all  $X^i$ , insert  $X_1^i = 1$  so that  $X \in \mathbb{R}^{n \times m}$  and  $1 \le j \le m$ . Initialize weight matrix  $W \in \mathbb{R}^{1 \times m}$  and thus a weight value is denoted as  $W_j$ .

We define the linear function in Equation (1).

$$Z^i = \sum_{j=1}^m W_j X_j^i \tag{1}$$

We define activation function as sigmoid function and thus the hypothesis function is given in Equation (2).

$$H(X^{i}) = G(Z^{i}) = \frac{1}{1 + e^{-Z^{i}}}$$
 (2)

Loss function is defined in Equation (3).

$$C(X^{i}) = -Y^{i} \log (H(X^{i})) - (1 - Y^{i}) \log (1 - H(X^{i}))$$
 (3)

Cost function is defined in Equation (4).

$$J(W) = \frac{1}{n} \sum_{i=1}^{n} C(X^{i})$$

$$= -\frac{1}{n} \sum_{i=1}^{n} [Y^{i} \log (H(X^{i})) + (1 - Y^{i}) \log (1 - H(X^{i}))]$$
(4)

The partial derivative of J(W) with respect to  $W_j$  is given as follows:

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Copyright 20XX ACM X-XXXXX-XX-X/XX/XX ...\$15.00.

$$\begin{split} & \frac{\partial J(W)}{\partial W_j} \\ &= \frac{1}{n} \sum_{i=1}^n \frac{\partial C(X^i)}{\partial W_j} \\ &= \frac{1}{n} \sum_{i=1}^n \left[ \frac{\partial Y^i \log (H(X^i))}{\partial W_j} + \frac{\partial (1 - Y^i) \log (1 - H(X^i))}{\partial W_j} \right] \end{split}$$

For simplicity, we calculate the equation separately:

$$\begin{split} &\frac{\partial Y^{i} \log \left(H(X^{i})\right)}{\partial W_{j}} \\ &= Y^{i} \frac{\partial \log \left(H(X^{i})\right)}{\partial H(X^{i})} \cdot \frac{\partial H(X^{i})}{\partial (Z^{i})} \cdot \frac{\partial (Z^{i})}{\partial W_{j}} \\ &= Y^{i} \frac{1}{H(X^{i})} \cdot H(X^{i})(1 - H(X^{i})) \cdot X_{j}^{i} \\ &= Y^{i} (1 - H(X^{i})) X_{i}^{i} \end{split}$$

$$\begin{split} &\frac{\partial (1-Y^i)\log (1-H(X^i))}{\partial W_j} \\ &= (1-Y^i)\frac{\partial \log (1-H(X^i))}{\partial (1-H(X^i))} \cdot \frac{\partial (1-H(X^i))}{\partial H(X^i)} \cdot \frac{\partial H(X^i)}{\partial (Z^i)} \cdot \frac{\partial (Z^i)}{\partial W_j} \\ &= (1-Y^i)\frac{1}{1-H(X^i)} \cdot (-1) \cdot H(X^i)(1-H(X^i)) \cdot (X^i_j) \\ &= (Y^i-1)H(X^i)X^i_j \end{split}$$

$$\therefore \frac{\partial J(W)}{\partial W_j} = -\frac{1}{n} \sum_{i=1}^n [Y^i (1 - H(X^i)) X_j^i + (Y^i - 1) H(X^i) X_j^i]$$
$$= \frac{1}{n} \sum_{i=1}^n (H(X^i) - Y^i) X_j^i$$

Thus, for all  $W_j$ , the update function is given as follows:

$$W_j = W_j - \frac{\gamma}{n} \sum_{i=1}^n (H(X^i) - Y^i) X_j^i$$

where  $\gamma$  is the learning rate given by users.

## 2. REGULARIZED LOGISTIC REGRESSION

The cost function of regularized logistic regression is given in Equation (5).

$$J(W) = \frac{1}{n} \sum_{i=1}^{n} C(X^{i}) + \frac{\lambda}{2n} \sum_{j=1}^{m} (W_{j})^{2}$$
 (5)

where  $\lambda$  is regularization ratio set by users.

The partial derivative of J(W) with respect to  $W_j$  is given as follows:

$$\begin{split} \frac{\partial J(W)}{\partial W_j} &= \frac{1}{n} \sum_{i=1}^n \frac{\partial C(X^i)}{\partial W_j} + \frac{\lambda}{2n} \frac{\partial \sum_{j=1}^m (W_j)^2}{\partial W_j} \\ &= \frac{1}{n} \sum_{i=1}^n (H(X^i) - Y^i) X_j^i + \frac{\lambda}{n} W_j \end{split}$$

However, we do not want to regularize the weight of the constant factor  $W_1$ . Thus, for all  $W_j$ , the update function is given as follows:

$$W_{j} = W_{j} - \frac{\gamma}{n} \sum_{i=1}^{n} (H(X^{i}) - Y^{i}) X_{j}^{i} \qquad (for \ j = 0)$$

$$W_{j} = W_{j} - \frac{\gamma}{n} [\sum_{i=1}^{n} (H(X^{i}) - Y^{i}) X_{j}^{i} + \lambda W_{j}] \quad (for \ j > 1)$$

$$W_j = W_j - \frac{\gamma}{n} [\sum_{i=1}^n (H(X^i) - Y^i) X_j^i + \lambda W_j]$$
 (for  $j > 1$ )