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

Given a data set $X = \{X_j^i | X_j^i \in \mathbb{R}^{n\times(m-1)}\}$ and a label set $Y = \{Y^i | Y^i \in \mathbb{R}^{n\times 1}\}$, where $1 \le i \le n$ and $2 \le j \le m$. For all X^i, insert $X_1^i = 1$ so that $X_j^i \in \mathbb{R}^{n\times m}$. Initialize weight matrix $W = \{W_j | W_j \in \mathbb{R}^{1\times m}\}$.

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).

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

Loss function is defined in Equation (3).

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

Cost function is defined in Equation (4).

$$\begin{aligned} J(W) &= \frac{1}{n}\sum_{i=1}^{n} C(X^i) \\ &= -\frac{1}{n}\sum_{i=1}^{n}[Y^i \log\left(H(X^i)\right) + (1-Y^i)\log\left(1-H(X^i)\right)] \end{aligned} \tag{4}$$

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

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

For simplicity, we calculate the equation separately:

$$\begin{aligned} &\frac{\partial Y^i \log\left(H(X^i)\right)}{\partial W_j} \\ &= Y^i \frac{\partial \log\left(H(X^i)\right)}{\partial H(X^i)} \cdot \frac{\partial H(X^i)}{\partial(-Z^i)} \cdot \frac{\partial(-Z^i)}{\partial W_j} \\ &= Y^i \frac{1}{H(X^i)} \cdot H(X^i)(1-H(X^i)) \cdot (-X_j^i) \\ &= -Y^i(1-H(X^i))X_j^i \end{aligned}$$

$$\begin{aligned} &\frac{\partial (1-Y^i)\log\left(1-H(X^i)\right)}{\partial W_j} \\ &= (1-Y^i)\frac{\partial \log\left(1-H(X^i)\right)}{\partial(1-H(X^i))} \cdot \frac{\partial(1-H(X^i))}{\partial H(X^i)} \cdot \frac{\partial H(X^i)}{\partial(-Z^i)} \cdot \frac{\partial(-Z^i)}{\partial W_j} \\ &= (1-Y^i)\frac{1}{1-H(X^i)} \cdot (-1) \cdot H(X^i)(1-H(X^i)) \cdot (-X_j^i) \\ &= (1-Y^i)H(X^i)X_j^i \end{aligned}$$

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

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

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

where λ is the learning rate given by users.