Functions of Linear Regression

Xueyuan Gong
Department of Computer and Information Science
University of Macau
Macau, China
{yb474530}@umac.mo

Given a data set $X \in \mathbb{R}^{n \times m}$ and a label set $Y \in \mathbb{R}^{n \times 1}$. Let two indices i and j be subject to $1 \leq i \leq n$ and $1 \leq j \leq m$ respectively. Thus, a data value is denoted as X^i_j and a label value is denoted as Y^i. Initialize weight matrix $W \in \mathbb{R}^{1 \times m}$ and bias b, thus a weight value is denoted as W_j.

Hypothesis function is defined in Equation (1).

$$H(X^i) = \sum_{j=1}^{m} W_j X^i_j + b \tag{1}$$

Loss function is defined in Equation (2).

$$C(X^i) = H(X^i) - Y^i \tag{2}$$

Cost function is defined in Equation (3).

$$\begin{aligned} J(W) &= \frac{1}{2n} \sum_{i=1}^{n} C^2(X^i) \\ &= \frac{1}{2n} \sum_{i=1}^{n} (\sum_{j=1}^{m} W_j X^i_j + b - Y^i)^2 \end{aligned} \tag{3}$$

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

$$\begin{aligned} \frac{\partial J(W)}{\partial W_j} &= \frac{1}{2n} \sum_{i=1}^{n} \frac{\partial C^2(X^i)}{\partial W_j} \\ &= \frac{1}{2n} \sum_{i=1}^{n} \frac{\partial C^2(X^i)}{\partial C(X^i)} \cdot \frac{\partial C(X^i)}{\partial H(X^i)} \cdot \frac{\partial H(X^i)}{\partial W_j} \\ &= \frac{1}{2n} \sum_{i=1}^{n} 2C(X^i) \cdot 1 \cdot X^i_j \\ &= \frac{1}{n} \sum_{i=1}^{n} (H(X^i) - Y^i) X^i_j \end{aligned}$$

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

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

where γ is the learning rate given by users.