Functions of Linear 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 $W = \{W_j | W_j \in \mathbb{R}^{1\times m}\}$.

Hypothesis function is defined in Equation (1).

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

Loss function is defined in Equation (2).

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

Cost function is defined in Equation (3).

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

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

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

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.