# Functions of Neural Network

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

## I. Neural Network

Given a data set $X \in \mathbb{R}^{n \times m_0}$ and a label set $Y \in \mathbb{R}^{n \times 1}$. Thus, a data value is denoted as $X_j^i$ and a label value is denoted as $Y^i$. Let two indices $i$ and $j$ be subject to $1 \leq i \leq n$ and $1 \leq j \leq m$ respectively.

Suppose the network has $l$ layers, where the size of each layer $L_k$ is denoted as $m_k$. Note the input/first layer $L_1$ accepts $X^i \in \mathbb{R}^{1 \times m_0}$ as input data. Initialize weight matrix list $W$ and bias list $b$ containing $l-1$ matrices and biases respectively, where each matrix $W^z \in \mathbb{R}^{m_k \times m_{k+1}}$ and each bias $b^z \in \mathbb{R}^{1 \times m_{k+1}}$

and biases $b$, thus a weight value is denoted as $W_j$.

We define the linear function in Equation (1).

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

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).

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

Cost function is defined in Equation (4).

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

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

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

For simplicity, we calculate the equation separately:

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

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

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

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

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

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

## II. Regularized Logistic Regression

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

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

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

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

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

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

$$
\begin{aligned}
W_j &= W_j - \frac{\gamma}{n}[\sum_{i=1}^{n}(H(X^i) - Y^i)X_j^i + \lambda W_j] \\
b &= b - \frac{\gamma}{n}\sum_{i=1}^{n}(H(X^i) - Y^i)X_j^i
\end{aligned}
$$