Derivation of Gradient Descent for Logistic Regression (2-class)

Derive $\mathbf{w} = \mathbf{w} - \frac{\alpha}{n} X_b^T(\mathbf{y} - \hat{\mathbf{y}})$

เรามี $Cost = -\frac{1}{n}\sum_{i=1}^{n}\left[y_i \log \hat{y}_i + (1-y_i)\log(1-\hat{y}_i)\right]$ และ $\mathbf{w} = \mathbf{w} - \alpha \nabla Cost$ โดยที่ $\mathbf{w} = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_p \end{bmatrix}$

$$\nabla Cost = \begin{bmatrix} \frac{\partial Cost}{\partial w_0} \\ \frac{\partial Cost}{\partial w_1} \\ \vdots \\ \frac{\partial Cost}{\partial w_p} \end{bmatrix}$$

พิจารณา $\frac{\partial Cost}{\partial w_d}$ เมื่อ $d \in \{0, 1, ..., p\}$ จะได้ว่า

$$\frac{\partial Cost}{\partial w_d} = -\frac{1}{n}\sum_{i=1}^{n} x_{i,d}(y_i - \hat{y}_i) \tag{1}$$

ซึ่งสามารถเขียนให้อยู่ในรูป matrix ได้ดังนี้

$$\frac{\partial Cost}{\partial w_d} = -\frac{1}{n}\begin{bmatrix} x_{1,d} & x_{2,d} & \cdots & x_{n,d} \end{bmatrix}\left(\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} - \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix}\right) \tag{2}$$

จาก $\nabla Cost$ เราสามารถเขียนให้อยู่ในรูปต่อไปนี้

$$\begin{bmatrix} \frac{\partial Cost}{\partial w_0} \\ \frac{\partial Cost}{\partial w_1} \\ \vdots \\ \frac{\partial Cost}{\partial w_p} \end{bmatrix} = -\frac{1}{n}\begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_{1,1} & x_{2,1} & \cdots & x_{n,1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1,p} & x_{2,p} & \cdots & x_{n,p} \end{bmatrix}\left(\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} - \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix}\right) \quad ; x_{i,0} = 1; i \in \{1, 2, ..., n\}$$

กำหนดให้

$$X_b = \begin{bmatrix} 1 & x_{1,1} & \cdots & x_{1,p} \\ 1 & x_{2,1} & \cdots & x_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n,1} & \cdots & x_{n,p} \end{bmatrix}, \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \hat{\mathbf{y}} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix}$$

นั่นคือ $\nabla Cost = -\frac{1}{n} X_b^T(\mathbf{y} - \hat{\mathbf{y}})$

$$\because \mathbf{w} = \mathbf{w} - \alpha \nabla Cost$$

ดังนั้น $\mathbf{w} = \mathbf{w} + \frac{\alpha}{n} X_b^T(\mathbf{y} - \hat{\mathbf{y}})$

Derive Equation 1

$$
\begin{aligned}
\frac{\partial Cost}{\partial w_d} &= \frac{\partial \frac{-1}{n}\sum_{i=1}^{n}\left[y_i \log \hat{y}_i + (1-y_i)\log(1-\hat{y}_i)\right]}{\partial w_d} \\
&= -\frac{1}{n}\sum_{i=1}^{n}\frac{\partial\left[y_i \log \hat{y}_i + (1-y_i)\log(1-\hat{y}_i)\right]}{\partial w_d} \\
&= -\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\partial y_i \log \hat{y}_i}{\partial w_d} + \frac{\partial (1-y_i)\log(1-\hat{y}_i)}{\partial w_d}\right] \\
&= -\frac{1}{n}\sum_{i=1}^{n}\left[y_i\frac{\partial \log \hat{y}_i}{\partial w_d} + (1-y_i)\frac{\partial \log(1-\hat{y}_i)}{\partial w_d}\right] \\
&= -\frac{1}{n}\sum_{i=1}^{n}\left[y_i\frac{\partial \log \hat{y}_i}{\partial \hat{y}_i}\cdot\frac{\partial \hat{y}_i}{\partial z_i}\cdot\frac{\partial z_i}{\partial w_d} + (1-y_i)\frac{\partial \log(1-\hat{y}_i)}{\partial(1-\hat{y}_i)}\cdot\frac{\partial(1-\hat{y}_i)}{\partial \hat{y}_i}\cdot\frac{\partial \hat{y}_i}{\partial z_i}\cdot\frac{\partial z_i}{\partial w_d}\right] \\
&= -\frac{1}{n}\sum_{i=1}^{n}\left[y_i\left(\frac{1}{\hat{y}_i}\right)\frac{\partial \hat{y}_i}{\partial z_i}x_{i,d} + (1-y_i)\left(\frac{1}{1-\hat{y}_i}\right)(-1)\frac{\partial \hat{y}_i}{\partial z_i}x_{i,d}\right] \\
&= -\frac{1}{n}\sum_{i=1}^{n}\left[\left(\frac{y_i}{\hat{y}_i}\right)\frac{\partial \hat{y}_i}{\partial z_i}x_{i,d} - \left(\frac{1-y_i}{1-\hat{y}_i}\right)\frac{\partial \hat{y}_i}{\partial z_i}x_{i,d}\right] \\
&= -\frac{1}{n}\sum_{i=1}^{n}\left[\left(\frac{y_i}{\hat{y}_i} - \frac{1-y_i}{1-\hat{y}_i}\right)\frac{\partial \hat{y}_i}{\partial z_i}x_{i,d}\right]
\end{aligned}
$$

เนื่องจาก

$$
\frac{\partial \hat{y}_i}{\partial z_i} = \hat{y}_i(1-\hat{y}_i) \tag{3}
$$

ดังนั้น

$$
\begin{aligned}
\frac{\partial Cost}{\partial w_d} &= -\frac{1}{n}\sum_{i=1}^{n}\left[\left(\frac{y_i}{\hat{y}_i} - \frac{1-y_i}{1-\hat{y}_i}\right)\hat{y}_i(1-\hat{y}_i)x_{i,d}\right] \\
&= -\frac{1}{n}\sum_{i=1}^{n}\left[y_i(1-\hat{y}_i)x_{i,d} - \hat{y}_i(1-y_i)x_{i,d}\right] \\
&= -\frac{1}{n}\sum_{i=1}^{n}\left[y_i x_{i,d} - y_i\hat{y}_i x_{i,d} - \hat{y}_i x_{i,d} + \hat{y}_i y_i x_{i,d}\right] \\
&= -\frac{1}{n}\sum_{i=1}^{n}\left[y_i x_{i,d} - \hat{y}_i x_{i,d}\right] \\
&= -\frac{1}{n}\sum_{i=1}^{n}\left[x_{i,d} y_i - x_{i,d}\hat{y}_i\right] \\
&= -\frac{1}{n}\sum_{i=1}^{n}x_{i,d}(y_i - \hat{y}_i)
\end{aligned}
$$

Derive Equation 2

จาก $\dfrac{\partial Cost}{\partial w_d} = -\dfrac{1}{n}\sum_{i=1}^{n} x_{i,d}(y_i - \hat{y}_i)$

พิจารณา $-\dfrac{1}{n}\sum_{i=1}^{n} x_{i,d}(y_i - \hat{y}_i)$

$$
\begin{aligned}
-\frac{1}{n}\sum_{i=1}^{n} x_{i,d}(y_i - \hat{y}_i) &= -\frac{1}{n}\left[x_{1,d}(y_1 - \hat{y}_1) + x_{2,d}(y_2 - \hat{y}_2) + \cdots + x_{n,d}(y_n - \hat{y}_n)\right] \\
&= -\frac{1}{n}\begin{bmatrix} x_{1,d} & x_{2,d} & \cdots & x_{n,d} \end{bmatrix}\begin{bmatrix} y_1 - \hat{y}_1 \\ y_2 - \hat{y}_2 \\ \vdots \\ y_n - \hat{y}_n \end{bmatrix} \\
&= -\frac{1}{n}\begin{bmatrix} x_{1,d} & x_{2,d} & \cdots & x_{n,d} \end{bmatrix}\left(\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} - \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix}\right)
\end{aligned}
$$

ดังนั้น $\dfrac{\partial Cost}{\partial w_d} = -\dfrac{1}{n}\begin{bmatrix} x_{1,d} & x_{2,d} & \cdots & x_{n,d} \end{bmatrix}\left(\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} - \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix}\right)$

Derive Equation 3

$$\begin{aligned}
\frac{\partial \hat{y}_i}{\partial z_i} &= \frac{\partial \frac{1}{1+e^{-z_i}}}{\partial(1+e^{-z_i})} \cdot \frac{\partial(1+e^{-z_i})}{\partial e^{-z_i}} \cdot \frac{\partial e^{-z_i}}{\partial(-z_i)} \cdot \frac{\partial(-z_i)}{z_i} \\
&= \frac{\partial(1+e^{-z_i})^{-1}}{\partial(1+e^{-z_i})} \cdot \frac{\partial(1+e^{-z_i})}{\partial e^{-z_i}} \cdot \frac{\partial e^{-z_i}}{\partial(-z_i)} \cdot \frac{\partial(-z_i)}{z_i} \\
&= -(1+e^{-z_i})^{-2}(0+1)e^{-z_i}(-1) \\
&= -\frac{1}{(1+e^{-z_i})^2}(1)e^{-z_i}(-1) \\
&= \frac{e^{-z_i}}{(1+e^{-z_i})^2} \\
&= \left(\frac{1}{1+e^{-z_i}}\right)\left(\frac{e^{-z_i}}{1+e^{-z_i}}\right) \\
&= \hat{y}_i(1-\hat{y}_i)
\end{aligned}$$