Convexity of Cross Entropy for Logistic Regression (2-class)

เราต้องการพิสูจน์ว่า cross entropy สำหรับ logistic regression (2-class) เป็น convex function

เรามี $Cost = -\frac{1}{n}\sum_{i=1}^{n}[y_i \log \hat{y}_i + (1-y_i)\log(1-\hat{y}_i)]$

การตรวจสอบการเป็น convex function สามารถทำได้โดยพิจารณา second derivative ของ function นั้น ๆ ถ้า second derivative มีค่ามากกว่าหรือเท่ากับ 0 ในทุกกรณีแล้ว หมายความว่า function นั้นเป็น convex function
จาก

$$\nabla^2 Cost = \begin{bmatrix} \frac{\partial^2 Cost}{\partial w_0^2} \\ \frac{\partial^2 Cost}{\partial w_1^2} \\ \vdots \\ \frac{\partial^2 Cost}{\partial w_p^2} \end{bmatrix}$$

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

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

พิจารณา $\frac{\partial^2 Cost}{\partial w_d^2}$ เมื่อ $d \in \{0, 1, ..., p\}$ จะได้ว่า

$$\frac{\partial^2 Cost}{\partial w_d^2} = \frac{1}{n}\sum_{i=1}^{n} x_{i,d}^2 \hat{y}_i(1-\hat{y}_i) \tag{2}$$

พิจารณา $x_{i,d}^2 \hat{y}_i(1-\hat{y}_i)$

เนื่องจาก $x_{i,d} \in \mathbb{R}$ ดังนั้น $x_{i,d}^2 \geq 0$

เนื่องจาก $\hat{y}_i \in (0, 1)$ ดังนั้น $\hat{y}_i(1-\hat{y}_i) > 0$

จะได้ว่า $\frac{1}{n}\sum_{i=1}^{n} x_{i,d}^2 \hat{y}_i(1-\hat{y}_i) \geq 0$
นั่นคือ

$$\nabla^2 Cost = \begin{bmatrix} \frac{\partial^2 Cost}{\partial w_0^2} \\ \frac{\partial^2 Cost}{\partial w_1^2} \\ \vdots \\ \frac{\partial^2 Cost}{\partial w_p^2} \end{bmatrix} \geq \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

สรุปได้ว่า cross entropy สำหรับ logistic regression (2-class) เป็น convex function

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

$$
\begin{aligned}
\frac{\partial^2 Cost}{\partial w_d^2} &= \frac{\partial}{\partial w_d}\left(\frac{\partial Cost}{\partial w_d}\right) \\
&= \frac{\partial}{\partial w_d}\left(-\frac{1}{n}\sum_{i=1}^{n} x_{i,d}(y_i - \hat{y}_i)\right) \\
&= \frac{\partial\left(-\frac{1}{n}\sum_{i=1}^{n} x_{i,d}(y_i - \hat{y}_i)\right)}{\partial w_d} \\
&= -\frac{1}{n}\sum_{i=1}^{n} \frac{\partial x_{i,d}(y_i - \hat{y}_i)}{\partial w_d} \\
&= -\frac{1}{n}\sum_{i=1}^{n} \frac{\partial [x_{i,d}y_i - x_{i,d}\hat{y}_i]}{\partial w_d} \\
&= -\frac{1}{n}\sum_{i=1}^{n} \left[\frac{\partial x_{i,d}y_i}{\partial w_d} - \frac{\partial x_{i,d}\hat{y}_i}{\partial w_d}\right] \\
&= -\frac{1}{n}\sum_{i=1}^{n} \left[0 - \frac{\partial x_{i,d}\hat{y}_i}{\partial w_d}\right] \\
&= -\frac{1}{n}\sum_{i=1}^{n} \left[-\frac{\partial x_{i,d}\hat{y}_i}{\partial w_d}\right] \\
&= \frac{1}{n}\sum_{i=1}^{n} \frac{\partial x_{i,d}\hat{y}_i}{\partial w_d} \\
&= \frac{1}{n}\sum_{i=1}^{n} x_{i,d}\frac{\partial \hat{y}_i}{\partial z_i}\cdot\frac{\partial z_i}{\partial w_d} \\
&= \frac{1}{n}\sum_{i=1}^{n} x_{i,d}\frac{\partial \hat{y}_i}{\partial z_i}x_{i,d} \\
&= \frac{1}{n}\sum_{i=1}^{n} x_{i,d}^2\frac{\partial \hat{y}_i}{\partial z_i}
\end{aligned}
$$

เนื่องจาก

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

ดังนั้น

$$\frac{\partial^2 Cost}{\partial w_d^2} = \frac{1}{n}\sum_{i=1}^{n} x_{i,d}^2\hat{y}_i(1 - \hat{y}_i)$$

Derive Equation 3

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