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

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

เรามี $Cost = -\frac{1}{n}\sum_{i=1}^{n}\sum_{c=1}^{k} y_{i,c}\log\hat{y}_{i,c}$

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

$$\nabla^2 Cost = \begin{bmatrix} \frac{\partial^2 Cost}{\partial w_{0,1}^2} & \frac{\partial^2 Cost}{\partial w_{0,2}^2} & \cdots & \frac{\partial^2 Cost}{\partial w_{0,k}^2} \\ \frac{\partial^2 Cost}{\partial w_{1,1}^2} & \frac{\partial^2 Cost}{\partial w_{1,2}^2} & \cdots & \frac{\partial^2 Cost}{\partial w_{1,k}^2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 Cost}{\partial w_{p,1}^2} & \frac{\partial^2 Cost}{\partial w_{p,2}^2} & \cdots & \frac{\partial^2 Cost}{\partial w_{p,k}^2} \end{bmatrix}$$

พิจารณา $\frac{\partial Cost}{\partial w_{d,s}}$ เมื่อ $d \in \{0,1,...,p\}, s \in \{1,2,...,k\}$ จะได้ว่า

$$\frac{\partial Cost}{\partial w_{d,s}} = -\frac{1}{n}\sum_{i=1}^{n} x_{i,d}(y_{i,s} - \hat{y}_{i,s}) \tag{1}$$

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

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

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

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

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

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

$$\nabla^2 Cost = \begin{bmatrix} \frac{\partial^2 Cost}{\partial w_{0,1}^2} & \frac{\partial^2 Cost}{\partial w_{0,2}^2} & \cdots & \frac{\partial^2 Cost}{\partial w_{0,k}^2} \\ \frac{\partial^2 Cost}{\partial w_{1,1}^2} & \frac{\partial^2 Cost}{\partial w_{1,2}^2} & \cdots & \frac{\partial^2 Cost}{\partial w_{1,k}^2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 Cost}{\partial w_{p,1}^2} & \frac{\partial^2 Cost}{\partial w_{p,2}^2} & \cdots & \frac{\partial^2 Cost}{\partial w_{p,k}^2} \end{bmatrix} \geq \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

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

Derive Equation 1

$$
\begin{aligned}
\frac{\partial Cost}{\partial w_{d,s}} &= \frac{\partial \frac{-1}{n} \sum_{i=1}^{n} \sum_{c=1}^{k} y_{i,c} \log \hat{y}_{i,c}}{\partial w_{d,s}} \\
&= -\frac{1}{n} \sum_{i=1}^{n} \frac{\partial \sum_{c=1}^{k} y_{i,c} \log \hat{y}_{i,c}}{\partial w_{d,s}} \\
&= -\frac{1}{n} \sum_{i=1}^{n} \frac{\partial [y_{i,1} \log \hat{y}_{i,1} + y_{i,2} \log \hat{y}_{i,2} + \cdots + y_{i,s} \log \hat{y}_{i,s} + \cdots + y_{i,k} \log \hat{y}_{i,k}]}{\partial w_{d,s}} \\
&= -\frac{1}{n} \sum_{i=1}^{n} \left[\frac{\partial y_{i,1} \log \hat{y}_{i,1}}{\partial w_{d,s}} + \frac{\partial y_{i,2} \log \hat{y}_{i,2}}{\partial w_{d,s}} + \cdots + \frac{\partial y_{i,s} \log \hat{y}_{i,s}}{\partial w_{d,s}} + \cdots + \frac{\partial y_{i,k} \log \hat{y}_{i,k}}{\partial w_{d,s}} \right] \\
&= -\frac{1}{n} \sum_{i=1}^{n} \left[y_{i,1} \frac{\partial \log \hat{y}_{i,1}}{\partial w_{d,s}} + y_{i,2} \frac{\partial \log \hat{y}_{i,2}}{\partial w_{d,s}} + \cdots + y_{i,s} \frac{\partial \log \hat{y}_{i,s}}{\partial w_{d,s}} + \cdots + y_{i,k} \frac{\partial \log \hat{y}_{i,k}}{\partial w_{d,s}} \right] \\
&= -\frac{1}{n} \sum_{i=1}^{n} \left[y_{i,1} \frac{\partial \log \hat{y}_{i,1}}{\partial \hat{y}_{i,1}} \cdot \frac{\partial \hat{y}_{i,1}}{\partial z_{i,s}} \cdot \frac{\partial z_{i,s}}{\partial w_{d,s}} + y_{i,2} \frac{\partial \log \hat{y}_{i,2}}{\partial \hat{y}_{i,2}} \cdot \frac{\partial \hat{y}_{i,2}}{\partial z_{i,s}} \cdot \frac{\partial z_{i,s}}{\partial w_{d,s}} + \cdots \right. \\
&\qquad \left. + y_{i,s} \frac{\partial \log \hat{y}_{i,s}}{\partial \hat{y}_{i,s}} \cdot \frac{\partial \hat{y}_{i,s}}{\partial z_{i,s}} \cdot \frac{\partial z_{i,s}}{\partial w_{d,s}} + \cdots + y_{i,k} \frac{\partial \log \hat{y}_{i,k}}{\partial \hat{y}_{i,k}} \cdot \frac{\partial \hat{y}_{i,k}}{\partial z_{i,s}} \cdot \frac{\partial z_{i,s}}{\partial w_{d,s}} \right] \\
&= -\frac{1}{n} \sum_{i=1}^{n} \left[y_{i,1} \left(\frac{1}{\hat{y}_{i,1}} \right) \frac{\partial \hat{y}_{i,1}}{\partial z_{i,s}} (x_{i,d}) + y_{i,2} \left(\frac{1}{\hat{y}_{i,2}} \right) \frac{\partial \hat{y}_{i,2}}{\partial z_{i,s}} (x_{i,d}) + \cdots \right. \\
&\qquad \left. + y_{i,s} \left(\frac{1}{\hat{y}_{i,s}} \right) \frac{\partial \hat{y}_{i,s}}{\partial z_{i,s}} (x_{i,d}) + \cdots + y_{i,k} \left(\frac{1}{\hat{y}_{i,k}} \right) \frac{\partial \hat{y}_{i,k}}{\partial z_{i,s}} (x_{i,d}) \right] \\
&= -\frac{1}{n} \sum_{i=1}^{n} \left[\frac{y_{i,1}}{\hat{y}_{i,1}} \frac{\partial \hat{y}_{i,1}}{\partial z_{i,s}} x_{i,d} + \frac{y_{i,2}}{\hat{y}_{i,2}} \frac{\partial \hat{y}_{i,2}}{\partial z_{i,s}} x_{i,d} + \cdots + \frac{y_{i,s}}{\hat{y}_{i,s}} \frac{\partial \hat{y}_{i,s}}{\partial z_{i,s}} x_{i,d} + \cdots + \frac{y_{i,k}}{\hat{y}_{i,k}} \frac{\partial \hat{y}_{i,k}}{\partial z_{i,s}} x_{i,d} \right]
\end{aligned}
$$

เนื่องจาก

$$\frac{\partial \hat{y}_{i,m}}{\partial z_{i,s}} = \hat{y}_{i,s}(1 - \hat{y}_{i,s}) \qquad \text{เมื่อ } m = s \tag{3}$$

$$\frac{\partial \hat{y}_{i,m}}{\partial z_{i,s}} = -\hat{y}_{i,m}\hat{y}_{i,s} \qquad \text{เมื่อ } m \neq s \tag{4}$$

ดังนั้น

$$
\begin{aligned}
\frac{\partial Cost}{\partial w_{d,s}} &= -\frac{1}{n}\sum_{i=1}^{n}\Bigg[\frac{y_{i,1}}{\hat{y}_{i,1}}(-\hat{y}_{i,1}\hat{y}_{i,s})x_{i,d} + \frac{y_{i,2}}{\hat{y}_{i,2}}(-\hat{y}_{i,2}\hat{y}_{i,s})x_{i,d} + \cdots \\
&\qquad + \frac{y_{i,s}}{\hat{y}_{i,s}}(\hat{y}_{i,s}(1-\hat{y}_{i,s}))x_{i,d} + \cdots + \frac{y_{i,k}}{\hat{y}_{i,k}}(-\hat{y}_{i,k}\hat{y}_{i,s})x_{i,d}\Bigg] \\
&= -\frac{1}{n}\sum_{i=1}^{n}[-y_{i,1}\hat{y}_{i,s}x_{i,d} - y_{i,2}\hat{y}_{i,s}x_{i,d} - \cdots + y_{i,s}(1-\hat{y}_{i,s})x_{i,d} - \cdots - y_{i,k}\hat{y}_{i,s}x_{i,d}] \\
&= -\frac{1}{n}\sum_{i=1}^{n}[(-y_{i,1}\hat{y}_{i,s} - y_{i,2}\hat{y}_{i,s} - \cdots + y_{i,s}(1-\hat{y}_{i,s}) - \cdots - y_{i,k}\hat{y}_{i,s})x_{i,d}] \\
&= -\frac{1}{n}\sum_{i=1}^{n}[(-y_{i,1}\hat{y}_{i,s} - y_{i,2}\hat{y}_{i,s} - \cdots + y_{i,s} - y_{i,s}\hat{y}_{i,s} - \cdots - y_{i,k}\hat{y}_{i,s})x_{i,d}] \\
&= -\frac{1}{n}\sum_{i=1}^{n}[(y_{i,s} - y_{i,1}\hat{y}_{i,s} - y_{i,2}\hat{y}_{i,s} - \cdots - y_{i,s}\hat{y}_{i,s} - \cdots - y_{i,k}\hat{y}_{i,s})x_{i,d}] \\
&= -\frac{1}{n}\sum_{i=1}^{n}[(y_{i,s} - (y_{i,1} + y_{i,2} + \cdots + y_{i,s} + \cdots + y_{i,k})\hat{y}_{i,s})x_{i,d}] \\
&= -\frac{1}{n}\sum_{i=1}^{n}[(y_{i,s} - (1)\hat{y}_{i,s})x_{i,d}] \\
&= -\frac{1}{n}\sum_{i=1}^{n}[(y_{i,s} - \hat{y}_{i,s})x_{i,d}] \\
&= -\frac{1}{n}\sum_{i=1}^{n}x_{i,d}(y_{i,s} - \hat{y}_{i,s})
\end{aligned}
$$

Derive Equation 2

$$
\begin{aligned}
\frac{\partial^2 Cost}{\partial w_{d,s}^2} &= \frac{\partial}{\partial w_{d,s}}\left(\frac{\partial Cost}{\partial w_{d,s}}\right) \\
&= \frac{\partial}{\partial w_{d,s}}\left(-\frac{1}{n}\sum_{i=1}^{n} x_{i,d}(y_{i,s}-\hat{y}_{i,s})\right) \\
&= \frac{\partial\left(-\frac{1}{n}\sum_{i=1}^{n} x_{i,d}(y_{i,s}-\hat{y}_{i,s})\right)}{\partial w_{d,s}} \\
&= -\frac{1}{n}\sum_{i=1}^{n}\frac{\partial x_{i,d}(y_{i,s}-\hat{y}_{i,s})}{\partial w_{d,s}} \\
&= -\frac{1}{n}\sum_{i=1}^{n}\frac{\partial[x_{i,d}y_{i,s}-x_{i,d}\hat{y}_{i,s}]}{\partial w_{d,s}} \\
&= -\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\partial x_{i,d}y_{i,s}}{\partial w_{d,s}}-\frac{\partial x_{i,d}\hat{y}_{i,s}}{\partial w_{d,s}}\right] \\
&= -\frac{1}{n}\sum_{i=1}^{n}\left[0-\frac{\partial x_{i,d}\hat{y}_{i,s}}{\partial w_{d,s}}\right] \\
&= -\frac{1}{n}\sum_{i=1}^{n}\left[-x_{i,d}\frac{\partial \hat{y}_{i,s}}{\partial w_{d,s}}\right] \\
&= \frac{1}{n}\sum_{i=1}^{n} x_{i,d}\frac{\partial \hat{y}_{i,s}}{\partial w_{d,s}} \\
&= \frac{1}{n}\sum_{i=1}^{n} x_{i,d}\frac{\partial \hat{y}_{i,s}}{\partial z_{i,s}}\cdot\frac{\partial z_{i,s}}{\partial w_{d,s}} \\
&= \frac{1}{n}\sum_{i=1}^{n} x_{i,d}\frac{\partial \hat{y}_{i,s}}{\partial z_{i,s}}x_{i,d} \\
&= \frac{1}{n}\sum_{i=1}^{n} x_{i,d}^2\frac{\partial \hat{y}_{i,s}}{\partial z_{i,s}}
\end{aligned}
$$

เนื่องจาก

$$\frac{\partial \hat{y}_{i,s}}{\partial z_{i,s}} = \hat{y}_{i,s}(1-\hat{y}_{i,s})$$

ดังนั้น

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

Derive Equation 3

$$
\begin{aligned}
\frac{\partial \hat{y}_{i,s}}{\partial z_{i,s}} &= \frac{\partial \left(\frac{e^{z_{i,s}}}{\sum_{c=1}^{k} e^{z_{i,c}}} \right)}{\partial e^{z_{i,s}}} \cdot \frac{\partial e^{z_{i,s}}}{\partial z_{i,s}} \\
&= \frac{\sum_{c=1}^{k} e^{z_{i,c}} \frac{\partial e^{z_{i,s}}}{\partial e^{z_{i,s}}} - e^{z_{i,s}} \frac{\partial \sum_{c=1}^{k} e^{z_{i,c}}}{\partial e^{z_{i,s}}}}{\left(\sum_{c=1}^{k} e^{z_{i,c}} \right)^2} \frac{\partial e^{z_{i,s}}}{\partial z_{i,s}} \\
&= \frac{\sum_{c=1}^{k} e^{z_{i,c}}(1) - e^{z_{i,s}}(1)}{\left(\sum_{c=1}^{k} e^{z_{i,c}} \right)^2} e^{z_{i,s}} \\
&= \frac{e^{z_{i,s}} \left(\sum_{c=1}^{k} e^{z_{i,c}} - e^{z_{i,s}} \right)}{\left(\sum_{c=1}^{k} e^{z_{i,c}} \right)^2} \\
&= \frac{e^{z_{i,s}}}{\sum_{c=1}^{k} e^{z_{i,c}}} \frac{\sum_{c=1}^{k} e^{z_{i,c}} - e^{z_{i,s}}}{\sum_{c=1}^{k} e^{z_{i,c}}} \\
&= \frac{e^{z_{i,s}}}{\sum_{c=1}^{k} e^{z_{i,c}}} \left(\frac{\sum_{c=1}^{k} e^{z_{i,c}}}{\sum_{c=1}^{k} e^{z_{i,c}}} - \frac{e^{z_{i,s}}}{\sum_{c=1}^{k} e^{z_{i,c}}} \right) \\
&= \hat{y}_{i,s}(1 - \hat{y}_{i,s})
\end{aligned}
$$

Derive Equation 4

$$
\begin{aligned}
\frac{\partial \hat{y}_{i,m}}{\partial z_{i,s}} &= \frac{\partial \left(\frac{e^{z_{i,m}}}{\sum_{c=1}^{k} e^{z_{i,c}}} \right)}{\partial e^{z_{i,s}}} \cdot \frac{\partial e^{z_{i,s}}}{\partial z_{i,s}} \\
&= \frac{\sum_{c=1}^{k} e^{z_{i,c}} \frac{\partial e^{z_{i,m}}}{\partial e^{z_{i,s}}} - e^{z_{i,m}} \frac{\partial \sum_{c=1}^{k} e^{z_{i,c}}}{\partial e^{z_{i,s}}}}{\left(\sum_{c=1}^{k} e^{z_{i,c}} \right)^2} \frac{\partial e^{z_{i,s}}}{\partial z_{i,s}} \\
&= \frac{\sum_{c=1}^{k} e^{z_{i,c}} (0) - e^{z_{i,m}} (1)}{\left(\sum_{c=1}^{k} e^{z_{i,c}} \right)^2} e^{z_{i,s}} \\
&= \frac{-e^{z_{i,m}} e^{z_{i,s}}}{\left(\sum_{c=1}^{k} e^{z_{i,c}} \right)^2} \\
&= \left(\frac{-e^{z_{i,m}}}{\sum_{c=1}^{k} e^{z_{i,c}}} \right) \left(\frac{e^{z_{i,s}}}{\sum_{c=1}^{k} e^{z_{i,c}}} \right) \\
&= -\hat{y}_{i,m} \hat{y}_{i,s}
\end{aligned}
$$