Convexity of Sum of Squared Errors for Linear Regression

เราต้องการพิสูจน์ว่า sum of squared errors สำหรับ linear regression เป็น convex function

เรามี $Cost = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$

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

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

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

$$\frac{\partial^2 Cost}{\partial w_d^2} = 2\sum_{i=1}^{n} x_{i,d}^2 \tag{2}$$

พิจารณา $x_{i,d}^2$

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

จะได้ว่า $2\sum_{i=1}^{n} x_{i,d}^2 \geq 0$

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 sum of squared error สำหรับ linear regression เป็น convex function

Derive Equation 1

$$
\begin{aligned}
\frac{\partial Cost}{\partial w_d} &= \frac{\partial \sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\partial w_d} \\
&= \sum_{i=1}^{n} \frac{\partial (y_i - \hat{y}_i)^2}{\partial (y_i - \hat{y}_i)} \cdot \frac{\partial (y_i - \hat{y}_i)}{\partial \hat{y}_i} \cdot \frac{\partial \hat{y}_i}{\partial w_d} \\
&= \sum_{i=1}^{n} 2(y_i - \hat{y}_i)(1)(-1)\frac{\partial \hat{y}_i}{\partial w_d} \\
&= \sum_{i=1}^{n} -2(y_i - \hat{y}_i)\frac{(w_0 + w_1 x_{i,1} + \cdots + w_d x_{i,d} + \cdots + w_p x_{i,p})}{\partial w_d} \\
&= \sum_{i=1}^{n} -2(y_i - \hat{y}_i)x_{i,d} \\
&= -2\sum_{i=1}^{n}(y_i - \hat{y}_i)x_{i,d} \\
&= -2\sum_{i=1}^{n} x_{i,d}(y_i - \hat{y}_i)
\end{aligned}
$$

Derive Equation 2

$$
\begin{aligned}
\frac{\partial^2 Cost}{\partial w_d^2} &= \frac{\partial}{\partial w_d}\left(\frac{\partial Cost}{\partial w_d}\right) \\
&= \frac{\partial}{\partial w_d}\left(-2\sum_{i=1}^{n} x_{i,d}(y_i - \hat{y}_i)\right) \\
&= \frac{\partial\left(-2\sum_{i=1}^{n} x_{i,d}(y_i - \hat{y}_i)\right)}{\partial w_d} \\
&= -2\sum_{i=1}^{n} \frac{\partial x_{i,d}(y_i - \hat{y}_i)}{\partial w_d} \\
&= -2\sum_{i=1}^{n} \frac{\partial[x_{i,d}y_i - x_{i,d}\hat{y}_i]}{\partial w_d} \\
&= -2\left[\frac{\partial x_{i,d}y_i}{\partial w_d} - \frac{\partial x_{i,d}\hat{y}_i}{\partial w_d}\right] \\
&= -2\sum_{i=1}^{n}\left[0 - \frac{\partial x_{i,d}\hat{y}_i}{\partial w_d}\right] \\
&= -2\sum_{i=1}^{n}\left[-\frac{\partial x_{i,d}\hat{y}_i}{\partial w_d}\right] \\
&= 2\sum_{i=1}^{n}\left[\frac{\partial x_{i,d}\hat{y}_i}{\partial w_d}\right] \\
&= 2\sum_{i=1}^{n}\left[x_{i,d}\frac{\partial \hat{y}_i}{\partial w_d}\right] \\
&= 2\sum_{i=1}^{n}\left[x_{i,d}\frac{w_0 + w_1x_{i,1} + \cdots + w_dx_{i,d} + \cdots + w_px_{i,p}}{\partial w_d}\right] \\
&= 2\sum_{i=1}^{n} x_{i,d}x_{i,d} \\
&= 2\sum_{i=1}^{n} x_{i,d}^2
\end{aligned}
$$