Derivation of Ridge Regression

Derive $\mathbf{w} = (X_b^T X_b + \lambda I)^{-1} X_b^T \mathbf{y}$

เรามี $Cost = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \lambda \sum_{j=0}^{p} w_j^2$ และเราต้องการหา $\mathbf{w} = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_p \end{bmatrix}$ ที่ทำให้ $Cost$ ต่ำที่สุด

จาก calculus เราทราบว่า $Cost$ ต่ำสุดเมื่อ $\nabla Cost = 0$ ดังนั้น

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

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

$$\sum_{i=1}^{n} x_{i,d}\hat{y}_i + \lambda w_d = \sum_{i=1}^{n} x_{i,d} y_i \tag{1}$$

ซึ่งสามารถเขียนให้อยู่ในรูป matrix ได้ดังนี้

$$\begin{bmatrix} x_{1,d} & x_{2,d} & \cdots & x_{n,d} \end{bmatrix} \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix} + \lambda w_d = \begin{bmatrix} x_{1,d} & x_{2,d} & \cdots & x_{n,d} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \tag{2}$$

จากสมการ

$$\begin{bmatrix} \frac{\partial Cost}{\partial w_0} \\ \frac{\partial Cost}{\partial w_1} \\ \vdots \\ \frac{\partial Cost}{\partial w_d} \\ \vdots \\ \frac{\partial Cost}{\partial w_p} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

เราจะสามารถเขียนให้อยู่ในรูปต่อไปนี้ได้

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_{1,1} & x_{2,1} & \cdots & x_{n,1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1,p} & x_{2,p} & \cdots & x_{n,p} \end{bmatrix} \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix} + \lambda \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_p \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_{1,1} & x_{2,1} & \cdots & x_{n,1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1,p} & x_{2,p} & \cdots & x_{n,p} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} ; x_{i,0} = 1; i \in \{1, 2, ..., n\}$$

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 $X_b^T \hat{\mathbf{y}} + \lambda \mathbf{w} = X_b^T \mathbf{y}$

$$\because \hat{\mathbf{y}} = X_b \mathbf{w} \tag{3}$$

จะได้

$$X_b^T X_b \mathbf{w} + \lambda \mathbf{w} = X_b^T \mathbf{y}$$

$$(X_b^T X_b + \lambda I)\mathbf{w} = X_b^T \mathbf{y}$$

ดังนั้น

$$\mathbf{w} = (X_b^T X_b + \lambda I)^{-1} X_b^T \mathbf{y}$$

Derive Equation 1

เนื่องจาก

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

และ

$$
\begin{aligned}
\frac{\partial \lambda \sum_{j=0}^{p} w_j^2}{\partial w_d} &= \lambda \frac{\partial \sum_{j=0}^{p} w_j^2}{\partial w_d} \\
&= \lambda \frac{\partial (w_0^2 + w_1^2 + \cdots + w_d^2 + \cdots + w_p^2)}{\partial w_d} \\
&= 2\lambda w_d
\end{aligned}
$$

ดังนั้น

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

Derive Equation 2

พิจารณา $\sum_{i=1}^{n} x_{i,d}\hat{y}_i$

$$\begin{aligned}\sum_{i=1}^{n} x_{i,d}\hat{y}_i &= x_{1,d}\hat{y}_1 + x_{2,d}\hat{y}_2 + \cdots + x_{n,d}\hat{y}_n \\ &= \begin{bmatrix} x_{1,d} & x_{2,d} & \cdots & x_{n,d} \end{bmatrix} \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix}\end{aligned}$$

พิจารณา $\sum_{i=1}^{n} x_{i,d}y_i$

$$\begin{aligned}\sum_{i=1}^{n} x_{i,d}y_i &= x_{1,d}y_1 + x_{2,d}y_2 + \cdots + x_{n,d}y_n \\ &= \begin{bmatrix} x_{1,d} & x_{2,d} & \cdots & x_{n,d} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}\end{aligned}$$

ดังนั้น จะได้ว่า

$$\sum_{i=1}^{n} x_{i,d}\hat{y}_i + \lambda w_d = \sum_{i=1}^{n} x_{i,d}y_i$$

$$\begin{bmatrix} x_{1,d} & x_{2,d} & \cdots & x_{n,d} \end{bmatrix} \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix} + \lambda w_d = \begin{bmatrix} x_{1,d} & x_{2,d} & \cdots & x_{n,d} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

Derive Equation 3

$$
\begin{aligned}
\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix} &= \begin{bmatrix} w_0 + w_1 x_{1,1} + \cdots + w_p x_{1,p} \\ w_0 + w_1 x_{2,1} + \cdots + w_p x_{2,p} \\ \vdots \\ w_0 + w_1 x_{n,1} + \cdots + w_p x_{n,p} \end{bmatrix} \\
&= \begin{bmatrix} 1 & x_{1,1} & \cdots & x_{1,p} \\ 1 & x_{2,1} & \cdots & x_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n,1} & \cdots & x_{n,p} \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_p \end{bmatrix}
\end{aligned}
$$

กำหนดให้ $\hat{\mathbf{y}} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix}, X_b = \begin{bmatrix} 1 & x_{1,1} & \cdots & x_{1,p} \\ 1 & x_{2,1} & \cdots & x_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n,1} & \cdots & x_{n,p} \end{bmatrix}, \mathbf{w} = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_p \end{bmatrix}$

ดังนั้น จะได้ว่า

$$\hat{\mathbf{y}} = X_b \mathbf{w}$$