# Mathematic behind Ridge

$$L = \sum_{i=1}^{n} (y_i - mx_i - b)^2 + \lambda m^2$$

$$\frac{\partial l}{\partial b} = 0, \quad \frac{\partial l}{\partial m} = 0$$

differentiation wrt b will remain same as in simple linear regression

$$b = \bar{y} - m\bar{x} \quad \text{①} \qquad , \bar{y} \to \text{mean}$$

$\bar{x} \to$ x-mean, $m \to$ slope.

$$\Rightarrow L = \sum_{i=1}^{n} (y_i - mx_i - \bar{y} + m\bar{x})^2 + \lambda m^2$$

(using ①)

$$\Rightarrow \frac{\partial l}{\partial m} \Rightarrow 2\sum_{i=1}^{n} (y_i - mx_i - \bar{y} + m\bar{x}) \cdot \frac{d}{dm}(y_i - mx_i - \bar{y} + m\bar{x})$$

$$+ \; 2\lambda m \cdot \frac{d m}{dm} = 0$$

$$\Rightarrow 2\sum_{i=1}^{n} [y_i - mx_i - \bar{y} + m\bar{x})(-x_i + \bar{x})] + 2\lambda m = 0$$

$$\Rightarrow -\cancel{2}\sum_{i=1}^{n} [y_i - mx_i - \bar{y} + m\bar{x})(x_i - \bar{x}) + \cancel{2}\lambda m = 0$$

$$\Rightarrow \lambda m - \sum_{i=1}^{n} [(y_i - \bar{y}) - m(x_i - \bar{x})](x_i - \bar{x}) = 0$$

$$\Rightarrow \lambda m - \sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x}) - m(x_i - \bar{x})^2 = 0$$

$$\lambda m - \sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x}) + m \sum_{i=1}^{n} (x_i - \bar{x})^2 = 0$$

$$\lambda m + m \sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})$$

$$m\left(\lambda + \sum_{i=1}^{n} (x_i - \bar{x})^2\right) = \sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})$$

$$\boxed{m = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2 + \lambda}}$$

Ridge regression m (slope)

If $\lambda = 0$, then m will be same as Simple linear regression

---

Ridge for higher dimension

$x_1 \quad x_2 \quad x_3 \; \ldots \; x_n \quad y$ (n+1 cols)

$w_1 \quad w_2 \quad w_2 \; \ldots \; w_n$ ⟷ coefficients.

m rows

$$L = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$ (for Simple Linear Regression)

For Multiple linear regression

Loss function $= (XW - Y)^T (XW - Y)$

$$Y = \begin{bmatrix} - \\ - \\ - \end{bmatrix}, \quad W = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_n \end{bmatrix}_{n \times 1}, \quad X = \begin{bmatrix} 1 & x_{11} & x_{12} & \ldots \\ \vdots & & & \\ 1 & x_{m1} & x_{m2} & \ldots \end{bmatrix}$$

m values

For Ridge ?

$$\boxed{L = (XW - Y)^T (XW - Y) + \lambda \| W^2 \|}$$

$$W = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_n \end{bmatrix}$$

$$\lambda w_0^2 + \lambda w_1^2 + \lambda w_2^2 + \lambda w_3^2 + \dots w_n^2$$

$$\lambda (w_0^2 + w_1^2 + \dots w_n^2)$$

$$W^T = [w_0 \; w_1 \; w_2 \dots w_n]$$

$$W^2 = W^T W \qquad [w_0 \; w_1 \dots w_n] \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_n \end{bmatrix}$$

can be rewrite as :-

$$L = (XW - Y)^T (XW - Y) + \lambda W^T W$$

$$L = [((XW)^T - Y^T)(XW - Y) + \lambda W^T W]$$

$$L = [(W^T X^T - Y^T)(XW - Y) + \lambda W^T W]$$

$$L = [(W^T X^T X W - W^T X^T Y - Y^T X W + Y^T Y) + \lambda W^T W]$$

Same

$$L = W^T X^T X W - 2 W^T X^T Y + Y^T Y + \lambda W^T W$$

$$\frac{\partial L}{\partial W} = ?$$

$$\Rightarrow 2X^TXW - 2X^TY + 0 + 2\lambda W = 0$$

$$\Rightarrow X^TXW + \lambda W = X^TY$$

$$\Rightarrow W(X^TX + \lambda I) = X^TY$$

$$\Rightarrow \boxed{W = (X^TX + \lambda I)^{-1} X^TY}$$