So, as $\lambda \uparrow$ $m \downarrow$. → due to this b will also affected & b will also change.

$\because (b = \bar{y} - m\bar{x})$

Ridge Regression for nD Data.

$$\boxed{Loss = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = (XW - y)^T (XW - y)}$$

	x_1	x_2	------	x_n	y
m rows	w_1	w_2	......	w_n	—

where,

m values, n values

$$y = \begin{bmatrix} - \\ - \\ - \\ - \\ - \end{bmatrix}, \quad w = \underset{(n \times 1)}{\begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_n \end{bmatrix}}, \quad X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1n} \\ 1 & x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & & & & \\ 1 & x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix}$$

Loss function with regularization term

$$L = (XW - y)^T (XW - y) + \lambda \|W^2\|$$

$\{\lambda w_0^2 + \lambda w_1^2 + \lambda w_2^2 - - - + \lambda w_n^2\}$

$\therefore \|W^2\| \simeq W^T W$

$$[w_0\ w_1 \cdots w_n] \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_n \end{bmatrix}$$

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

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

$\because (a-b)^T = a^T - b^T$

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

$\because (ab)^T = b^T a^T$

$$= W^T X^T X W - \underbrace{W^T X^T y - y^T X W}_{\text{same}} + y^T y + \lambda W^T W$$

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

$$\frac{\partial L}{\partial W} \Rightarrow 2X^T X W - 2X^T y + 0 + 2\lambda W = 0$$

$$\Rightarrow X^T X W + \lambda W = X^T y$$

$$\Rightarrow (X^T X + \lambda I) W = X^T y$$

$$\boxed{W = (X^T X + \lambda I)^{-1} X^T y}$$

$$\boxed{W = (X^T X)^{-1} X^T y}$$

normal without regularization.

Extra due to regularization.