RIDGE REGRESSION USING GRADIENT DESCENT

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

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

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & x_{13} & \cdots & x_{1n} \\ 1 & x_{21} & x_{22} & x_{23} & \cdots & x_{2n} \\ \vdots & & & & & \\ 1 & x_{m1} & x_{m2} & x_{m3} & \cdots & x_{mn} \end{bmatrix} \quad Y = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_m \end{bmatrix}$$

$$W = \begin{bmatrix} W_0 \\ W_1 \\ \vdots \\ W_n \end{bmatrix}$$

If $W_0 . W_1 \cdots W_n$ (parameters)
then we have to take update

$$W_0 = W_0 - \eta \frac{\partial L}{\partial W_0}, \quad W_1 = W_1 - \eta \frac{\partial L}{\partial W_1} \cdots$$

$$W_{new} = W_{old} - \eta \boxed{\frac{\Delta L}{\Delta W}} \rightarrow \text{gradient} \rightarrow \begin{bmatrix} \frac{\partial L}{\partial W_0} \\ \frac{\partial L}{\partial W_1} \\ \vdots \end{bmatrix}$$

$$= \frac{1}{2}(W^T X^T - Y^T)(XW - Y) \cdot + \frac{1}{2}\lambda W^T W$$

$$\frac{1}{2}\Big[W^T X^T X W - \underbrace{W^T X^T Y - Y^T X W}_{\text{same}} + Y^T Y \Big] + \frac{1}{2}\lambda W^T W \Big]$$

$$\frac{1}{2}\Big[W^T X^T X W - \underset{2W^T X^T Y}{\cancel{2Y^T X W}} + Y^T Y \Big] + \frac{1}{2}\lambda W^T W$$

$$\frac{\partial l}{\partial W} = \frac{1}{2}\Big[2X^T X W - \underset{2W^T X^T Y}{\cancel{2Y^T X}} \Big] + \frac{1}{2} 2\lambda W \quad (\text{differentiating})$$

$$\Rightarrow \boxed{X^T X W - \underset{X^T Y}{\cancel{Y^T X}} + \lambda W = \frac{\partial l}{\partial W}}$$

$$W = \begin{array}{c} W_0 \; W_1 \; \ldots \\ [0, 1 \ldots\ldots 1] \end{array} \quad (\text{Initializing})$$

epochs.

$$\boxed{W = W - \eta \frac{\partial l}{\partial W}}$$