Ridge Regression. Theory.

1. Destignation

$a = (a_1 \ \dots \ a_l)$ - prediction

$y = (y_1 \ \dots \ y_l)$ - target

$$X = \begin{pmatrix} 1 & x_{11} & \dots & x_{1d} \\ 1 & & \dots\dots\dots & \\ 1 & & \dots\dots\dots & \\ 1 & x_{l1} & \dots & x_{ld} \end{pmatrix}$$ - features

$$w = \begin{pmatrix} w_0 \\ w_1 \\ \dots \\ w_d \end{pmatrix}$$ - weights

$a = Xw = np.dot(X, w)$ - prediction

$Q(w, X) = \frac{1}{2l} \| Xw - y \|^2 + \frac{\lambda}{2} \| w \|^2$ - loss function

2. Standartization

Math	Python
$\mu_j = \frac{1}{l} \sum_{i=1}^{l} x_i^j$	$\mu = np.mean(X, axis = 0)$
$\sigma_j = \sqrt{\frac{1}{l} \sum_{i=1}^{l} \left(x_i^j - \mu_j \right)^2}$	$\sigma = np.std(X, axis = 0)$
$X := \frac{X - \mu}{\sigma}$	

3. Optimization

$Q(w, X) \to \min$

$$\frac{\partial Q}{\partial w} = \frac{1}{l} (a - y) \cdot \frac{\partial a}{\partial w} + \lambda w = \frac{1}{l} (a - y) X + \lambda w$$

$$w := w - \alpha \frac{\partial Q}{\partial w}$$

4. Analytical solution

$$Q = \| Xw - y \|^2 + \lambda \| w \|^2 \rightarrow \min$$

$$Q = \left(Xw - y \right)^T \left(Xw - y \right) + \lambda w^T w$$

$$\frac{\partial Q}{\partial w} = X^T \left(Xw - y \right) + \lambda w = 0$$

$$X^T Xw - X^T y + \lambda w = 0$$

$$w = \left(X^T X + \lambda I \right)^{-1} X^T y$$