$$\mathcal{L} = \left[\frac{1}{2}\frac{1}{N}\sum_{n=1}^{N}\left(\vec{w}\cdot\vec{x}_n - y_n\right)^2\right] + \left[\frac{1}{2}\lambda\underbrace{\sum_{d=1}^{D}(w_d)^2}_{\|\ \|_n}\right] \quad 3.1.b$$

$\nabla_{\vec{w}}\mathcal{L} = ?$

$$\frac{\partial\mathcal{L}}{\partial w_{1}}_{\hookrightarrow d} = \cancel{2}\frac{1}{\cancel{2}}\frac{1}{N}\cdot\sum_{n=1}^{N} x_{n1 \hookrightarrow d}\left(\vec{w}\cdot\vec{x}_n - \vec{y}_n\right)^1 + \lambda\cdot\frac{\cancel{2}}{\cancel{2}}\cdot w_{1 \hookrightarrow d} \quad (1)$$

$$\boxed{\vec{\nabla}_{\vec{w}}\mathcal{L} = \frac{1}{N}\sum_{n=1}^{N}\vec{x}_n\left(\underbrace{\vec{w}\cdot\vec{x}_n}_{\sum^{D}_{d}} - y_n\right) + \lambda\vec{w}} \quad (2)$$

$$= \frac{1}{N}\left(\underbrace{\underset{D}{W}\cdot\underset{D,N}{X^T}} - \underset{N}{Y}\right)\cdot\underset{N,D}{X} + \lambda W \quad (3)$$

$$w^*: \vec{\nabla}_{\vec{w}}\mathcal{L} = \vec{0} \iff \frac{1}{N}\vec{w}X^TX - YX + \lambda\vec{w} = \vec{0} \quad (4)$$

$$\vec{w}\left(\frac{1}{N}X^TX + \lambda I_{d_D}\right) = YX \quad (5)$$

$$\vec{w} = YX\cdot\left(\frac{1}{N}X^TX + \lambda I_{d_D}\right)^{-1} \quad (6)$$

$D = 1$:

$$w = YX\left(\frac{1}{N}X^TX + \lambda\right)^{-1} \quad (7)$$

$$w \xrightarrow[\lambda\to 0]{} YX\left(\frac{1}{N}X^TX\right)^{-1} \quad \text{back to regression}$$

$$w \xrightarrow[\lambda\to\infty]{} \frac{YX}{\frac{1}{N}X^TX + \lambda} \longrightarrow 0 \quad \text{weights are crushed towards 0.}$$

3.2 Gradient Descent Step for Ridge Regression, $\vec{\nabla}_{\vec{w}}\mathcal{L}_{Ridge} = \vec{\nabla}_{\vec{w}}\mathcal{L} + \lambda\ldots$

$$\vec{w} \to \vec{w} - \eta\cdot\vec{\nabla}_{\vec{w}}\mathcal{L}_{Ridge} = \vec{w} - \eta\cdot\vec{\nabla}_{\vec{w}}\mathcal{L}_{(\lambda=0)} - \eta\cdot\lambda\cdot\vec{w}$$

$$= w\underbrace{(1 - \eta \cdot \lambda)}_{\in [0,1]} - \eta \vec{\nabla}_{\vec{w}} \mathcal{L}_{(\lambda=0)}$$

$$\left(\text{because } \begin{array}{l} \eta \in [0,1] \\ \lambda > 0 \end{array}\right)$$

$$= \vec{w} \cdot \underbrace{\alpha}_{\in [0,1]} - \eta \vec{\nabla}_{\vec{w}} \mathcal{L}$$

3.3 Lasso : first encounter, naive approach