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

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

$$\frac{\partial\mathcal{L}}{\partial w_1} = \frac{2}{2}\frac{1}{N}\cdot\sum_{n=1}^{N} x_{n1}\left(\vec{w}\cdot\vec{x}_n - \vec{y}_n\right)^1 + \lambda\cdot\frac{2}{2}\cdot w_1 \quad (1)$$

(1d) (1d)

$$\nabla_{\vec{w}}\mathcal{L} = \frac{1}{N}\sum_{n=1}^{N}\vec{x}_n\left(\vec{w}\cdot\vec{x}_n - y_n\right) + \lambda\vec{w} \quad (2)$$

$$= \frac{1}{N}\left(w\cdot X^T - Y\right)\cdot X + \lambda w \quad (3)$$

$(D,1)\ (D,N)\ (N,1)\ (N,D)$

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

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

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

$D = 1$:

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

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

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

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

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

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

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

3.3 Lasso: first encounter, naive approach

$$\frac{\partial}{\partial w_d}(|w_d|) = \text{sign}(w_d) = \begin{cases} 1 & \text{if } w_d > 0 \\ 0 & \text{if } w_d = 0 \\ -1 & \text{if } w_d < 0 \end{cases}$$

→ actually with pseudo-gradients, one can choose $\frac{\partial |x|}{\partial x} \in [-1, 1]$ when $x = 0$

$$\vec{\nabla}_{\vec{w}} \mathcal{L}_{\text{lasso}} = \vec{\nabla}_{\vec{w}} \mathcal{L}_{(\lambda=0)} + \lambda \underbrace{\vec{\nabla}_{\vec{w}} |\vec{w}|}_{\downarrow ?}$$

$$\frac{\partial}{\partial w_1} |\vec{w}| = \frac{\partial}{\partial w_1} \sum_{d=1}^{D} |w_d| = \frac{\partial}{\partial w_1} |w_1| + 0 + 0 + \dots + 0$$

$$= \text{sign}(w_1)$$

$$\vec{\nabla}_{\vec{w}} |\vec{w}| = \begin{pmatrix} \text{sign}(w_1) \\ \text{sign}(w_2) \\ \vdots \\ \text{sign}(w_D) \end{pmatrix} = \overrightarrow{\text{sign}}(\vec{w})$$

Solve for w: $\vec{\nabla}_{\vec{w}} \mathcal{L}_{\text{lasso}} \overset{!}{=} \vec{0}$

$$\Leftrightarrow \vec{0} = X^T(Xw - Y) + \lambda \vec{\text{sign}}(\vec{w})$$

$$X^T X \vec{w} + \lambda \vec{\text{sign}}\, \vec{w} = X^T Y$$

3.3.6 GD step! $\vec{w} \leftarrow \vec{w} - \eta \vec{\nabla}_{\vec{w}} \mathcal{L}_{\text{lasso}}$

$$w_1 = w_1 - \dots$$

$$w_2 = w_2 - \dots$$