Regularización

(1) Ridge - L2 $\longrightarrow$ $J(w) = \frac{1}{n}\sum_{i=1}^{n} (y_i - w^T x_i)^2 + \lambda \underbrace{\|w\|^2}_{w^T w}$ } L2

(2) LASSO - L1 $\longrightarrow$ $J(w) = \frac{1}{n}\sum_{i=1}^{n} (y_i - w^T x_i)^2 + \lambda \sum_{j=1}^{p} |w_j|$ } L1

Solución Cerrada

ANALÍTICO

(1) ECM $\longrightarrow \min_{\theta} \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{g}_2(x_i))^2 \longrightarrow$ RL : $\underbrace{\hat{y} = w^T x_i}_{\text{solución cerrada}}$

(2) MV $\longrightarrow \min_{\theta} \frac{1}{n} \sum_{i=1}^{n} -\ln(f_\theta(y/x_i))$

(3) Bayesiano $\longrightarrow \min_{\substack{\theta \\ V.A}} \frac{1}{n} \sum_{i=1}^{n} -\ln(f_\theta(y/x_i)) - \underbrace{\ln f_\theta(\theta)}_{\text{distribución a priori}}$

NUMÉRICA

$$\boxed{\nabla_w J(w) = 0}$$

Gradiente Descendente

(0) $\bar{w} \longleftarrow$ random(m,1)

$\bar{w} \Rightarrow m \times 1$

$x \Rightarrow n \times m$

$y \Rightarrow n \times 1$

Forward
$$\begin{cases} (1)\ \hat{y} = Xw \\ (2)\ e = y - \hat{y} \end{cases}$$

(3) Calcular gradiente $\bar{\nabla}_w J(w)$ $\quad J(w) = \frac{1}{n}\sum_{i=1}^{n}(y_i - w^T x_i)^2 \} \in \mathbb{R}^1$

(4) $\bar{w} = \bar{w} - \underset{\sim}{\alpha}\, \bar{\nabla}_w (J(w)) \longrightarrow \nabla_w (J(w)) \to m \times 1$

learning rate $\qquad\qquad \mathbb{R}$

(3) $\nabla_w \underbrace{\frac{1}{n}\sum_{i=1}^{n}(y_i - \underbrace{w^T x_i}_{x_1 w_1 + x_2 w_2 + \dots x_m w_m})^2}_{(1)}$

$$\nabla_w J_w = \begin{bmatrix} \frac{\partial J_w}{\partial w_1} \\ \vdots \\ \frac{\partial J_w}{\partial w_m} \end{bmatrix}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\begin{bmatrix} \frac{\partial (1)}{\partial w_1} \\ \vdots \\ \frac{\partial (1)}{\partial w_m} \end{bmatrix} = \frac{1}{n}\sum_{i=1}^{n}\begin{bmatrix} -2(y_i - w^T x_i)x_{i1} \\ \vdots \\ \underbrace{-2(y_i - w^T x_i)x_{im}}_{\mathbb{R}^1} \end{bmatrix}$$

$$= \frac{1}{n}\sum_{i=1}^{n} \underbrace{-2\overbrace{(y_i - w^T x_i)}^{\beta}}_{\mathbb{R}^1} \, \underbrace{x_i}_{\mathbb{R}^m}$$

$$y_1 = (x_{11}w_1 + x_{12}w_2 + \dots x_{1m}w_m) \in \mathbb{R}^1 \quad (\beta)$$

$$\underbrace{\underbrace{\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}}_{\mathbb{R}^{n\times 1}} - \underbrace{\begin{bmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{bmatrix}}_{\mathbb{R}^{n\times m}} \underbrace{\begin{bmatrix} w_1 \\ \vdots \\ w_m \end{bmatrix}}_{\mathbb{R}^{m\times 1}}}_{\mathbb{R}^{n\times 1}} * \underbrace{\begin{bmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{bmatrix}}_{\mathbb{R}^{n\times m}}$$

$$(\beta_{(1)} \; \beta_{(2)} \cdots \beta_{(m)}) \begin{bmatrix} x_{11} \\ x_{12} \\ \vdots \\ x_{1m} \end{bmatrix}$$

$$= (\beta x_{11} \;\; \beta x_{12} \cdots \beta x_{1m})$$

$$\mathbb{R}^{1\times m} \quad (i = 1)$$

$$\mathbb{R}^{n\times m} \quad (\theta)$$

$\nabla_w J(w)$ np.sum(θ, axis=0) $\longrightarrow \mathbb{R}^{1\times m}$

$$\underbrace{\bar{w}}_{m\times 1} = \bar{w} - \left(\alpha \cdot -\frac{2}{n}\,\text{np.sum}(\theta, \text{axis}=0)\right)$$