EE660

Oct 10, 2018

Discussion 8

Today:

- Project Q & A
- Exercises

Ex. Murphy 13.6

$X^T X = I$

$J(w) = RSS(w) + reg(w)$

$RSS(w) = \|Xw - y\|_2^2 = y^T y - 2w^T X^T y + \underbrace{w^T X^T X w}_{\|w\|_2^2}$

$reg(w) = 0$ for OLS

$reg(w) = \lambda_2 \|w\|_2^2$ for ridge reg.

$reg(w) = \lambda_1 \|w\|_1$ for lasso

$$\frac{dRSS}{dw} = -2X^T y + 2w$$

$$\boxed{\frac{\partial RSS}{\partial w_K} = -\underbrace{2 X_{:,K}^T y}_{c_K} + 2w_K}$$

column K of X

OLS:

$$\hat{w}_K = \frac{c_K}{2}$$

Ridge regression:

$$\frac{\partial \|w\|_2^2}{\partial w_K} = 2w_K$$

$$\frac{\partial J(w)}{\partial w_K} = -c_K + 2w_K(1+\lambda_2)$$

$$\hat{w}_K^r = \frac{c_K}{2(1+\lambda_2)}$$

Lasso:

$$\frac{\partial J(w)}{\partial w_K} = -c_K + 2w_K + \lambda_1 \frac{d|w_K|}{dw_K}$$

Subderivative:

$$a = \lim_{x \to x_0^+} \frac{f(x) - f(x_0)}{x - x_0} \ , \quad b = \lim_{x \to x_0^-} \frac{f(x) - f(x_0)}{x - x_0}$$

The subdifferential of f at x_0 $\partial f(x)\big|_{x_0} \in [a, b]$

$$\frac{d|w_k|}{dw_k} = \begin{cases} 1 \ , & w_k > 0 \\ [-1, 1] \ , & w_k = 0 \\ -1 \ , & w_k < 0 \end{cases}$$

Use this in $\dfrac{\partial J(w)}{\partial w_k}$

$$\frac{\partial J(w)}{\partial w_k} = \begin{cases} 2w_k - c_k + \lambda_1 & w_k > 0 \\ [2w_k - c_k - \lambda_1, \ 2w_k - c_k + \lambda_1] & w_k = 0 \\ 2w_k - c_k - \lambda_1 & w_k < 0 \end{cases}$$

$$\hat{w}_k^{\ell}(c_k) = \begin{cases} \dfrac{c_k - \lambda_1}{2} & c_k > \lambda_1 \\ 0 & -\lambda_1 < c_k < \lambda_1 \\ \dfrac{c_k + \lambda_1}{2} & c_k < -\lambda_1 \end{cases}$$

Soft threshold: $\hat{w}_k^{\ell} = \operatorname{sign}(c_k) \max\left(0, \dfrac{|c_k| - \lambda_1}{2}\right)$

$$w_{reg} = \underset{w}{\operatorname{argmin}} \underbrace{E_{in}(w) + \lambda \|w\|^2}_{E_{aug}(w)} \quad \circledast$$

$$w_{lin} = \underset{w}{\operatorname{argmin}} \; E_{in}(w)$$

Let's assume $\|w_{lin}\|^2 < \|w_{reg}\|^2$

$$E_{in}(w_{lin}) \leq E_{in}(w_{reg})$$

Using the assumption

$$E_{in}(w_{lin}) + \lambda \|w_{lin}\|^2 < E_{in}(w_{reg}) + \lambda \|w_{reg}\|^2$$

$$\Downarrow$$

Contradiction

$$\Uparrow$$

$$E_{aug}(w_{lin}) < E_{aug}(w_{reg})$$

$\circledast$ $E_{in}(w_{reg}) + \lambda \|w_{reg}\|^2 \leq E_{in}(w) + \lambda \|w\| \quad \forall w$