September 16 2024.

Pg 18 $\mathcal{J}_{reg} = \mathcal{J}(w) + \lambda R(w) = \mathcal{J}(w) + \frac{1}{2}\sum_j w_j^2$

L_2 norm of w, or $\|w\|_2^2$

$\frac{\partial \mathcal{J}}{\partial w_j} > 0 \Rightarrow$ increase w_j $\Rightarrow$ increase $\mathcal{J}$ $\Rightarrow$ you should decrease w_j

$\frac{\partial \mathcal{J}}{\partial w_j} < 0 \Rightarrow$ increase w_j $\Leftarrow$ decrease $\mathcal{J}$ $\Rightarrow$ you should increase w_j

$\Rightarrow$ when gradient $\nabla \mathcal{J}(w_j) < 0 \Rightarrow$ increase w_j } going towards the opposite

$> 0 \Rightarrow$ decrease w_j } direction of the gradient

until we find the critical point ($\nabla \mathcal{J}(w_j) = 0$)

opposite direction of the gradient

Update rule: $w_j \longleftarrow w_j - \alpha \frac{\partial \mathcal{J}}{\partial w_j}$ or $w_j \longleftarrow w_j - \alpha \cdot \nabla \mathcal{J}(w_j)$

$$w \leftarrow w - \frac{1}{N}\sum_i^N \alpha \cdot (y^i - t^i) \cdot x^i$$

derived in previous lectures

$$w \leftarrow w - \frac{\alpha}{N}\sum_i^N (y^i - t^i)\, x^i$$

$$w \leftarrow w - \frac{\alpha \cdot \partial}{\partial w}(\mathcal{J} + \lambda R(w)) \cdot \alpha = w - \alpha \cdot \left(\frac{\partial \mathcal{J}}{\partial w} + \lambda \cdot \frac{\partial R}{\partial w}\right)$$

$$= w - \alpha \cdot \left(\frac{\partial \mathcal{J}}{\partial w} + \lambda \cdot \frac{\partial}{\partial w} \cdot \frac{1}{2}\sum_j w_j^2\right)$$

$$= w - \alpha \cdot \left(\frac{\partial \mathcal{J}}{\partial w} + \lambda \cdot \frac{1}{2} \cdot 2 \cdot \underbrace{[w_1, \dots, w_D]}_{\hookrightarrow w}\right) = w - \alpha \cdot \left(\frac{\partial \mathcal{J}}{\partial w} + \lambda w\right)$$

$$= w - \alpha \cdot \frac{\partial \mathcal{J}}{\partial w} - \alpha \cdot \lambda \cdot w = w - \alpha \cdot w \cdot \lambda - \alpha \cdot \frac{\partial \mathcal{J}}{\partial w}$$

$$= (1 - \alpha\lambda)\, w - \alpha \cdot \frac{\partial \mathcal{J}}{\partial w}$$