Regularizations and loss functions

<u>L2 regularization -</u>

We have $l_R(\omega) = \underline{l(\omega)} + \frac{\alpha}{2}\|\omega\|_2^2$

$\|\omega\|_2^2 = \omega_1^2 + \omega_2^2 + \cdots + \omega_n^2$

Let us suppose <u>optimal</u> value is ω^*

$\boxed{\nabla l(\omega) = 0} \rightarrow \omega^*$

$$\underline{l(\omega)} = l(\omega^*) + \underbrace{(\omega-\omega^*)\nabla l(\omega^*)} + \underbrace{\frac{(\omega-\omega^*)^T H(\omega-\omega^*)}{2}}$$

$\underbrace{\nabla l(\omega^*) = 0}$ ✓

$$\Rightarrow \underline{l(\omega)} = l(\omega^*) + \boxed{\frac{(\omega-\omega^*)^T H(\omega-\omega^*)}{2}}$$

$$\nabla_\omega (\omega-\omega^*)^T H(\omega-\omega^*) = H(\omega-\omega^*)$$

Now, $\nabla l(\omega) = \underbrace{\nabla l(\omega^*)} + \underbrace{H(\omega-\omega^*)}$

$$\Rightarrow \boxed{\nabla l(\omega) = H(\omega-\omega^*)}$$ ✓

✓ $l_R(\omega) = \underbrace{l(\omega)} + \frac{\alpha}{2}\|\omega\|_2^2$

Now, $\nabla l_R(\omega) = \nabla l(\omega) + \underline{\alpha\omega}$

$= \underline{H\omega - H\omega^*} + \alpha\omega$

$= \underline{(H+\alpha I)}\omega - \underline{H\omega^*} \rightarrow$

$(H+\alpha I)\omega = H\omega^*$

$\underline{\omega} = (H+\alpha I)^{-1} H\underline{\omega^*}$

$\omega = H^{-1} H\omega^*$

$= \omega^*$

for extremum points; $\nabla l_R(\omega) = 0$

$$\Rightarrow \boxed{\omega = \underbrace{(H+\alpha I)^{-1}} H\underline{\omega^*}}$$ ✓

<u>L1 regularization -</u>

We have $\underline{l_R(\omega)} = l(\omega) + \underline{\alpha\|\omega\|_1}$

$\underbrace{\alpha\left[\,|\omega_1| + |\omega_2| + \cdots + |\omega_n|\,\right]}$

$\nabla l_R(\omega) = \underbrace{\nabla l(\omega)} + \alpha.\underbrace{sign(\omega)}$

1 if $\omega_1 > 0$

-1 if $\omega_1 < 0$

Now for $\nabla l_R(\omega) = 0$.

$$\boxed{H(\omega-\omega^*) + \alpha\, sign(\omega) = 0}$$

$$\boxed{H(w - w^*) + \alpha\, \text{sign}(w) \quad = \quad}$$

$$\Rightarrow \quad w = H^{-1}\left[-\alpha\, \text{sign}(w) + H w^*\right]$$

$$\Rightarrow \quad w = H^{-1} H w^* - \alpha H^{-1} \text{sign}(w)$$

h_{ii} = diag of H

$$\Rightarrow \quad w = w^* - \alpha H^{-1} \text{sign}(w)$$

$$\Rightarrow \quad w = \begin{cases} \max\left(w^* - \dfrac{\alpha}{h_{ii}}, 0\right) & \text{if } w \geq 0 \\ \min\left(w^* + \dfrac{\alpha}{h_{ii}}, 0\right) & \text{if } w < 0 \end{cases}$$

We consider $l(w) = w_1^2 - 6w_1 + w_2^2 - 4w_2 + 13$

$\nabla l(w)$

We have, $\dfrac{\partial l}{\partial w_1} = \underbrace{2w_1 - 6}$ & $\dfrac{\partial l}{\partial w_2} = 2w_2 - 4$

$\dfrac{\partial l}{\partial w_1} = 0$

The optimal values are $w_1^* = 3$ & $w_2^* = 2$.

If l2 regularization is to be obtained, we calculate the hessian

$$H = \begin{pmatrix} \dfrac{\partial^2 l}{\partial w_1^2} & \dfrac{\partial^2 l}{\partial w_2 \partial w_1} \\ \dfrac{\partial^2 l}{\partial w_1 \partial w_2} & \dfrac{\partial^2 l}{\partial w_2^2} \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

$(H + \alpha I)^{-1} H w^*$

$H + \alpha I$

We get l2 regularized value of w as

$$\ldots^{-1} \begin{pmatrix} 2 & 0 \end{pmatrix} \begin{pmatrix} 3 \end{pmatrix}$$

We get ℓ_2 regularization ...

$$w = \begin{pmatrix} 2+\alpha & 0 \\ 0 & 2+\alpha \end{pmatrix}^{-1} \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 3 \\ 2 \end{pmatrix}$$

$H + \alpha I$

$$w^* = \begin{pmatrix} w_1^* \\ w_2^* \end{pmatrix}$$

We consider $\alpha = 0.3$ to get

$$\underline{\underline{w}} = \begin{pmatrix} 2.61 \\ 1.74 \end{pmatrix}$$

$$\begin{pmatrix} 3 \\ 2 \end{pmatrix} \quad \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}$$

We have $w_1 > 0$ & $w_2 > 0$. If ℓ_1 regularization was to be applied,

$$w_1 = \max\left\{3 - \frac{0.3}{2}, 0\right\} = \underline{\underline{2.85}} \qquad \left\{w_1^* - \frac{\alpha}{h_{11}}, 0\right\}$$

$$w_2 = \max\left\{2 - \frac{0.3}{2}, 0\right\} = \underline{\underline{1.85}} \qquad \left\{w_2^* - \frac{\alpha}{h_{22}}, 0\right\}$$