L2-Regularization

$$M = \frac{1}{2}\sum_{i}^{N}\left(t_i - W^T\phi(x_i)\right)^2 + \frac{\lambda}{2}W^TW$$

-> loop over the elements of S

train -> over train subset

[validate -> complete MSE of val subset

└-> select best λ as

$$L(w) = \frac{1}{2}\sum_{i}^{N}\left(t_i - W^T\phi(x_i)\right)^2 + \frac{\lambda}{2}W^TW$$

{ L(w) is convex function }

for a function to be convex it should satisfy:

$$\alpha L(w_1) + (1-\alpha)L(w_2) \geq L(\alpha w_1 + (1-\alpha)w_2)$$

$$\forall \alpha \in 0 \leq \alpha \leq 1$$

$$\cancel{W_{optima} = (\lambda I + \phi^T\phi)^{-1}\phi\phi^T t}$$

$$\boxed{W_{optimal} = (\lambda I + \phi^T\phi)^{-1}\phi^T t}$$

$E_d(w_i$

$$E_d(w_i) = \frac{1}{2}(t_i - w^T\phi(x_i))^2$$

$$\nabla E_d(w_i) = -\underbrace{(t_i - w^T\phi(x_i))}\underbrace{\phi(x_i)}_{\text{vector}}$$

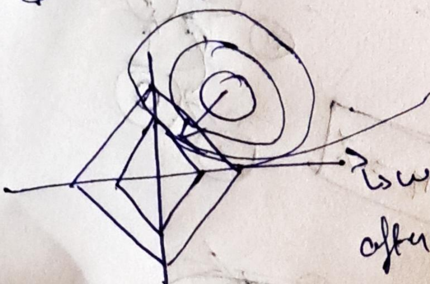
$$E_w(w) = \frac{1}{2}w^Tw; \quad \nabla_w E_w(w) = w$$

$$w^{(\tau+1)} = w^{\tau} + \eta(t_i - w^{(\tau)T}\phi_n)\phi_n - \eta_\lambda w^{(\tau)}$$

$$L(w) = \frac{1}{2}\sum_{i=1}^{N}[t_i - w^T\phi(x_i)]^2 + \lambda\underbrace{\sum_{i=1}^{D}(w_i)}_{L_1 \text{ norm of } w}$$

parse solution when one of $w=0$

→ L_1 Regula. is used when we want to make some weight 0.



W_{optima} (location)

→ w_1 for different λ.

after $\lambda > \lambda_c$ → critical λ

one of w will be 0

L_2 → Ridge Regression

L_1 → LASO.