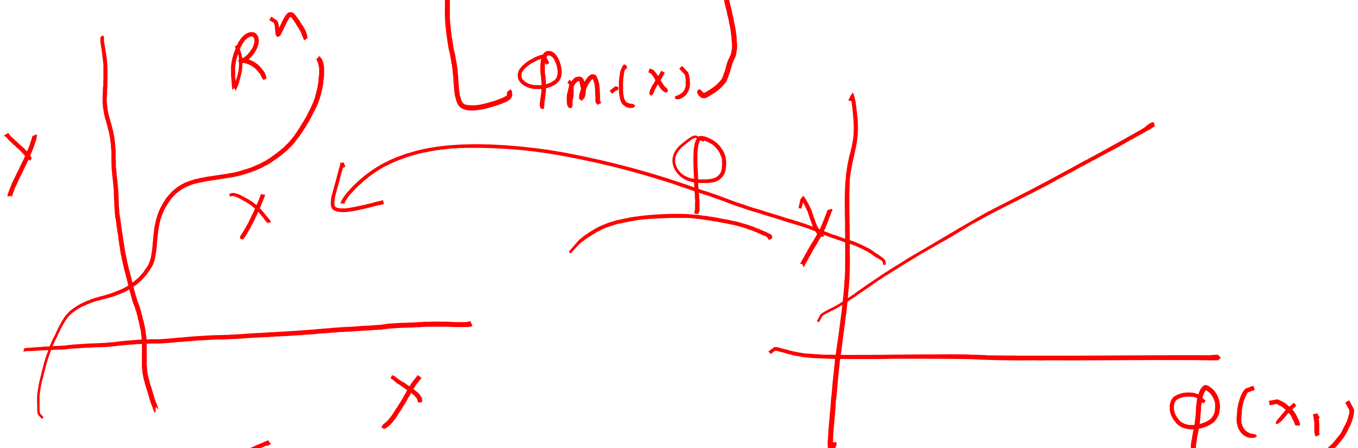
$$\min_{w} \frac{\lambda}{2} w^T w + \frac{1}{2} \sum_{i=1}^{N} \left(y_i - \left(w^T \varphi(x_i) \right) \right)^2$$

$$m >> N$$

$$\varphi(x_i) = \begin{bmatrix} \varphi_0(x) \\ \varphi_1(x) \\ \vdots \\ \varphi_m(x) \end{bmatrix}$$

$$(w^T w) \le k$$



$w^T x + b$

$w^T \varphi(x) + b$

$$f(x) = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ \\ w_m \end{bmatrix}^T \begin{bmatrix} \phi_0(x) \\ \phi_1(x) \\ \phi_2(x) \\ \\ \phi_m(x) \end{bmatrix}$$

$$\begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}^T \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$w \in R^n$

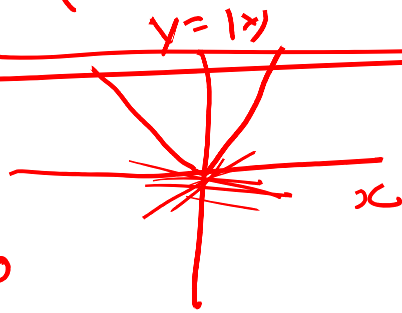
$$\|w\|_1 = |w_1| + |w_2| + \dots |w_m|$$

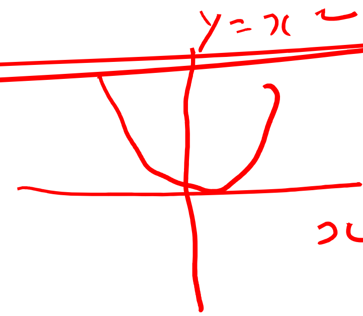
$$\min_{w} \frac{\lambda}{2}\|w\|_1 + \frac{1}{2}\sum_{i=1}^{N}\left(y_i - w^T\phi(x_i)\right)^2$$

$1 \quad x > 0$

$-1 \quad x < 0$

$[-1, 1] \quad x = 0$





Defination Sub-gradients

A sub-gradient of a convex func f at $x \in R^m$ is an func $g \in R^m$ such that

$$f(y) \geq f(x) + g^T(y-x) \quad \forall\, y \in R^m$$

$$\delta((\|w\|_1)) = g$$

$$g[i] = \begin{cases} \text{sign } w[i] & \text{if } w[i] \neq 0 \\ \text{any element in } [-1, 1] & \text{if } w[i] = 0 \end{cases}$$

$$J(w) = \boxed{\frac{\lambda}{2}||w||_1 + \frac{1}{2}\sum_{i=1}^{N}\left(y_i - w^T\varphi(x_i)\right)^2}$$

$$\varphi(x_i) = \begin{bmatrix} \varphi_0(x) \\ \varphi_1(x) \\ \vdots \\ \varphi_m(x) \end{bmatrix}$$

$$\delta(J(w)) = \lambda g(i) - \frac{1}{2}\sum_{i=1}^{N}\left(y_i - w^T\varphi(x_i)\right)\varphi(x_i)$$

Algorithm I

Input :- Training set T and tolerence ϵ

Initialize $w^0 \in R^{m+1}$

Repeat

$$w^{(k+1)} = w^{(k)} - \eta\,\delta(J(w))$$

Until $||\delta(J(w^{(k)}))|| < \epsilon$

Stochastic Subgradient descent method

Input :- Training set T, max_iter

Initialize $w^0 \in R^{m+1}$

for $i = 1$ to max_iter

Draw a random subset B of size k from T. (training set)

$$w^{(k+1)} = w^{(k)} - \eta_i \, \delta\left(J_B\left(w^{(k)}\right)\right)$$

end for

where $\delta(J_B(w)) = \lambda g - \sum_{(x_i, y_i) \in B} (y_i - w^T \varphi(x_i)) \, \varphi(x_i)$