feb 14 2020

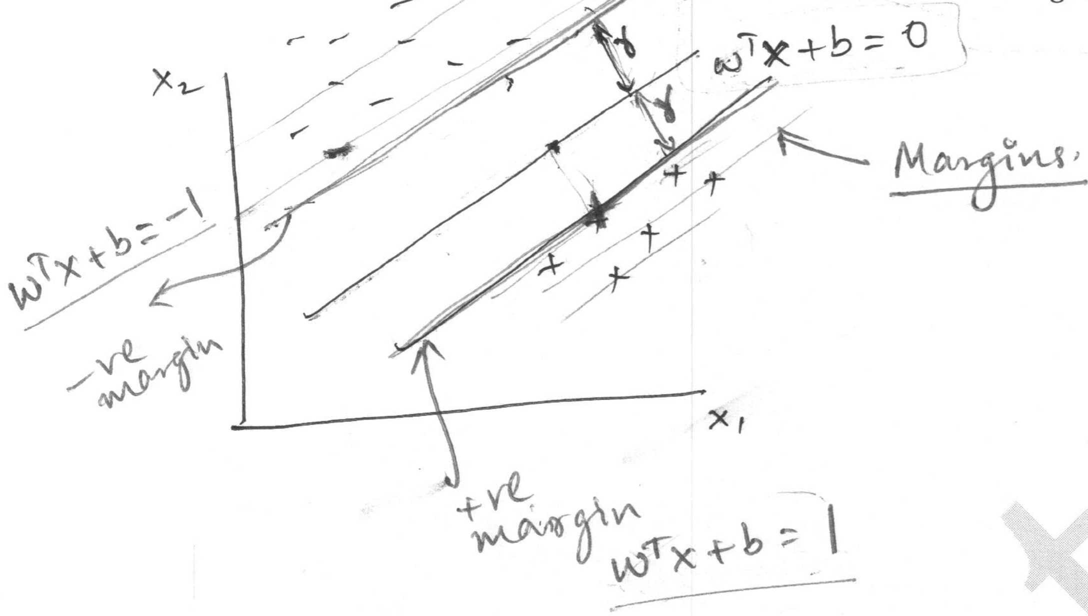
W → vector

w_1	w_2	------	w_D

$$|W| = |w_1| + |w_2| + \cdots + |w_D| \quad \text{l1 norm}$$

$$\|W\| = \sqrt{w_1^2 + w_2^2 + \cdots + w_D^2} \rightarrow \text{l2 norm}$$

$$\|W\|_p = \left(w_1^p + \cdots + w_D^p\right)^{1/p} \rightarrow \text{lp norm}$$

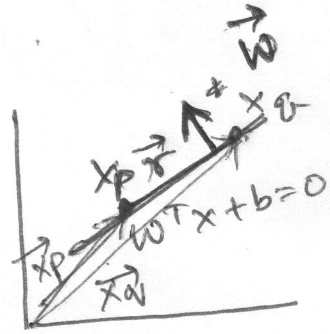


$$W \rightarrow \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \qquad \|W\| = \left(\sum_{i=1} w_i^2\right)^{1/2} \qquad \|W\| = \sqrt{w_1^2 + w_2^2}$$

$$W^T W = \|W\|^2$$

$$\gamma = \frac{1}{\|W\|}$$

$$\text{Size of margin} = 2\gamma = \frac{2}{\|W\|}$$



$$\vec{x_q} = \vec{x_p} + \vec{\gamma}$$

$$\vec{\gamma} = (\vec{x_q} - \vec{x_p})$$

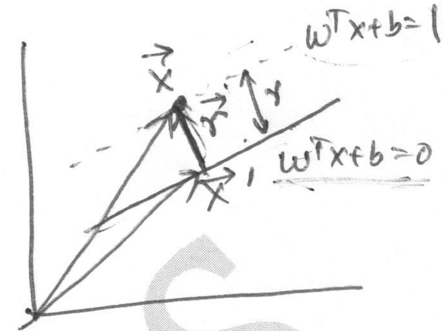
$$\vec{w}^T \vec{\gamma} = \vec{w}^T(\vec{x_q} - \vec{x_p})$$

$$= \vec{w}^T \vec{x_q} - \vec{w}^T \vec{x_p}$$

$$= (\vec{w}^T \vec{x_q} + b) - (\vec{w}^T \vec{x_p} + b)$$

(Implies that $\vec{w}$ is $\perp$ to the line) $= 0$

Size of the margin is $\frac{2}{\|w\|}$



$$\vec{\gamma} = \gamma \frac{\vec{w}}{\|w\|}$$

$$\vec{x} = \vec{x}' + \vec{\gamma}$$

$$= \vec{x}' + \frac{\gamma \vec{w}}{\|w\|}$$

$$\vec{w}^T \vec{x} + b = 1$$

$$\vec{w}^T \vec{x}' + \frac{\vec{w}^T (\gamma \vec{w})}{\|w\|} + b = 1$$

$$\vec{w}^T \vec{x}' + b = 0$$ because $\vec{x}'$ lies on the boundary.

$$\gamma \frac{\vec{w}^T \vec{w}}{\|w\|} = 1 \qquad \vec{w}^T \vec{w} = \|w\|^2$$

$$\gamma \|w\| = 1 \qquad \gamma = \frac{1}{\|w\|} \leftarrow \text{Proof.}$$

Size of Margin $= \frac{2}{\|w\|}$

SVM formulation:

$$\min_{w,b} \frac{\|w\|^2}{2}$$

$$\text{s.t. } y_i (w^T x_i + b) \geq 1 \quad \forall i = 1, \ldots, N$$

$f(w_1, w_2) = w_1^2 + 3w_1^3 + 2w_1 + 7$

$w_1 + w_2 \leq 1 \longrightarrow w_1 + w_2 - 1 \leq 0$

$L(w,\alpha) = (w_1^2 + 3w_1^3 + 2w_1 + 7) + \alpha(w_1 + w_2 - 1)$

s.t. $\alpha \geq 0$

Let $\delta = \log \alpha$ $\qquad \alpha = \exp(\delta)$

"log-barrier"

$$\underset{w,b,\alpha}{L_p} = \frac{\|W\|^2}{2} + \sum_{i=1}^{N} \alpha_i [1 - y_i(W^T x_i + b)]$$

We will first minimize this w.r.t w & b.

$$\frac{\partial L_p}{\partial W} = W + \sum_{i=1}^{N}(-\alpha_i y_i x_i) \qquad \frac{\partial}{\partial W}\|W\|^2 = 2W$$

$$\frac{\partial}{\partial W} W^T W = 2W$$

Setting to 0

$$\boxed{\frac{\partial L_p}{\partial W} = 0 \Rightarrow W = \sum_{i=1}^{N} \alpha_i y_i x_i}$$

$$\frac{\partial L_p}{\partial b} = -\sum_{i=1}^{N} \alpha_i y_i = 0$$

$$\sum_{i=1}^{N} \alpha_i y_i = 0$$

$$w = \sum_{i=1}^{N} \alpha_i y_i x_i$$

$$= \alpha_1 y_1 x_1 + \alpha_2 y_2 x_2 + - - - - - + \alpha_N y_N x_N$$

$$\boxed{w = \sum_{i \text{ is a s.v}} \alpha_i y_i x_i}$$

For a new test point, x^*

$w^T x^* + b$

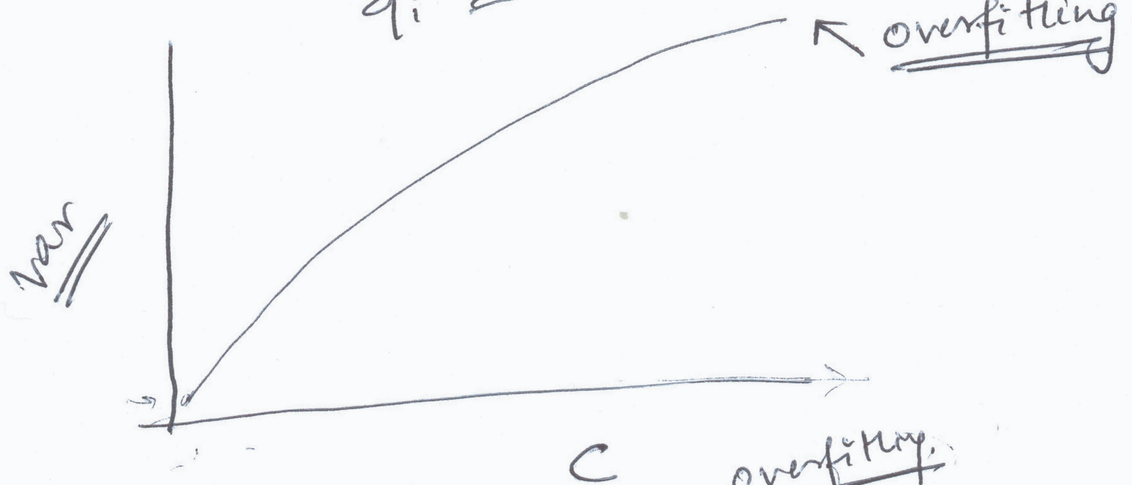
$$\boxed{\left(\sum_{x_i \text{ is a s.v}} \alpha_i y_i x_i^T x^*\right) + b}$$

Non-separable Case SVM

$$\min \quad \frac{\|W\|^2}{2} + C \sum_{i=1}^{N} \xi_i$$

$$\text{s.t} \quad \left[y_i (W^T x_i + b) \geq 1 - \xi_i \right] \quad \forall i$$

$$\xi_i \geq 0$$



$\neq$ $\boxed{\text{Good Performance on Training Data} \not\Rightarrow \text{Good performance on Test Data}}$

generalization

[Probably Approximately Correct (PAC)]