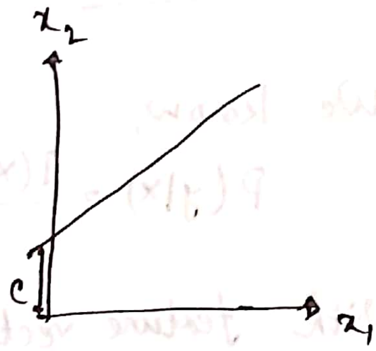
Support Vector Machines



The standard eqn of a straight line!

$$y = \overset{\to\ \text{slope}}{m}x + \underset{\to\ \text{intercept in } y \text{ axis}}{c}$$

$\downarrow$

$$x_2 = mx_1 + c$$

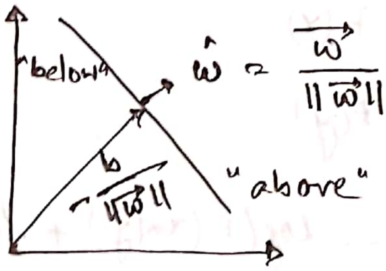
Or, $w_1x_1 + w_2x_2 + b = 0$

$$\Rightarrow x_2 = \underbrace{-\frac{w_1}{w_2}}_{\text{slope}} x_1 \underbrace{- \frac{b}{w_2}}_{\text{intercept}}$$

Or we can say,

$$\boxed{w^T x + b = 0}$$

If $\left\{ \begin{array}{l} w \text{ is a vector which contain } w_1 \text{ and } w_2 \\ x \text{ is another vector which contains } x_1 \text{ and } x_2 \end{array} \right.$



Now, for a point x^*

if $w^T x^* + b = 0$

that means the point is on the line cause it is satisfying our straight line eqn.

If $w^T x^* + b > 0$ then x^* is "above" the line

If $w^T x^* + b < 0$ then x^* is "below" the line.

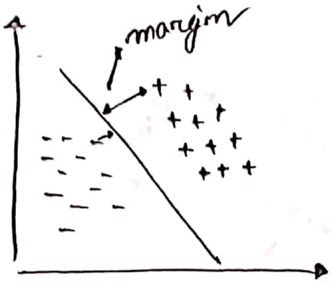
For classification! $\text{sign}(w^T x^* + b)$ is what we need.

Motivation!

we want $|w^T x^* + b|$ to be large

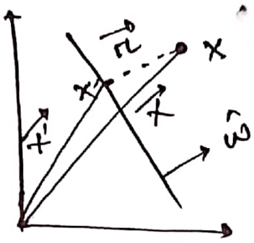
Or,

we want $y^*(w^T x^* + b)$ to be large



margin,

$$\gamma = \frac{w^T x + b}{\|w\|}$$



Here,

$$\vec{r} = \vec{x} - \vec{x'} \quad \text{------} \quad (1)$$

For x' : $w^T x' + b = 0$ ------ (2)

$\therefore \vec{r}$ is parallel to $\hat{w}$

So, let $|\vec{r}|$ be γ

$$\Rightarrow \vec{r} = \gamma \hat{w} = \frac{\gamma \vec{w}}{\|\vec{w}\|}$$

Now from eqn (1)

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

And for eqn (2)

$$\vec{w}^T \left(\vec{x} - \frac{\gamma \vec{w}}{\|\vec{w}\|} \right) + b = 0$$

$$\Rightarrow \vec{w}^T \vec{x} - \frac{\gamma \vec{w}^T \vec{w}}{\|\vec{w}\|} + b = 0$$

$$\vec{w}^T \vec{w} = \|w\|^2$$

$$\|w\|^2 = \sqrt{w_1^2 + w_2^2} \Leftarrow \text{For 2 dimension}$$

$$\Rightarrow \vec{w}^T \vec{x} - \gamma \|w\| + b = 0$$

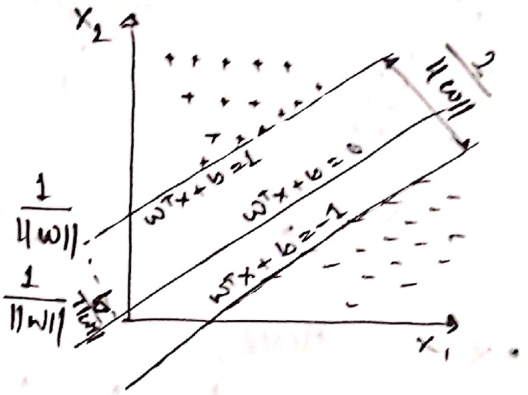
$$\therefore \gamma = \frac{\vec{w}^T \vec{x} + b}{\|w\|}$$

If $\vec{x}$ is below the line :

$$\gamma = -\left(\frac{w^T \vec{x} + b}{\|w\|} \right)$$

in general, $\gamma = y\left(\frac{w^T \vec{x} + b}{\|w\|} \right)$

margin will be, $\frac{2}{\|w\|}$



For +ve points, $\gamma = \frac{w^Tx+b}{\|w\|}$

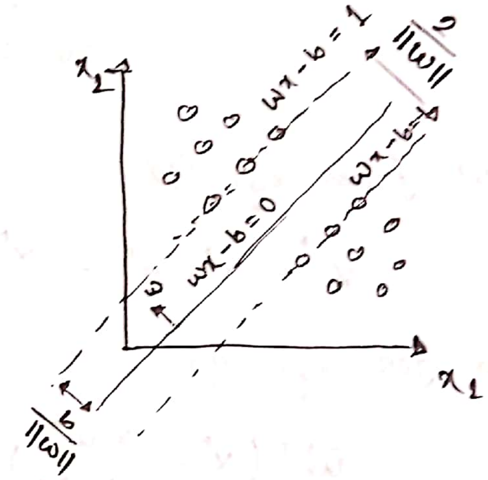
For +ve points on margin $\gamma = \frac{1}{\|w\|}$

For -ve points on margin $\gamma = \frac{1}{\|w\|}$

$\therefore$ Total margin $= \frac{2}{\|w\|}$

Technical

Linear Model:



$$wx - b = 0$$

$$w_i x_i - b \geq 1 \text{ if } y_i = 1$$

$$w \cdot x_i - b \leq -1 \text{ if } y_i = -1$$

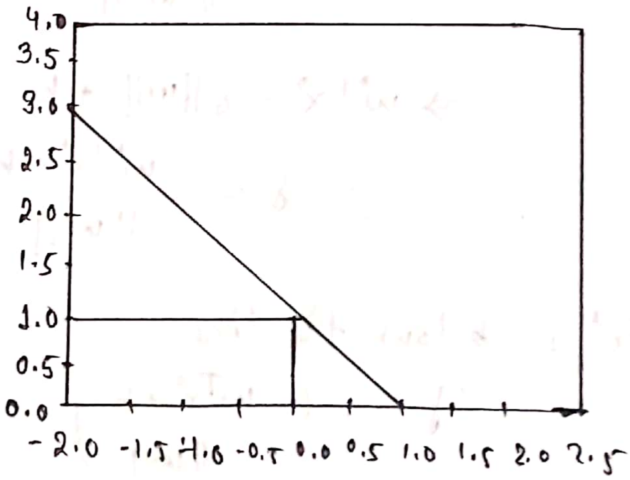
$$\Rightarrow y_i (w \cdot x_i - b) \geq 1$$

Cost Function:

Hinge Loss:

$$l = \max(0, 1 - y_i(w \cdot x_i - b))$$

$$l = \begin{cases} 0 & \text{if } y \cdot f(x) \geq 1 \\ 1 - y \cdot f(x) & \text{otherwise} \end{cases}$$



Add Regularization:

$$J = \lambda \|w\|^2 + \frac{1}{n} \sum_{i=1}^{n} \max(0, 1 - y_i (w \cdot x_i - b))$$

if $y_i \cdot f(x) \geq 1$:

$$J_i = \lambda \|w\|^2$$

else:

$$J_i = \lambda \|w\|^2 + 1 - y_i (w \cdot x_i - b)$$

Gradients:

if $y_i \cdot f(x) \geq 1$:

$$\frac{dJ_i}{dw_k} = 2\lambda w_k$$

$$\frac{dJ_i}{db} = 0$$

else:

$$\frac{dJ_i}{dw_k} = 2\lambda w_k - y_i \cdot x_i$$

$$\frac{dJ_i}{db} = y_i$$

Update Rule:

For each training sample x_i:

$$w = w - \alpha \cdot dw$$

$$b = b - \alpha \cdot db$$