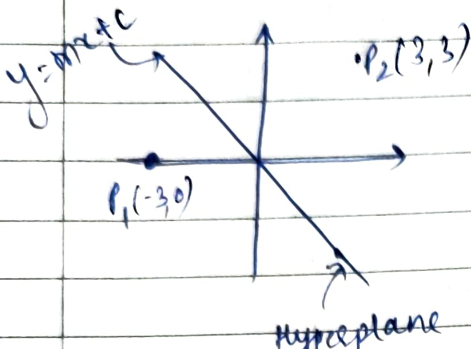
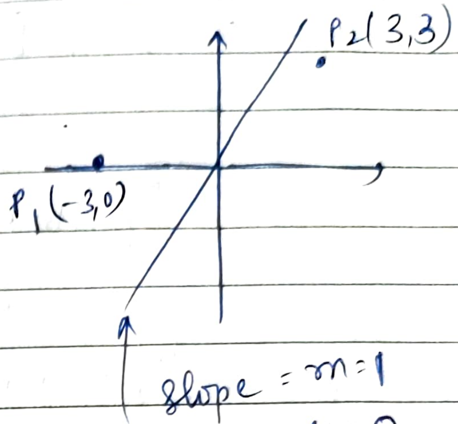
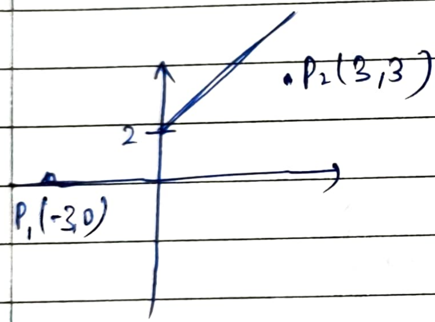
SVM cont. →



let ↳ slope m=-1
Intercept c=0



Slope = m=1
Intercept = 0



Intercept=2

lets consider 1st case $\quad m=-1$, $c=0$

let me store these two parameters in w.

w → parameters of line

$$\underbrace{(m, c)}_{\text{weight \& bias}} = (-1, 0)$$

$P_1 = (-3, 0)$

$w^T x = \begin{bmatrix} -1 \\ 0 \end{bmatrix} [-3, 0]$

Taking transpose of w, so shape can be matched for matrix multiplication.

$= 3$
positive

So we can, all the points lying below the hyperplane, $w^T x$ will have positive value.

lets take $P_2 = (3, 3)$

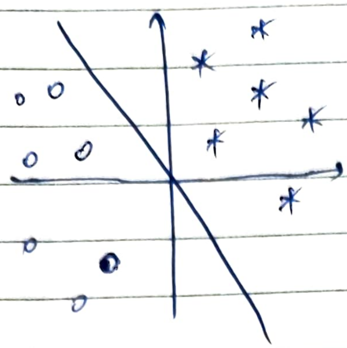
$w^T x = \begin{bmatrix} -1 \\ 0 \end{bmatrix} [3, 3]$

$= -3 \Rightarrow$ negative

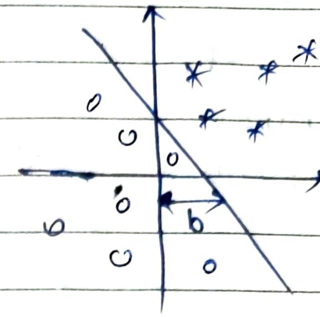
So, we can infere from here is all the points which lie above the hyperplane (right side of hyperplane), $w^T x$ value will be Negative

→ So here, $w^T x$ works as label for the datapoints, i.e. whethere the data point belongs to positive class or negative class.

In Some of the cases, hyperplane may not pass through the origin. ⇒



Label = $w^T x$

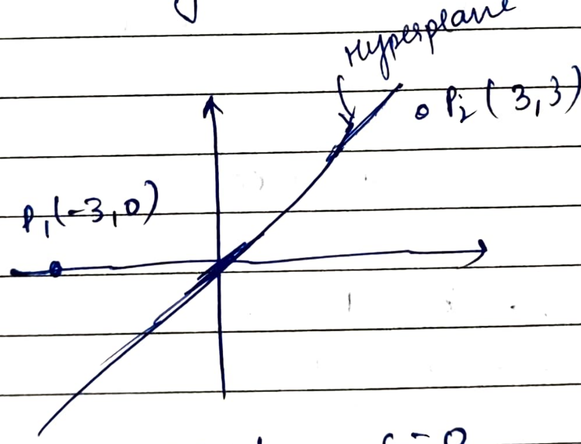


Label = $w^T x + b$

It is Important to note that →

~~This is not generalized.~~

consider →



slope $m = 1$, $c = 0$

$w = (1, 0)$

$P_1(-3, 0)$ ⇓

$w^T x$

$= \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} -3, & 0 \end{bmatrix}$

$= -3$ (neg)

$P_2(3, 3)$ ⇓

$w^T x$

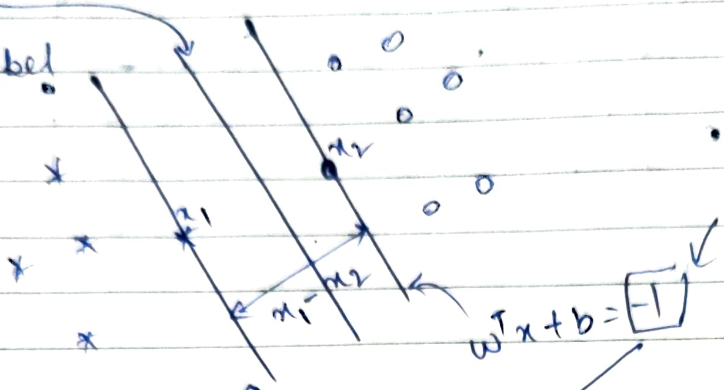
$\begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 3, & 3 \end{bmatrix}$

$= 3$ (pos)

so Here value left to the hyperplane will be in class pos.
& right ———— neg.

optimization for max. margin →

Hyperplane $w^T x + b = \text{label}$



$w^T x + b = \boxed{-1}$ ← negative

$w^T x + b = \boxed{1}$ ← positive

This can be any value we are not interested in magnitude. We want to see, its positive or negative.

$x_1 - x_2$ = Margin distance

$$w^T x_1 + b = 1$$
$$w^T x_2 + b = -1$$
$$\underline{- \quad\quad - \quad\quad -}$$
$$w^T (x_1 - x_2) = 2$$

w^T ← It's a vector

divide by $\|w\|$ {both the side}
↑ norm of w.

vector cancelled with norm of vector

$$\Rightarrow \frac{\cancel{w^T}}{\|w\|}(x_1 - x_2) = \frac{2}{\|w\|}$$

$$\Rightarrow x_1 - x_2 = \frac{2}{\|w\|} \quad \text{(Margin)}$$

So, $x_1 - x_2 = \frac{2}{||w||}$

→ we want to maximize our margin so as to optimize our support vector machine model.

So, Label can have positive value or negative value so we can simplify this and write this as:

Let y_i be the label of each datapoint.

$$y_i = \begin{cases} -1, & w^T x_1 + b \leq -1 \text{ (any neg.)} \\ 1, & w^T x_1 + b \geq 1 \text{ (any positive)} \end{cases}$$ LABEL

and, we know

$$x_1 - x_2 = \frac{2}{||w||}$$ MARGIN

we want to maximize margin, so we can write it as.

$$\boxed{\begin{array}{l} \max\left(\frac{2}{||w||}\right) \text{ such that,} \\ y_i = \begin{cases} -1, & w^T x_1 + b \leq -1 \\ 1, & w^T x_1 + b \geq 1 \end{cases} \end{array}}$$

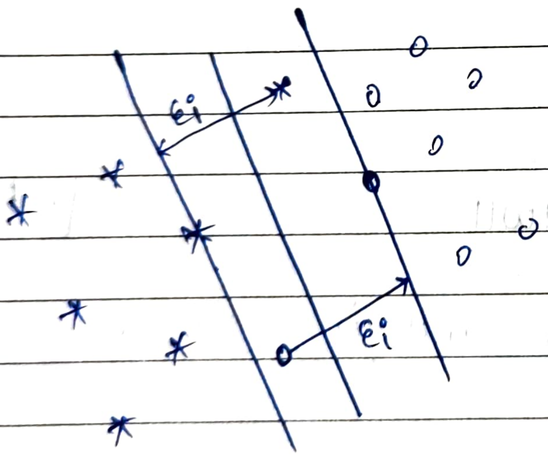
I can rewrite this as →

$$\min\left(\frac{2}{\|w\|}\right) + C * \sum \varepsilon_i$$

$C \rightarrow$ no. of errors
$\varepsilon \rightarrow$ Error magnitude

→ As we want to increase our margin, we can not keep increasing margin value, otherwise it will overfit, so we have to stop somewhere.

So here we are allowing the model to make some errors.



C = no. of errors
ε = magnitude of error

so $\underbrace{C * \sum \varepsilon_i}$
tolerance for error.