Support Vector classifier indepth Maths

① Eqn of line, plane, hyperplane.

$y = mx + c$

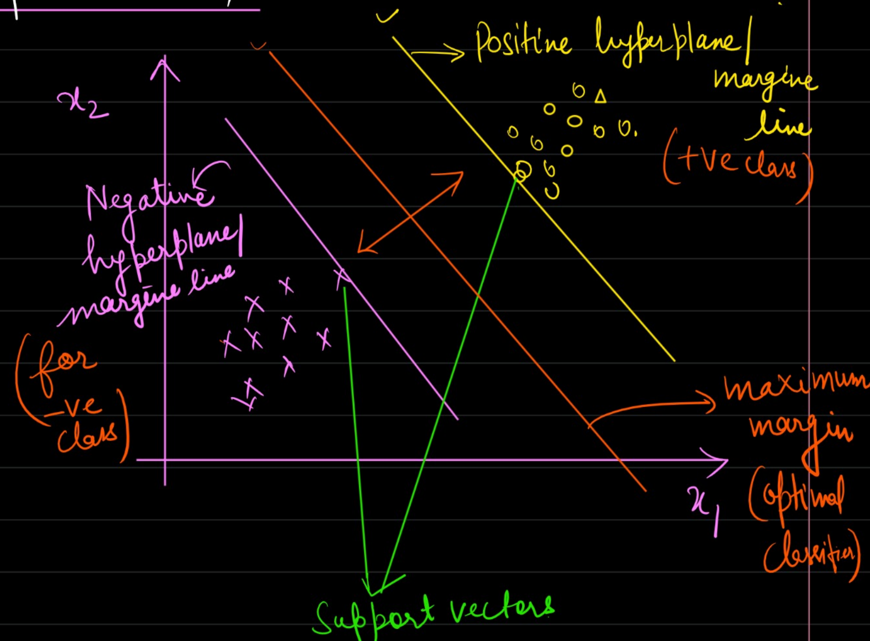
$y = \theta_0 + \theta_1 x_1$

$ax + by + c = 0$

$by = -ax - c$

$y = \boxed{\frac{-a}{b}} x - \boxed{\frac{c}{b}}$

$y = \overset{m}{m}x + \underline{c}$



in morethan 3d

$y = \theta_0 + \theta_1 x_1$

$y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \dots \theta_n x_n \longrightarrow$ hyperplane

$y = b + w_1 x_1 + w_2 x_2 + w_3 x_3 + \dots w_n x_n$

b - bias
w - weights

$y = b + w_1 x_1 + w_2 x_2 + w_3 x_3 + w_4 x_4$

$w = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{bmatrix} \qquad x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$

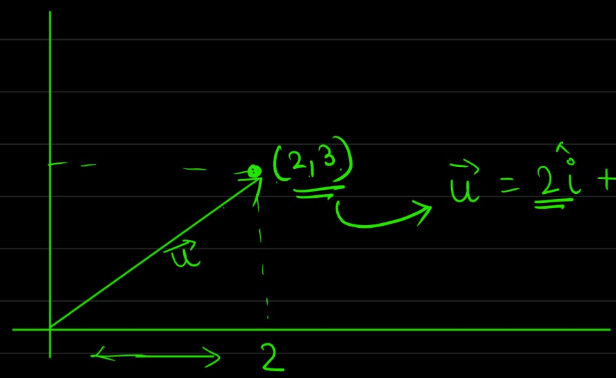
$\underbrace{w^T}_{w^T} = [w_1 \ w_2 \ w_3 \ w_4] \cdot \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}}_{x} \Rightarrow w_1 x_1 + w_2 x_2 + w_3 x_3 + w_4 x_4 + b$

$$\boxed{y = w^T x + b} \quad \left(\begin{array}{l} y = mx + c \\ ax + by + c = 0 \end{array}\right)$$

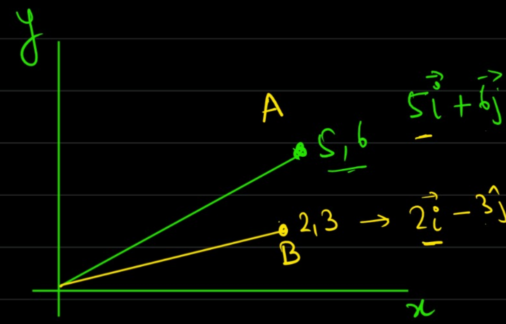
$$\Downarrow$$

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

(2)



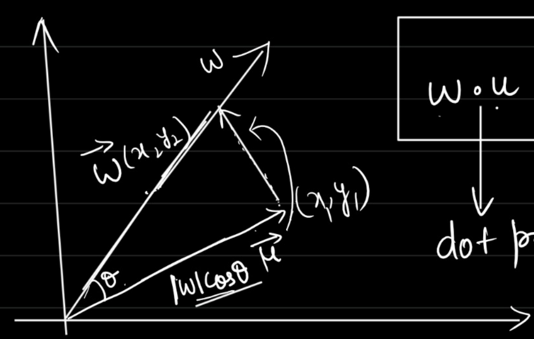
$\vec{u} = 2\hat{i} + 3\hat{j}$



(3) Vector Subtraction $\vec{A} - \vec{B}$

$$(5-2)\vec{i} + (6-3)\vec{j} = 3\hat{i} + 3\vec{j}$$

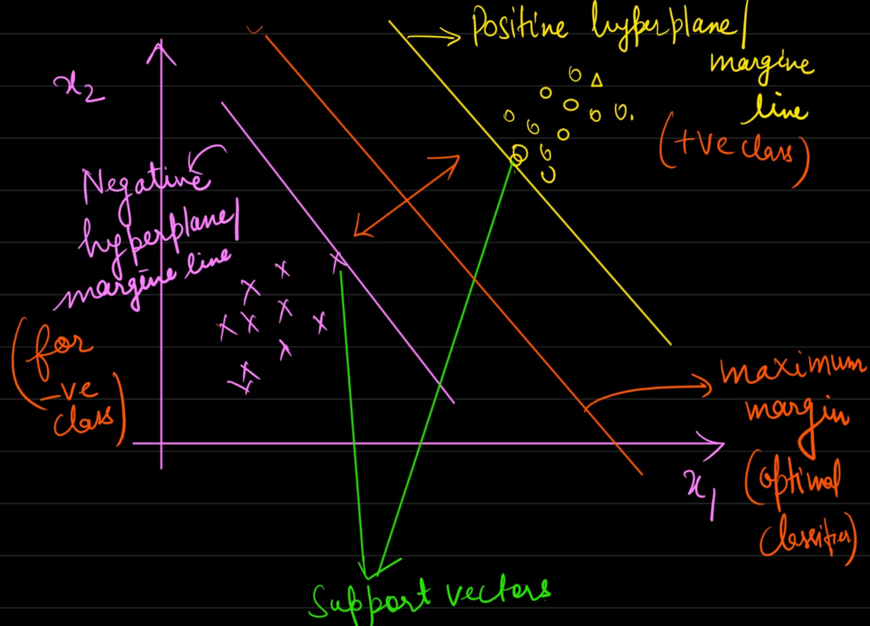
(4) Dot product of vectors.



magnitude of w, magnitude of μ

$$\boxed{w \cdot u = |w| \cos\theta \cdot |\mu|}$$

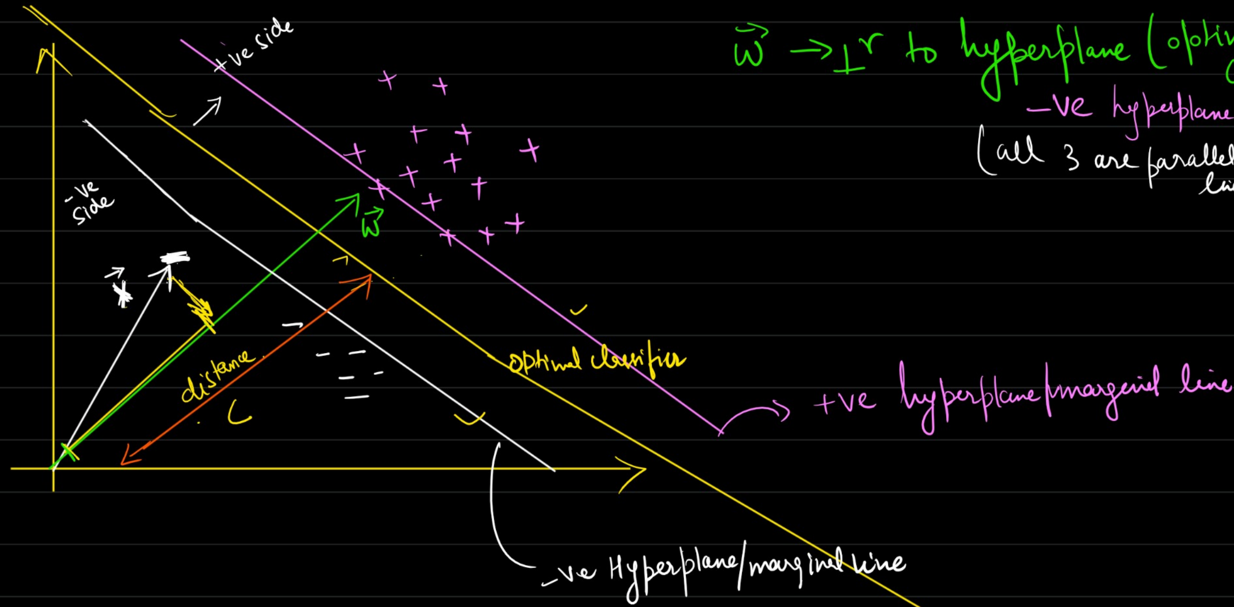
dot product means projection of $\vec{u}$ on $\vec{w}$



$\vec{w} \rightarrow \perp^r$ to hyperplane (optimal classifier)

-ve hyperplane / +ve hyperplane

(all 3 are parallel lines)

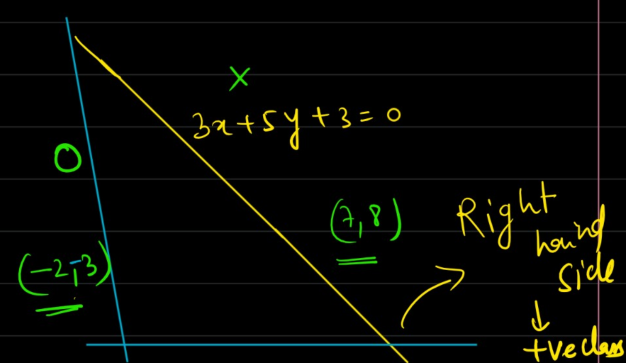


Project is dot product.

$\vec{X} \cdot \vec{w} = (c)$ distance (the point lies on decision boundary)

$\vec{X} \cdot \vec{w} > (c)$ distance (positive class)

$\vec{X} \cdot \vec{w} < (c)$ distance (negative class)

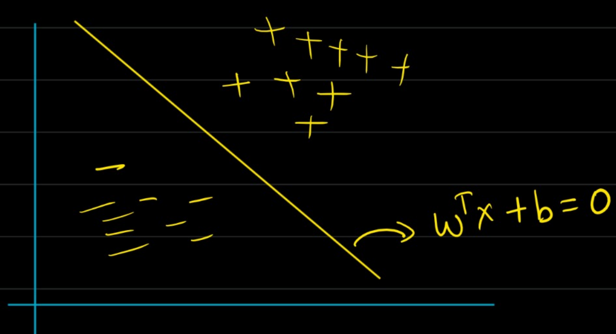


$(7,8) \Rightarrow 3\times7 + 5\times8 + 3 > 0$

$(-2,-3) \Rightarrow 3\times-2 + 5\times-3 + 3 < 0$

-ve side (left side)

-ve class.



$\vec{X} \cdot \vec{w} \geq c$

$\vec{X} \cdot \vec{w} - c \geq 0$

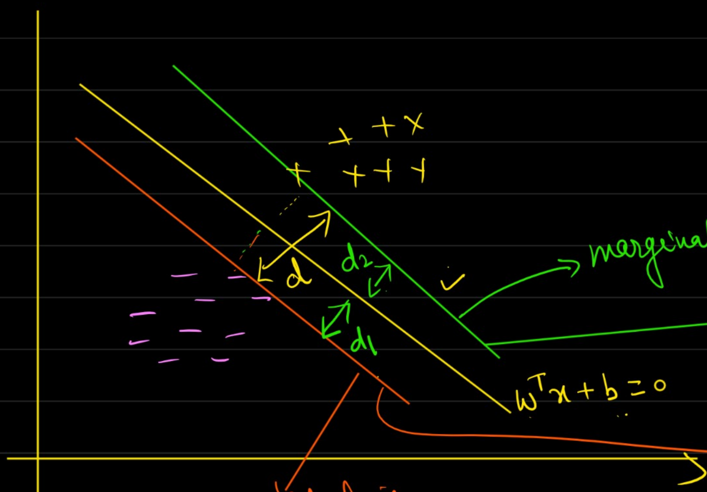
$-c$ as b

$\vec{X} \cdot \vec{w} + b \geq 0 \longrightarrow$ +ve point

$$\vec{x}\cdot\vec{w} + b \geq 0$$

$$y = \begin{cases} +1 & \text{if } \vec{x}\cdot\vec{w} + b \geq 0 \\ -1 & \text{if } \vec{x}\cdot\vec{w} + b < 0 \end{cases}$$

$\rightarrow$ d is maximum



$\rightarrow$ marginal line for +ve class

$\rightarrow w^T x + b = 1$

$\rightarrow w^T x + b = -1$

marginal line for -ve class

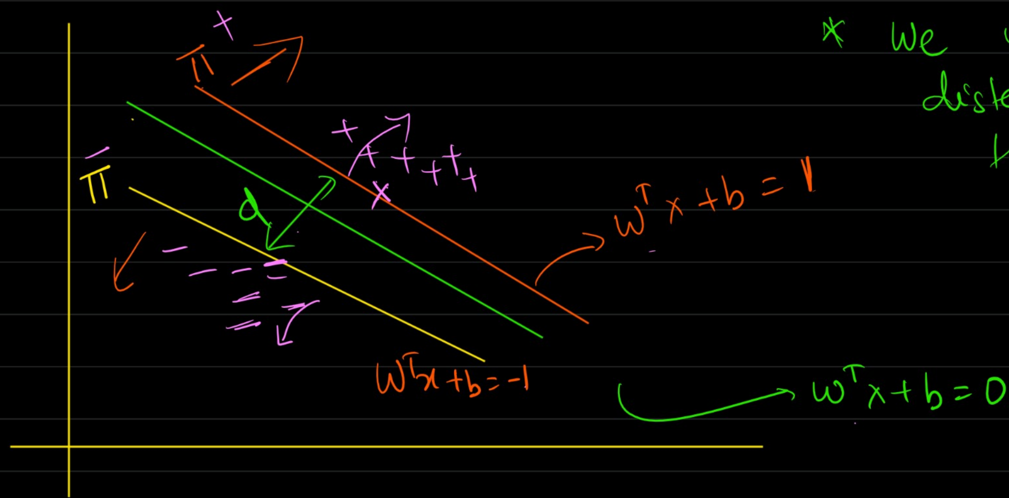
$$\begin{cases} \text{for positive class } +1 \\ \text{for negative class } -1 \end{cases}$$

* Why equal (both 1)? $\rightarrow$ d_1 & d_2 should be equidistant (optimal line should pass through center of margin.

* Why only 1? $\rightarrow$ It doesn't make a difference

Both will be same line. $\longleftarrow \begin{cases} 2x + y = 1 \\ 2x + y = 2 \end{cases}$

→ even if we multiply the whole equation with some other number the line doesn't change



* we want calculate distance (d) such that no positive or negative point can cross the margin line

$$\left.\begin{array}{ll} +1 \quad \text{for +ve class dps} & \vec{w}^T\cdot\vec{x} + b \geq 1 \\ -1 \quad \text{for -ve class datapoints} & \vec{w}^T\cdot\vec{x} + b \leq -1 \end{array}\right\}$$

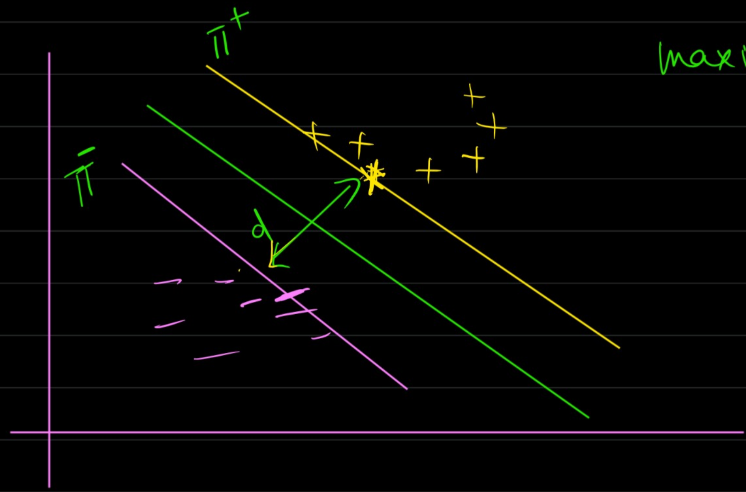
we want to maximize 'd' such that this constraint holds true

for +ve class → $\underset{+1}{y_i}\,(\underbrace{\vec{w}^T\cdot\vec{x} + b}_{+ve}) \geq 1$

for -ve class = $y_i(\vec{w}^T\vec{x} + b) \leq -1$

$= +1(\vec{w}^T\cdot\vec{x} + b) \geq +1$

$= y_i(\vec{w}^T\cdot\vec{x}_i + b) \geq 1$



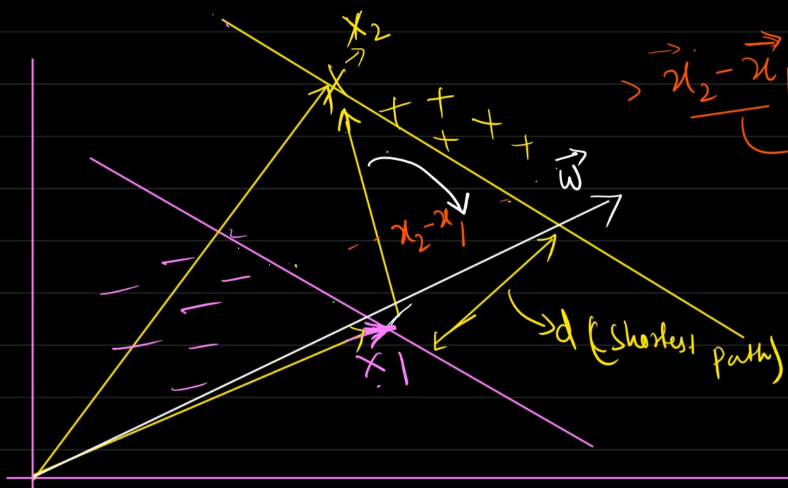
maximise d

$$y_i(\vec{w}^T \cdot \vec{x} + b) \geq 1$$

for support vector

$$y_i(\vec{w}^T \cdot \vec{x} + b) = 1$$

equality because Support Vectors falls on marginal hyperplane.



→ $\vec{x}_2 - \vec{x}_1$

↳ b/w vectors x_2, x_1

→ To get shortest distance we need a unit vector ⊥r to all marginal hyperplane

→ Projection of $\vec{x}_2 - \vec{x}_1$ on Unit Vector $\vec{w}$ to get d

$$d = (\vec{x}_2 - \vec{x}_1) \cdot \frac{\vec{W}}{\|W\|}$$

$$\frac{x_2 \cdot \vec{w} - \vec{x}_1 \cdot w}{\|W\|} \quad \text{—— (1)}$$

(x_1 & x_2 are support Vectors, they lie on marginal hyperplane)

Since $x_1 x_2$ are Support Vectors, it should follow

$$y_i(\vec{w} \cdot \vec{x} + b) = 1$$

for +ve class = $y_i = 1$ for x_1

$$1 \times (w^T x_1 + b) = 1$$

$$\vec{w}^T \vec{x}_1 = 1 - b \quad \text{—— (2)}$$

for -ve class $-1(w^T x_2 + b) = 1$

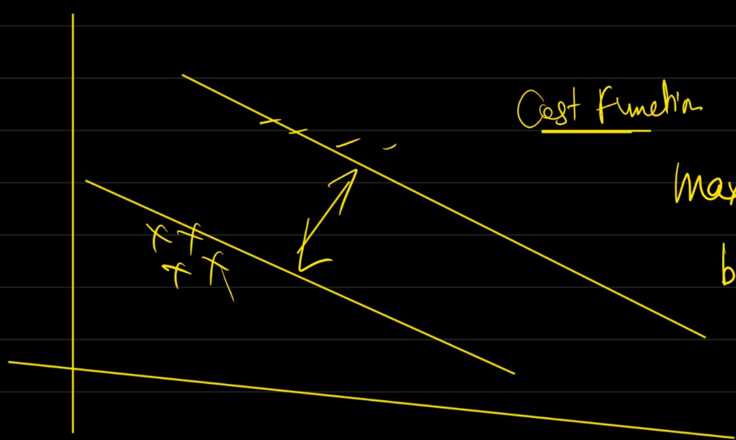
$$\vec{w}^T \cdot \vec{x}_2 = -b - 1 \quad \text{—— (3)}$$

Putting eqn ② & ③ in ①

$$\frac{(1-b)-(-b-1)}{\|W\|} = \frac{2}{\|W\|} = d$$

$$\underset{(w,b)}{\text{maximise}} \ \frac{2}{\|W\|} \quad \text{such that}$$

$$y_i\left(\vec{W}^T \cdot \vec{X} + b\right) \geq 1$$



Cost Function

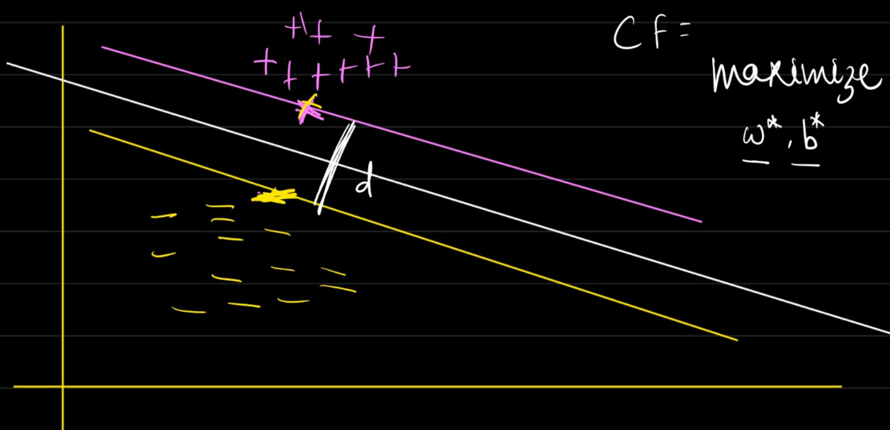
maximize $\frac{2}{\|W\|}$

by changing $\underline{w, b}$ (intercept / coeffs)

* Modified Cost fn for Hardmargin SVC

minimise $\frac{\|W\|}{2}$ by varying w & b

Constraint $y_i(w^T x_i + b) \geq 1$



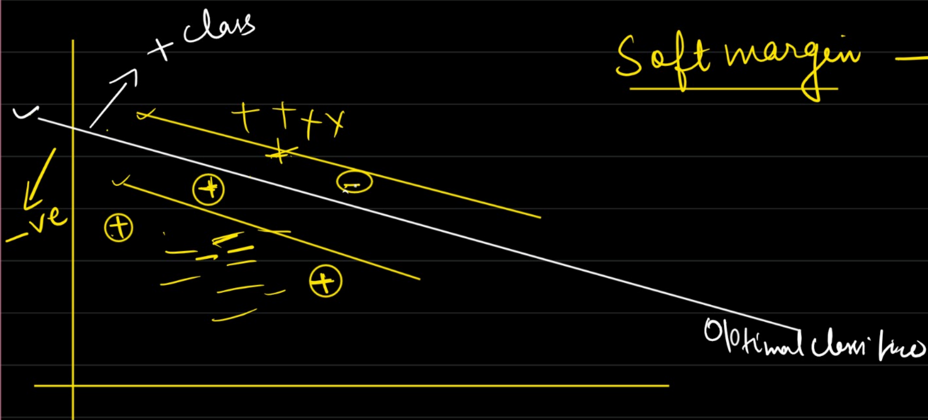
CF = $\underset{w^*, b^*}{\text{maximize}} \ \frac{2}{\|W\|}$

Constraint $y_i\,(w^T x_i + b) \geq 1$

$$\max f(x) \Leftrightarrow \min \frac{1}{f(x)}$$

$$CF = \min_{w^*b^*} \frac{\|w\|}{2}$$

$$y_i(w^T x_i + b) \geq 1$$



Soft margin — No of datapoint you want to sacrifice / error datpoint / misclassified datapoint because such distinguishable scenario is not possible, there will be some overlapping datapoints.

→ Error dp <<< correctly classified dp's

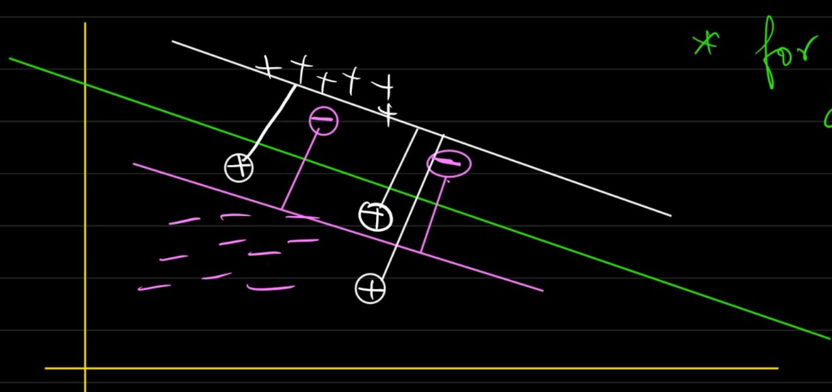
CF for soft margin SVC

$$\underset{w^*b^*}{\text{minimise}} \frac{\|w\|}{2} + C \sum_{i=1}^{n} \xi_i \quad \text{Such that } y_i(w^T x_i + b) \geq 1$$

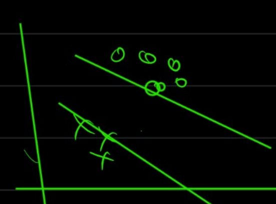
ξ → Zeta

C ↓ no of misclassified dp's (hyperparameter).

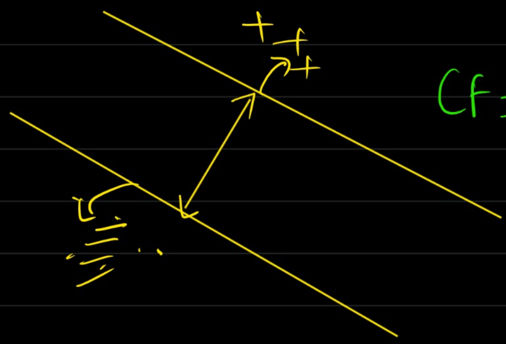
ξ → is the distance of all misclassified dp's to correct marginal plane.



* for all correctly classified dp's ⇒ hard margin $\xi = 0$



→ higher the d ⇒ distance b/w the hyperplanes of two class, lower the error



$$CF = \underbrace{\frac{\|W\|}{2}}_{\text{SVM error}} + C \underbrace{\sum_{i=1}^{n} \xi_i}_{\text{classification error}}$$

→ Hinge loss

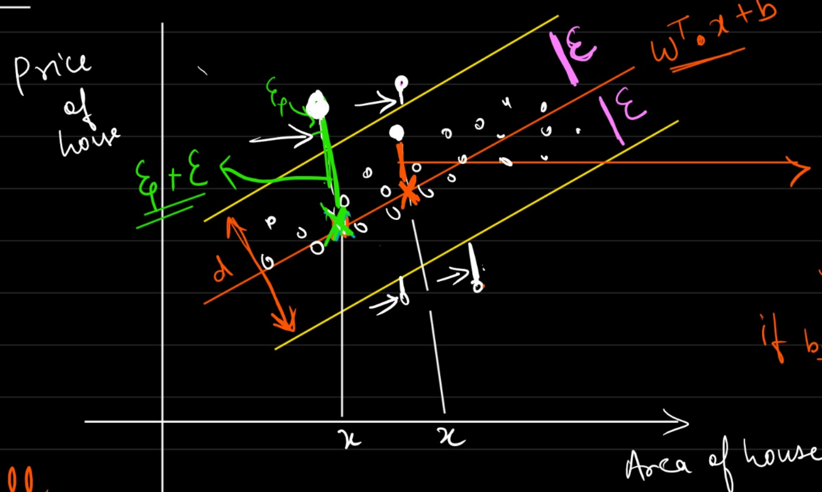
Such that $y_i(w^T x_i + b) \geq 1$

$$\boxed{C = \frac{1}{\lambda}}$$

Support Vector Regressor

SVC

$$\min_{w,b} \frac{\|W\|}{2} + C \underbrace{\sum_{i=1}^{n} \xi}_{\text{hinge loss}}$$



$y_{act} - y_{pred}$

⇓

$y_{act} - w^T x + b$

if $b = 0$ ⇓

$y_{act} - w^T x_i$

→ We want all our dp's surrounding best line.

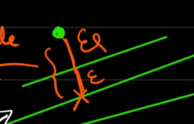
⇓

⇒ All the data points should be in the boundary of marginal plane.

All the dp's $w^T x + b + \varepsilon$ — $w^T x + b - \varepsilon$

$$CF = \min_{w,b} \frac{\|W\|}{2} + C \sum_{i=1}^{n} (\xi_i + \varepsilon)$$

as less as possible



Constraints $|y_i - w^T x_i| \leq \varepsilon + \xi_i$

$y_{act} - y_{pred}$ should be lesser than $\varepsilon + \xi_i$

* All the dp's will not be in between marginal plane.

$$\Downarrow$$

ξ is the distance of dp's from its correct marginal plane.

$$\Downarrow$$

ξ should be minimum.

$C \sum_{i=1}^{n} \xi$ — Zeta

ε — Epsilon

$\Big\} \rightarrow$ hyperparameter