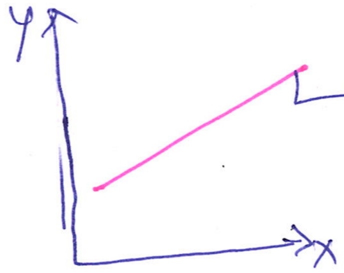
→ Support vector Machins (svm)

1. classification → SVC → Support vector classifier
2. Regression → SVR → support vector Regression



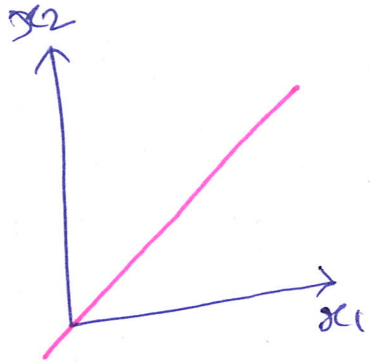
$y = mx + c$

$y = \beta_0 + \beta_1 x$

$\Downarrow$

$ax + by + c = 0 \Rightarrow y = \underbrace{\boxed{-\frac{a}{b}}}_{\text{Coefficient}} x - \underbrace{\boxed{\frac{c}{b}}}_{\text{Intercept}}$

If, I have



$ax_1 + bx_2 + c = 0$

or

$w_1 x_1 + w_2 x_2 + b = 0$ [b/or $\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} [x_1, x_2]$]

or

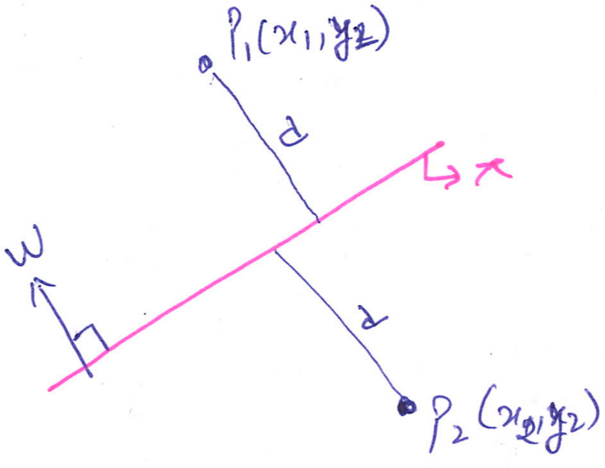
$w^T x + b = 0$

If line is passing through origin

$b = 0$.

then, $\boxed{w^T x = 0}$

Now, If I've a line



To find the distance of Point to the plane.

$\boxed{d = \frac{w^T P_1}{\|w\|}}$

$d \Rightarrow$ distance of a point from Plane

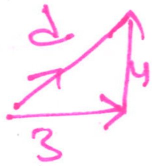
* Unit Vector ⇒ Vector which has magnitude 1

for example,

$d = \sqrt{9+16}$

$d = \sqrt{25}$

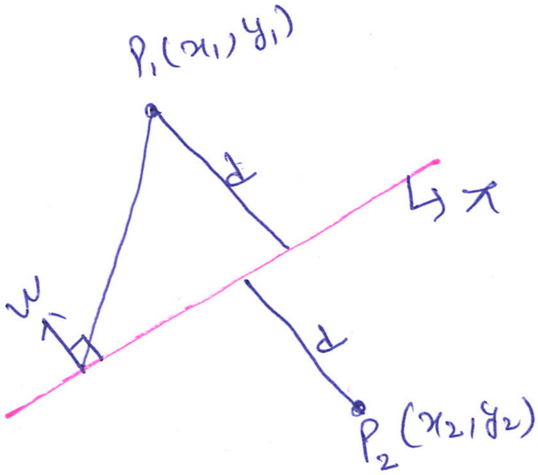
$\boxed{d = 5}$



$\hat{d} = \frac{d}{||d||}$

If we expand, $d = \frac{W^T P_1}{||W||}$

Case I, $\Rightarrow \boxed{||W|| \, ||P_1|| \cos\theta}$



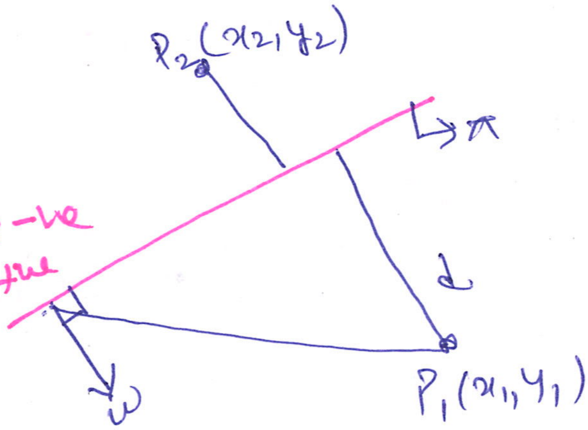
Now, the angle between vector w & point P_1 will be always less than 90°.

So,

$\boxed{d = ||w|| \, ||P_1|| \cos\theta \text{ will always +ve}}$

Any point ~~above~~ above plane π, the distance will always +ve

& " " below " ", " " " " -ve.



Case II

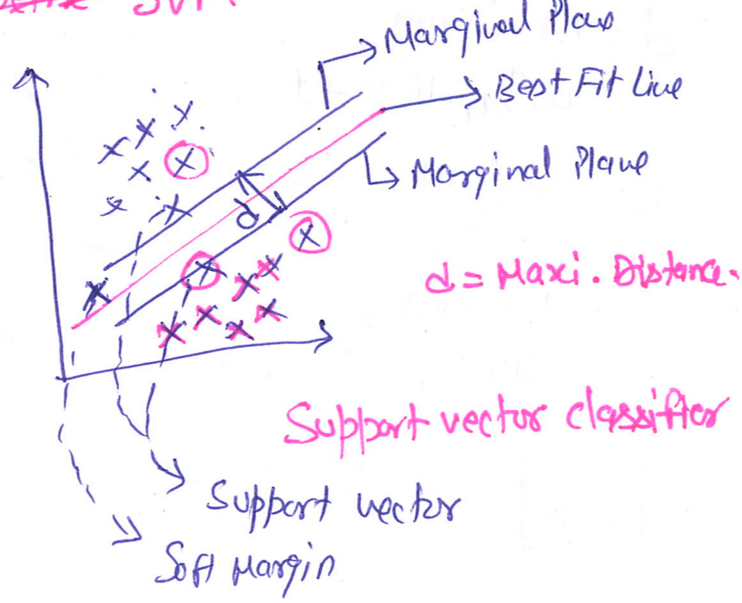
Any point above plane π, the distance will be -ve

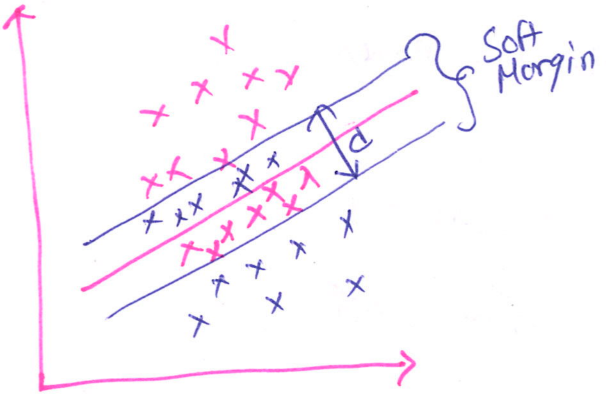
& " " below " ", " " " " +ve

→ Geometric Intution Behind ~~Geometric~~ SVM

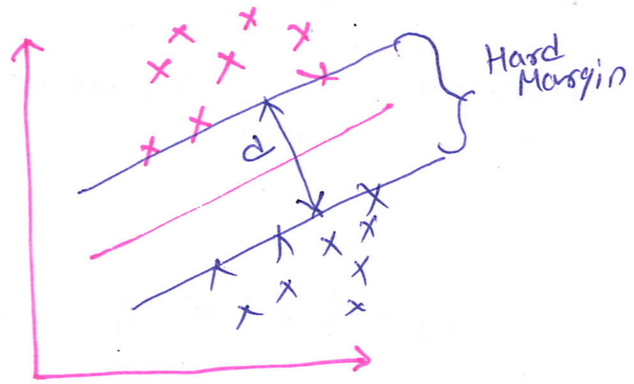
The aim of GI SVM is to find the two planes along with the BEST FIT Line Such that the distance b/w that will be maximum.

with no errors ⇒ Hard Margin
with errors ⇒ Soft margin.



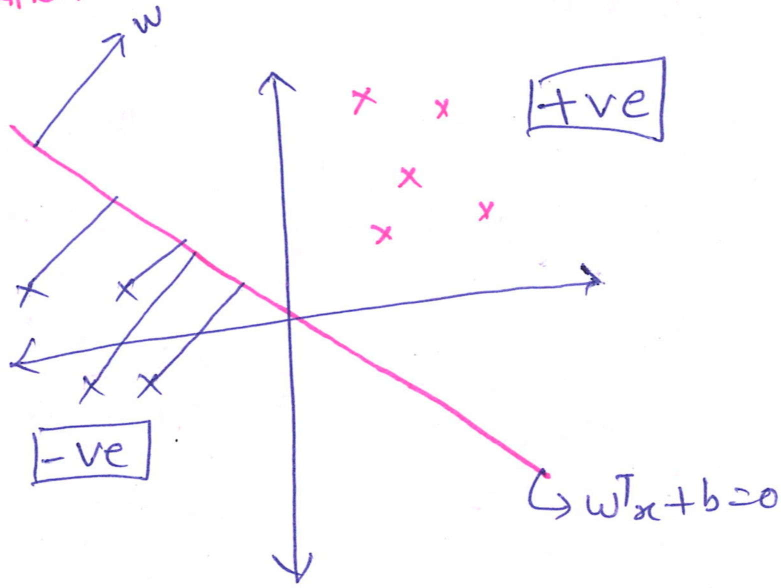


(A)



(B)

→ SVM Mathmatical Intuition



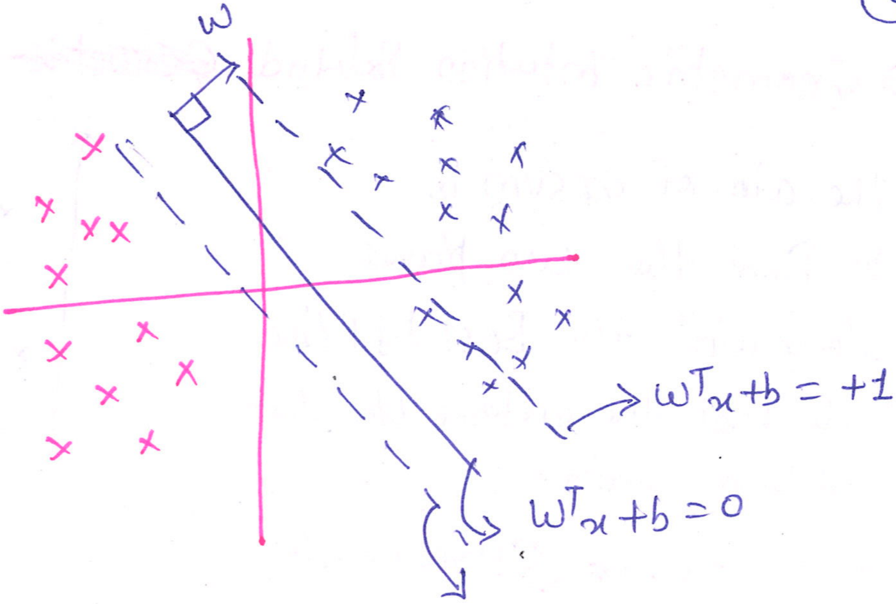
$d = \text{maximum}$

$w^T x_1 + b = +1$

$\underset{(-)}{w^T} \; \underset{(-)}{x_2} + b = \underset{(+)}{-1}$

$\dfrac{w^T(x_1 - x_2)}{\|w\|} \Rightarrow \boxed{\dfrac{2}{\|w\|}}$

$\boxed{\beta_0, \beta_1}$



Cost function $\Rightarrow$

$\boxed{\underset{w,b}{\text{maximize}} = \dfrac{2}{\|w\|}} \Rightarrow$ Distance b/w Marginal Planes

Constant such that $y_i \begin{cases} 1 & w^T x + b \geq 1 \\ -1 & w^T x + b \leq -1 \end{cases}$

For all Correct Points

$\boxed{\text{Constraints} \rightarrow y_i * (w^T x + b) \geq 1}$

$\underset{w,b}{\text{Maximize}} \dfrac{2}{\|w\|} \Rightarrow \boxed{\underset{w,b}{\text{minimize}} \dfrac{\|w\|}{2}} \Rightarrow$ Loss function $\Downarrow$ Minimize

$$\text{Cost Function} = \underset{w,b}{\text{Min}} \frac{\|w\|}{2} + \boxed{C_i \sum_{i=1}^{w} \xi_i}$$

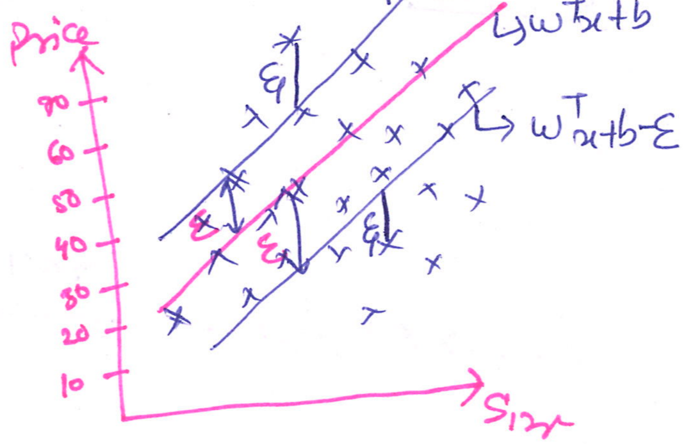
$=$ Hinge Loss

Soft Margin

where, C_i = How many points we can ignore for mis classification = Hyperparameter

ξ_i = Submission of the distance of the incorrect data points from Marginal Plane.

② Support vector Regression.



Cost function $\Rightarrow \min_{w,b} \frac{\|w\|}{2} + \boxed{C_i \sum_{i=1}^{n} \varepsilon_i}$

Hinge Loss

$\varepsilon \Rightarrow$ Marginal error

$\varepsilon_i \Rightarrow$ Error above the margin.

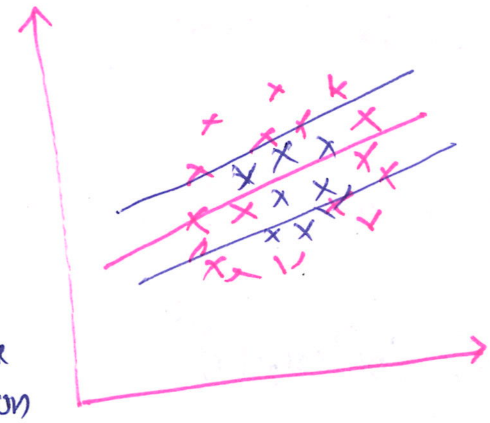
Constraint MAE

$$\boxed{|y_i - w_i x_i| \le \varepsilon + \varepsilon_i}$$

will svm impacted by outliers? $\Rightarrow$ yes

we need to preform standardization? $\Rightarrow$ yes

③ SVM kernel

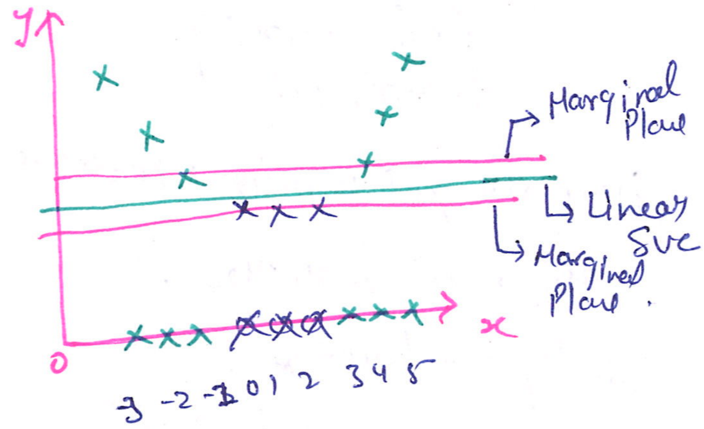


SVM kernel ⟶ Transformation ⟶ increase Dimension of Data.

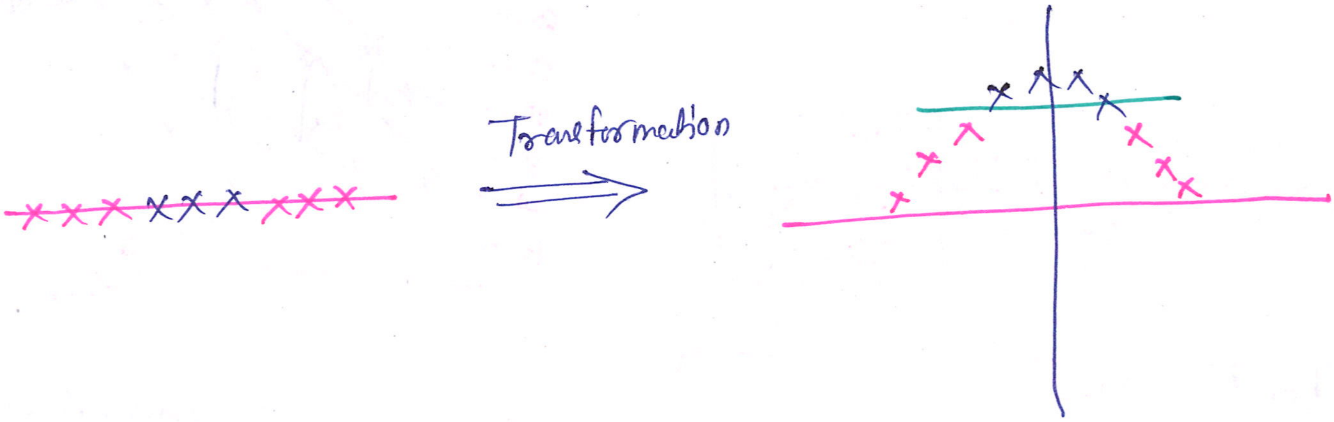
Transformation ⇓ (These are Mathematical Formula)

Transformation takes Place in SVM kernel from 1d to 2d.

$$\boxed{y = x^2}$$



Example:



There are 3 types of SVM kernal.

- → Polynomial kernel
- → RBF kernel
- → Sigmoid kernel.

(1) Polynomial kernel

$$f(x_1, x_2) = (x_1^T \cdot x_2 + 1)^d$$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \cdot \begin{bmatrix} x_1 & x_2 \end{bmatrix}$$

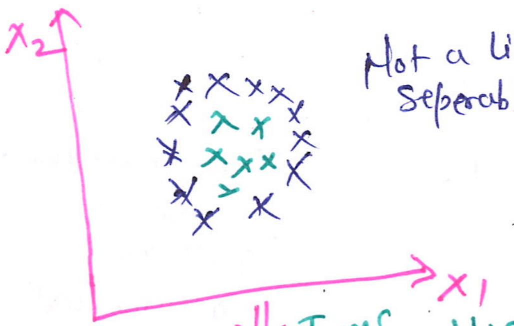
$$\begin{bmatrix} x_1^2 & x_1 x_2 \\ x_1 x_2 & x_2^2 \end{bmatrix}$$

Here we get 3 unique features.

x_1^2, x_2^2 & $x_1 x_2$.

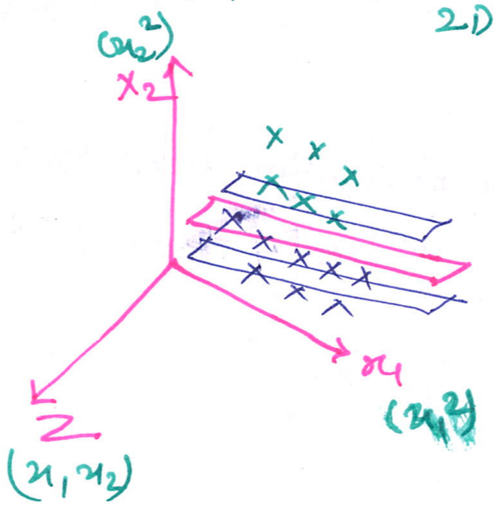
Initially we have

x_1, x_2 & y (o/p).



Not a linear separable data.

⇓ Transformation from 2D to 3D



→ Ro·c And AUC

For any algorithm the threshold is super for perfect.

Example,

y	$\hat{y}$	$\hat{y}(0)$	$\hat{y}(0.2)$	$\hat{y}(0.4)$
1	.8	1	1	1
0	.96	1	1	1
1	.4	1	1	0
1	.3	1	1	0
0	.2	1	1	0
1	.7	1	0	0
		1	1	1

Confusion Matrix For $\hat{y}(0)$

Predicted \ Actual	1	0
1	4	2
0	0	0

① $\boxed{TPR = \frac{TP}{TP+FN}}$

$$TPR = \frac{4}{4+0}$$

$$\boxed{TPR = 1}$$

② $\boxed{FPR = \frac{FP}{FP+TN}} = \frac{2}{2+0} = 1$

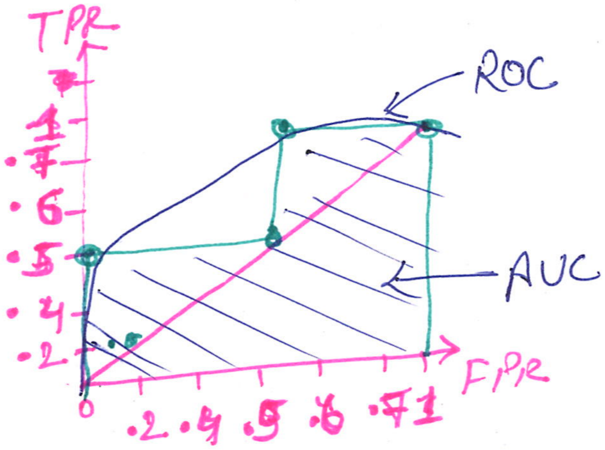
Now, we will Plot the dia.

Confusion Matrix for $\hat{y}(0.2)$

4	1
0	1

Confusion Matrix for $\hat{y}(0.4)$

2	1
2	1



Higher the area under the curve better the model Performance.