Date:— 29th October, 2022

(47)

Machine Learning (Day-5)

Agenda

1) SVC (Support Vector Classifier)
2) SVR (Support vector Regressor)

*) Support Vector Machine (SVM):—

⇒ Support Vector Machine or SVM is one of the most popular Supervised Learning Algorithms, Which is Used for Classification as well as Regression Problems. However, primarily, it is used for Classification Problems in Machine Learning.

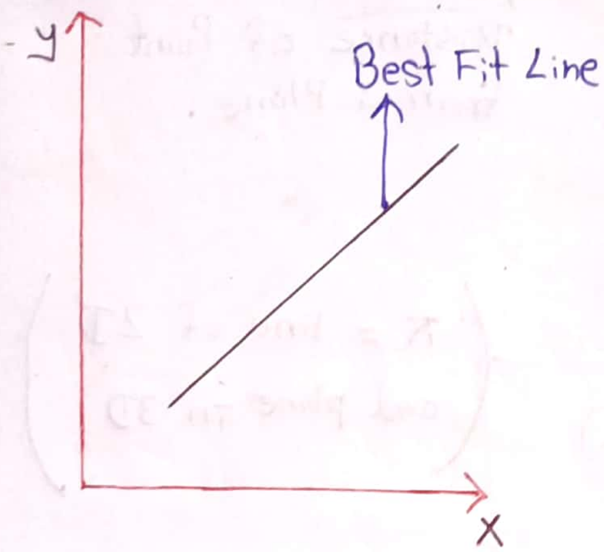
Note:— The goal of the SVM Algorithm is to create the best line or decision boundary that can Segregate n-dimensional Space into classes So that we can easily put the new data point in the correct category in the future. This best decision Boundary is called a hyperplane.

*) It can Solve both Classification and Regression Problem:—

a) **Classification:—** SVC (Support Vector Classifier)

b) **Regression:—** SVR (Support Vector Regressor)

✪) Some basics :-



✪) Equation of Line :-

$y = mx + c$ **OR**

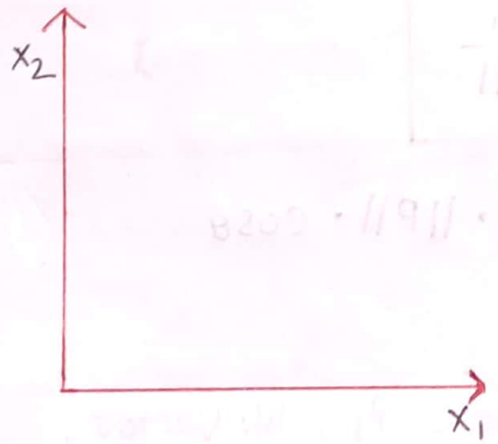
$y = B_0 + B_1 x$ **OR**

$ax + by + c = 0$

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

coefficient

Intercept



$ax_1 + bx_2 + c = 0$

$w_1 x_1 + w_2 x_2 + b = 0$

$\therefore W^T x + b = 0$

$\therefore$ If Line passes through origin then, $W^T x = 0$

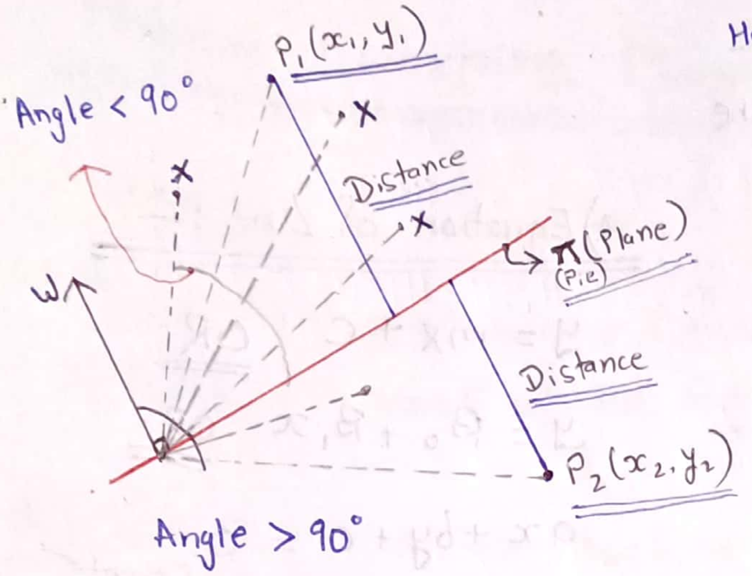
✪) Matrix Multiplication :-

$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \begin{bmatrix} x_1 & x_2 \end{bmatrix}$

$W^T x + b = 0$ ($W^T x$: W Transpose x)

$\therefore$ Equation of Line passing through origin is :-

$$\boxed{W^T x = 0}$$



Here, We have to find the Distance of Point from a Plane.

(π = line of 2D and plane in 3D)

★) Distance of a Point to the Plane.

$$\boxed{\text{Distance } (d) = \frac{W^T P_1}{\|W\|}}$$

$$= \|W\| \cdot \|P\| \cdot \cos\theta$$

Note:- Where,

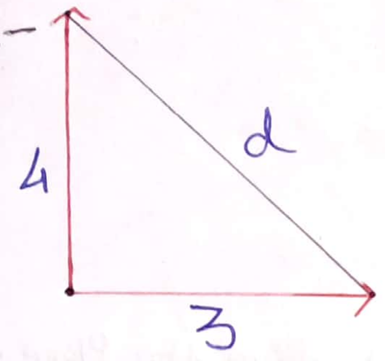
$W^T P_1$: W Transpose P_1, W: Vector,

$\|W\|$: magnitude of W.

P.T.O

★) **Unit Vector** :– A vector which has a Magnitude of 1 is basically called Unit Vector.

Eg:–



Now, (Hypotheuse) :–

The largest Side which is opposite to the Right-angled (90 degree) is Known as the Hypotenuse.

$$d = \sqrt{3^2+4^2}$$
$$= \sqrt{25}$$
$$\therefore d = 5$$

Where, Vector, $\hat{d} = \dfrac{d}{\|d\|}$ → Magnitude.

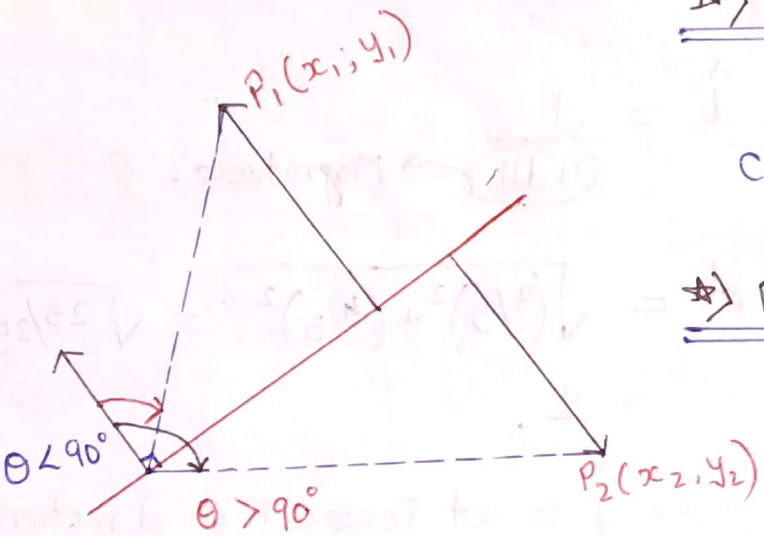
$$(3/5, 4/5) = d = \sqrt{(3/5)^2 + (4/5)^2} = \sqrt{25/25}$$
$$= 1.$$

∴ Unit Vector is a way to get Focused on direction not on magnitude.

P.T.O

Upward Vector :-

$$\boxed{\begin{aligned} d &= \frac{W^T P_1}{\|W\|} \\ d &= \|W\| \cdot \|P_1\| \cdot \cos\theta \end{aligned}}$$



Point above the Plane :-

as, $\theta < 90^\circ$

$\cos\theta$ will always be +ve.

Point below the Plane ;-

as, $\theta < 90^\circ$

$\cos\theta$ will always be -ve.

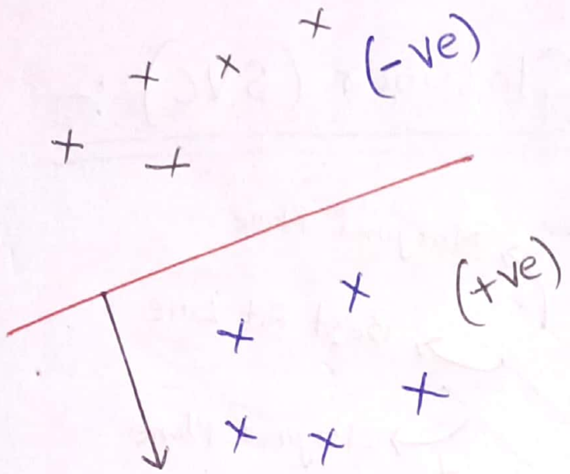
Note :-

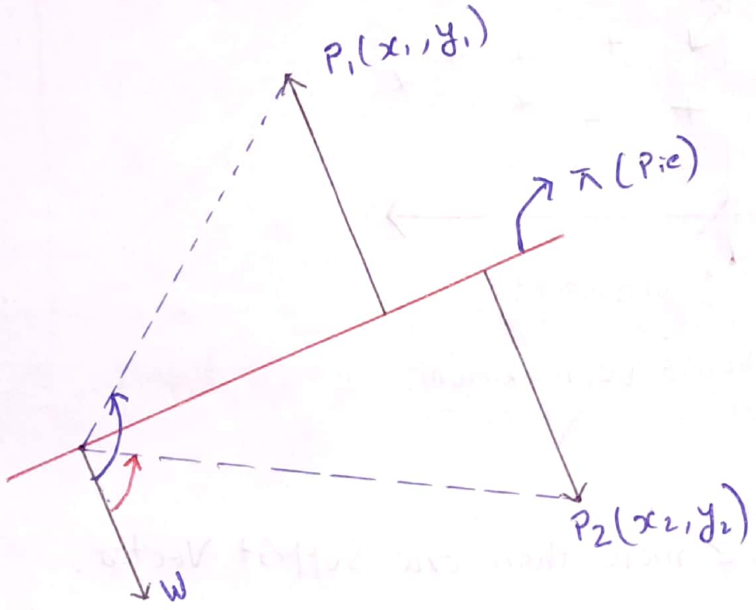
ⓐ If any point Falling above the plane, then θ must be Less than 90°. (+ve Value)

ⓑ If any Point falling below the Plane, then θ must be Greater than 90°. (-ve Value).

P.T.O

*) Downward Vector :-





Note :-

*) Point below the plane,

$\theta < 90°$

$\cos\theta$ will always be "+ve"

*) Point above the plane,

$\theta > 90°$

$\cos\theta$ will always be "-ve"

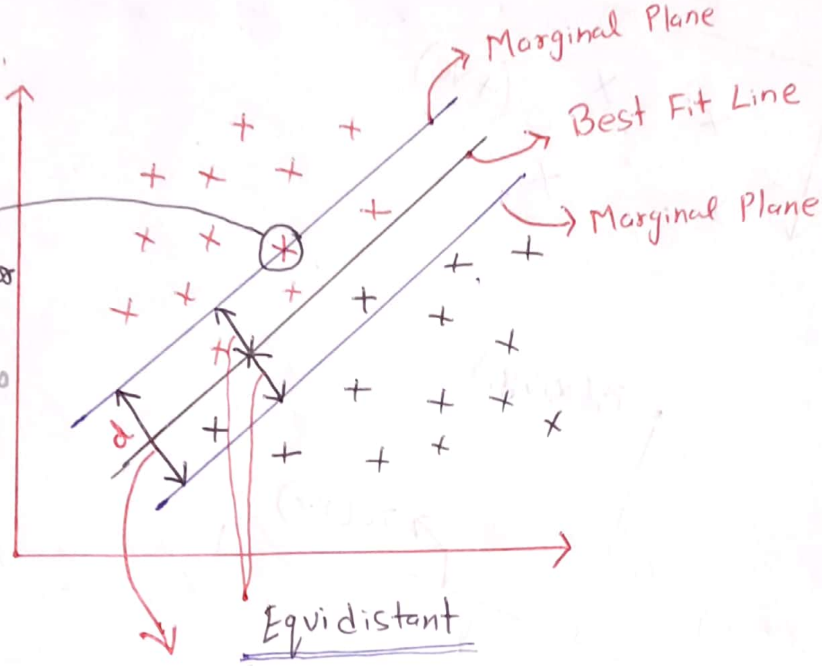
P.T.O

*) Geometric Intuition Behind Support Vector Machine :-

Support Vector Classifier (SVC) :-

Support Vector

{Point which are crossed by Marginal Plane}



Equidistant

d should be maximum.

Note :- You can have more than one Support Vector.

*) There will be many Possible hyperplanes that Separate different Classes.

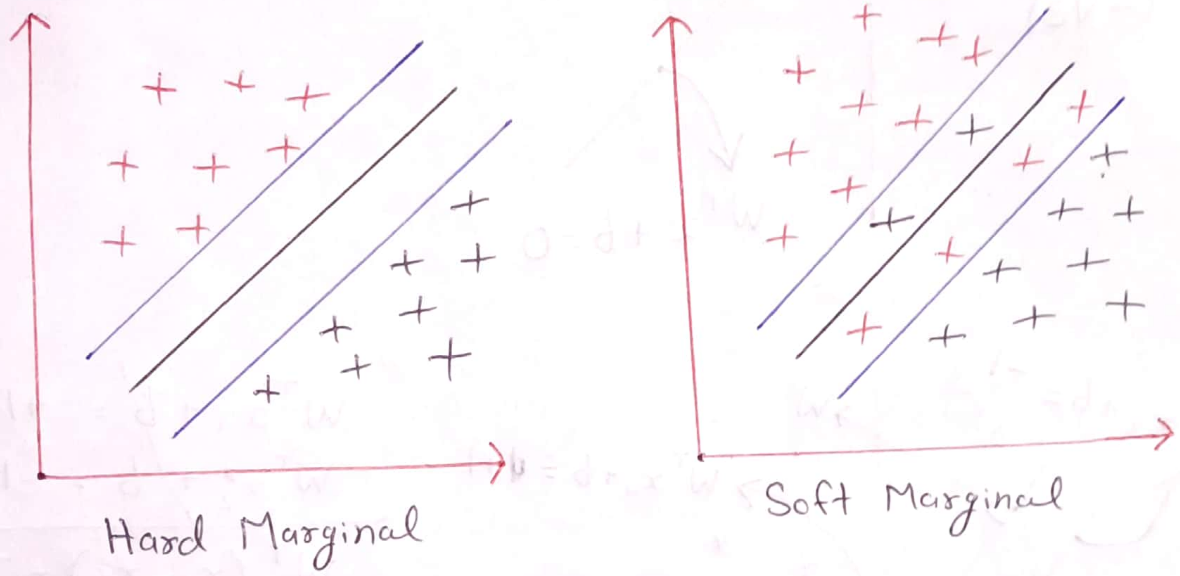
We have learnt in LR that the Probability of a Point belonging to any class given at very close to the hyperplane will be close to 0.5.

P.T.O

So, we want a hyperplane that Seperate (+ve) Pts and (-ve) Pts as far away as Possible.

↳ Key idea of SVM,

Such hyperPlane is called Margin-Maximizing Plane.



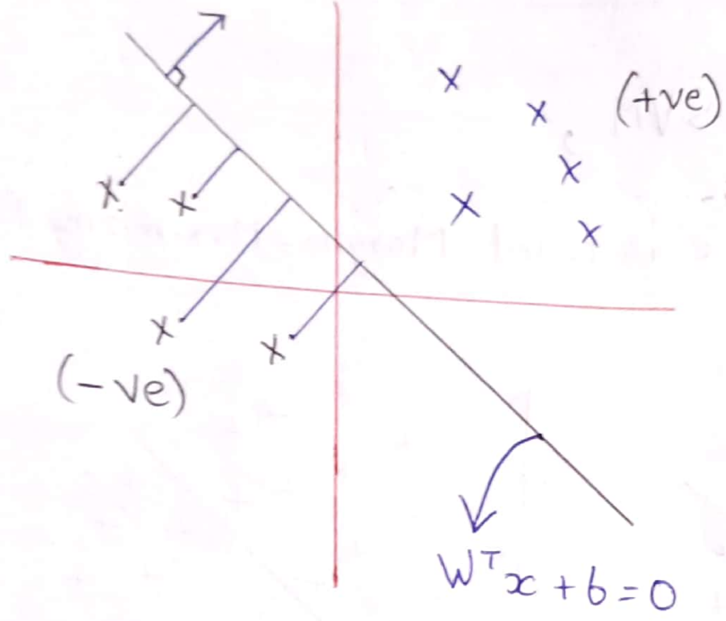
*) Marginal Plane :— Line Passes through Nearest Points.

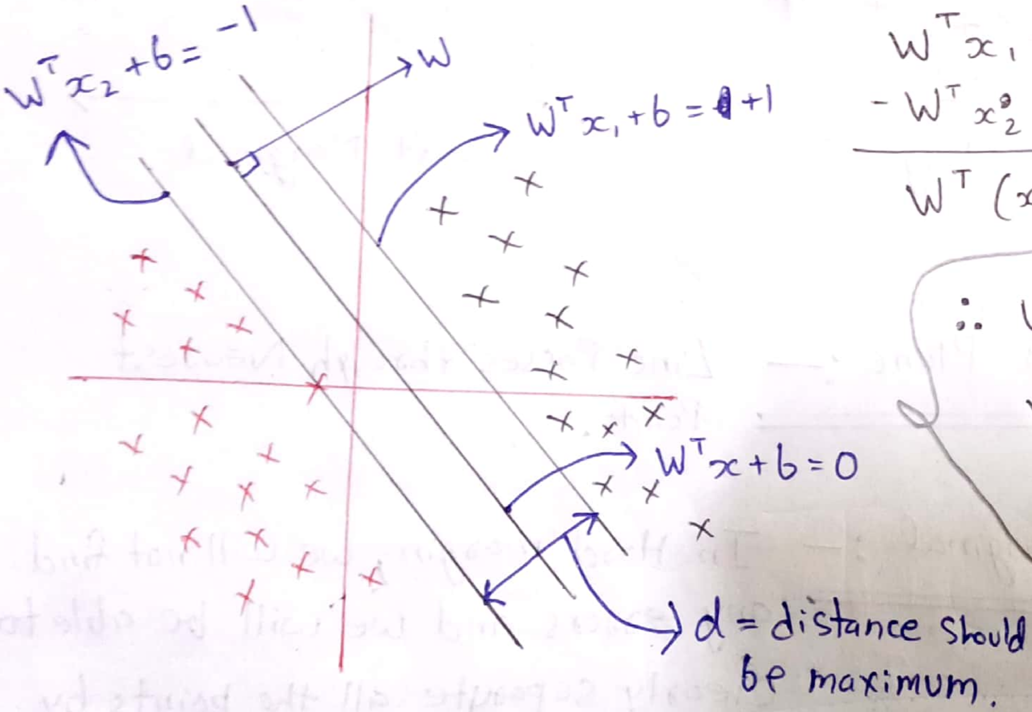
*) Hard Marginal :— In Hard margin, we will not find any errors and we will be able to clearly separate all the points by using the marginal Plane.

∴ But, In Real World, there will be many Overlapping with many Errors, So Marginal Plane Line will be called — Soft Margin.

*) Marginal Plane Should be Equidistance from Best Fit — Line.

✪) SVM Mathematical Intuition :-





$$
\begin{aligned}
W^T x_1 + \cancel{b} &= +1 \\
- W^T x_2 + \cancel{b} &= -1 \\
\hline
W^T (x_1 - x_2) &= 2
\end{aligned}
$$

∴ Unit Vector of W,

$$\frac{W^T (x_1 - x_2)}{\|W\|} = \frac{2}{\|W\|}$$

$$\boxed{\therefore \text{distance } (d) = \frac{W^T (x_1 - x_2)}{\|W\|} = \frac{2}{\|W\|}}$$

Cost Function :-

We have to maximize the value of $\frac{2}{\|w\|}$

by changing the value of w, b.

$\frac{2}{\|w\|} \rightarrow$ Distance between Marginal Plane.

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

For all Classified Correct Points,

$$\boxed{\text{Constraint} \longrightarrow y_i \times (w^T x + b) \geq 1}$$

$$\underset{w,b}{\text{Maximize}} \frac{2}{\|w\|} \Rightarrow \underset{w,b}{\text{minimize}} \frac{\|w\|}{2} \Rightarrow \text{Loss Function}$$

⇓

minimize

* Loss Function Focus on Minimizations.

P.T.O

✯) Cost Functions :-

minimize $\frac{\|w\|}{2}$ by Changing W, b.

$$\boxed{\text{Min } \frac{\|w\|}{2} + C_i \sum_{i=1}^{w} \varepsilon_i}$$

Hinge Loss For ~~Function~~ Soft Margin

Where,

C_i = How many points we can ignore for mis- Classification.

Hyper Parameter.

ε_i (Eta) : Summation of the distance of Incorrect datapoints from the Marginal Plane.

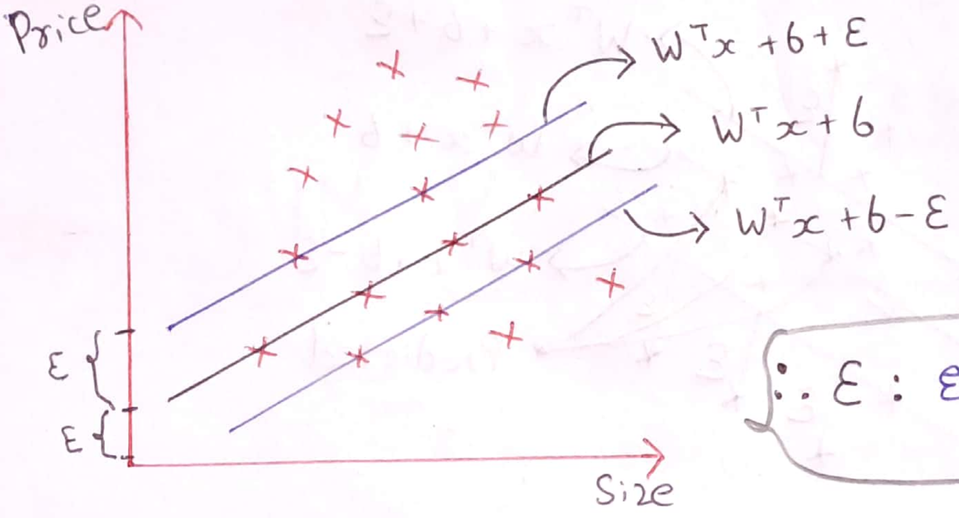
P.T.O

*) Support Vector Regressor (SVR) :-

⇒ Support Vector Regressor (SVR) uses the same Principle as SVM, but for Regression Problems.

*) Problem Statement :-

Based on the Size of the House, we have to predicts Price of the House.



∴ ε : Epsilon → Marginal Error.

*) Cost Functions :-

$$\underset{w,\, b}{\text{Minimize}} \ \frac{\|w\|}{2} + \boxed{C_i \sum_{i=1}^{w} \varepsilon_i} \rightarrow \text{Hinge Loss}$$

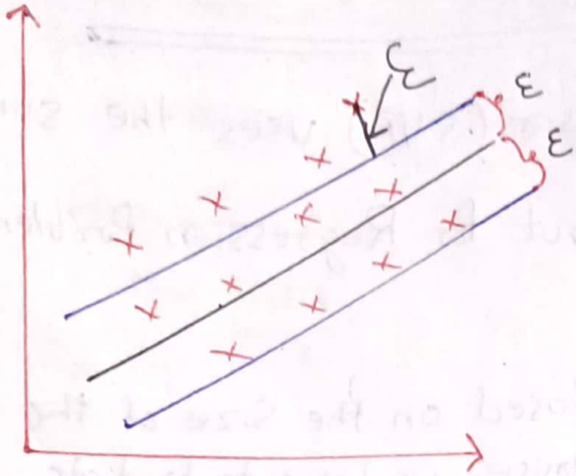
MAE

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

y_i → Truth Point, $w_i x_i$ → Predicted Point, ε → Epcilon, ε_i → Eta of

ε = Margin of Error (to decide Original Plane)

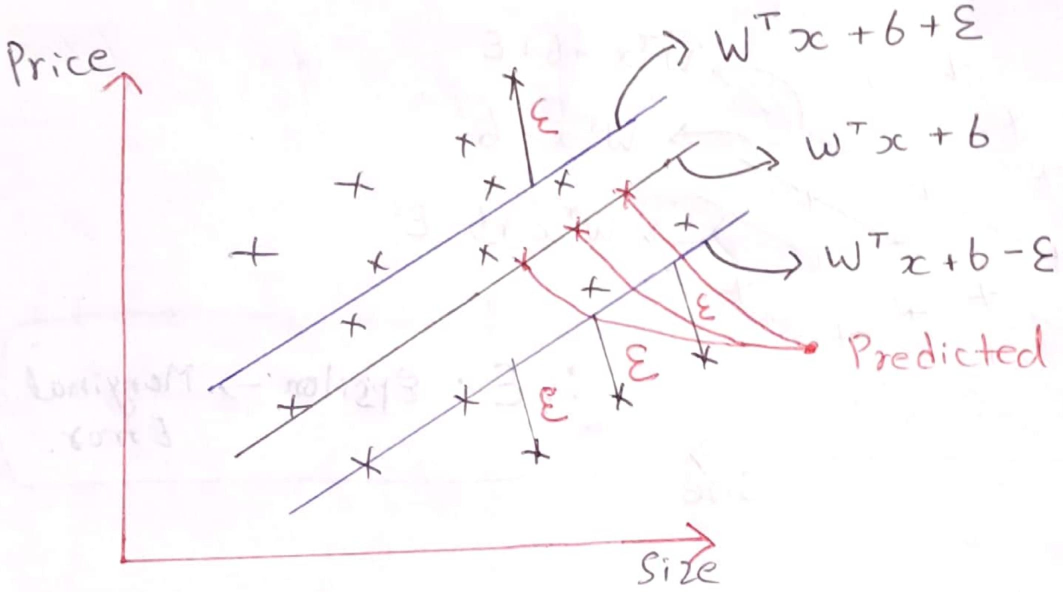
ε_i = Error above the Margin.



Hyper Parameter :-

→ Keep adjusting ε to get best Margin

→ We can't say incorrect point in Regressor.



In Regressor, No Complete Incorrect Value, because, it will be continuous Value.

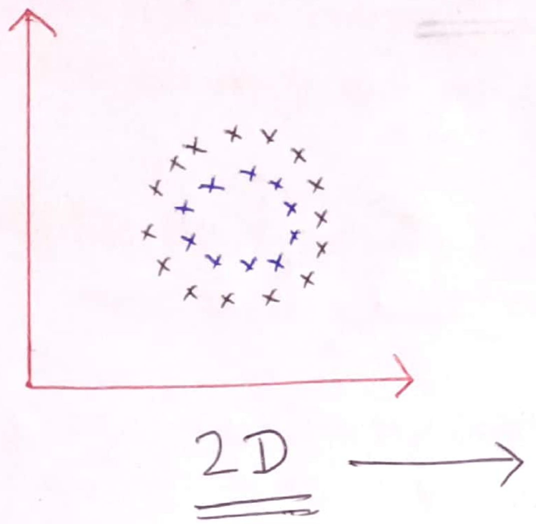
Q) Is SVM Impacted by the Outliers?

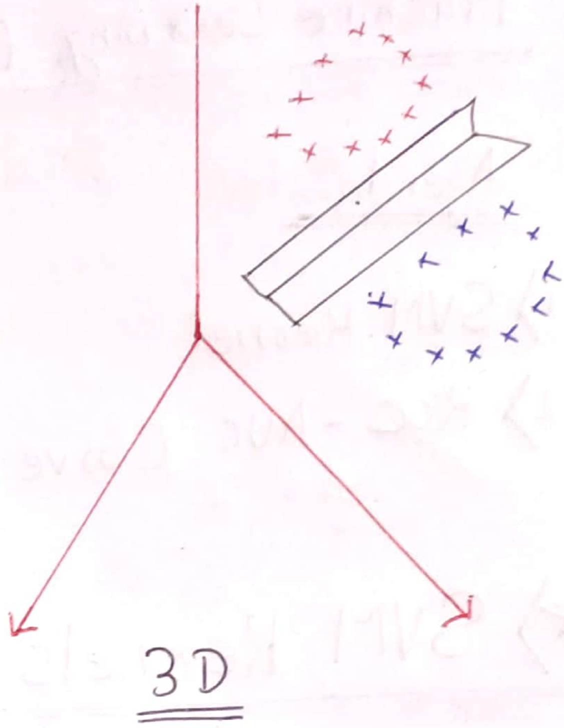
⇒ Yes, SVM is impacted by the outliers.

Q) Does Standardization is need in SVM?

⇒ Yes, we need to Perform Normalization and Standardization.

✳) SVM Kernel :-





✳) SVM Kernels :-

P.T.O