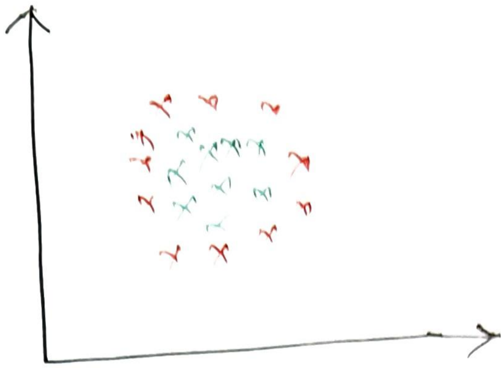
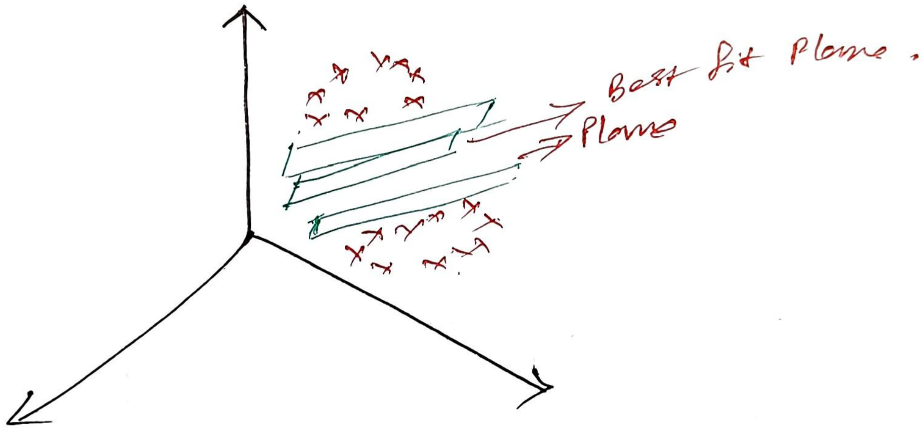
SVM Kernel



→ we can't use SVC here, result will be bad.

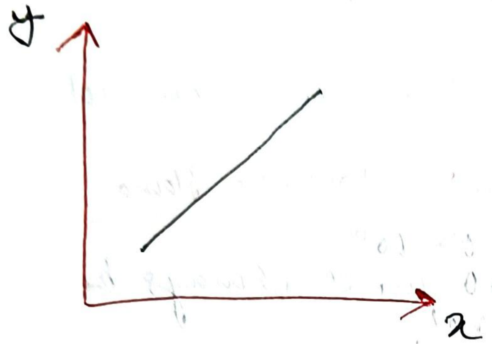
→ In SVM Kernel, we first convert the 2D data points to 3-D points. $\text{2-D} \xrightarrow{\text{Transformation}} \text{3-D}$



29-oct-2022

Support Vector Machines (SVM)

- We can solve both classification and regression problem.
- classification → SVC (Support Vector classification)
  Regression → SVR (Support Vector Regressor)

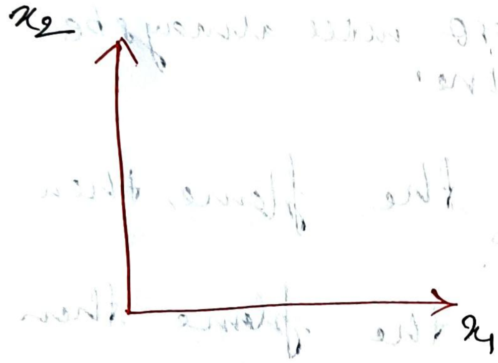


$y = mx + c$

$y = \beta_0 + \beta_1 x$

$ax + by + c = 0$

$y = \underbrace{\boxed{-\frac{a}{b}}}_{\text{Coefficient}} x - \underbrace{\boxed{\frac{c}{b}}}_{\text{Intercept}}$



$ax_1 + bx_2 + c = 0$

$w_1x_1 + w_2x_2 + b = 0$

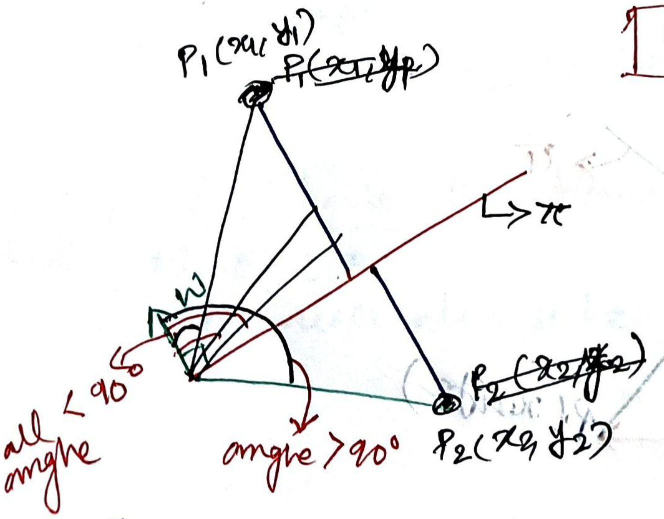
$W^Tx + b = 0$ → if this line passes through origin then $b = 0$

Matrix Multiplication

$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} [x_1 \ \ x_2]$

Eqn of line passes through origin

$\boxed{W^Tx = 0}$



π = line in 2-D and plane in 3-D

W = vector

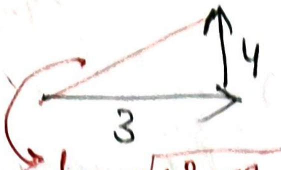
Distance of a point to the plane,

$\boxed{d = \frac{W^T P_1}{\|W\|}} \boxed{= \|W\| \, \|P_1\| \cos\theta}$

we need unit vector that's why divided by magnitude of W.

→

(3,4)



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

$\boxed{\text{Unit vector, } \hat{d} = \frac{d}{\|d\|}}$ → Magnitude

$\left(\frac{3}{5}, \frac{4}{5}\right) \Rightarrow d = \sqrt{\left(\frac{3}{5}\right)^2 + \left(\frac{4}{5}\right)^2} = 1$

→ Unit vector is a way to get focused on direction not on magnitude.

Upward Vector

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

Point below the Plane

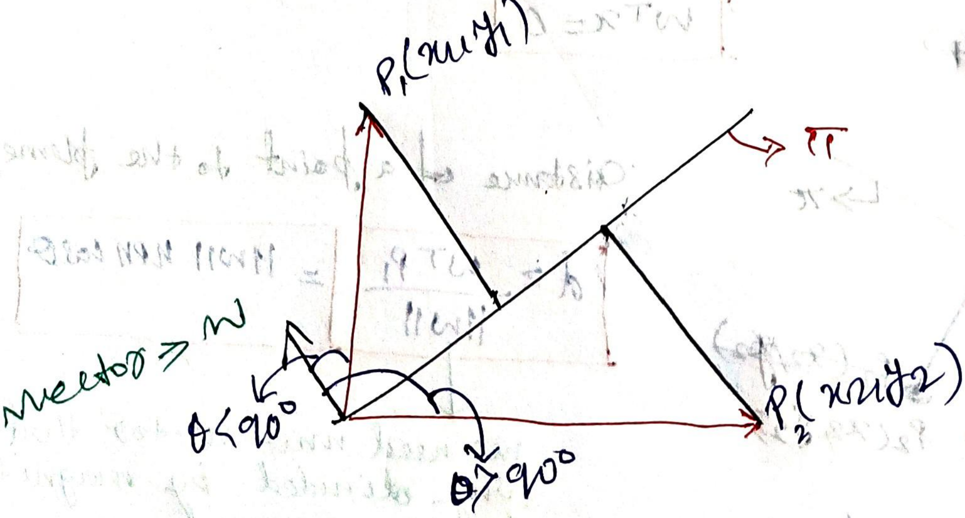
as $\theta > 90°$
$\cos\theta$ will always be (−ve).

Point above the Plane

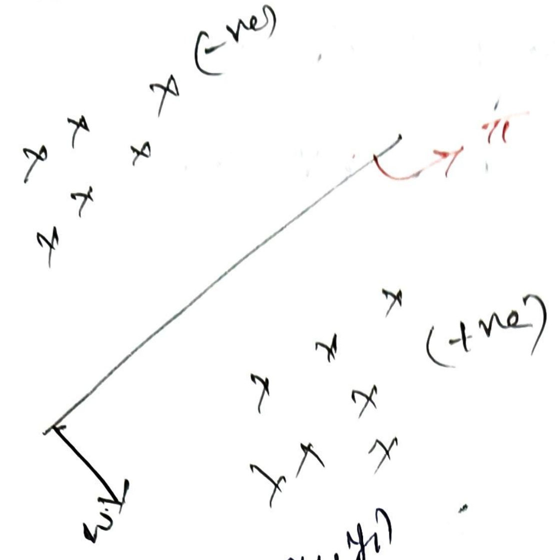
as $\theta < 90°$
$\cos\theta$ will always be (+ve)

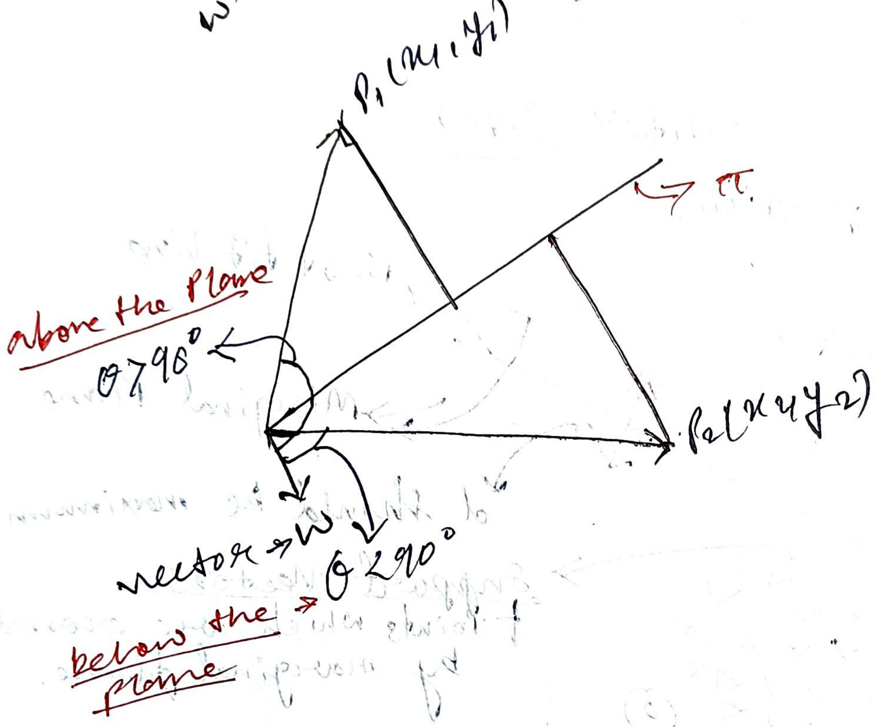
→ If any point falling above the plane, then θ will be less than 90°.

→ If any point falling below the plane then θ will be greater than 90°.



→ Downward Vectors





Point below the plane

as $\theta < 90^\circ$

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

Point above the plane

as $\theta > 90^\circ$

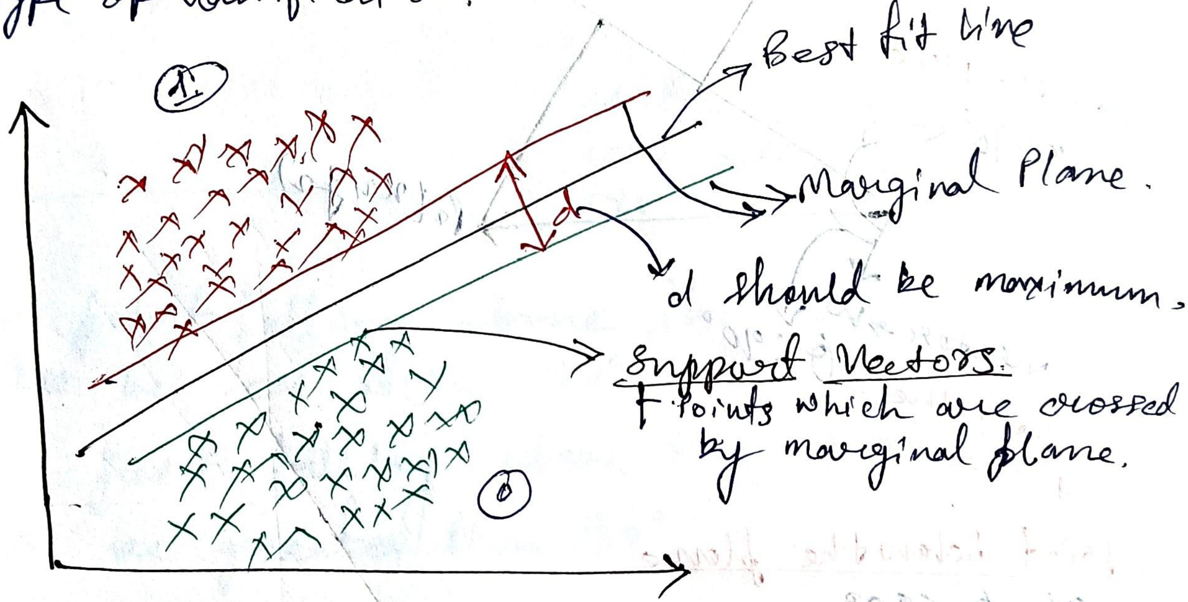
$\cos\theta$ will always be '−ve'.

Geometric Intuition Behind Support Vector Machine

Logistic Regression (Binary classification)

Support Vector Classifier (SVC)

→ Any type of classification.



Marginal Plane

- Line passes through nearest point.

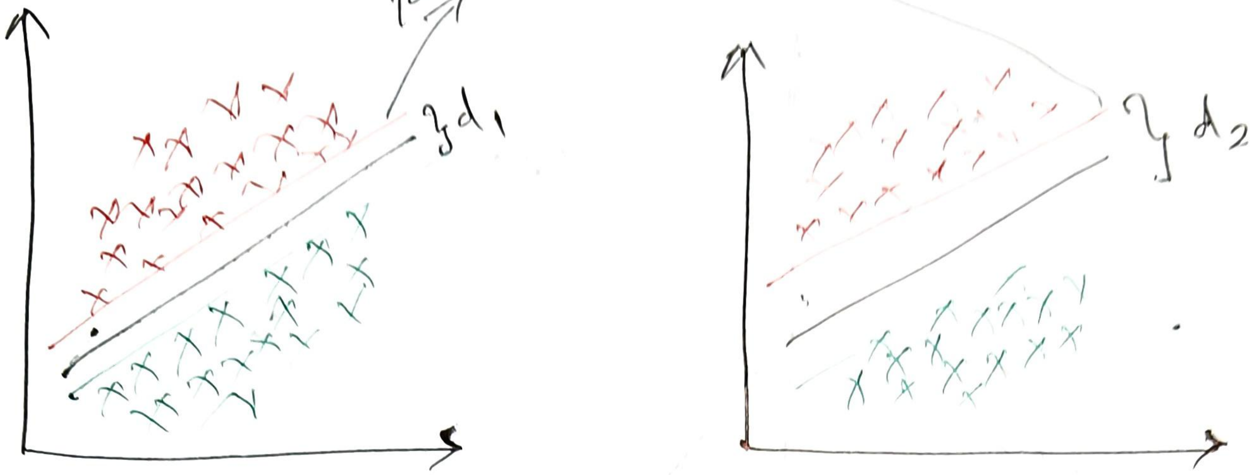
Hard Margin

- Less error. we will able to separate all the points without any error.

Soft Margin

- Having error. we so will not get marginal plane without any error.

Hard Margin (without error)

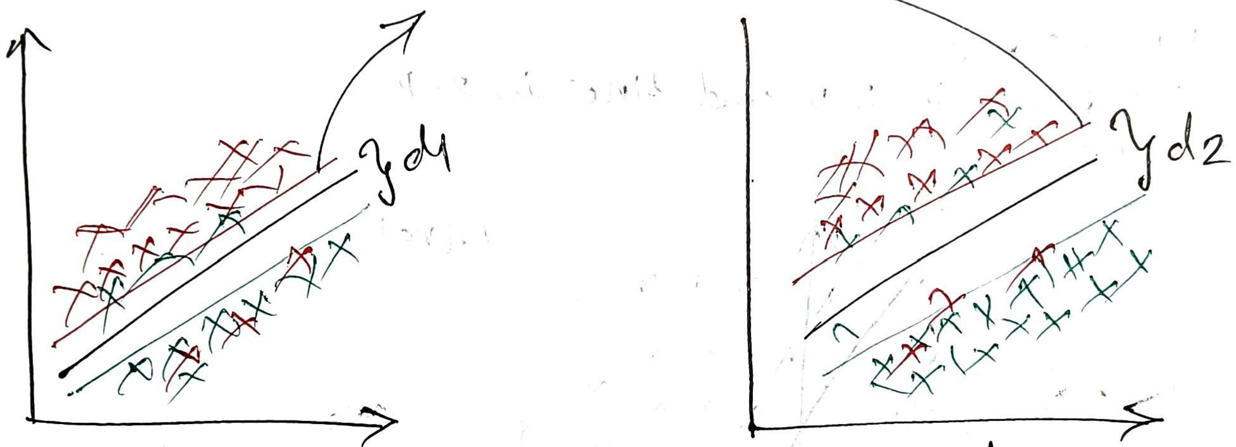


Bad

$d_2 > d_1$

Good

Soft Margin (with Error)



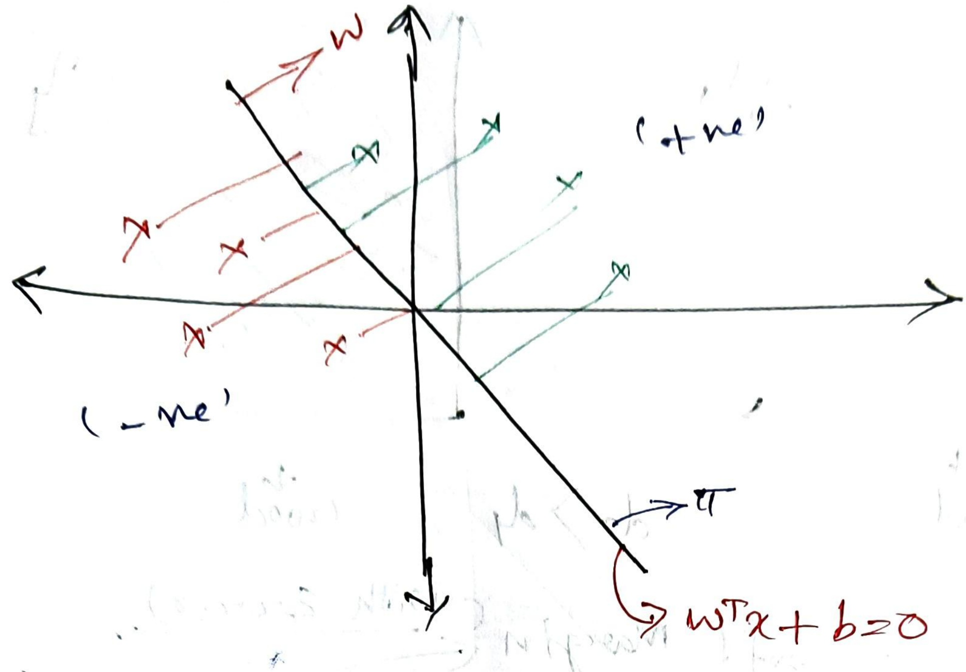
Bad

$d_2 > d_1$

Good

→ Marginal Plane should be equi-distance from best fit line.

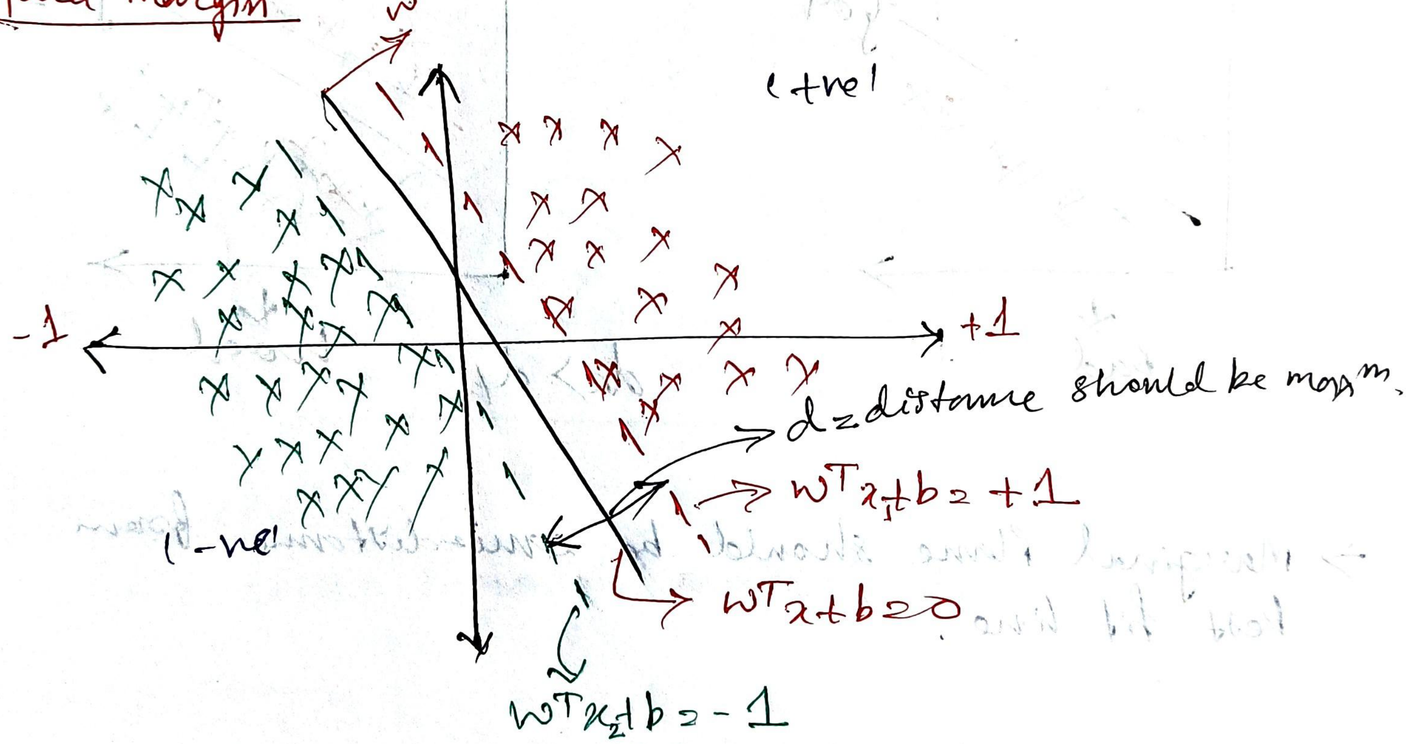
SVM Mathematical Intuition



w = vector

π = Plane in 3-D and line in 2-D

Hard Margin



$$w^T x_1 + b = +1 \quad \text{—— (1)}$$

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

(1) - (2) $\quad (-) \quad (-) \quad (+)$

$$w^T (x_1 - x_2) = 2$$

Unit vector of w,

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

cost function

Maximize, $\frac{2}{||w||}$ → Distance b/w Marginal Plane
w, b

change it to get maximum distance

constraint such that

condition for all correct classified Points

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

For all classified correct points

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

Maximize $\frac{2}{||w||}$ by changing w, b

Minimize $\frac{||w||}{2}$ by changing w, b

Both are same.

→ Loss function focus on minimization

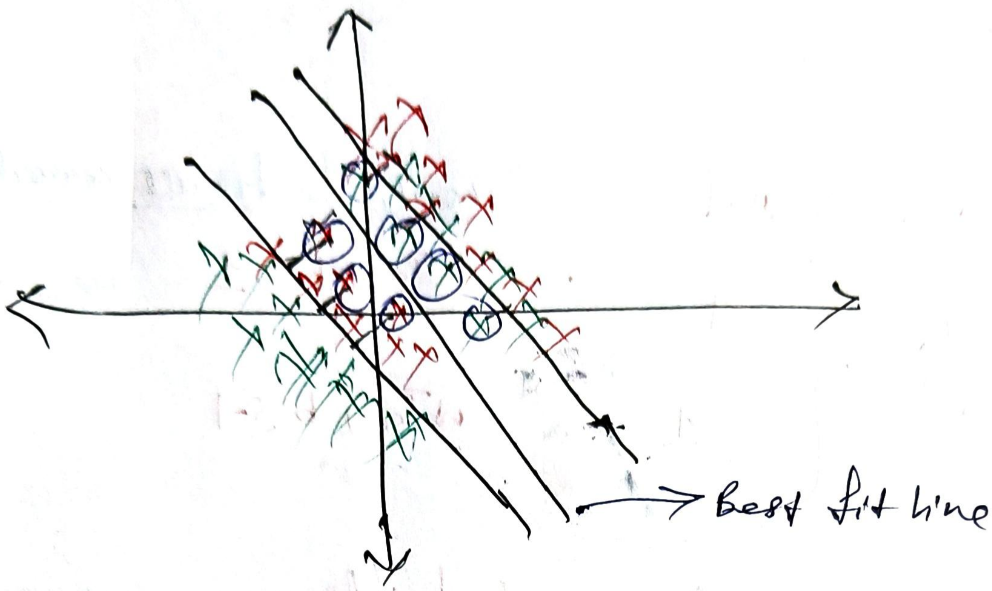
cost function

Min $\frac{||w||}{2}$ by changing w, b

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

Hinge loss for soft margin

ζ_i = { How many points we can ignore for missclassification.

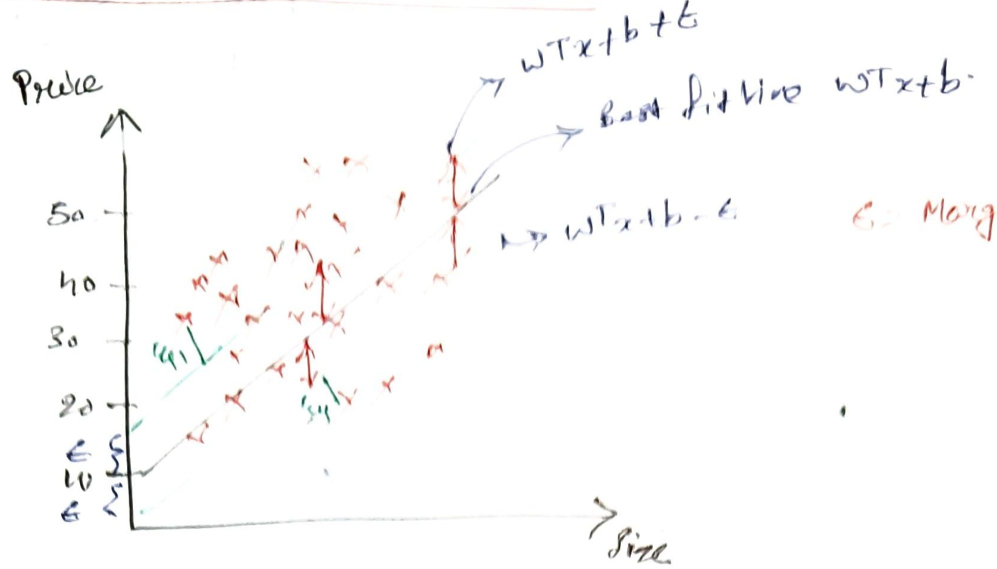


ζ_i = 7 points are miss classified → Ignore.

ζ_i = Summation of the distance of the incorrect data points from the marginal plane.

7 points distance from the marginal plane.

SVR (Support Vector Regressor)



ε = Marginal Error

Cost function

Minimize $\frac{\|W\|}{2} + \boxed{C_i \sum_{i=1}^{n} |\xi_i|}$

by changing w, b

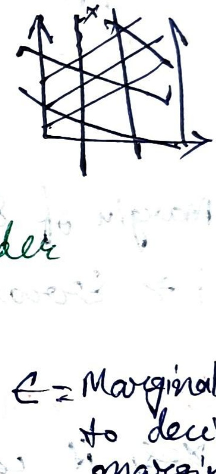
Hinge loss

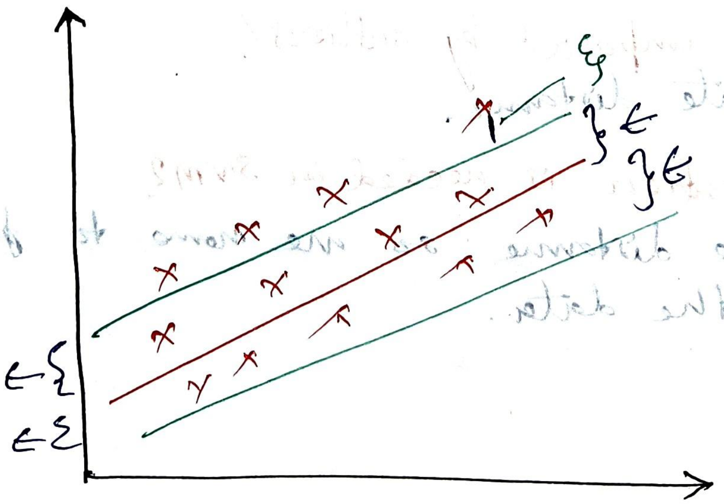
constraint

$$\boxed{|y_i - w_i x_i| \le \boxed{\epsilon + |\xi_i|}}$$

diff b/w truth and the predicted point

value should be under the this value



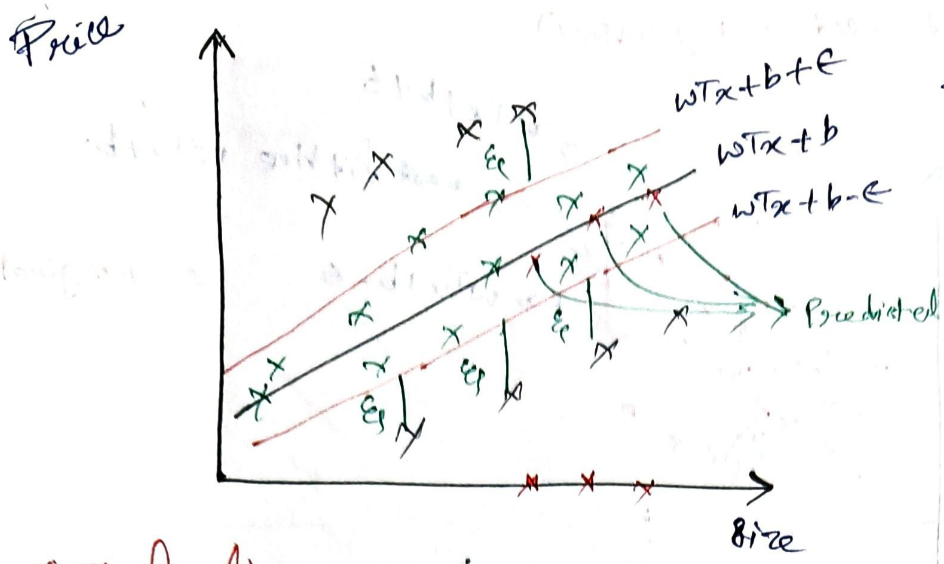


ε = Marginal Error to decide marginal plane

Hyperparameter

→ Keep adjusting ξ_i to get best margin.

→ We can't say incorrect point in regressor.



In Regressor
- No complete incorrect value, because it will be continuous value.

Hyperparameter
- $\epsilon \rightarrow$ Marginal Error

$\xi_i \rightarrow$ Distance of point from marginal plane

Cost Function

$$\min_{w,b} \frac{\|w\|}{2} + \boxed{C \sum_{i=1}^{n} |\xi_i|}$$

$\downarrow$

Hinge loss

Constraint

$$\boxed{|y_i - w_i x_i| \leq \epsilon + |\xi_i|}$$

$\epsilon \rightarrow$ Margin of Error

$\xi_i \rightarrow$ Error above the margin.

Que Will SVM get impacted by outliers?

Ans Yes, it calculate distance.

Que Does standardization is needed in SVM?

Ans Yes, it involve distance so we have to perform scaling of the data.