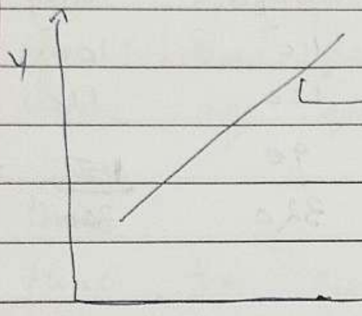
SUPPORT VECTOR MACHINES (SVM)

* SVM can solve both Classification & Regression Problem.

① Classification ⎫ → SVC → Support Vector Classification
② Regression ⎭ → SVR → Support Vector Regression



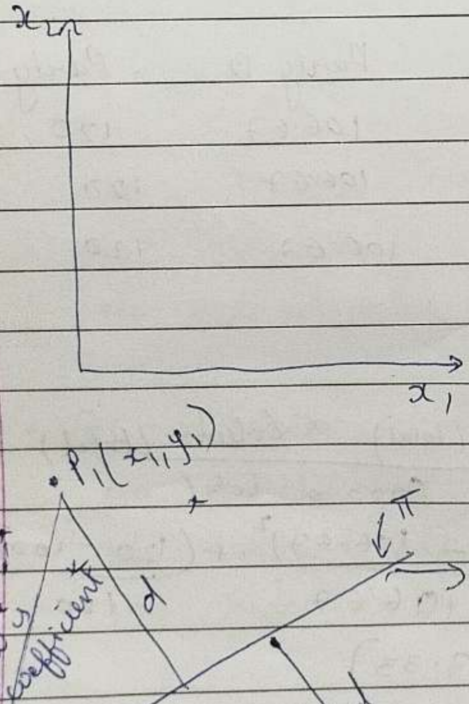
$\hookrightarrow y = mx + c$

$y = \beta_0 + \beta_1 x$

$\Downarrow$

$ax + by + c = 0 \Rightarrow \boxed{y = \left(-\frac{a}{b}\right) x - \left(\frac{c}{b}\right)}$ → Intercept

coefficient



$ax_1 + bx_2 + c = 0$

$w_1 x_1 + w_2 x_2 + b = 0$

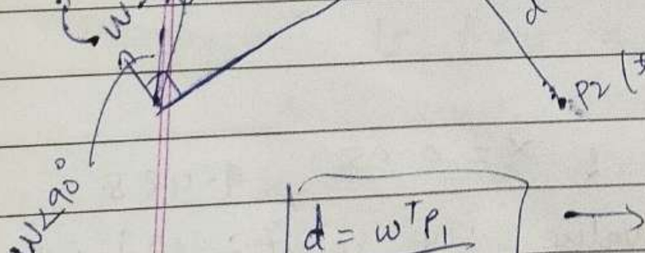
$w^T x + b = 0$

Multiplication

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

w^T → Transpose

$\boxed{w^T x = 0}$ ⟹ Eqn. of line passing through origin



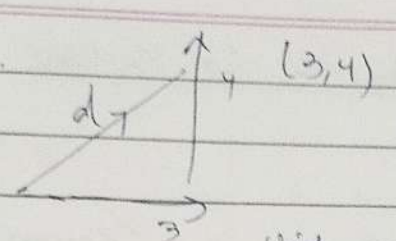
→ This can be plane or a line.

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

→ Dividing by $|W|$ to get unit vector.

Unit Vector → Vector which has magnitude of 1.

Eg:-



Hypotenuse = $\sqrt{9+16} = \sqrt{25}$

$d = 5$

Unit Vector → $\hat{d} = \frac{d}{|d|}$

to make it unit vector $= \left(\frac{3}{5}, \frac{4}{5}\right)$

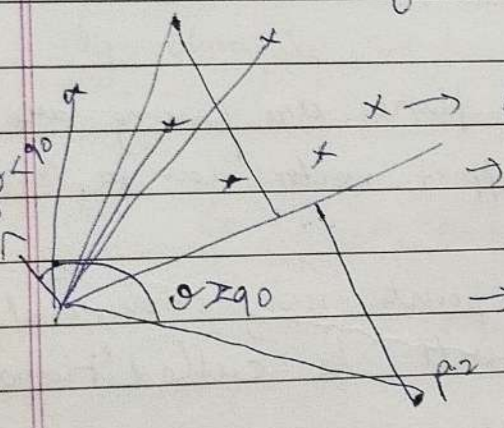
$= \sqrt{\frac{9}{25} + \frac{16}{25}}$

$= \sqrt{1} = 1$

$$d = \frac{w^T p_1}{|w|} \Rightarrow |w||p_1| \cos\theta$$

(Magnitude of w (|w|))

If $\theta < 90^\circ$ value will be +ve.



x → Points

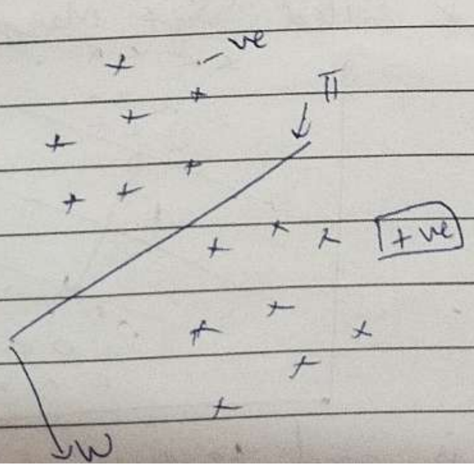
→ If points are above the plane distance is +ve.

→ For points below the plane distance value is -ve.

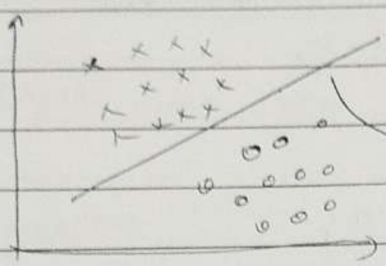
$$d = \frac{w^T p_2}{|w|} = |w||p_2| \cos\theta \Rightarrow \boxed{-ve}$$

S	A
T	C

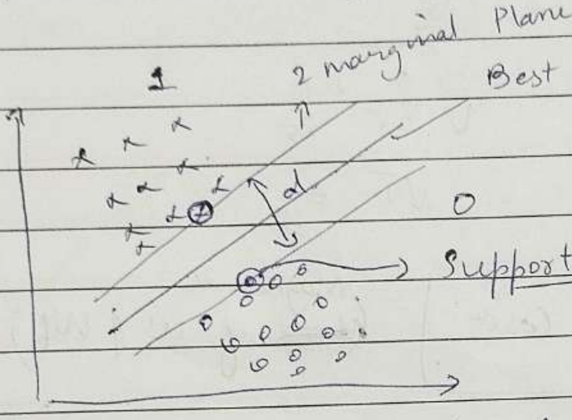
Scenario.



Geometric Intuition Behind Support Vector Machine



Aim of Logistic Regression
→ To get Best fit line



Support Vector Classifier

Here d should be max

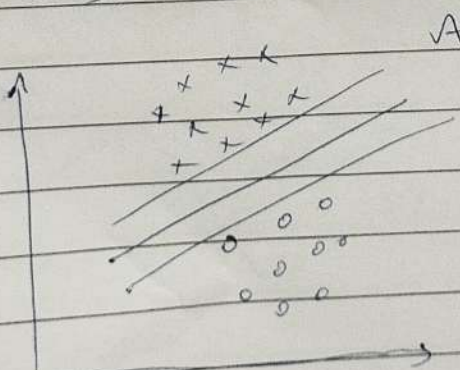
Aim → Along with the best fit line there are 2 marginal planes where d would be maximum.

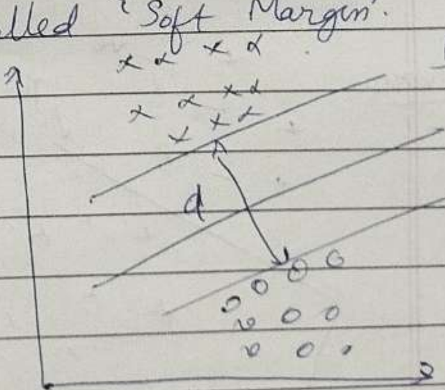
→ Point through which 2 marginal planes are passing are support vectors. More than 1 support vector can be there.

→ Marginal planes are equidistant.

→ If we clearly separate all the points using marginal planes, best fit line and there are no errors it will be called 'Hard Margin'.

→ If there are some errors b/w marginal planes & best fit lines ~~then~~ it will be called 'Soft Margin'.

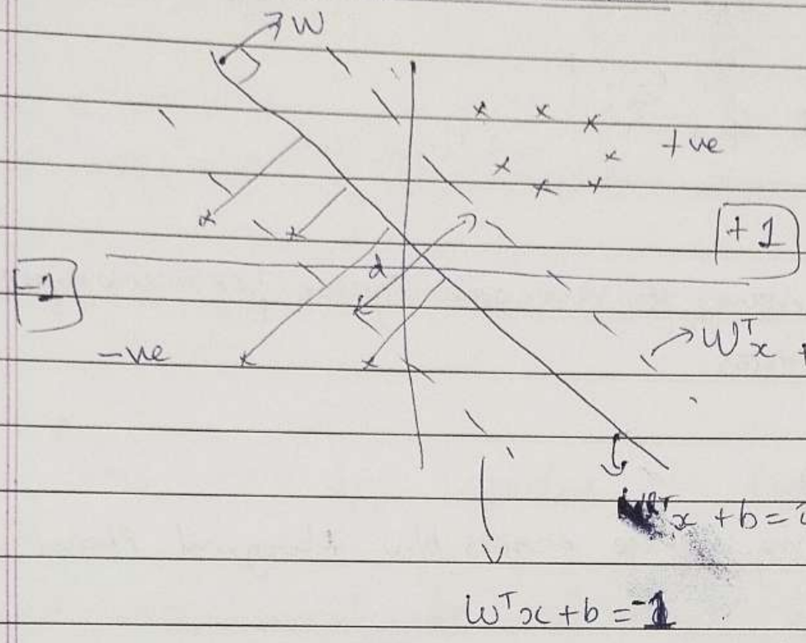




B is good because d is greater.

PAGE

(2) SVM Mathematical Intuition



d ⇒ Maximum

$w^Tx_1 + b = +1$

$-\ (w^Tx_2 + b = -1)$

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

~~unit vector we divi~~

To get magnitude of w

Cost function

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

by changing w, b

o/p → condition for all correct classified points

Constraint such than $y_i \begin{cases} 1 & x^Tx + b \geq 1 \\ -1 & x^Tx + b \leq -1 \end{cases}$

* for linear algebra -1 & +1 fits best

For all corret points,

$\boxed{\text{Constraints} \rightarrow y_i \times (w^Tx + b) \geq 1}$

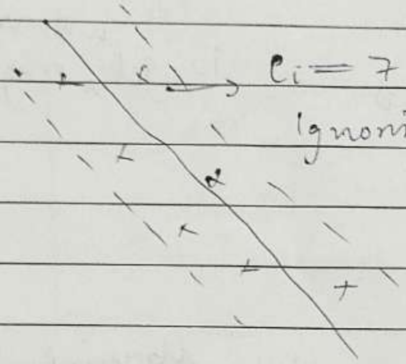
$y_i \rightarrow$ Truth points

$\underset{w,b}{\text{Maximize}}\ \frac{2}{|w|} \Rightarrow \boxed{\underset{w,b}{\text{Minimize}}\ \frac{|w|}{2}}$ Because for ever Cost fn. we are trying to minimize everyw

Cost function Soft Margin $\xi \Rightarrow$ Etta

$$\min_{w,b} \frac{|w|}{2} + \boxed{C_i \sum_{i=1}^{w} \xi_i} \Rightarrow \text{Hinge loss}$$

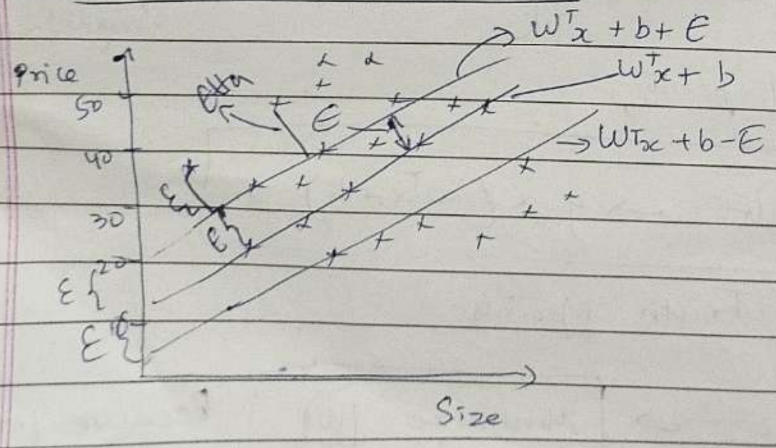
$C_i \rightarrow$ How many pt. we can ignore for misclassification
It is hyperparameter.



$\rightarrow C_i = 7$
Ignoring these errors b/w Marginal Planes

$\xi \rightarrow$ Etta is another hyperparameter.
Summation of the distance of incorrect data points from the marginal Plane.

② SUPPORT VECTOR REGRESSION $\epsilon \rightarrow$ Epsilon (Marginal Error)



Cost fn

$$\min_{w,b} \frac{|w|}{2} + \boxed{C_i \sum_{i=1}^{n} \xi_i} \Rightarrow \text{Hinge loss}$$

Constraint

$$|y_i - w_i x_i| \le \epsilon + \sum_{i=1}^{n} \xi_i$$

Diff b/w Truth & Predicted Pt.

$\xi_i \rightarrow$ Distance b/we pt & marginal plane.

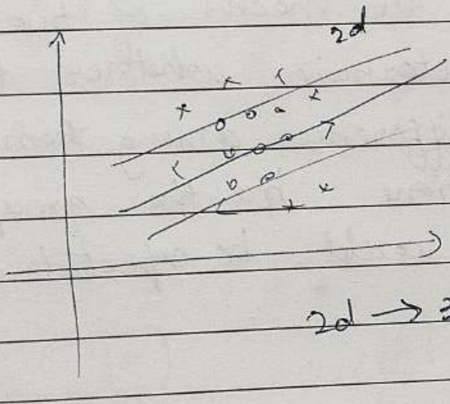
Q. Will SVM get Impacted by Outliers.

Ans Yes

Q: Do we need to normalize / Standardize data.

Ans Yes

* SVM Kernel



→ we would not be able to solve this problem using SVC or SVR.

→ Here we use SVM Kernel

2d → 3D