Support Vector Machine (SVM)

Reg.

- Supervised learning Algo.
- Classification Problems O/P is categorized value
- " $y = f(x)$

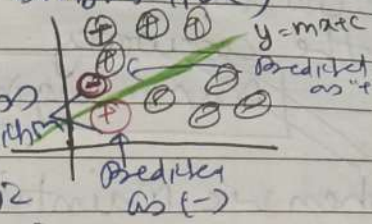
dependent variable → Unknown value → Categorized Data

x → independent variable

→ { Mathematical functions / Hyper Planes }

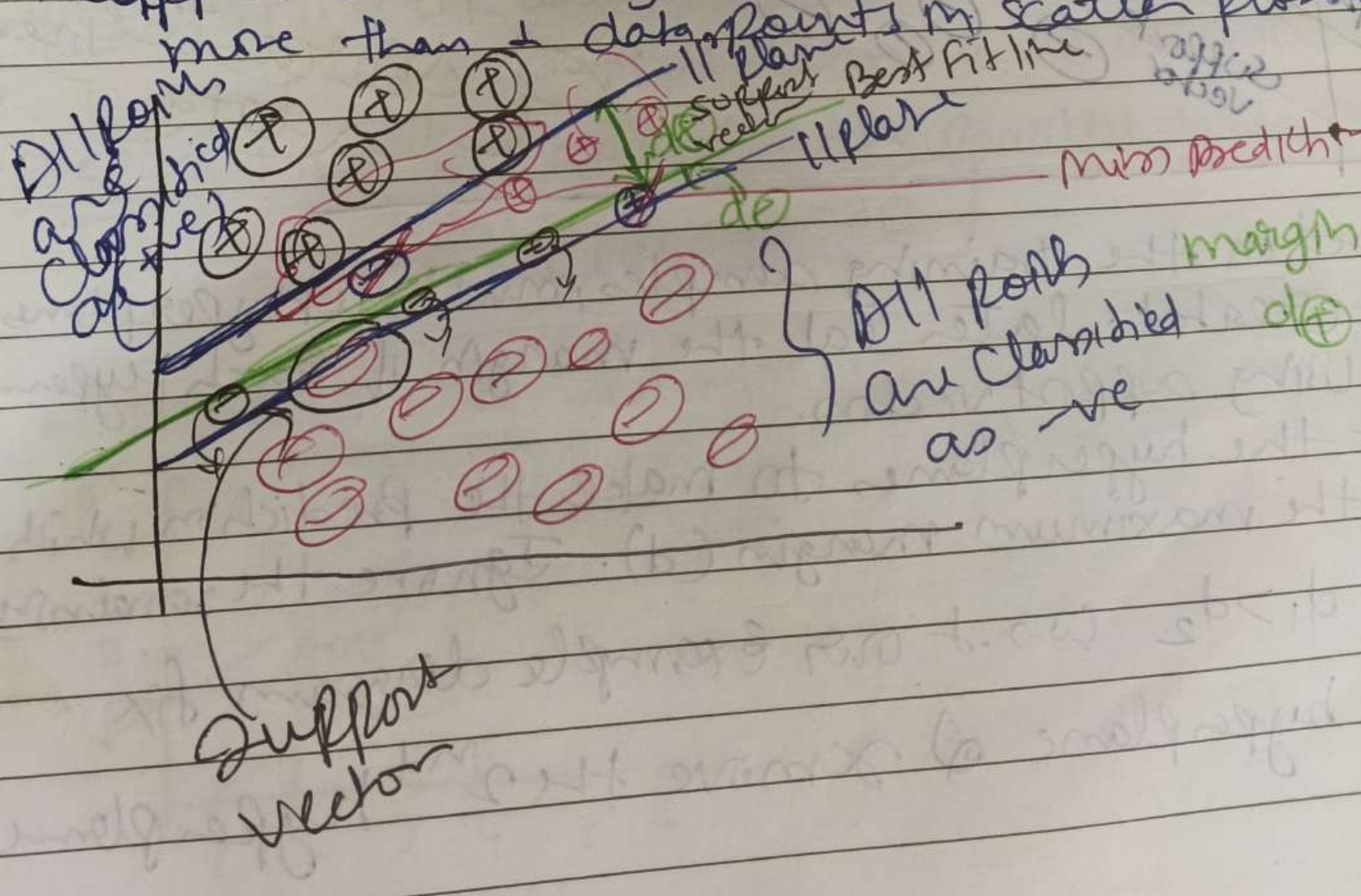
- SVM is expansion of a logistic Regression used to classify the data points with more accuracy by creating multiple hyperplanes with Maximum Margin

→ In the logistic Regression data points classified on the single hyper plane (single straight line)
Accuracy is low (y=mx+c)



(5) To handle the above issue, improvement is required in the logistic Regression classifier called SVM Classifier
(6) In SVM Classifier, two parallel planes are created by connecting the +ve points & -ve points respectively w.r.to original hyperplane
(7) To which +ve points the parallel plane is moving through points are called as the support vectors of the plan.
Similarly, to which -ve points ∥^el planes is moving & points are called as the support vectors of plane

→ Support vectors may be a single data point (or) more than 1 data points in scatter plot.



$\frac{2}{||w||}$ max & $y_i(w^Tx+b) > 0$

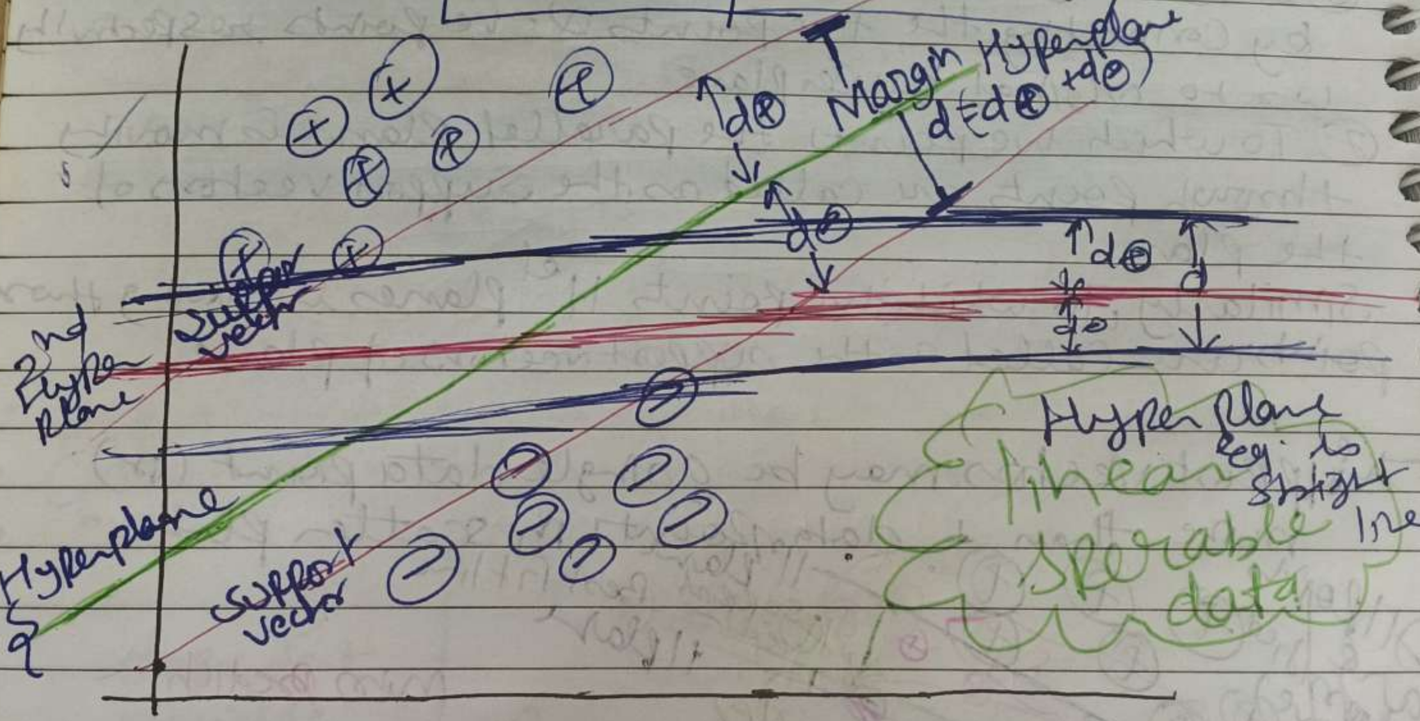
$w^Tx+b>0$ $\hat{y}=1$ (true), $\hat{y}=-1$

Goal of SVM

→ Goal of SVM is "Maximize the margin(d)" to improve the accuracy.

$$\boxed{\text{Max margin}(d) = d_{\oplus} + d_{\ominus}}$$

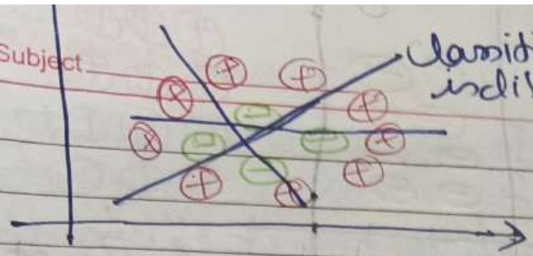
· When we maintain the max margin then miss prediction are less ∴ Model become a generalized model (Good model) i.e, $\boxed{\text{Low Bias \& Low Variance}}$



· Based on the training samples, multiple hyperplane are created. later cal. the margin of each hyperplane using support vectors.

· Select the hyperplanes to make the prediction which having the maximum margin (d). Ignore the remaining

eg. If $d_1 > d_2$ w.r.t our example diagram fix first hyperplane & remove the 2nd hyperplane



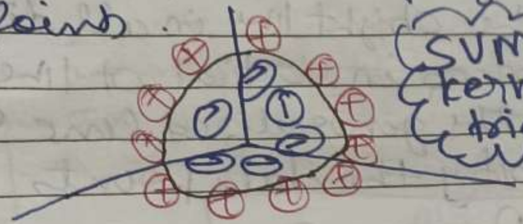
Classification is difficult

{Nonlinear sperable data}

Hyperplane is not a straight line eq.

- When the data is Nonlinear separable then SVM kernal trick is used to convert the lower dimension data into a higher dimension. So, Nonlinear polynomial is used to classify data points.

ie,



{SVM kernal trick} → 2D data 3D etc

SVM Algoritham [Maths in SVM]

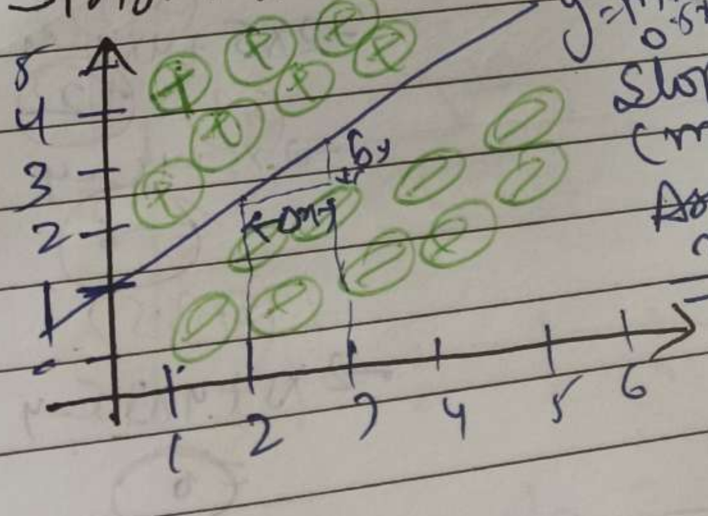
Step1 → Find the hyperplan w.r.t the training samples {Here, hyperplane eq. is straight line eq. Becoz data is assumed as 2D with linear separable}.

Step2 → Find the distance b/w the Hyperplane and the +ve & -ve support vectors ie d_{+} & d_{-} calcul

Step3 → Find the margin b/w the parallel planes ie,

$$\boxed{\text{Margin}_{(d)} = d_{+} + d_{-}}$$

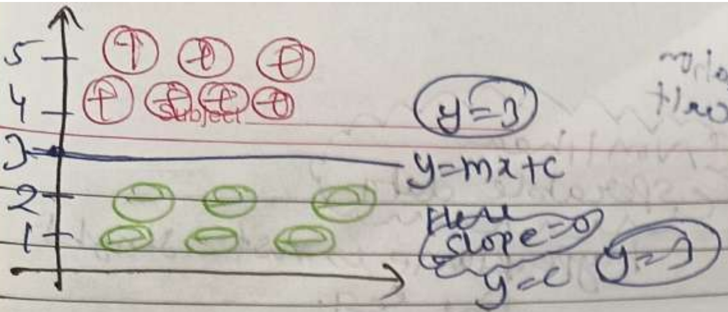
- Stright line eq. used as a hyperplane eq. ie y

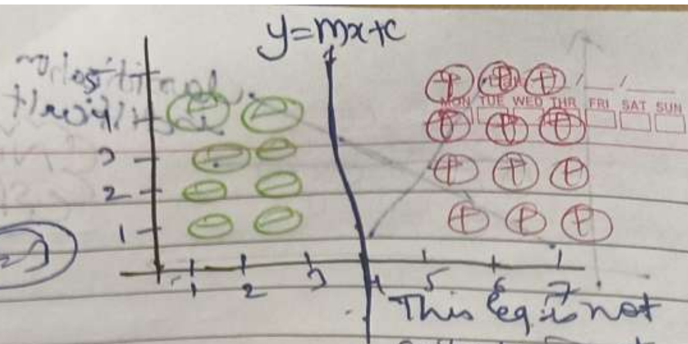


$y = mx + c$
$0.5x + 1$

Slope (m) $= \frac{\Delta y}{\Delta x}$

Assume,
$m = 0.5$





This eq is not suitable is not straight line

$\boxed{\text{slope}(m) \cong \infty}$

• When we draw the straight line in all directions, y=mx+c eq. is not suitable becoz in vertical st-line slope is approach to '∞' so, use the generalized line eq. to draw the hyperplane, to classify the data points

$$\boxed{Ax+By+c=0}$$

A & B are the co-cofficients of the co-ordinates

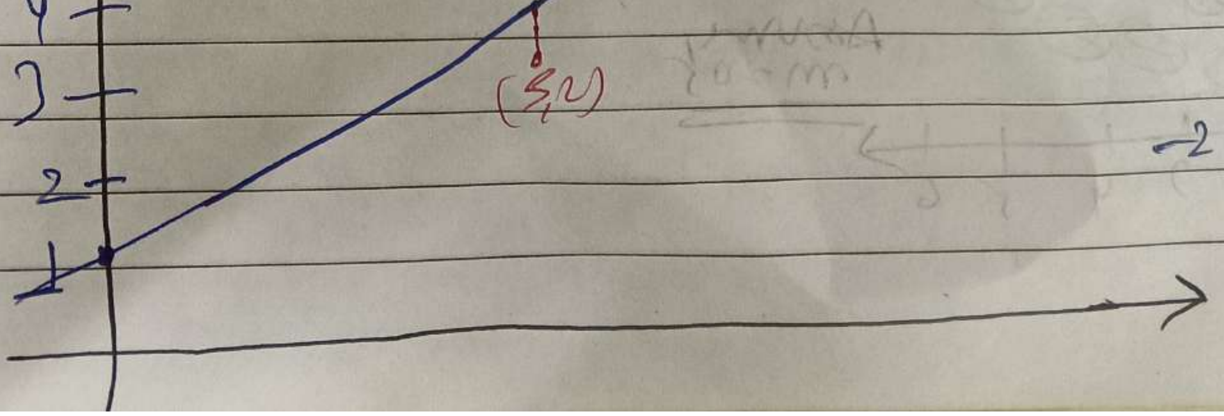
intercept

Generalized line eq. is $Ax+By+c=0$

$$By = -Ax - c$$

$$y = \underbrace{-(A/B)}_{\text{slope}}x - \underbrace{c/B}_{\text{intercept}}$$

Assume $(-2x+4y-4=0)$



(5,4)
$-2\times5+4\times4-4$
$+2$

(5,2)
$-2\times5+4\times2-4$
$=-6$

(5,3.5)
$-2\times5+4\times3.5-4$
$=0$

... the plane (the)

... then we can say that data

When the line eq. O/P is ... then we can say that data point in on top of the plane so, classify the point as ...

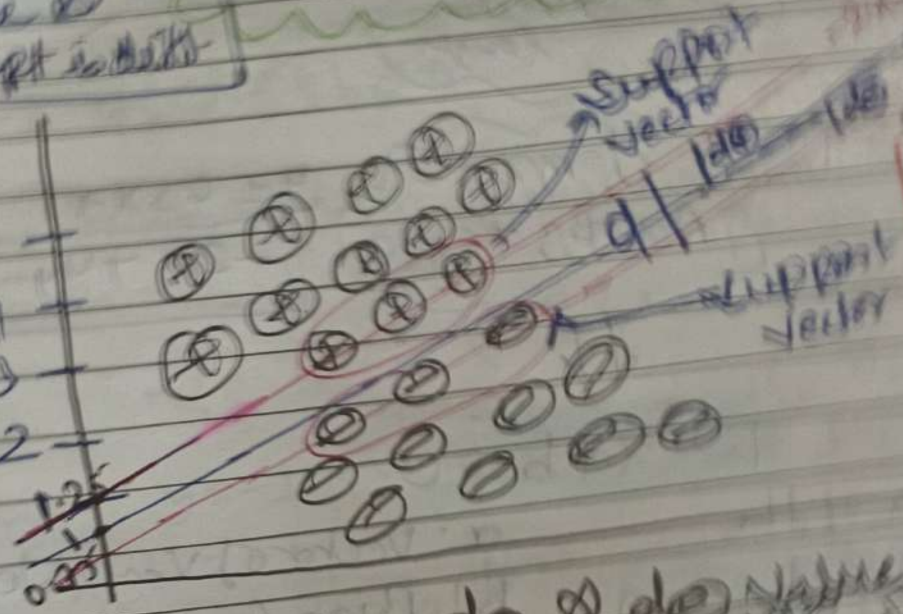
When the line eq. O/P is -ve then we can say that data point is on bottom of the plane so, classify the point as ...

$2x+4y-4=0 \longrightarrow$ original plane

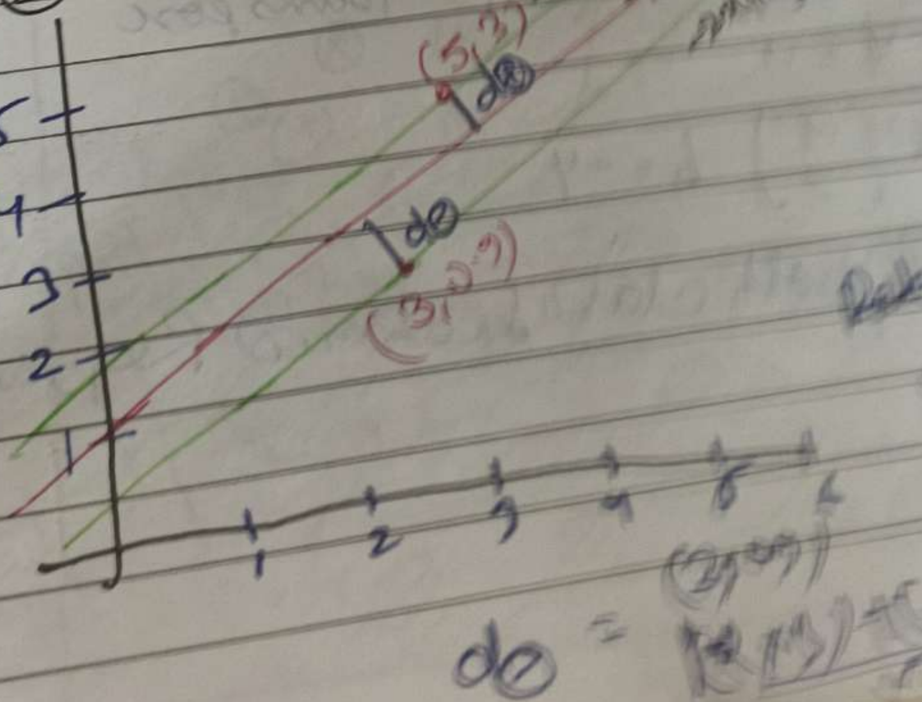
$2x+4y-4=+1 \longrightarrow$ +ve plane

$2x+4y-4=-1 \longrightarrow$ -ve plane

$\boxed{\text{Same value of slope \& intercept is diff}}$



Step 2 Find the d_1 & d_2 value



$d_2 =$

Subject

Step 3 Find the Margin b/w the ||el planes

$$\text{Margin}(d) = \frac{|C_2 - C_1|}{\sqrt{A^2+B^2}}$$ $w^T x + b$

$-2x+4y-4 = 1$ $-2x+4y-4 = -1$

$-2x+4y-5=0$ $-2x+4y-3=0$

$C_1 = -5$ $C_2 = -3$

Max

$$\frac{2}{\|w\|} = \frac{|C_2 - C_1|}{\sqrt{A^2+B^2}} = \frac{2}{\|w\|} = \frac{2}{\|w\|}$$

vector length

Straight line eq.

$y = mx + c$ $y = 0.5x + 1$

$Ax + By + c = 0$ $\boxed{-2x+4y-4=0}$

$$\Downarrow$$

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

w: It a Vector of the Co-efficients of a Co-ordinates ie $w: \begin{bmatrix} A \\ B \end{bmatrix}$

x: Vector of Variable

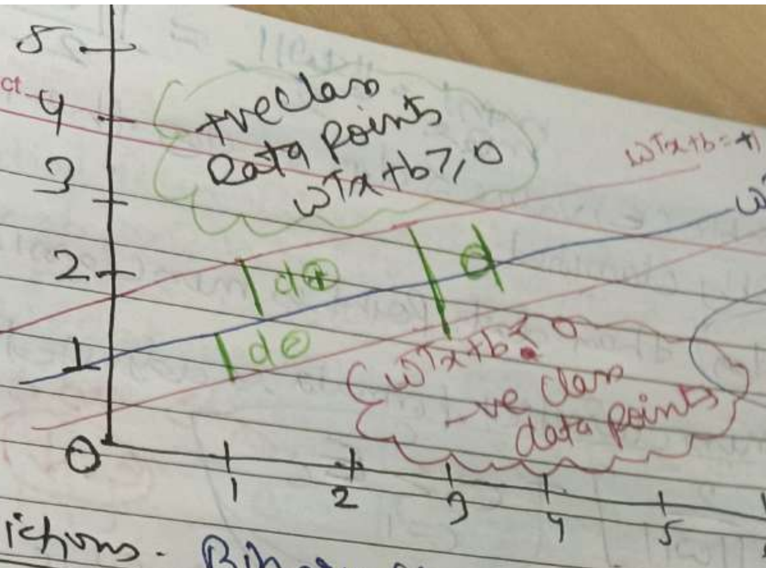
b: Bias (intercept)

T: Transpose

$-2x+4y-4=0$

$w = \begin{bmatrix} -2 \\ 4 \end{bmatrix}$ $x: \begin{bmatrix} x \\ y \end{bmatrix}$ $b = -4$

) SVM Can be used both classification & Regression Problem



$$d = \frac{2}{\sqrt{A^2+B^2}} = \frac{2}{||\omega||}$$

· Predictions - Binary Class Prediction

Class label $\left[\begin{matrix}1\\-1\end{matrix}\right.$

$\hat{y} = 1$ if $\omega^T x + b \geq 0$
$\hat{y} = -1$ if $\omega^T x + b \leq 0$

$\hat{y} \left[\begin{matrix}1 & +ve\\-1 & -ve\end{matrix}\right.$

SVM uses Soft Margin technique for outliers or non-linear separable data

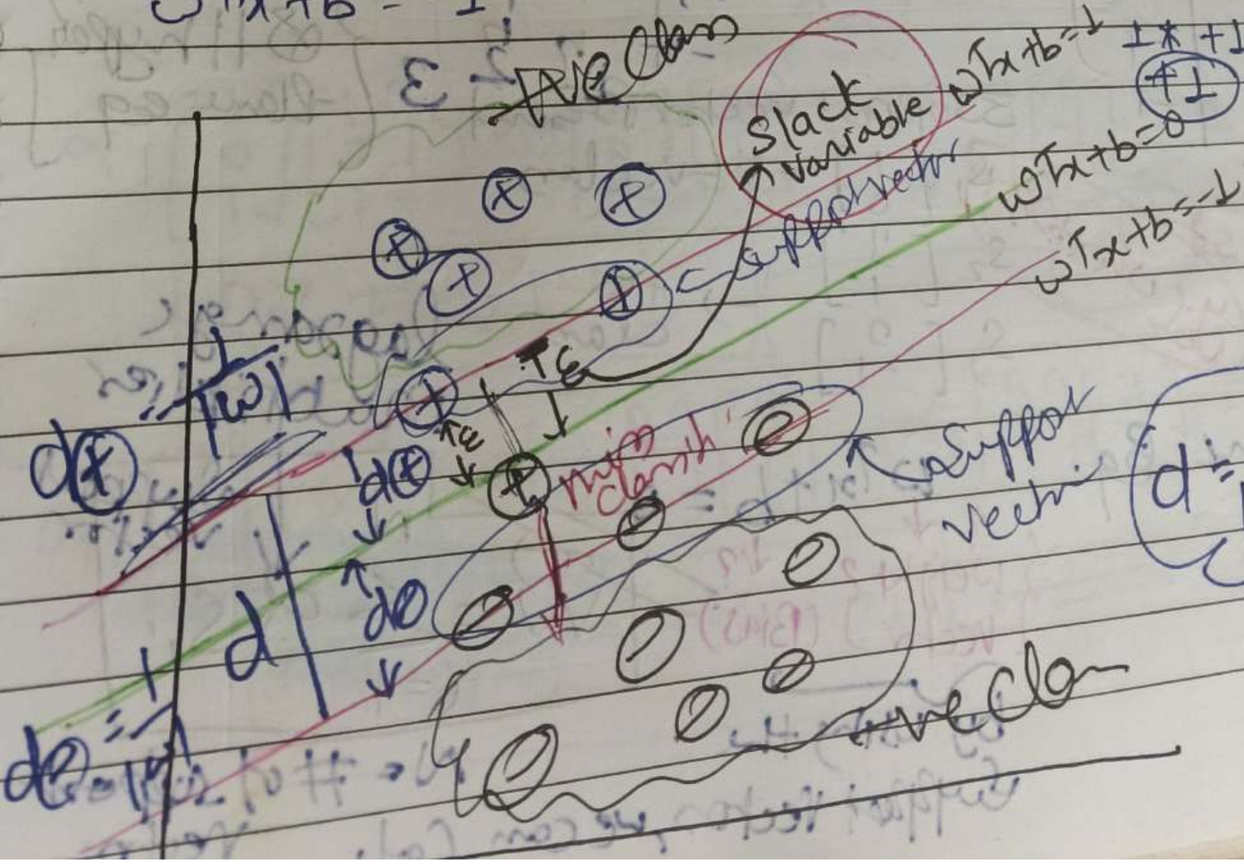
Assume that, training data set contain 4 samples with first 2 samples as +ve class & next 2 samples as -ve class.

$y = \{1, 1, -1, -1\}$

Prediction Model
$y_i(\omega^T x + b) \geq 1$

1^{st} point $y = 1$ +ve class
$\omega^T x + b = +1$
$1 * +1$
$+1$

3^{rd} point $y = -1$ -ve class
$\omega^T x + b = -1$



$$d = \frac{2}{||\omega||}$$

$min \frac{1}{2} \frac{||w||}{2} = \frac{||w||^2}{2}$

① If slack variable (ϵ) value $\leq d_{+}$ or d_{-} than data points are correctly classified.

② $\epsilon > d_{+}$ or d_{-} than data point is misclassified.

To allow the misclassified formula is adjusted as

$$\boxed{Max\left[\frac{2}{||w||}\right] + C\sum_{i=1}^{n} \epsilon_i}$$

as $C \uparrow$ $\epsilon_i \downarrow$

C : # of misclassifications

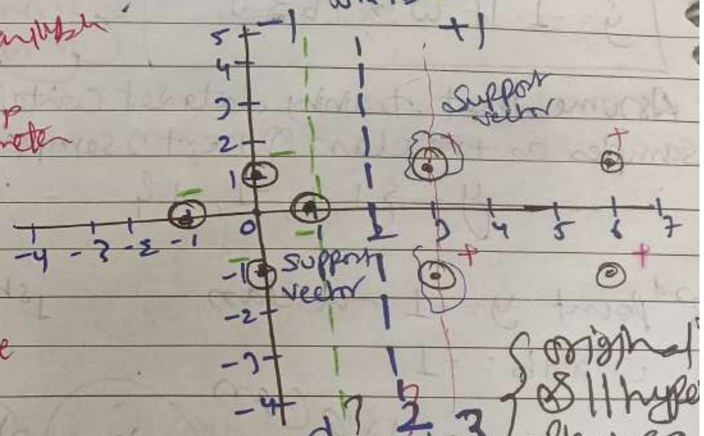
ϵ : Slack Variable (Distance b/w the data points & its ||el plane)

$$\boxed{y_i(w^Tx+b) \geq 1-\epsilon_i}$$

Sample Data → i/p variables

x_1	x_2	y
3	1	1
3	-1	1
6	1	1
6	-1	1
1	0	-1
0	1	-1
0	-1	-1
-1	0	-1

y: o/p parameter; first four rows: +ve; last four rows: -ve



{ original & || hyperplane eq }

3 Support vector present

$S_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ → -ve Class

$S_2 \begin{bmatrix} 3 \\ 1 \end{bmatrix}$, $S_3 \begin{bmatrix} 3 \\ -1 \end{bmatrix}$ → +ve Class

$min \frac{1}{2}||w||^2$ s.t. $y_i(w^Tx+b) \geq 1$

Hyperplane Eq.

$$\underbrace{w}_{\text{weight vector}}{}^T x + \underbrace{b}_{\text{(Bias)}} = 0$$

by using the support vectors, we can cal.

lagrange multiplier

$$\boxed{W = \sum_{i=1}^{N} \alpha_i S_i}$$

Support vectors

N = # of support vec

• Initial Bias value is '1'. If Bias is '0' then plane moves through origin '0' So, this plane doesn't classify the data points clearly.

Support vectors with Bias are

$S_1 : \begin{bmatrix}1\\0\end{bmatrix}$ Adding Bias $\begin{bmatrix}1\\0\\1\end{bmatrix}$ → -ve class Support vector

$S_2 : \begin{bmatrix}3\\1\end{bmatrix}$ Bias $\begin{bmatrix}3\\1\\1\end{bmatrix}$ → +ve class Support vector

$S_3 : \begin{bmatrix}3\\-1\end{bmatrix}$ $\begin{bmatrix}3\\-1\\1\end{bmatrix}$ → +ve class Support vector

• Here, 3 Eq. are formed w.r.t the above support vectors $(\vec{S_1}, \vec{S_2}, \vec{S_3})$. Each support vector contain the weight (w)

- 'α' is language multiplier. It is '0' when the data Point is not a support vector. It is Non zero when the Data point is a support vector.

Here $\vec{S_1} \rightarrow \alpha_1$ -ve class $S_1 = \alpha_1 \vec{S_1}\vec{S_1} + \alpha_2 \vec{S_1}\vec{S_2} + \alpha_3 \vec{S_1}\vec{S_3} = -1$

$\vec{S_2} \rightarrow \alpha_2$ +ve $S_2 = \alpha_1 \vec{S_2}\vec{S_1} + \alpha_2 \vec{S_2}\vec{S_2} + \alpha_3 \vec{S_2}\vec{S_3} = +1$

$\vec{S_3} \rightarrow \alpha_3$ +ve $S_3 = \alpha_1 \vec{S_3}\vec{S_1} + \alpha_2 \vec{S_3}\vec{S_2} + \alpha_3 \vec{S_3}\vec{S_3} = +1$

$$\alpha_1 \begin{bmatrix}1\\0\\1\end{bmatrix}\begin{bmatrix}1\\0\\1\end{bmatrix} + \alpha_2 \begin{bmatrix}1\\0\\1\end{bmatrix}\begin{bmatrix}3\\1\\1\end{bmatrix} + \alpha_3 \begin{bmatrix}1\\0\\1\end{bmatrix}\begin{bmatrix}3\\-1\\1\end{bmatrix} = -1$$

$$\boxed{2\alpha_1 + 4\alpha_2 + 4\alpha_3} \quad \text{(1)}$$

$$\alpha_1 \begin{bmatrix}3\\1\\1\end{bmatrix}\begin{bmatrix}1\\0\\1\end{bmatrix} + \alpha_2 \begin{bmatrix}3\\1\\1\end{bmatrix}\begin{bmatrix}3\\1\\1\end{bmatrix} + \alpha_3 \begin{bmatrix}3\\1\\1\end{bmatrix}\begin{bmatrix}3\\-1\\1\end{bmatrix} = 1$$

$$\boxed{4\alpha_1 + 11\alpha_2 + 9\alpha_3 = +1} \quad \text{(2)}$$