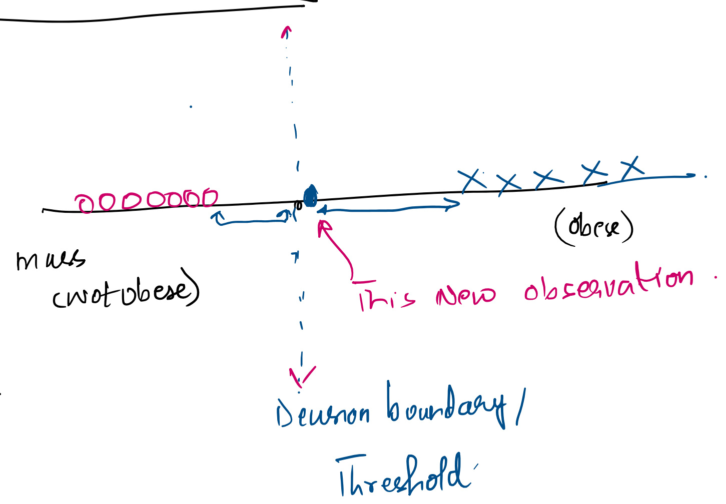
SVM [Support Vector Machine]

Consider >



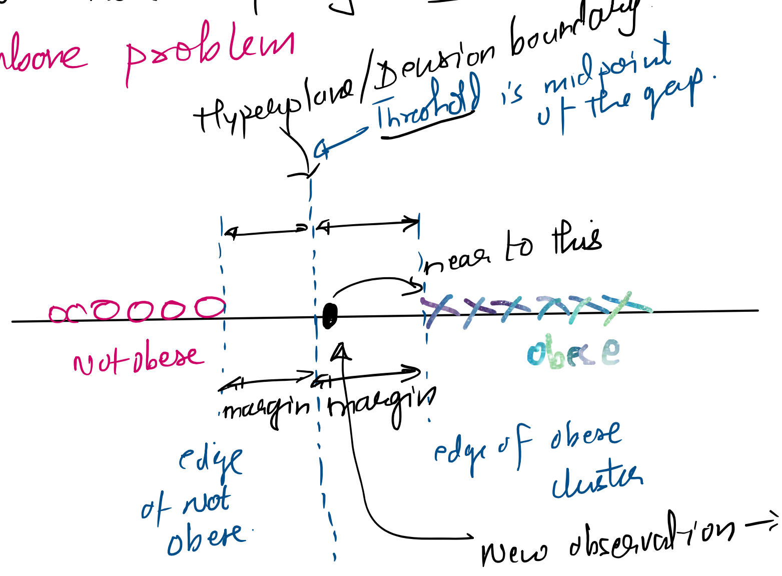
* Here the new observation will be classified as obese.
Eventhough the new observation is more near to Not obese as compared to obese.

* The threshold here is pretty lame.

To Solve the above problem

Consider >

We will consider threshold at midpoint of the gap betn the edges of clusters



Here the margin > It is gap betn threshold (Decision Boundary) and the edge of the cluster

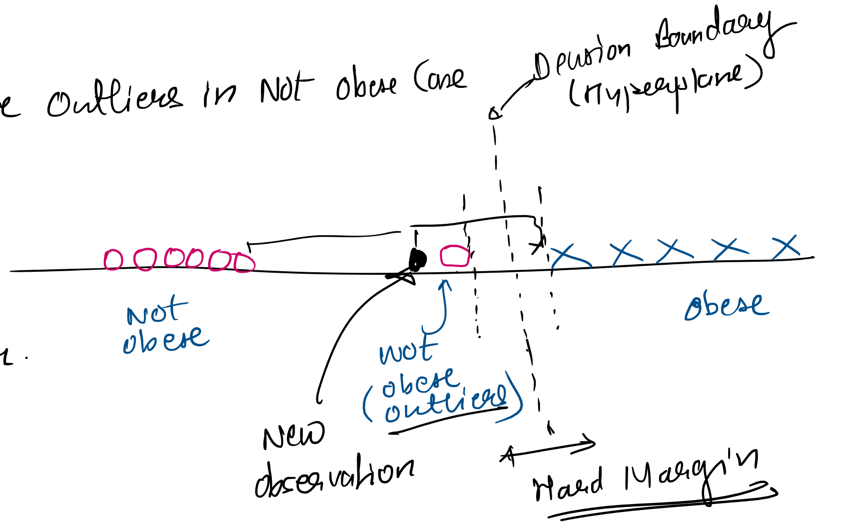
→ Here the margin is maximum.

→ Known as Maximum Margin classifier.

→ Here In above Case there is No Misclassification

→ Also there are no <u>Outliers</u>

<u>Consider</u> → Here we have Outliers in Not obese Case



* If we consider Margin as the edge of the cluster considering the Outlier.

* Here the new observation will be classified as <u>Not obese</u>, eventhough it is near to <u>obese</u>.

(Hard Margin)

* Here the <u>Maximum Margin Classifier</u> is very sensitive to <u>outlier</u>.

* Here Eventhough we have outlier, however all the given data points are correctly classified → <u>low Bias</u>.

* But for the <u>new observation</u> it incorrectly classifies ⇒ <u>High Variance</u>

<u>Note</u> → <u>With Hard Margin</u> →

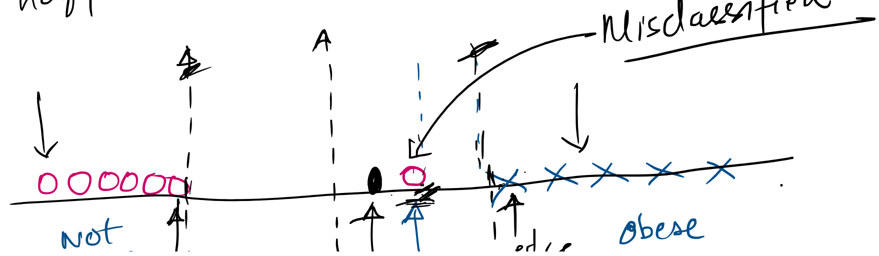
→ <u>Misclassification</u> <u>Not Allowed</u>

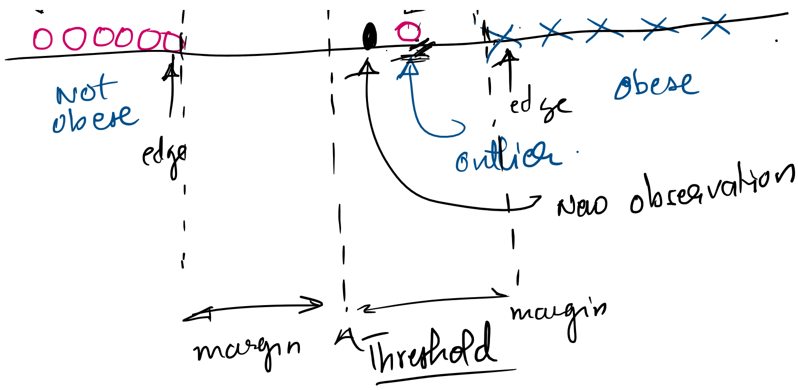
→ All Data points Correctly Classified ⇒ <u>Low Bias</u>

→ New Observation Incorrectly Classified ⇒ <u>High Variance</u>

<u>Now</u>: We can do better (To make the threshold Insensitive to Outlier)

→ For this to happen we must allow Misclassification.





Soft Margin

* Here since while considering edge of cluster, we have allowed Misclassification, so the Margin now known as Soft Margin

Note

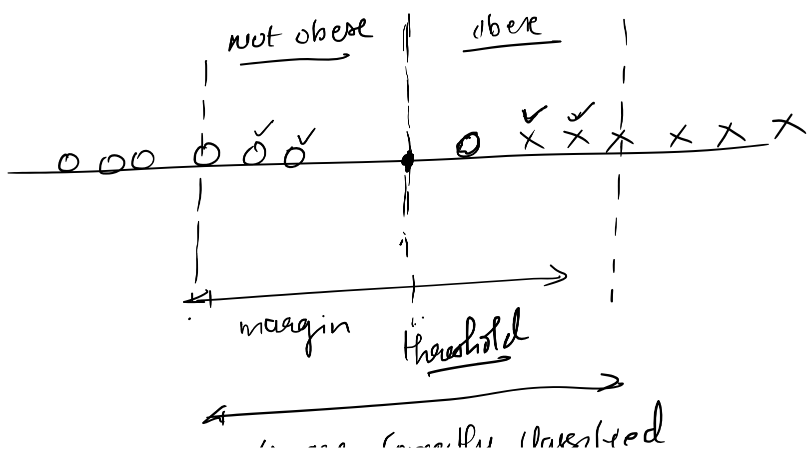
* With Soft Margin ⇒
  - → Misclassification is Allowed
  - → All given data points are not correctly classified (outliers). → High Bias
  - → New observations are correctly classified ⇒ Low Variance

Soft Margin ⇒ When Misclassification is allowed then the distance betn threshold and given observation is known as soft margin.

* ~~For~~ above case we can have many soft margins.

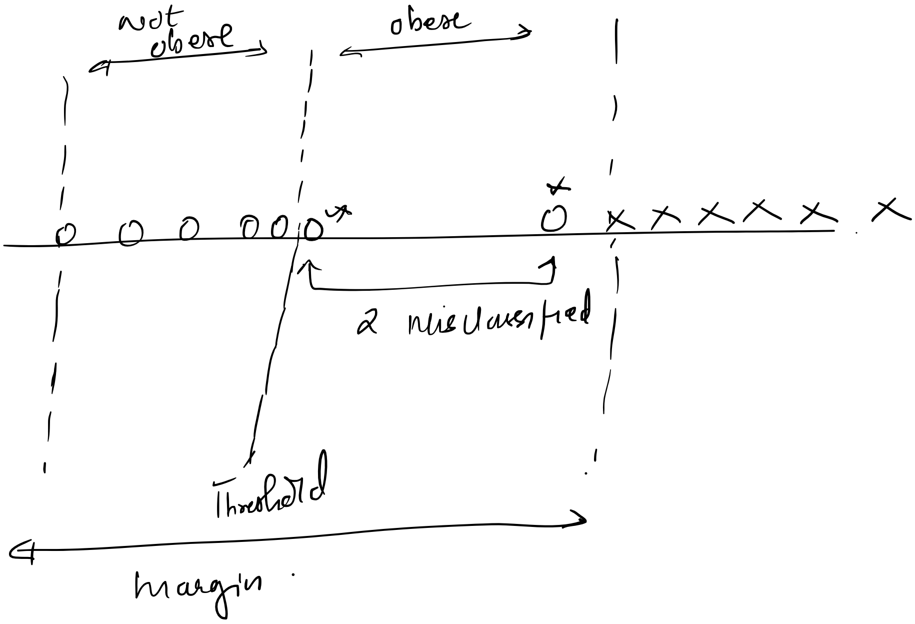
Ex

Consider Case I.



4 are Correctly classified
1 is Incorrectly classified.

Case 2)



* We will determine which soft margin is better by observing how many Misclassification each soft margin Allows.

VImp.

When we use Soft Margin to determine the location of threshold then we are using "Soft Margin Classifier" Also known as "Support Vector Classifier"

* Support Vector Classifier → Main Aim is to create a Decision Boundary with allowed Misclassification

-> It allows small amount of Overlapping

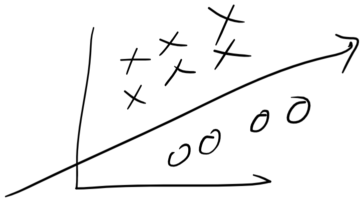
But if the Dataset is as follows →

* Here there is very high level of overlapping

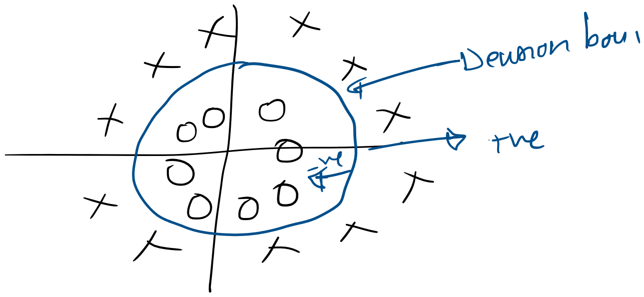
* Here there is very high level of overlap

* Here we cannot have linear separator.

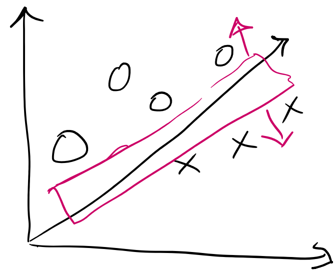
Also → [For 1-D data Decision Boundary is a point
For 2-Data " " is a line
For 3-D Data " " is a plane.]



Also



Here also Linear Decision Boundary Not Possible.

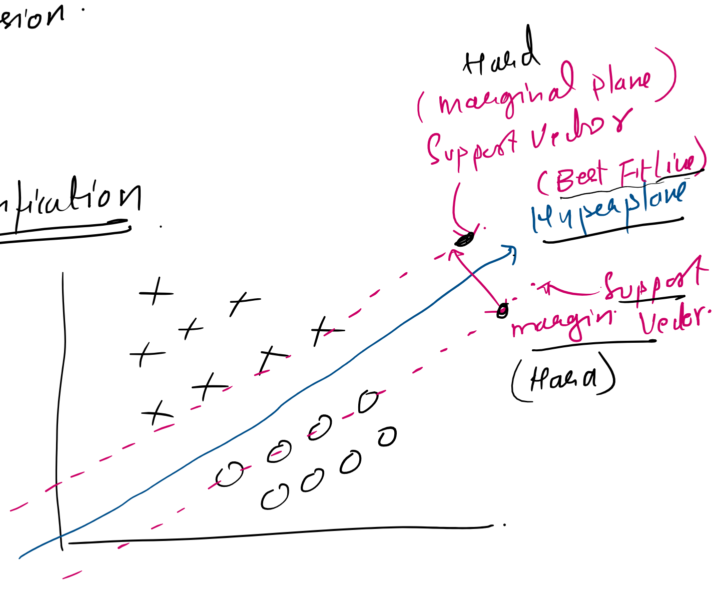


Note

SVM → ML Algorithm
→ Supervised
→ Task → Classification / Regression.

* Support Vector Classifier for Classification.

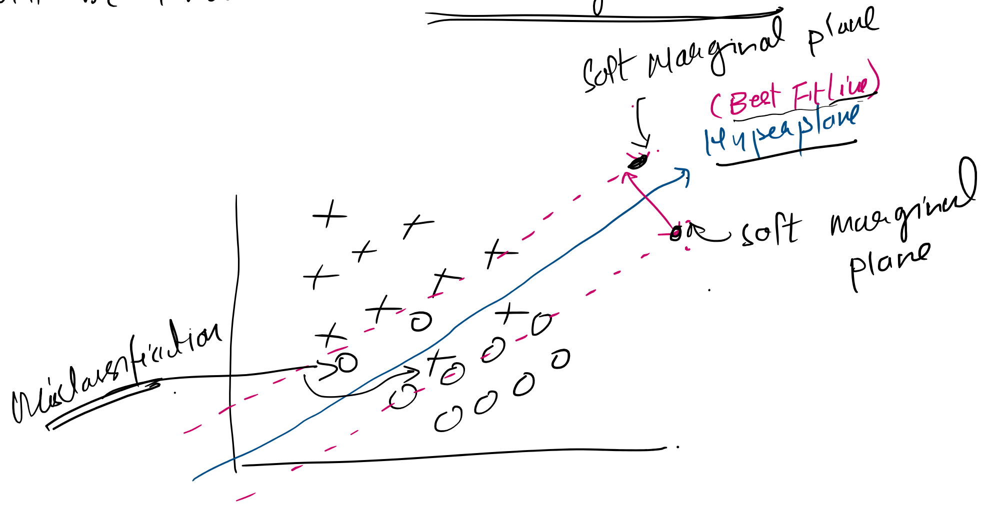
* Consider Binary Classification →



AIM →
Create a Hyperplane and Marginal plane such that the Margin is Maximum →.

* If we get perfect Marginal plane & Hyperplane [No Misclassification] then we can linearly separate the Data points.
This Margin will be known as Hard Marginal Plane

* If there are Misclassification allowed then the planes will be known as "Soft Marginal Plane"



Note Consider Hyperplane ⇒ Here 2-D Data
Hyperplane will be line.

Eq:

$$y = mx + c$$

or $y = \theta_1 x_1 + \theta_0$

or $y = w_1 x_1 + b$

For multilinear Regression ⇒ Many features $x_1\ x_2\ x_3 \dots x_n$

Each feature will have weight $w_1\ w_2\ w_3 \dots w_n$

$$y = w_1 x_1 + w_2 x_2 + w_3 x_3 + \dots w_n x_n + b.$$

$$y = w^T X + b$$

$w^T = [w_1\ w_2 \dots w_n]$

$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$

actual value (y)

predicted ($w^T X$)

Consider ⇒

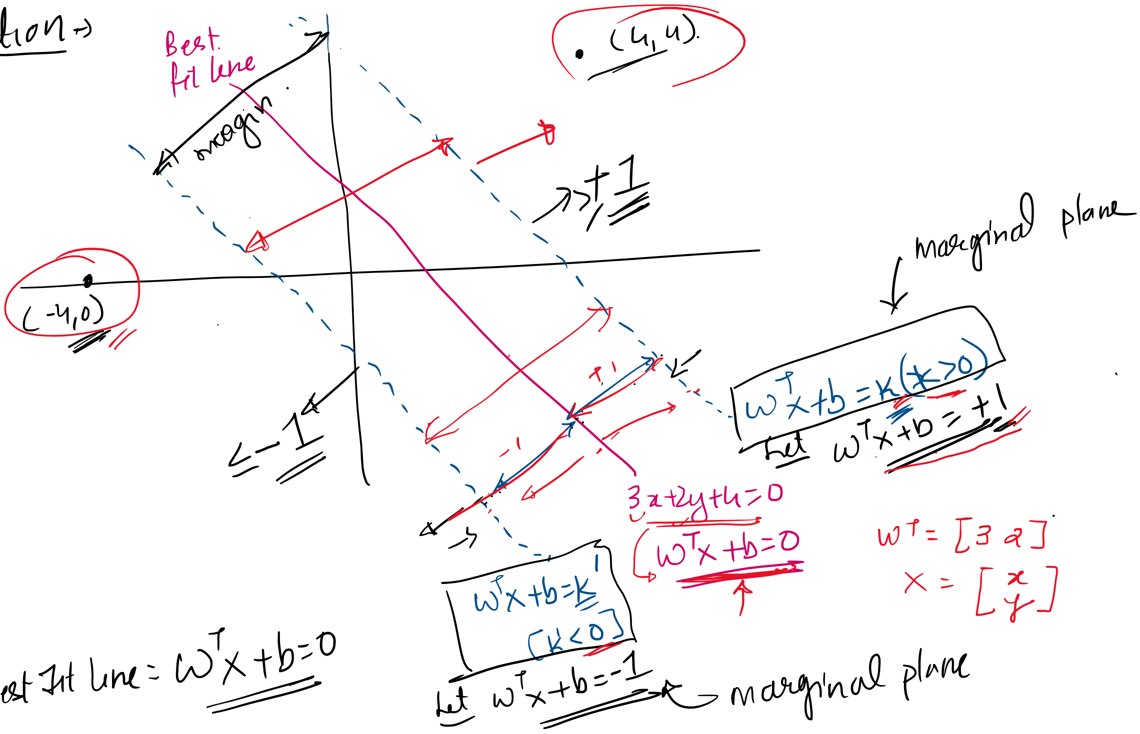
SVM For Classification →

Consider

$Eq^n \Rightarrow 3x+2y+4=0$

put $(-4,0) = 3(-4)+2(0)+4$
$= -8 < 0$

$(4,4) = 3(4)+2(4)+4$
$= 24 > 0$



The Hyperplane / Best fit line $= w^Tx+b=0$ Let $w^Tx+b=-1$ ← marginal plane

Let the Marginal plane be $w^Tx+b=+1$ [on +ve side]
And $w^Tx+b=-1$ [on -ve side]

Our Aim is to draw two Marginal planes (+ve & -ve side) and need to Ensure the distance (Margin) is Maximum.

* We want to find distance betn the Marginal Planes.

Lets find difference :

$$\begin{aligned} w^Tx_1 + b &= +1 \\ -\ w^Tx_2 + b &= -1 \\ \hline w^T(x_1-x_2) &= 2 \end{aligned}$$

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

Here w = slope (coefficient) < magnitude, direction } Vector.

To Convert w^T into Vector ie $\vec{w}^T$

To Convert w^T into Vector i.e $\vec{w}^T$

dividide by $||w||$

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

difference (margin) value.

$\therefore$ To Maximize the Margin →

$$\boxed{\underset{(w,b)}{\text{Maximize}} = \frac{2}{||w||}}$$

predicted

Constraints → y_i Actual/observed $= \begin{cases} +1 & \text{when } w^Tx + b \geq 1 \\ & (\text{point lies outside (above) of } w^Tx+b=+1) \\ -1 & \text{when } w^Tx + b \leq -1 \\ & (\text{point lies below of } w^Tx+b=-1) \end{cases}$

The Constraints are for Correctly classified Data point

$\therefore$ Final Constraints for S.V classifier.

$$\boxed{y_i * (w^Tx + b) \geq 1}$$ For Correctly classified Data point

y_i: observed (actual); (w^Tx+b): Predicted value.

$y_i = +1$

$w^Tx + b = +1$

$\therefore$ To Maximize $= \dfrac{2}{||w||}$ subjected to $y_i * (w^Tx + b) \geq 1$

∴ To Maximize = $\frac{2}{||w||}$ subjected to ...

Also can be written as

* To Minimize = $\frac{||w||}{2}$ Subjected to $y_i * (w^T x + b) \geq 1$

only for Correctly Classified Data points.

Here we have not Considered for Misclassification

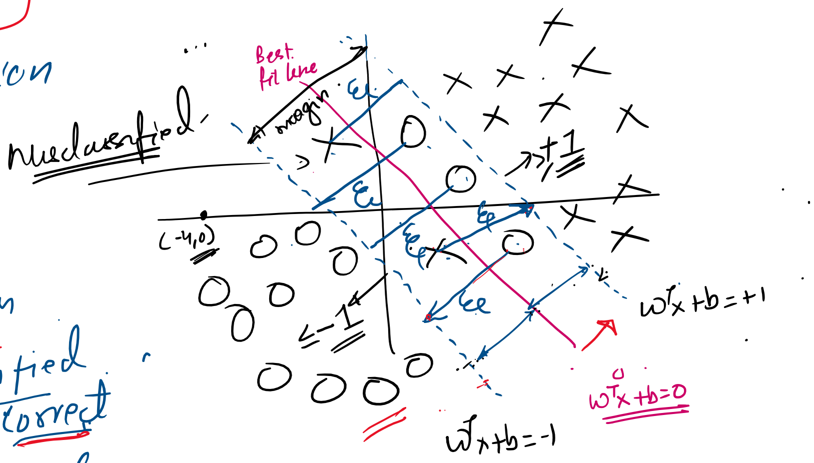
But in real world there will be always Misclassification.

∴ To Consider for Misclassification we have to use

"Hyper parameters"

We allow Misclassification

Now we will use two Hyperparameter



ξ_i ⇒ Eta ⇒ distance betⁿ the Misclassified point and correct Marginal plane

C_i ⇒ No of Allowed Misclassified points

Now Final Cost Fⁿ (For All data points) [Correctly & Incorrectly Classified]

* Objective ⇒

$$\text{To Minimize}_{(w,b)} \; \frac{||w||}{2} + \boxed{C_i \sum_{i=1} \xi_i} \qquad \text{Constraint } y_i * (w^T x + b) \geq 1$$

Correctly identified

for Misclassification Hyperparameters.

... label Cost Fⁿ for SVG used for Classification

Correctly Identified Misclassification.

The total cost F^n. for SVG used for classification

[For Allowed Misclassification].