For example purpose we used two tree with the first two values of "Exp" column. Actually all the values will create it's own tree and we will calculate variance Reduction from each of them and find the greatest variance Reduction value.

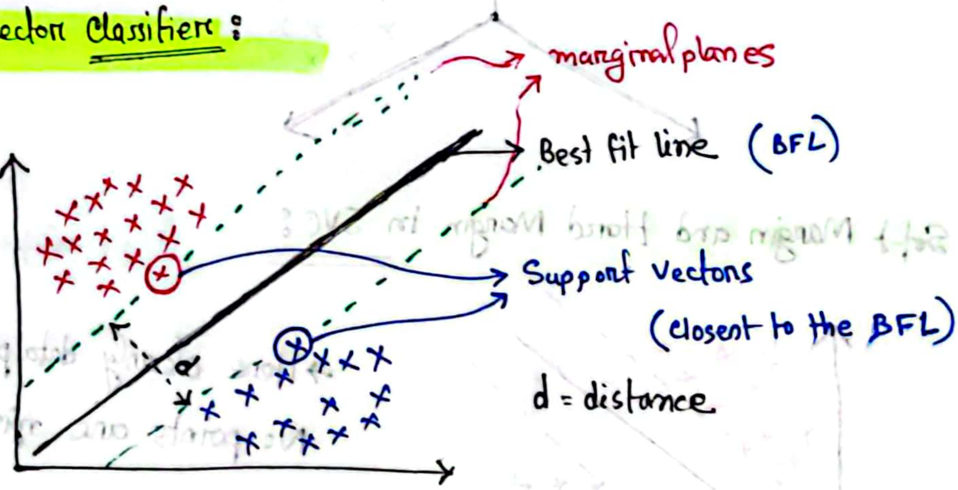
"Information gain" will be calculated like before (Decision Tree Classifier)

Support Vector Machine Algorithm

① Support vector Classifier (SVC) | classification

② Support Vector Regressor (SVR) | Regression

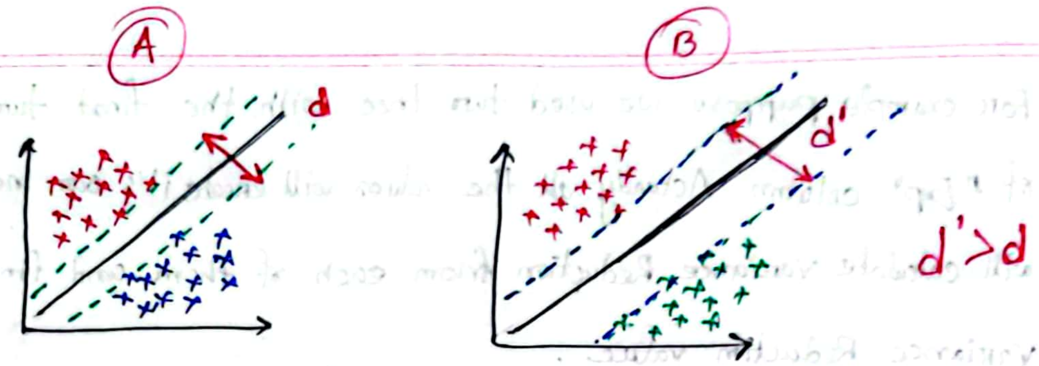
① Support Vector Classifier:



In SVC we not only create the best fit line, we also create two marginal planes which go through the nearest data of the best fit lines.

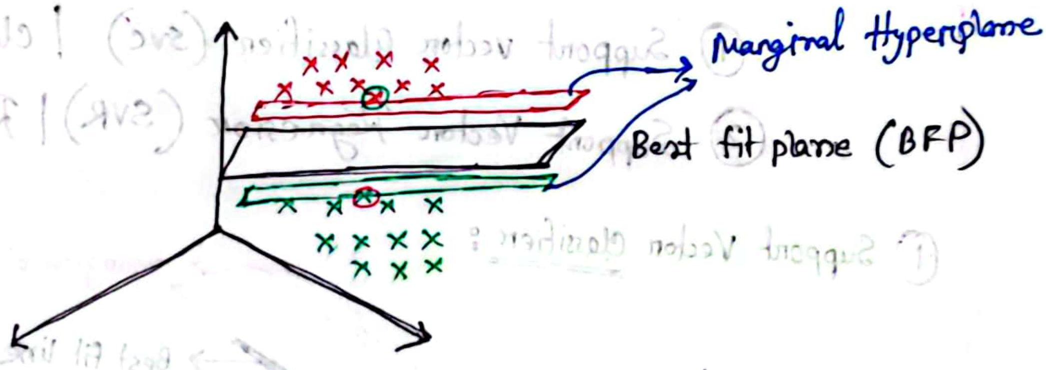
The marginal lines can't go outside of the nearest data and the distance of best fit line and marginal lines should be maximum.

CamScanner

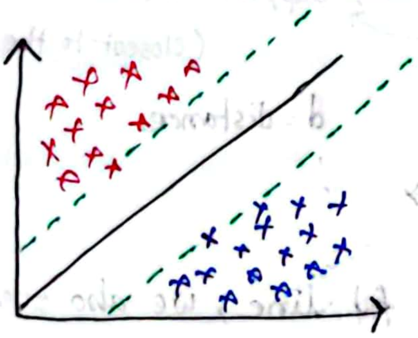


We always will take the SVC whose distance of marginal planes are maximum. Suppose, in the above scenario we will choose (B) SVC

for 3D it will look something like below →



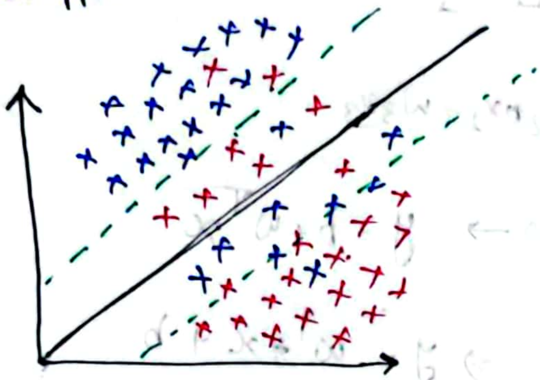
Soft Margin and Hard Margin in SVC:



→ Here clearly data points are seperated
No points are missclassified

→ So we call this Hard Margin

But Hard Margin is rare in real because in most cases datasets are overlapped.



→ Some datapoints are misclassified

→ So this called Soft Margin

So, datapoints can overlapped. In this case, although we know that there is overlapping (some data are crossing their marginal line and even entering into different class area) we ignore this because we are able to classifying most of the data.

Mathematical Intution of SVC :-

Line equation $y = mx + c$

or, ~~h(θ) =~~ $h_\theta(x) = \theta_0 + \theta_1 x_1$

Also we can write, $ax + by + c = 0$

For Multiple linear Regression,

$$y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3$$

or, we can write, $y = b + w_1 x_1 + w_2 x_2 + w_3 x_3$

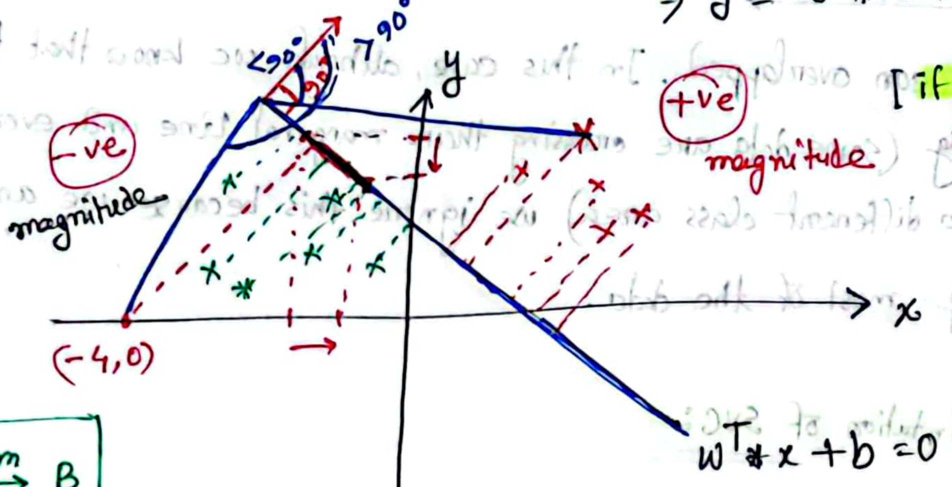
Here, $W = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}$ $X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$

$$w^T = \begin{bmatrix} w_1 & w_2 & w_3 \end{bmatrix} * x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

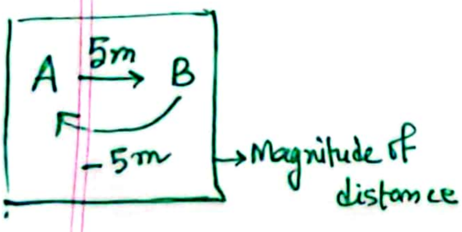
output $\rightarrow$ $w_1 x_1 + w_2 x_2 + w_3 x_3$

So, we can write the formula as $\rightarrow$ $y = b + w^T x$

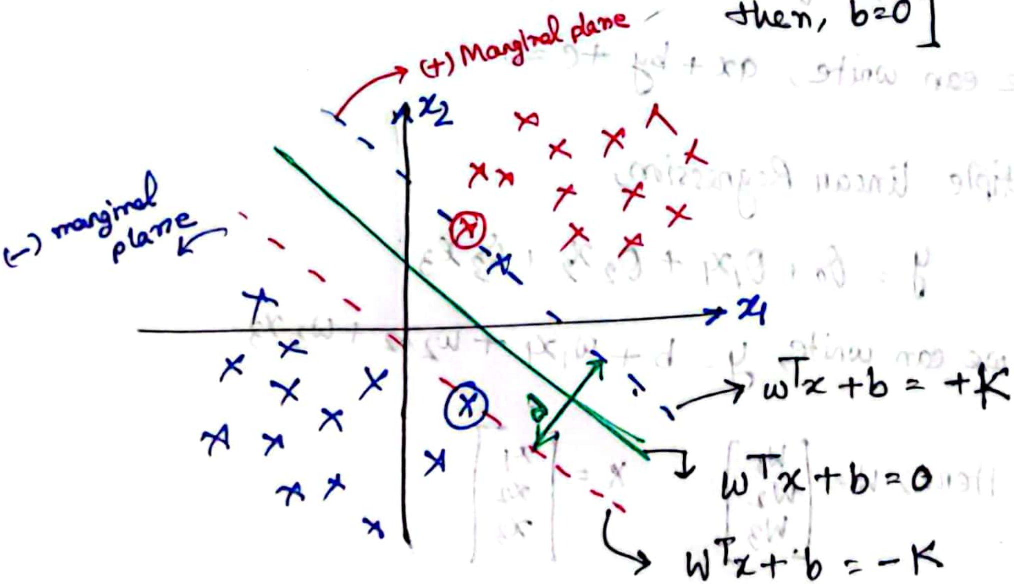
$$\Rightarrow y = w^T * x + b$$



[if you can't relate, rewatch video 252 again]



[$w^T x = 0$
↳ if line would go through origin
then, $b = 0$]



Let say, $K=1$, So, d will be the subtraction of the two marginal plane →

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

$$\Rightarrow \frac{w^T(x_1-x_2)}{\|w\|} = \boxed{\frac{2}{\|w\|}} \rightarrow$$ Distance between Marginal planes.

Aim of our Cost Function:

To maximize → $\frac{2}{\|w\|}$ (Distance between marginal planes)

constraint such that $y_i \begin{cases} +1, & \text{if } w^T x + b \geq 1 \quad \text{[The upper right datapoints]} \\ -1, & \text{if } w^T x + b \leq -1 \quad \text{[The lower left data points]} \end{cases}$

For all correct classified datapoints →

$$\boxed{y_i * [w^T x + b] \geq 1}$$

Modified cost function of SVC: (Soft Margin)

To minimize → $\frac{\|w\|}{2}$

constraint such that $y_i \begin{cases} +1, & \text{if } w^T x + b \geq 1 \\ -1, & \text{if } w^T x + b \leq -1 \end{cases}$

CamScanner

For, soft margin case,

$$\text{cost function} = \min_{w,b} \frac{\|W\|}{2} + \boxed{C_i \sum_{j=1}^{w} \xi_j}$$

[ξ = ita]

C_i = hyperparameters [How many points we can consider (accept) for misclassification]

This whole equation called "Hinge Loss"

[We mostly get soft margin datapoints in real life where we can't properly seperate them using marginal planes. They overlap.]

ξ_j = Summation of the distance of incorrect data points from Marginal planes