25/04/23

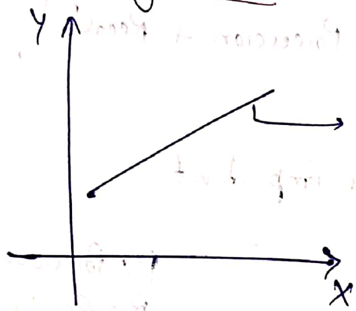
M.L Day-5

Support Vector Machines (SVM)

① Classification
② Regression

$\left.\begin{array}{l}\rightarrow \text{SVC} \rightarrow \text{Support Vector classifier} \\ \rightarrow \text{SVR} \rightarrow \text{Support Vector Regressor}\end{array}\right\} \Rightarrow$ Intuition.

Eqn of Straight Line :→



Coefficient

$y = mx + c$

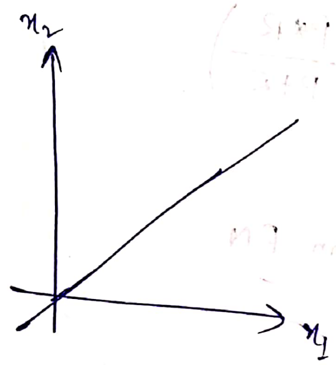
$y = \beta_0 + \beta_1 x$

$\Downarrow$

$ax + by + c = 0 \Rightarrow \boxed{y = \boxed{-a/b}\, x - \boxed{c/b}}$

Coefficient

Intercept



$: ax_1 + bx_2 + c = 0$

$w_1 x_1 + w_2 x_2 + b = 0$

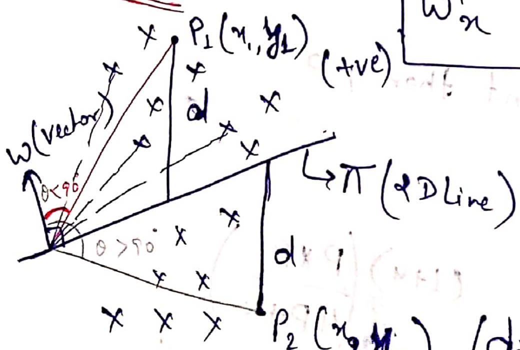
↳ coefficient

$W^T x + b = 0$, $\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} [x_1, x_2] = 0$

If line is passing through origin then $\boxed{b = 0}$

$\boxed{W^T x = 0}$ → Eqn of a line passing through origin.

Distance of a Point to the plane



+ve value

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

(distance of point P_1 from the line)

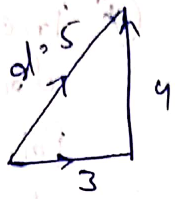
magnitude to get the Unit Vector

Cos θ when θ < 90° is always +ve i.e 0 to 1.

$W \rightarrow$ Vector of Coefficient.

(*) Hence any point above the plane will have distance in +ve since $\theta < 90°$ and $\cos\theta$ will lie b/w 0 to 1.

Example 1 → $P(3,4)$



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

$$\boxed{\hat{d} = \frac{d}{\|d\|}}$$

Unit Vector (magnitude is ~~same~~ changing but direction will be same)

Magnitude

Now, $\left(3/5,\ 4/5\right)$

$\sqrt{(3/5)^2 + (4/5)^2}$

$= \sqrt{\frac{9+16}{25}} = \sqrt{\frac{25}{25}} = \sqrt{1} = 1$ ← unit vector.

Unit vector :→ Vector which has a magnitude of 1.

(*) Below the plane $\theta > 90°$. Hence any point which lies below the plance will make angle greater than 90° with w.

If $\theta > 90°$ then $\cos\theta < 0$ i.e $\cos\theta$ will be -ve.

Hence- In case Vector facing upward

$$d = \frac{w^T P_2}{\|w\|} = \|w\|\,\|P_2\|\cos\theta \Rightarrow \boxed{-ve} \quad \boxed{\theta > 90°}$$ ← Below the Plane

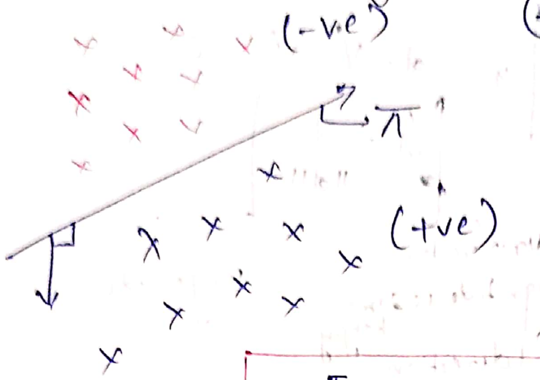
$$d = \frac{w^T P_1}{\|w\|} = \|w\|\,\|P_1\|\cos\theta \Rightarrow \boxed{+ve} \quad \boxed{\theta < 90°}$$ ← Above the Plane

Conclusion

(*) Any line which falls above the plane and form an angle of <90° with Vector. w. then the distance of the plane with the point is +ve.

(*) ||'y Any line which falls below the plane and form an angle >90° with vector w then the distance of the plane with the Point is -ve.

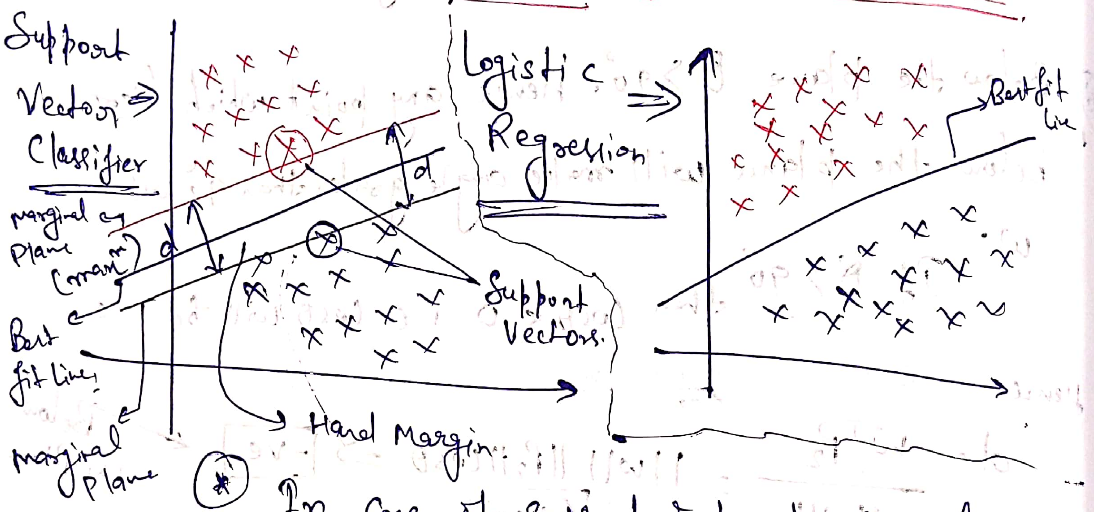
In Case Vector facing downwards



(*) In case if the Vector facing downward in that case Points above plane having distance -ve. and Points below plane will have distance +ve.

$$\boxed{\frac{W^T P}{\|W\|} = \text{Distance of a point to a plane}}$$

Geometric Intution Behind Support Vector Machine



(*) In Case of Support Vector classifier along with the Best fit line we draw the Marginal plane in such a way that distance b/w the Marginal Plane should be maximum.

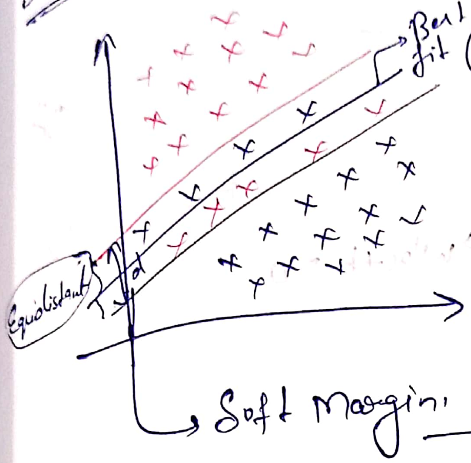
(*) If distance is max^m means we are clearly able to divide the data points.

(*) Marginal line :→ Marginal line is similar to the best fit line which passes through the nearest data points.

(*) The point or nearest category through which the Marginal Plane is passing is called as support vectors. Bcz theare nearest Point / Categories are helping to create the marginal plane. We can have more than one support vector.

(*) Hard margin :→ When we are able to clearly seperate the all the points using marginal plane and there is no error then i.e called Hard margin.

But In real world we don't get clearly seperable margin and i.e called soft margin.

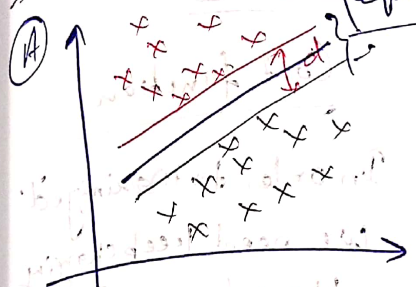


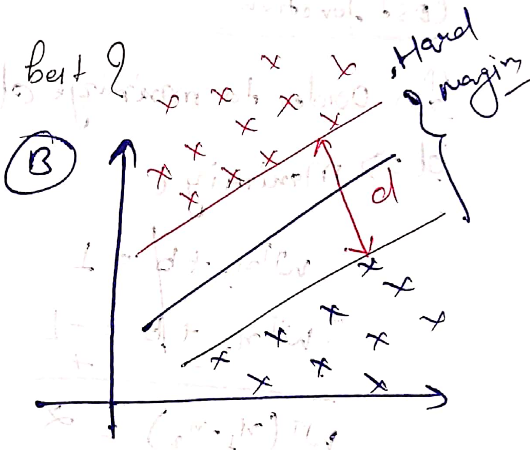
(*) Marginal Plane should be equidistant from the Best fit line.

with error ⇒ $\boxed{\text{Soft margin}}$

Soft margin.

Q: Out of this two which one is best ?





'd' is small Hence not that good.

This is Best Since distance 'd' is more.

$\boxed{\text{distance should be max}^m}$

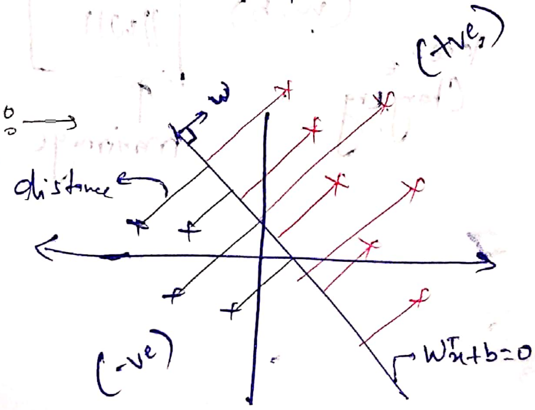
with no error ⇒ $\boxed{\text{Hard margin.}}$

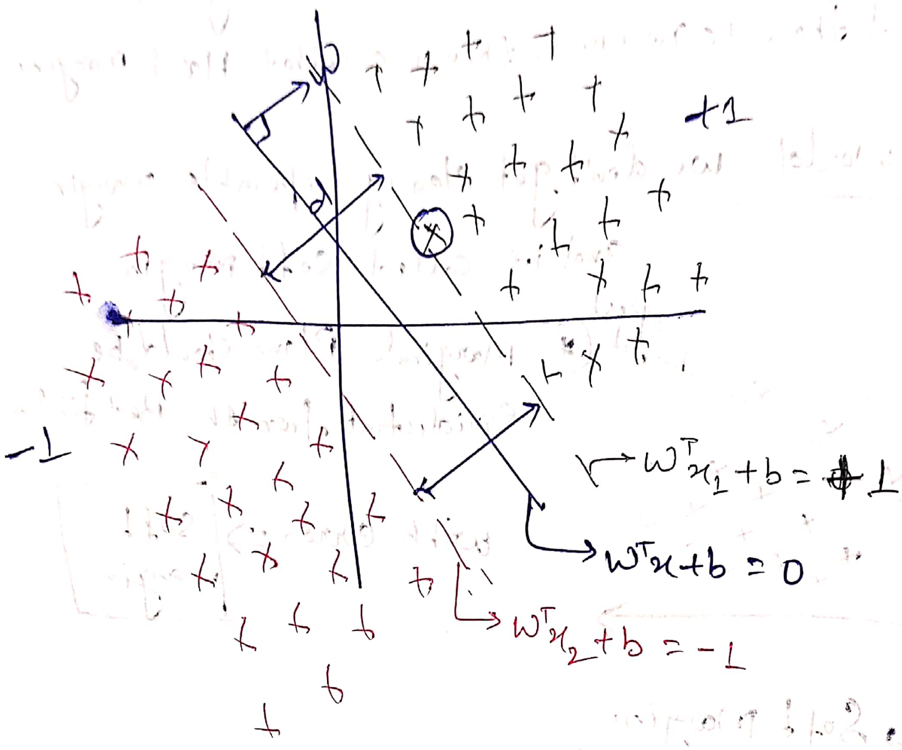
(*) This Marginal Plane should be equidistant from the best fit line.

SVM Mathematical Intution :→

(*) Distance b/w point and the best fit line below the line will be −ve and above the line will be +ve.

Since below the line $\theta > 90°$.
above the line $\theta < 90°$.





Cost function

In order to maximize d.

$d \Rightarrow$ Maximize

$$
\begin{array}{r}
w^T x_1 + b = 1 \\
-\; w^T x_2 + b = -1 \\
\hline
w^T (x_1 - x_2) = 2
\end{array}
$$

$\Rightarrow$ In order to get unit vector

{unit vector} $\boxed{\dfrac{w^T (x_1 - x_2)}{\|w\|} = \boxed{\dfrac{2}{\|w\|}}}$ $\rightarrow$ If we get this max^m means 'd' will be max.

Hence

Cost function

(d) In order to maximize 'd' we need keep changing the value of w, b.

$$\underset{(w,b)}{\text{Maximize}} \; \boxed{\frac{2}{\|w\|}} \rightarrow \text{Distance b/w Marginal Plane}$$

(w, b) → keep Changing

$\|w\|$ ↑ maximize

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

Truth Points

for all correct classified Points.

$\Updownarrow$

for all correct Points

for both +ve and -ve Points.

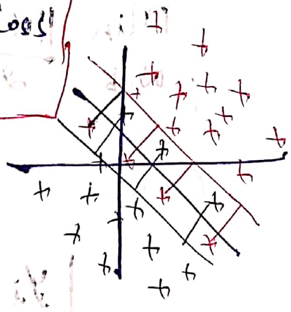
$$\boxed{\text{Constraint} \rightarrow y_i * (w^T x + b) \geq 1}$$

$$\Rightarrow \underset{(w,b)}{\text{Maximize}} \frac{2}{\|w\|} \Rightarrow \boxed{\underset{(w,b)}{\text{Minimize}} \frac{\|w\|}{2}} \Rightarrow \text{Loss function} \Downarrow \text{minimizing}$$

Cost function : →

Hyperparameter

$$\boxed{\underset{(w,b)}{\min} \frac{\|w\|}{2} + \boxed{C_i \sum_{i=1}^{n} \eta_i}} \Rightarrow \text{Hinge Loss}$$



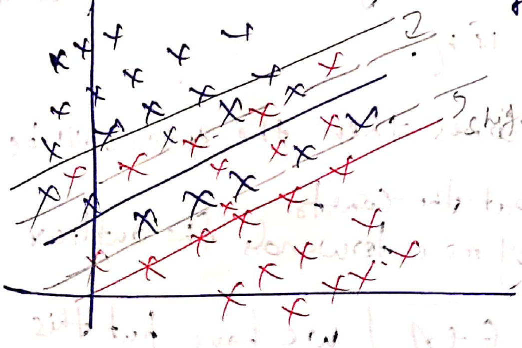
Where C_i = {How many points we can ignore for misclassification}

$\sum_{i=1}^{n} \eta_i$ = Summation of the distance of the incorrect data points from the marginal plane.

7 incorrect data Points.

$C_i = 7$

$\eta_i = \Sigma$

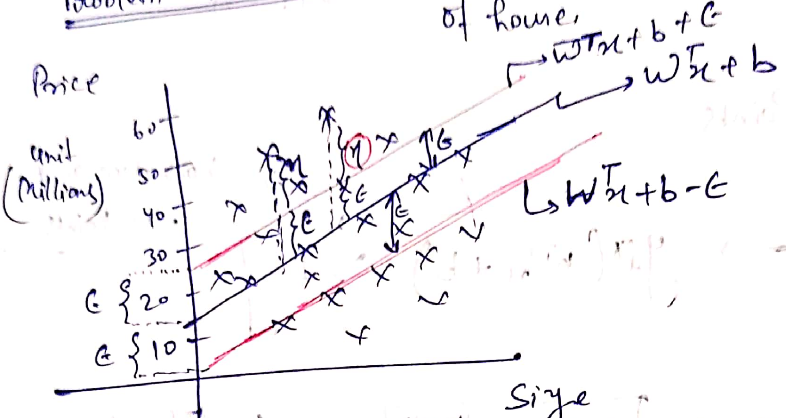


Here, Initially C and η value is very high

If we try to restrict 'C' and 'η' then the marginal plane will shift and it value will decrease

Support Vector Regression :→

Problem statement :→ Based on size of house predict the price of house.



Where
$\epsilon \rightarrow$ Marginal Error

(*) Marginal error is the difference b/w Marginal plane & Best fit line.

(*) We Construct the Marginal plane with ϵ i.e Marginal Error means this much error is fine.

Cost function :→

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

Constraint

$$\boxed{|\underbrace{y_i}_{\text{Actual}} - \underbrace{w_i x_i}_{\hat{y}\text{ (Predicted)}}| \le \underbrace{\epsilon + |\eta_i|}_{\text{Hyperparameter}}}$$ → MAE

Where, ϵ = Marginal Error
η_i = Error above the Margin

(*) Points which lies outside the marginal plane their distance from the marginal plane is η.

(*) Suppose if an outlier comes then the error will be more ~~but we have put the consts~~ and plane may move towards the outlier but since $\boxed{\text{error} \le \epsilon + \eta}$ we have put this constraint our marginal plane will not move w.r.t outlier.

(*) There is no incorrect point inside Regression since Regression is having Continious Values.

(*) In Regression problem we can't predict the exact right o/p there will be some difference hence there is no Incorrect points.

(*) The points which are inside the marginal plane those distance will be always less than ϵ.

for that $\boxed{|y_i - W_i x_i| \leq \epsilon}$ — when Points Inside Marginal Plane

Model is Performing well

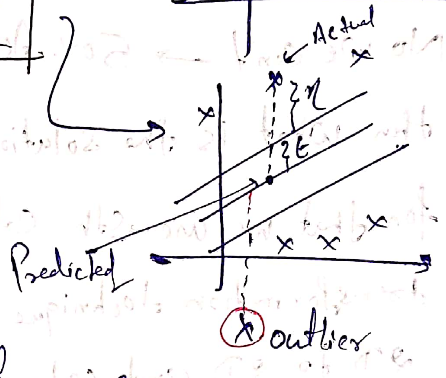
Marginal Plane is Perfect

(*) Points which are outside the Marginal Plane for that we add η along with the ϵ

Hence,

$$\boxed{|y_i - W_i x_i| \leq \epsilon + \eta_i}$$

(*) Hence it says that if the difference b/w actual & Predicted Point is less than equal to $\epsilon + \eta_i$ then also our Marginal plane is right.



ϵ = Marginal Error

η = Error outside Margin

(*) If η value is very-very high means if there is an outlier then it may become greater than 'error' which is $|y_i - W_i x_i|$ which is wrong. ~~So~~ So we need to do Hyperparameter tuning. So we need to continiously change η value with ϵ such that our points are near to the marginal plane ~~doesn't change~~.

(*) In Case of outlier the Algorithm will automatically change the Marginal Plane so that $|error| \leq \epsilon + \eta$

Q Will SVM get impacted by outliers?

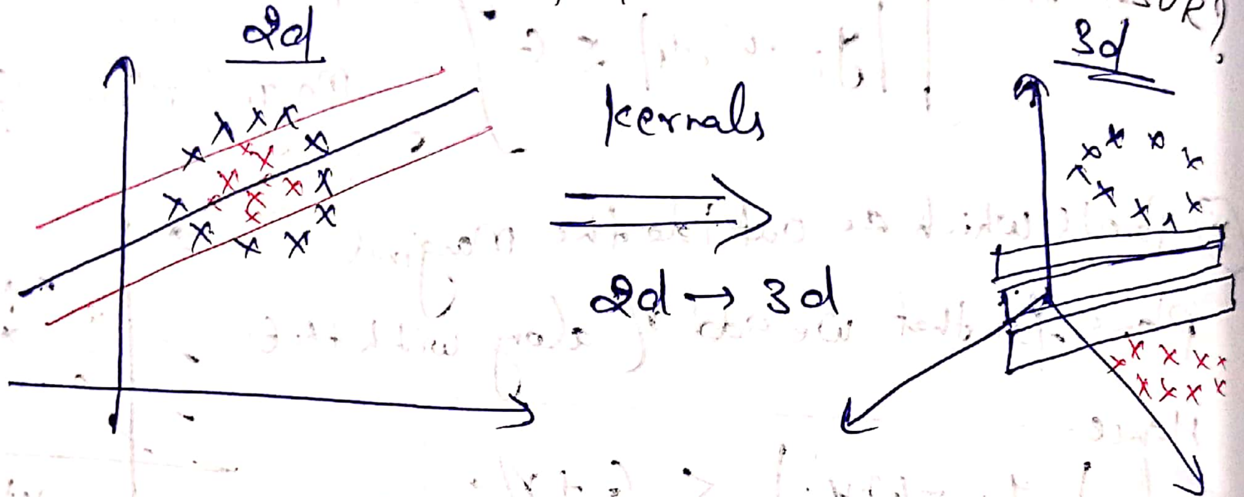
YES → Marginal Plane will get shifted till accuracy is good

Q Do we need to normalize our data points like standardization

YES

SVM kernal :→

Q Can we solve this kind of problem with SVM or SVR?



No we can't → 50% Accuracy.

(x) then what is the solution?

for that we use sth called "kernal" which is a simple transformation technique which converts the plane into 2D to 3D and seperates the data points ~~and then~~ raise the bar and then apply the plane and solve the problem.