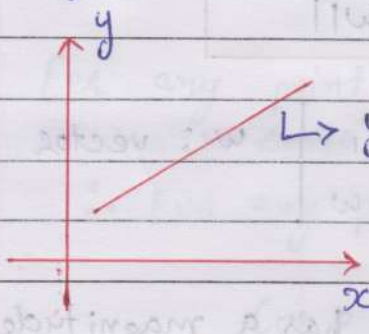
MACHINE LEARNING DAY 5

29/10/2022

Support Vector Machine (SVM)

It can solve both classification and Regression problem.

① Classification → SVC → Support vector classifier
② Regression → SVR → Support vector Regressor

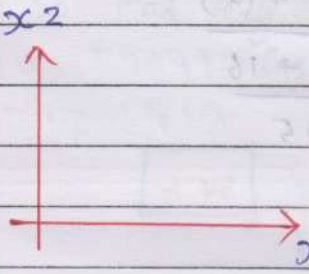


↳ $y = mx + c$ {Equation of Straight Line}

$y = \beta_0 + \beta_1 x \Rightarrow ax + by + c = 0$

$$y = \boxed{\frac{-a}{b}}x - \boxed{\frac{c}{b}}$$

Coefficient ← → Intercept

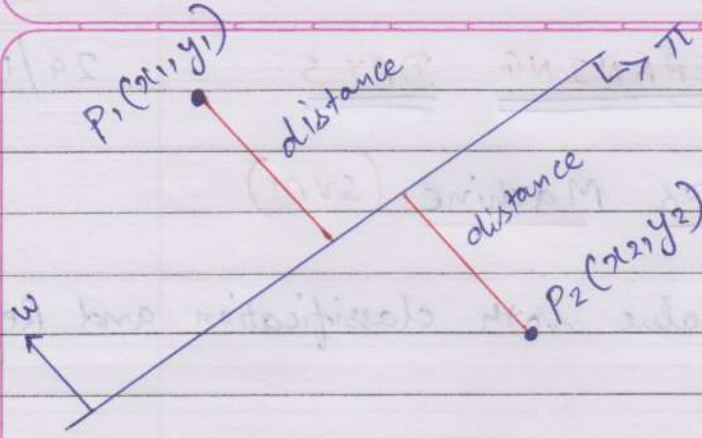


$$ax_1 + bx_2 + c = 0$$
$$w_1 x_1 + w_2 x_2 + b = 0$$
$$w^T x + b = 0$$

{$w^T x$: w Transpose x}

∴ Equation of line passing through Origin is:

$$\boxed{w^T x = 0}$$



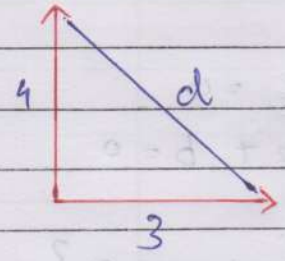
We have to find the distance of point from the plane.

Equation:

$$\text{distance}(d) = \frac{w^T p_1}{\|w\|}$$

where $w^T p_1$: w Transpose p_1 | w : vector
$\|w\|$: Magnitude of w

Unit vector: A vector which has a magnitude of 1 is basically called unit vector.



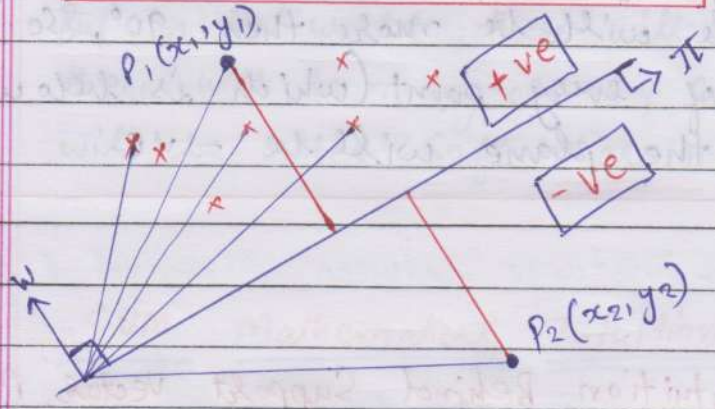
$$\text{Now } d = \sqrt{(3)^2 + (4)^2}$$
$$= \sqrt{9 + 16}$$
$$= \sqrt{25}$$
$$d = 5$$

$$\text{Now } \sqrt{\frac{(3)^2}{(5)^2} + \frac{(16)^2}{(5)^2}} = \sqrt{\frac{9}{25} + \frac{16}{25}} = \sqrt{\frac{25}{25}}$$

$$\therefore \hat{d} = \frac{d}{\|d\|} = \sqrt{\frac{25}{25}} = 1$$

Here $\hat{d}$: Unit vector
$\|d\|$: Magnitude of d

$$w^T p_1 \Rightarrow \|w\| \|p\| \cos\theta$$



For any point above the place the value of $\theta < 90^\circ$

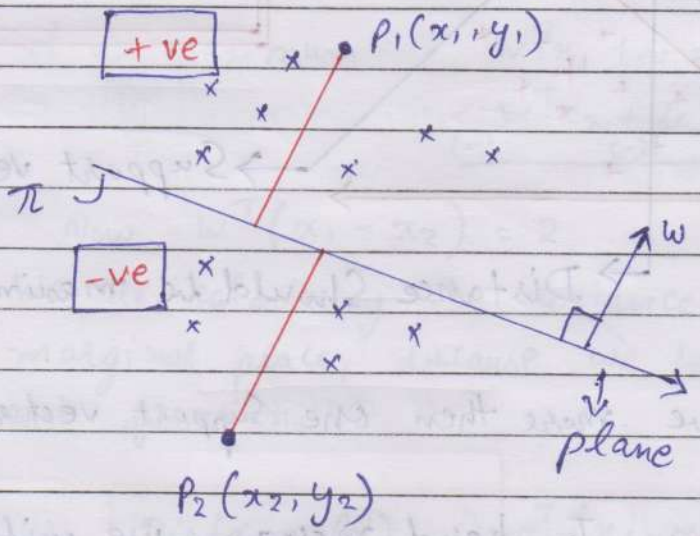
$\therefore$ For any point above the plane

$$\cos\theta \{0 - 90\} \Rightarrow +ve \text{ number}$$

Similarly any point below the line, the angle will be greater then 90°

So For any point below the line

$$\cos\theta \{> 90^\circ\} \Rightarrow -ve \text{ number}$$



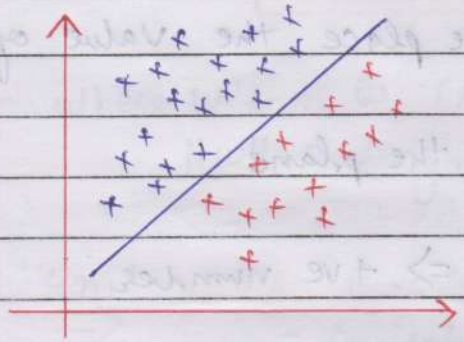
Note

If points are above the plane, that means θ will be $0 - 90^\circ$, so the distance of every point from the plane will be +ve value.

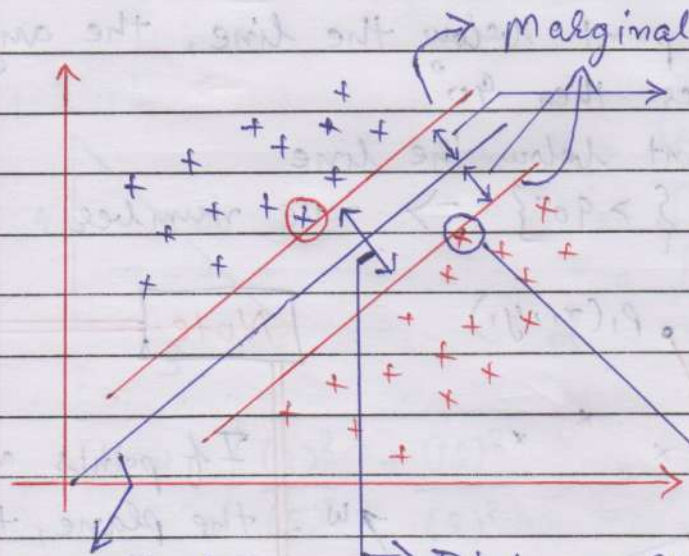
Note

Similarly if the points are below the place, then the angle will be more then 90°, so the the distance of every point (which is below the peane) from the plane will be -ve.

★ Geometric Intuition Behind Support Vector Machine



⇒ Linear Regression.



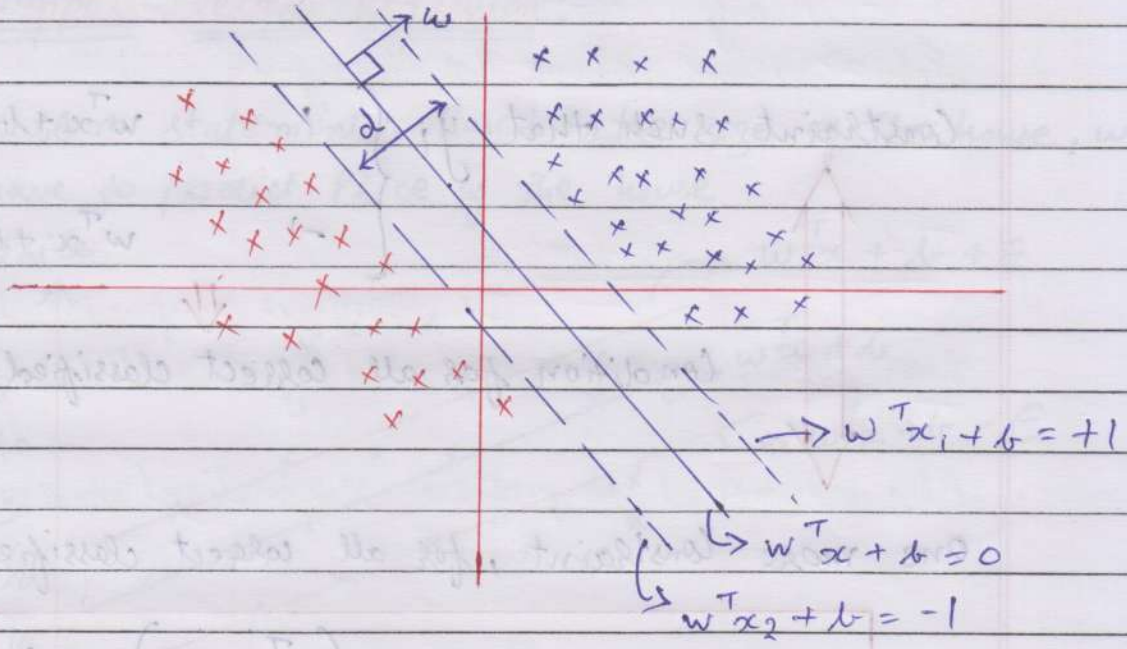
= You can have more then one Support vectors also.

Hard Margin:- In hard margin, we will not find any errors and we will be able to clearly separate all the points by using the

marginal plane.

But in real world, there will be many overlapping, there will be many errors, so marginal plane lines will be called Soft Margin

✱ SVM Mathematical Intuition :-



d => maximum

$$w^T x_1 + b = +1$$

$$w^T x_2 + b = -1$$

$$(-) \quad (-) \quad (+)$$

Now $w^T (x_1 - x_2) = 2$

We are calculating the distance (d) between the marginal place, because we have to calculate the cost function.

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

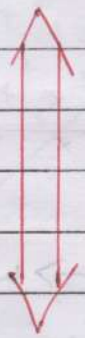
Cost Function

We have to maximise the value of $\frac{2}{||w||}$

By changing the values of w, b

$\frac{2}{||w||}$ $\Rightarrow$ Distance between Marginal Planes.

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



$\Downarrow$

Condition for all correct classified points.

One more constraint, for all correct classified points

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

$$\underset{w,b}{\text{Maximize}} \ \frac{2}{||w||} \Rightarrow \underset{w,b}{\text{Minimize}} \ \frac{||w||}{2} \Rightarrow \text{Loss Function}$$

$\Downarrow$

Minimize

Cost Function

Soft Margin

$$\underset{w,b}{\text{Minimize}} \ \frac{||w||}{2} + C_i \sum_{i=1}^{w} \zeta_i \Rightarrow \text{Hinge Loss}$$

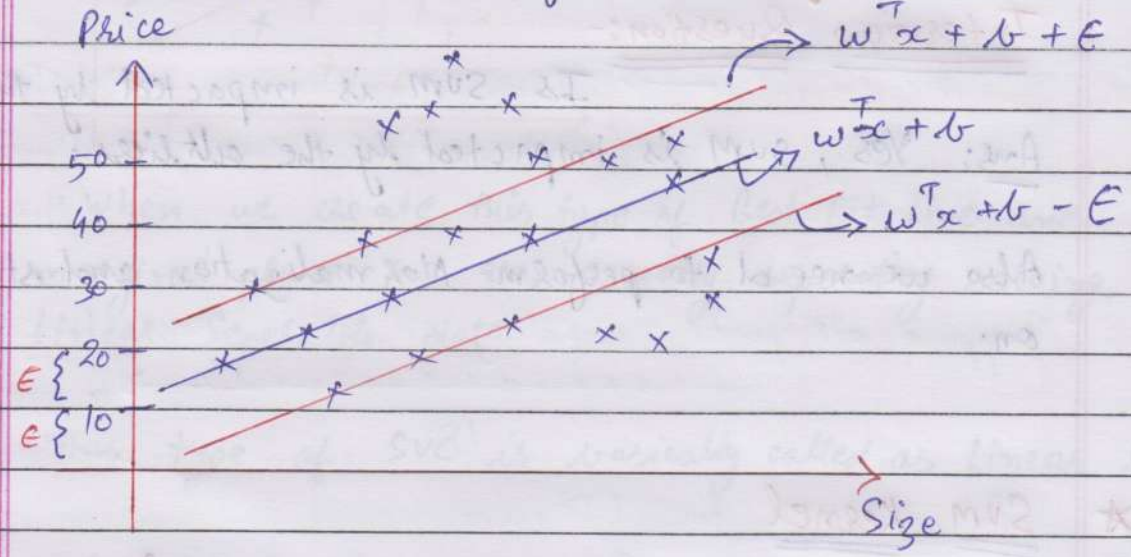
Where:

C_i : How many points we can ignore for misclassification

ζ_i (Eta): Summation of the distance of the incorrect data points from the marginal plane.

★ Support vector Regression :-

Problem statement :- Based on the size of the house, we have to predict Price of the house



ϵ : Epsilon = Marginal Error.

Cost Function

$$\underset{w,b}{\text{Minimize}} \quad \frac{\|w\|}{2} + \boxed{C_i \sum_{i=1}^{w} \zeta_i} \Rightarrow \text{Hinge Loss}$$

Constraint

$$|y_i - w_i x_i| \leq \epsilon + \xi_i$$

y_i → Truth point
$w_i x_i$ → Predicted point
ϵ → Epsilon
ξ_i → Eta of i

$\epsilon \Rightarrow$ Margin of Error
$\xi i \Rightarrow$ Error above the margin.

Interview Question:-

Is SVM is impacted by the outliers?

Ans: Yes, SVM is impacted by the outliers.

Also we need to perform Normalization and standardization.

★ SVM Kernel