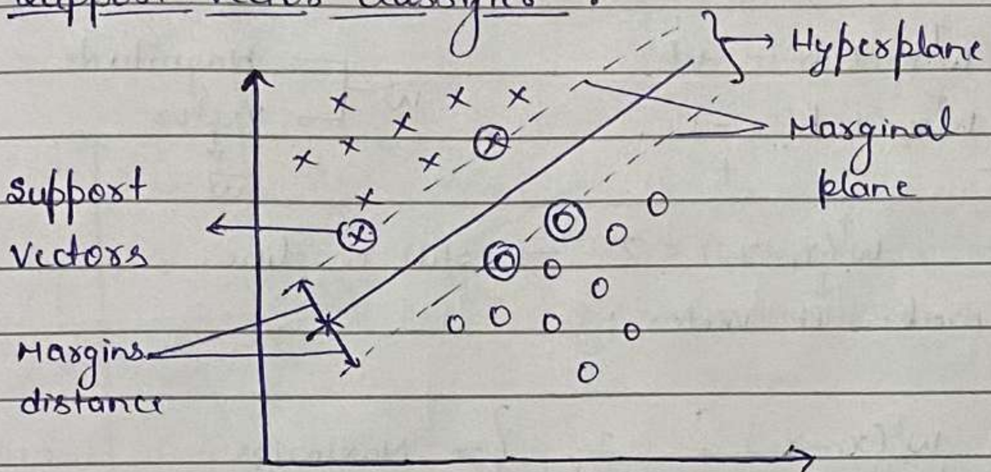
Support Vector Machines

→ It is a supervised machine learning algorithm which is used to solve both classification and regression problem by finding an optimal hyperplane.

* Support Vector classifier :

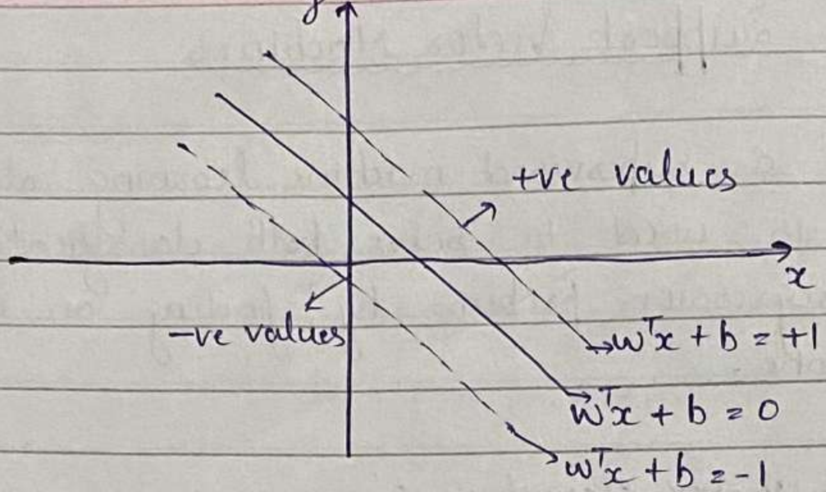


Note : A hyperplane having the maximum marginal distance is considered to be optimal/best hyperplane.

→ Equations of straight line :

- $y = mx + c$
- $y = \theta_0 + \theta_1 x = \theta^T x$
- $y = w_0 + w_1 x = w^T x$
- $ax + by + c = 0$

All equations are same.



$$w^Tx_1 + b = +1$$
$$w^Tx_2 + b = -1$$
$$- \quad - \quad +$$

$w \rightarrow$ Magnitude
$w \rightarrow$ vector $\rightarrow \vec{w}$

$$w^T(x_1 - x_2) = 2 \rightarrow \text{still a line.}$$

To make it vector :

$$\frac{w^T(x_1 - x_2)}{||w||} = \frac{2}{||w||} \Big\} \Rightarrow \text{Maximize}$$

$\therefore \underset{(w,b)}{\text{Maximize}} = \frac{2}{||w||} \Rightarrow$ Marginal plane distance

constraint : $y \begin{cases} +1 & w^Tx + b \geq 1 \\ -1 & w^Tx + b \leq -1 \end{cases}$

For all accurate datapoints :

$$\boxed{y * (w^Tx_i + b) \geq 1} \rightarrow \text{constraint.}$$

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

same $\rightarrow$

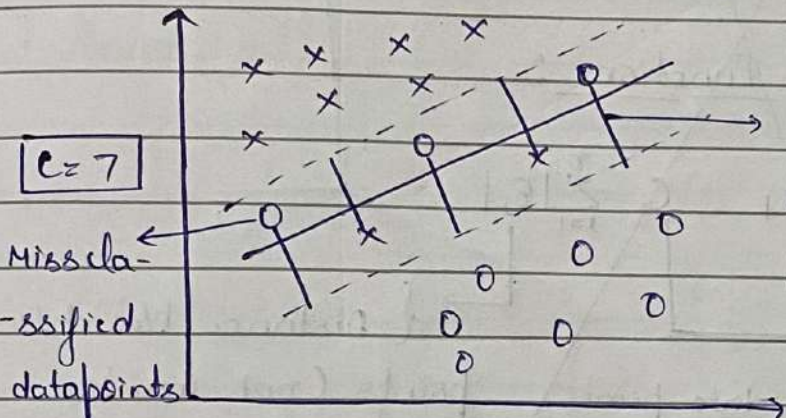
→ Cost Function :

$$\underset{(w, b)}{\text{Minimize}} = \frac{||w||^2}{2} + C_i \sum_{i=1}^{n} \xi_i$$

{ How many points we can avoid misclassification

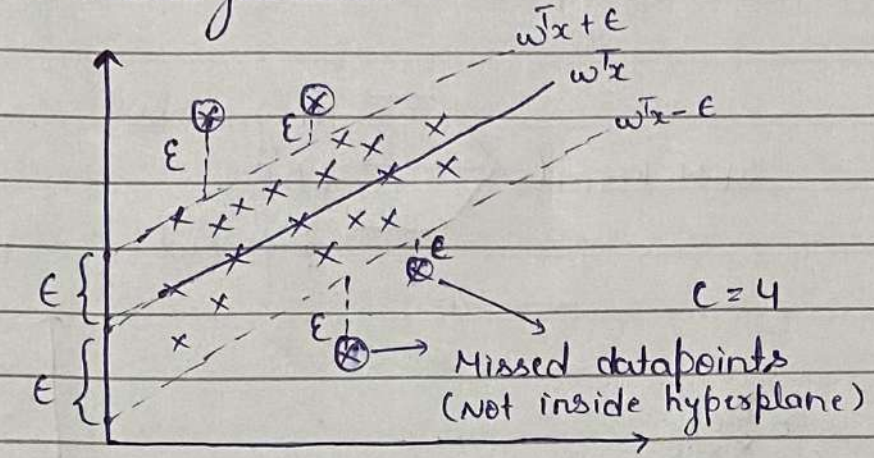
Eta

{ Summition of distance of missclassified points from marginal plane }



Note : when a hyperplane perfectly separate the data points without any misclassification, then the margin is called hard margin.
when hyperplane allows some misclassification, then the margin is called soft margin.

→ Support Vector Regression :



• Constraint :

$$|y - w^Tx| \leq \epsilon + \varepsilon$$

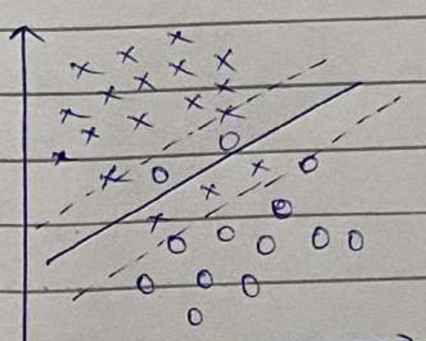
• Cost Function :

$$\underset{(w,b)}{\text{Minimize}} \; \frac{\|w\|^2}{2} + C_i \sum_{i=1}^{n} |\varepsilon_i|$$

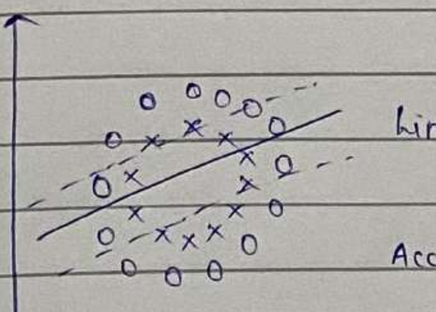
C_i → { How many data points we can adjust outside hyperplane

$\sum_{i=1}^{n} |\varepsilon_i|$ → { Distance b/w data points (not inside hyperplane) and marginal plane }

* SVM Kernals :



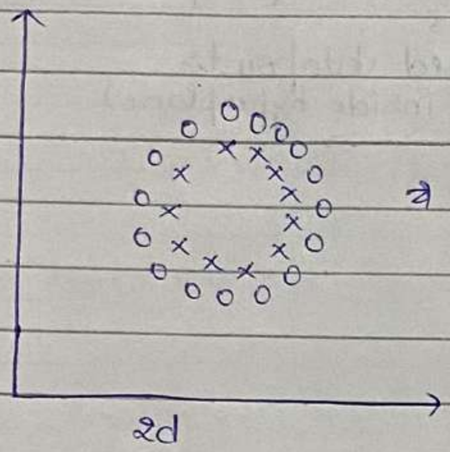
Linear SVC



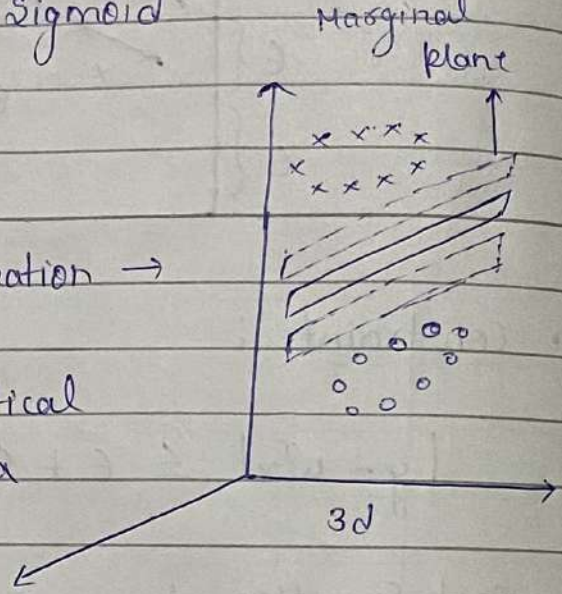
Linear SVC
↓
×
Accuracy ↓↓

Note: In this case, we use SVM kernels.

SVM Kernels → Polynomial
SVM Kernels → RBF
SVM Kernels → Sigmoid



→ Transformation →
↓
Mathematical formula



Eg: Linear SVC → x → Transformation →
↓
$y = x^2$