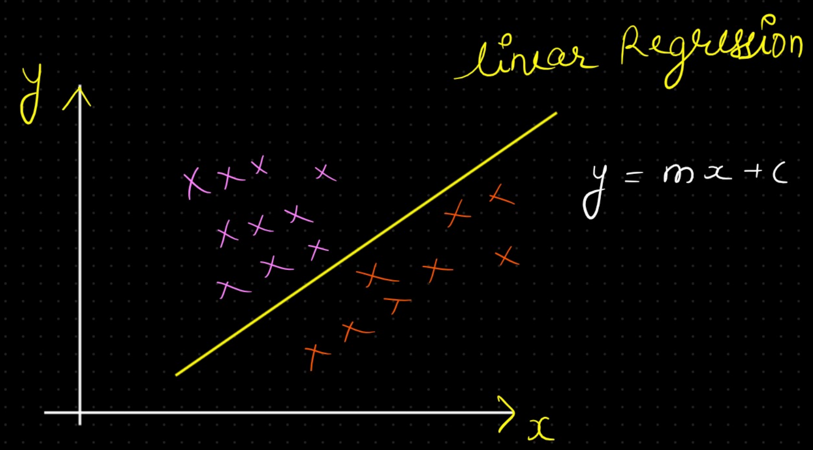
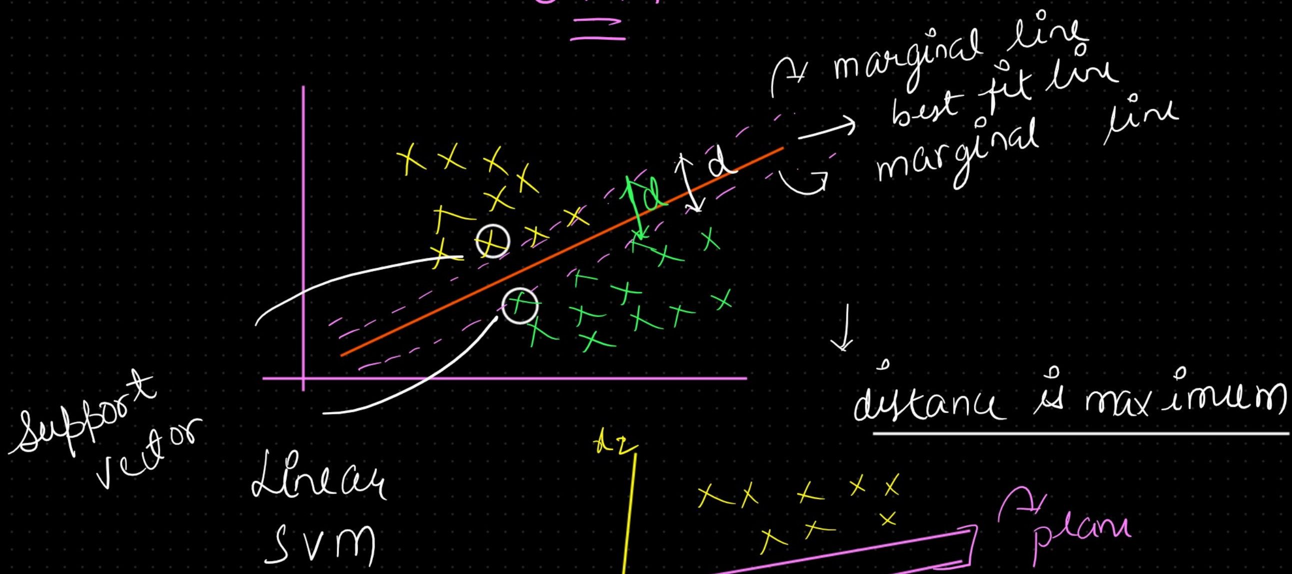
Support Vector Machine (SVM)

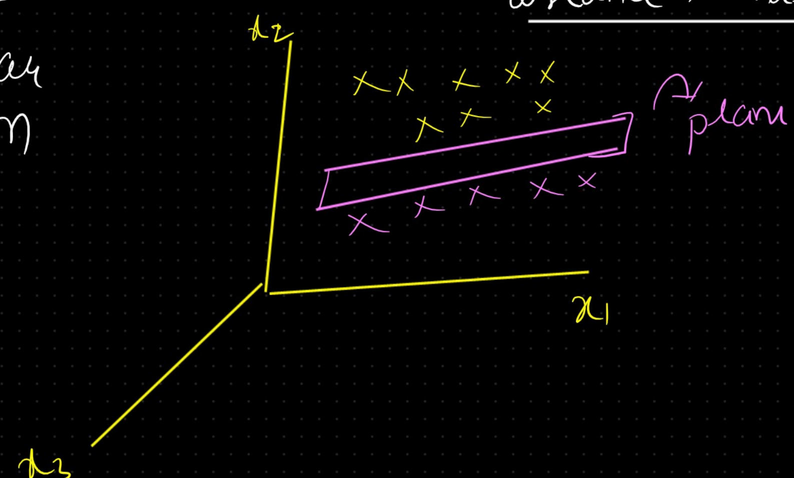
Support vector Machines ML Algorithm (SVM)

→ Support Vector Machine (SVM) is a supervised machine learning algorithm used for classification and regression tasks.
→ SVM works by finding the hyperplane that best separates data points into different classes while maximizing the margin between the classes.

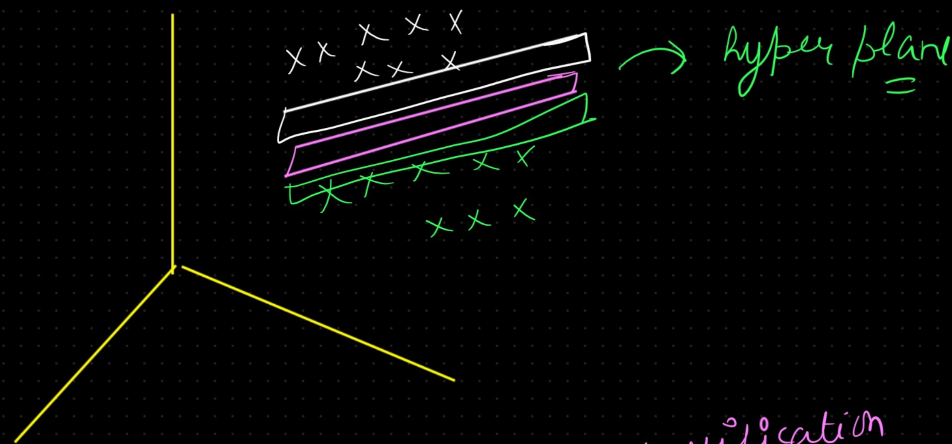


SVM





SVM - 3D

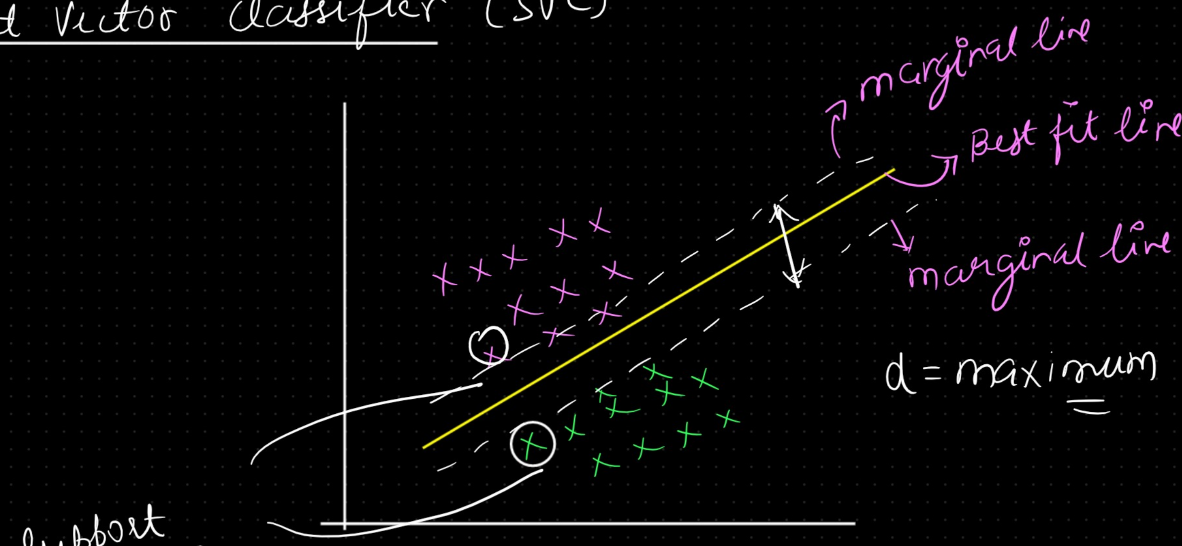


classification

1. SVC (Support Vector Classifier)
2. SVR (Support Vector Regressor)

Regression

Support Vector Classifier (SVC)



Linear SVM →

Equation of straight line

$$y = mx + C$$

or

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots +$$

$$y = \beta_0 + [\underline{w_1 x_1 + w_2 x_2 + w_3 x_3}]$$

$$w = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} \qquad x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$w^T = \begin{bmatrix} w_1 & w_2 & w_3 \end{bmatrix} \qquad w^T \cdot x$$

$$w^T \cdot x = [w_1 x_1 + w_2 x_2 + w_3 x_3]$$

$$\boxed{y = w^T x + b} \Rightarrow y = mx + C$$

$$y = 0 \qquad \boxed{w^T \cdot x + b = 0}$$

where → w → weight vector

x → input feature vector

b → bias term or intercept

The equation of the hyperplane —

$$\boxed{f(x) = w_0^T x + b}$$

Aim: SVM aims to maximize the margin b/w the hyperplane and the nearest data points (support vectors) from both classes

→ The margin is defined as the perpendicular distance from the hyperplane to the nearest data point.

Cost function

$$d \Rightarrow \frac{\|w\|^2}{2}$$ ⇒ Distance b/w Marginal plane.

Vector → dirⁿ
Vector → magnitude

constraint

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

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

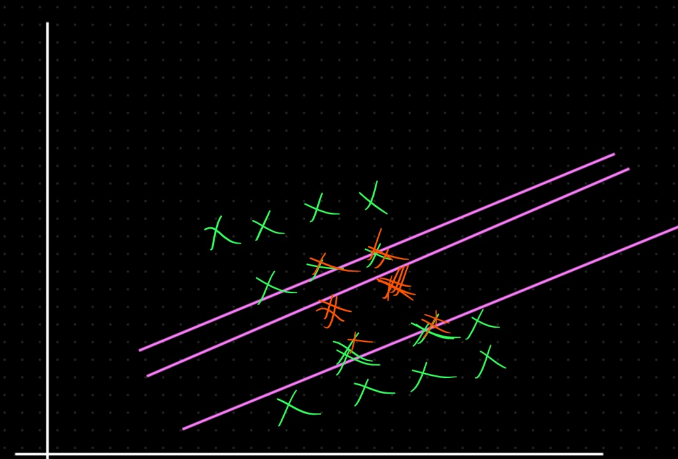
Hard Margin Vs Soft Margin

Hard margin → None of the datapoint are misclassified.

Soft margin

↓

Some data points are misclassified [Error]



Modified cost func^n soft Margin (SVC)

$$\min_{w,b} \frac{\|w\|}{2} + \boxed{C_i \sum_{i=1}^{n} \xi_i}$$

→ ita

hinge loss

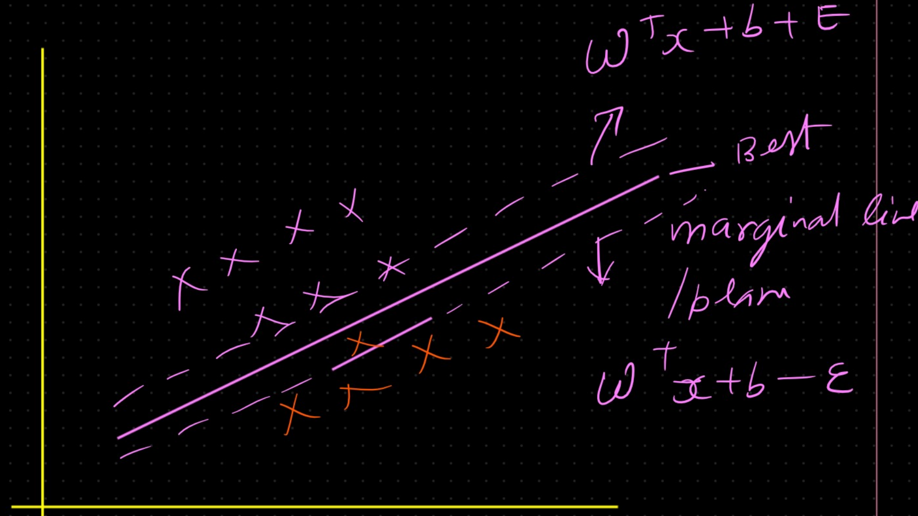
$C = \frac{1}{\lambda}$

Hinge Loss → Hinge loss is the loss function used in SVM, which penealizes missclassification. [Error]

$$\boxed{\max(0, 1 - y_i(w^T \cdot x_i + b))}$$

Support vector Regressor (SVR)

$E \rightarrow$ Marginal Error

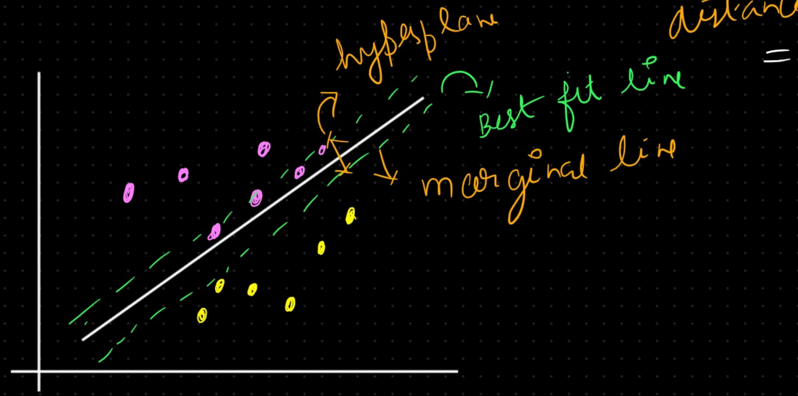


Kernel → f(x)

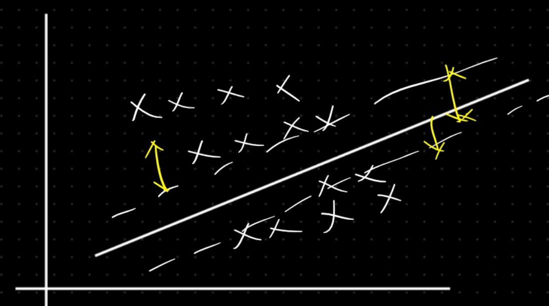
Non - Linear SVM (Kernel SVM)

In cases where the data is not linearly separable, SVM can be extended using the kernel trick. The kernel function maps the input data into a higher-dimensional space where it can be linearly separable

distance ↑

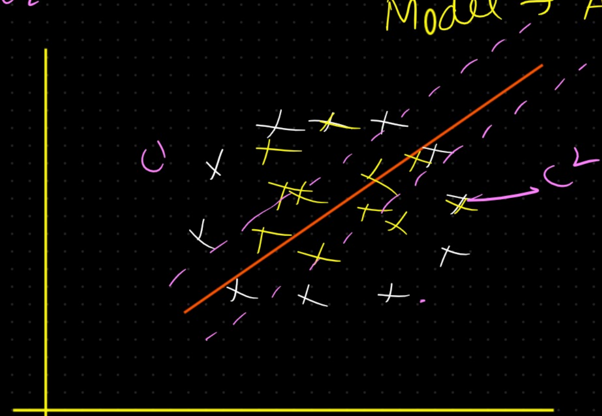


linear SVM



Non- linear

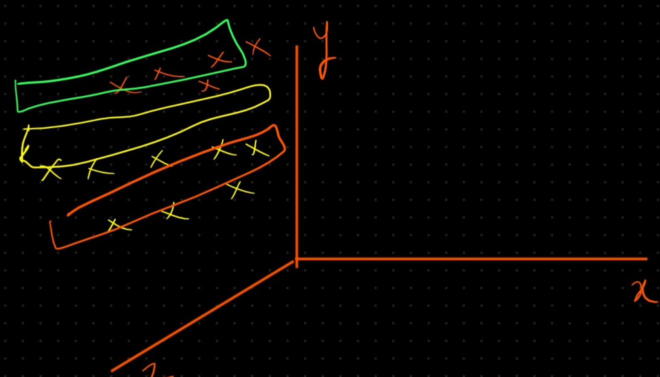
Model → Acc ↓ ↓



⇓ 2D ⟷ Kernel function

⇓

3D



Equation of a Hyperplane in Linear SVM

$$\boxed{w^T \cdot x + b = 0}$$

Non-linear SVM Equation

$$f(x) = \sum_{i=1}^{n} \alpha_i y_i \underbrace{K(x, x_i)}_{\text{Kernel function}} + b$$

Commonly Used Kernels:

1. Linear Kernel
2. Polynomial Kernel
3. RBF (Radial Bias Func^n) / Gaussian kernel
4. Sigmoid kernel