SVM

Supervied Learning

1. Linear regression
   - Ridge Reg.
   - Lasso Reg
   - elastic req.

   } Regurlization

2. Logistic Regression

- Optimization

- Gradient Decent

3. SVM
   - Regression
   - classification

4. DT
   - Req
   - class

5. Eensemble techniques
   1. bagging → RF
   2. Boosting → AB, GB, XG, Catboost
   3. Stacking → Archetecture
   4. Cascading → Architecture

6. KNN, Naive bayes
   - KNN → Distance based approch
   - Naive bayes → Probability based approach

- HyperParameter tuning
  1. Grid search CV
  2. Random " "
  3. optuna
  4. hyperopt

ML

1. Regression
2. Classification

target variable

1. Num ← Req.
2. Cat ← Classification
   - binary ← 2 class
   - multiclass ← More than 2 class

Evalution

Regression → R^2, Adj R^2

MSE, RMSE, MAE

Classification → Confusion matrix

- Accuracy, Precision, Recall, F-BScore
- PR curve
- Roc-Auc curve

Unsupervised

Clustering

- K-Mean, K-Mean++
- Hirarchical
- DBScan

PCA → DR

Stats

LDA

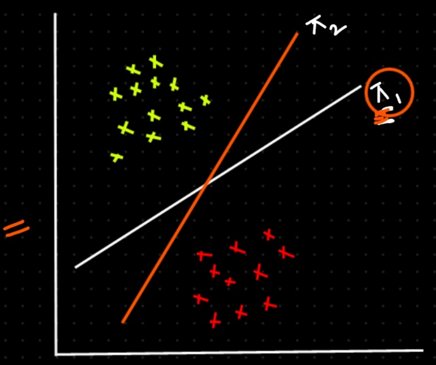
t-SNE

{Support vector} machine [1990]

↳ 2005 | 2006

- Robust
- capacable
- classification, Regression

— ML community

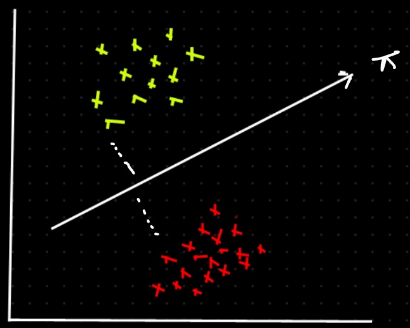


SVM ⇒ π_1, π_2 this is called hyperplane

it is classifying the data

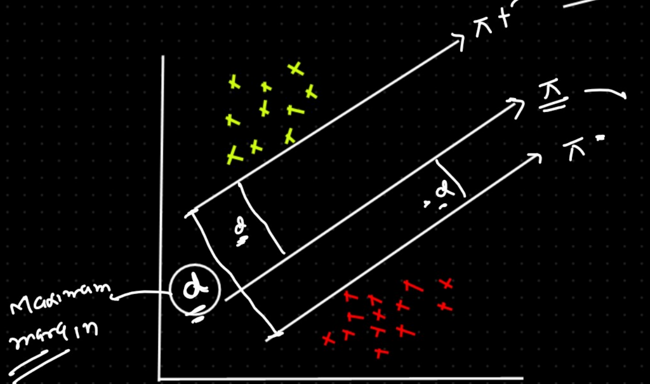
Prob stat

Aim. ⇒ We have to find to the that hyper plane which classify the point as widely as possible.



⇒ SVM ⇒ Margin.

that hyperplane → maximize margin

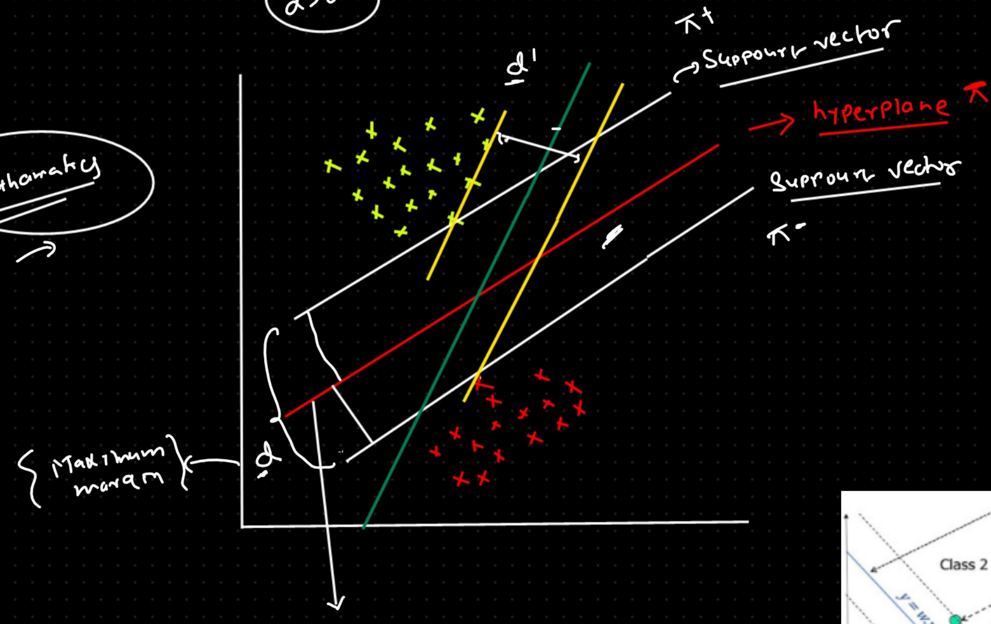


SVM

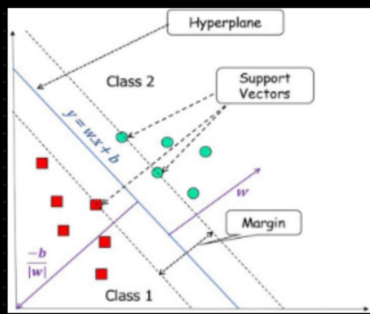
⇒ Mathmatics

$d > d'$

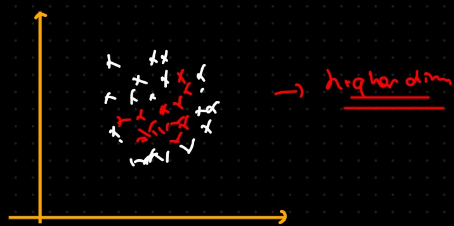
Mathamatics



{Maximum margin hyperplane}



① Support vector machine → Linear seprable
→ {Non Linear sep} → kernel trick



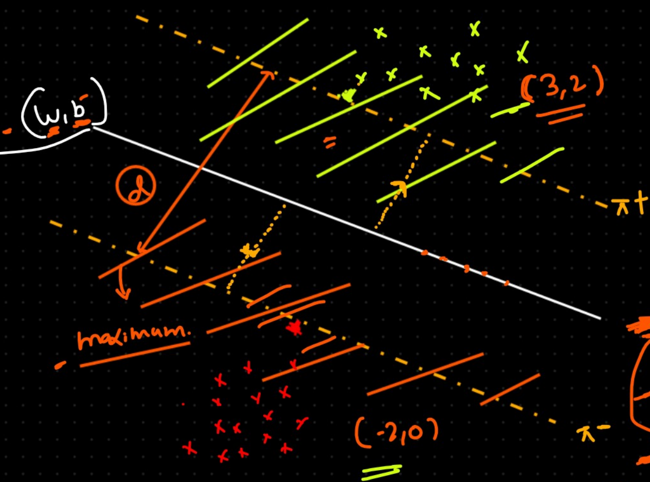
② Robust For outlier

③ Reg & classification

$\boxed{ax + by + c = 0}$

$\boxed{Y = mx + c}$ ← Lin reg

decision rule, decision boundry



$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix}^T \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \leftarrow [w_1, w_2] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

$\begin{bmatrix} a \\ b \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$

$\boxed{w_1 x_1 + w_2 x_2}$

T → transpose

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

$\boxed{ax + by + c = 0}$

$\boxed{w_1 x_1 + w_2 x_2 + c = 0}$

2D

$W^T x + b > 0 \rightarrow +ve$

$W^T x + b < 0 \Rightarrow -ve$

$ax + by + c = 0$

$2x + 3y + c = 0 \Rightarrow (3,2)$

$2 \times (3) + 3 \times (2) + 3$

$6 + 6 + 3 = 15$

$2x + 3y + c = 0$

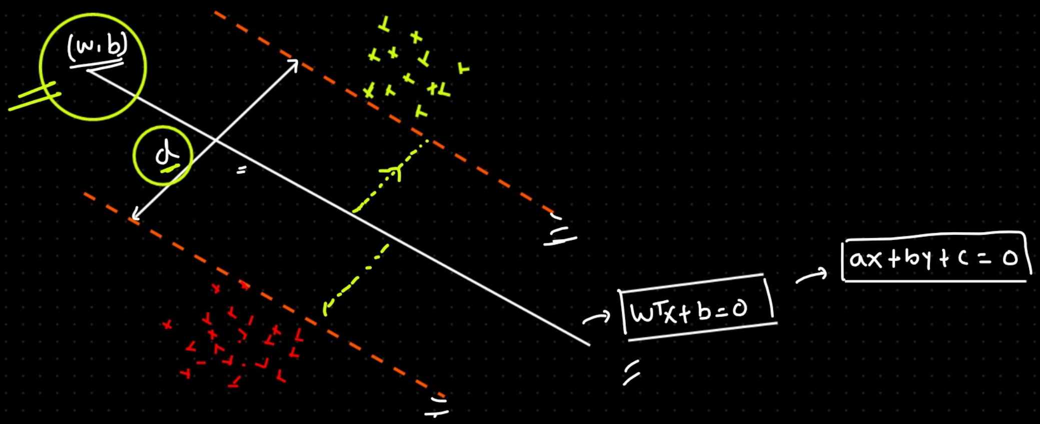
$2(-2) + 3(0) + 3$

$-4 + 0 + 3 = -1$

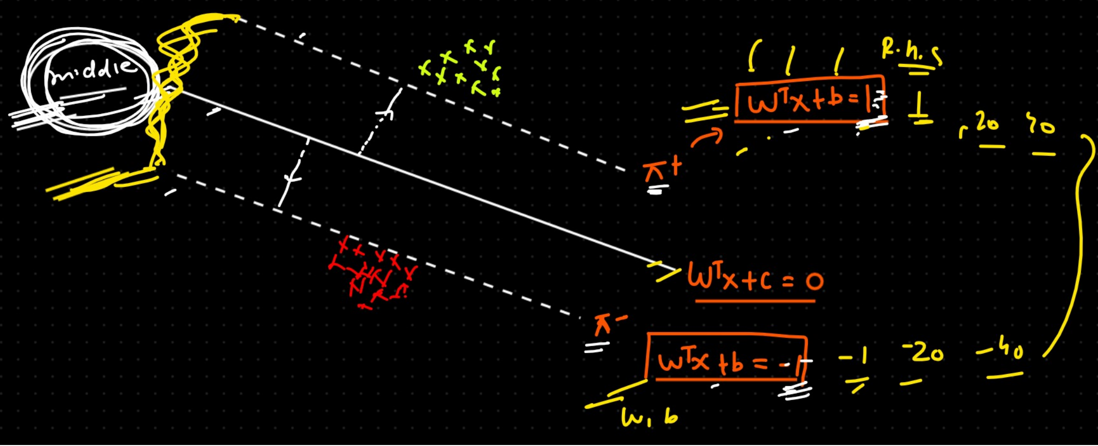
$(3,2)$

$(2x + 3y + 3 = 0)$

$\{ (-2,0) \}$



$$\hat{y} \begin{cases} +1 & \text{if } W^T x + b > 0 \\ -1 & \text{if } W^T x + b < 0 \end{cases}$$



question → ① why, take +1 & -1 ⇒

② another number (-20, +20 | -30, +30 | -40, +40)

+1, +1

w, b

w,b LHS

10

0.1

$$\begin{cases} 2x+3y+3 = 1 \\ 2\cdot x+3y+3 = 0 \\ 2x+3y+3 = -1 \end{cases}$$

R.HS

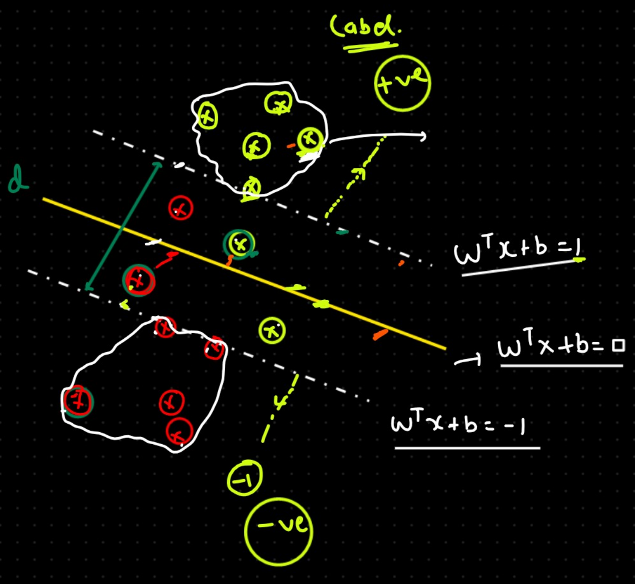
Constraint, d, soft classifier → practical h.y

rule

kernel trick

Support vector

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



$$y_i * (w^T x + b) \geq 1$$

$$\boxed{y_i * (w^T x + b) = +ve}$$

D.P

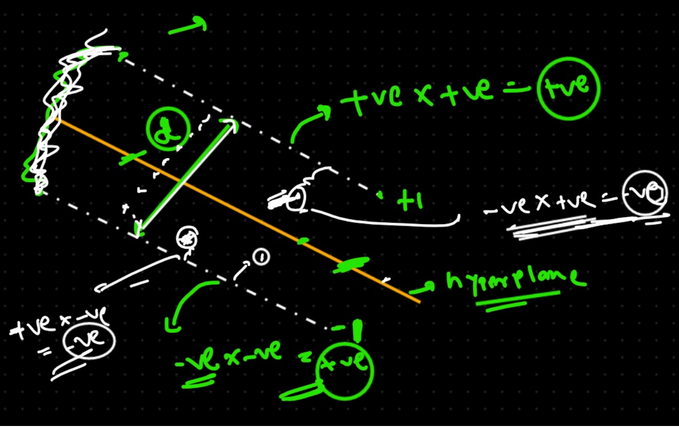
+ve * +ve = +ve ← correct classification

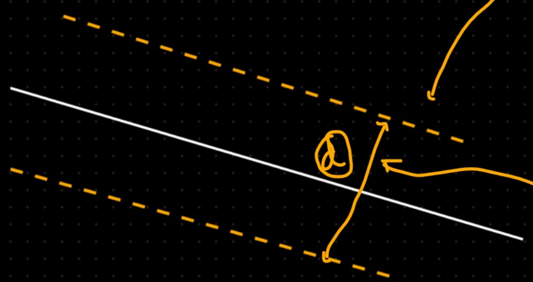
-ve * -ve = +ve ← correct classified

⇒ -ve * +ve = -ve ⇒ ?

⇒ +ve * -ve = -ve ⇒ ?

incorrect





$y_i(w^Tx+b) = +ve$

↓

Correct classification

$\boxed{y_i(w^Tx+b) \geq 1}$

$$\boxed{d = (\vec{x}_2 - \vec{x}_1) \cdot \frac{\vec{w}}{\|w\|}}$$

$$= \frac{\vec{x}_2 \cdot w - \vec{x}_1 \vec{w}}{\|w\|} \Rightarrow \frac{(1-b) - (-b-1)}{\|w\|}$$

$$\Rightarrow \frac{1 - \cancel{b} + \cancel{b} + 1}{\|w\|}$$

$$\Rightarrow \frac{2}{\|w\|} \rightarrow d \rightarrow \text{hard classifier}$$

$y_1(wx_i + b) = 1$

$1(\vec{w} \cdot \vec{x}_2 + b) = 1$ ← $+\pi$ above support vec

$\boxed{\vec{w} \cdot \vec{x}_2 = (1-b)}$

$-1(\vec{w}\,\vec{x}_1 + b) = 1$ ← lower support $-\pi$

$\boxed{\vec{w} x_1 = -b-1}$

↓

$\frac{2}{\|w\|}$

$w^Tx + b = +1$

$w^tx + b = -1$

10

-10

$+1, 20$

$-1, -20$

w, b

$\frac{20}{\|w\|}$

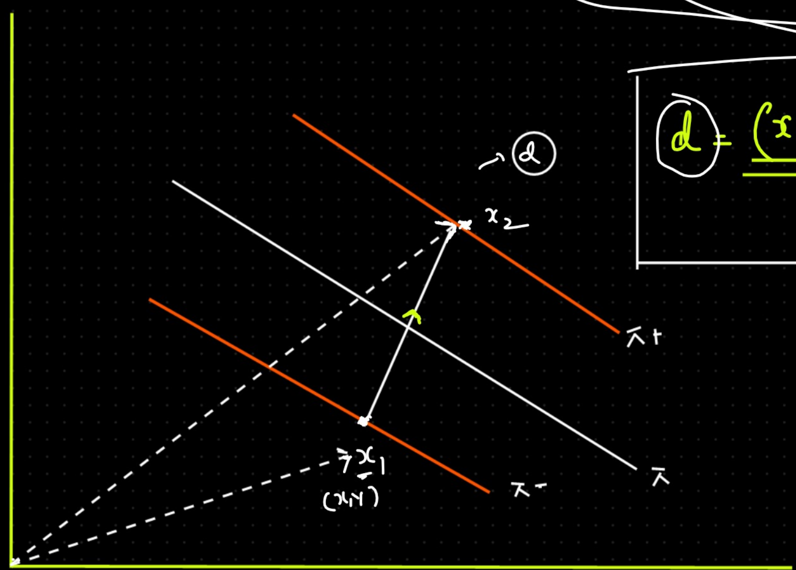
L.h.s

r.h.s

w, b

w, b

$$d = (x_2 - x_1) \cdot \frac{\vec{W}}{\|W\|}$$



$$Y_i(W^T X + b) = +ve$$

$$d = \frac{2}{\|W\|}$$

1st, 2nd

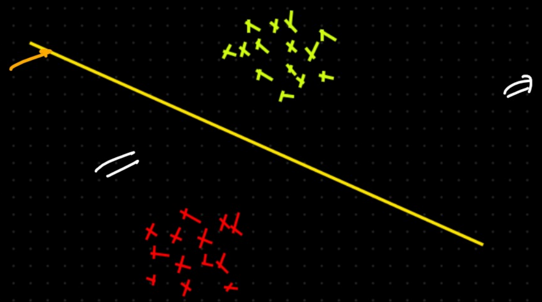
SVC

hard classifier

① Hard classifier

② Soft classifier

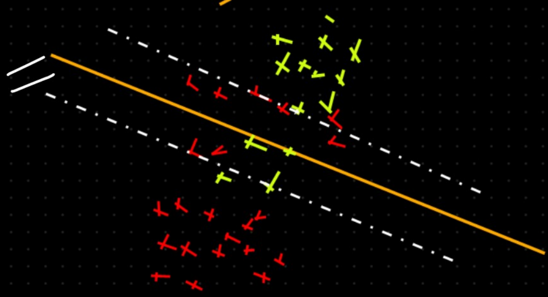
①



Hinge loss $d =$

②

missclassification



$$\left\{ \begin{array}{l} \underset{(w,b)}{\operatorname{argmax}} \left\{ \frac{2}{\|w\|} \right\} \\ \\ y_i * (w^T x_i + b) = +ve \end{array} \right\}$$

$2 \rightarrow \frac{1}{2} = 0.5 \rightarrow 0.001$

$100 \rightarrow \frac{1}{100}$

reverse minimum

$\left\{ \Rightarrow \max f(x) \Rightarrow \min \frac{1}{f(x)} \right\}$

$= \underset{}{\operatorname{argmax}} \frac{2}{\|w\|}$

$$\underset{(w,b)}{\operatorname{argmin}} \frac{\|w\|}{2} + C \sum_{i=1}^{n} \xi_i$$

zeta

hingeloss

$\Rightarrow$ softclassifier

SVC

Cost

$\left\{ \begin{array}{l} \text{Margin} \\ \text{(maximum margin)} \\ \text{(margin error)} \end{array} \right.$ $\left\{ \text{classification error} \right\}$

$$\underset{(w,b)}{\operatorname{argmin}} \frac{\|w\|}{2} + C \sum_{i=1}^{n} \xi_i \Rightarrow \text{hingeloss}$$

zeta

Margin error

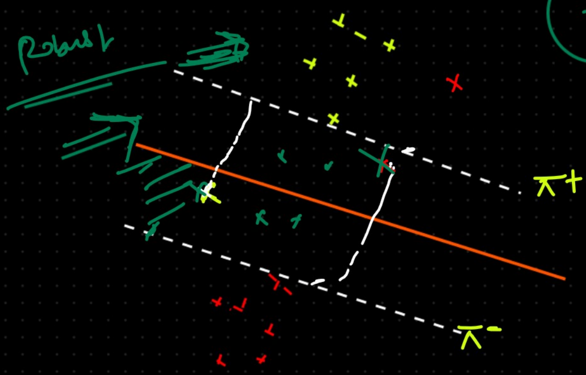
$\downarrow$

margin

classification error

how much miss classified point you will consider

Robust



real $\rightarrow \dfrac{\|w\|}{2} + C\sum_{i=1}^{n} \xi_i = \xi$ → 0

Hard classifier

hyperparameter

5

regulate

soft-margin support vector machine

$$\min \quad \tfrac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{\ell}\xi_i$$

$$\text{s.t.} \quad y_i(\mathbf{w}'\mathbf{x}_i - b) \geq 1 - \xi_i \quad \forall i = 1\ldots\ell$$

$$\xi_i \geq 0$$

best model min

1

$$\left\{ \frac{\|w\|}{2} + C\sum_{i=1}^{n} \xi_i \right\} \rightarrow \text{large}$$

Dist — error

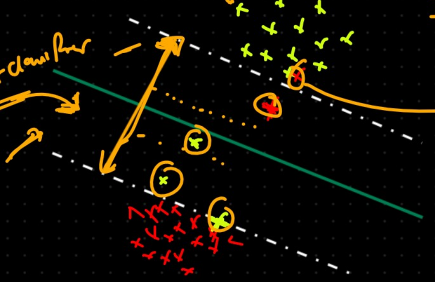
$\sum K$ dist

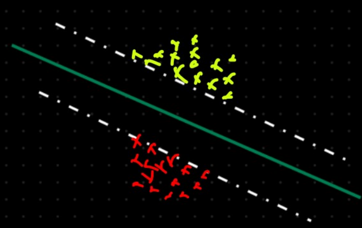
2

Hard classifier

real time

soft classifier





hyperparameter tuning

Kernel trick, Practical of SVC

↓

30 min

start