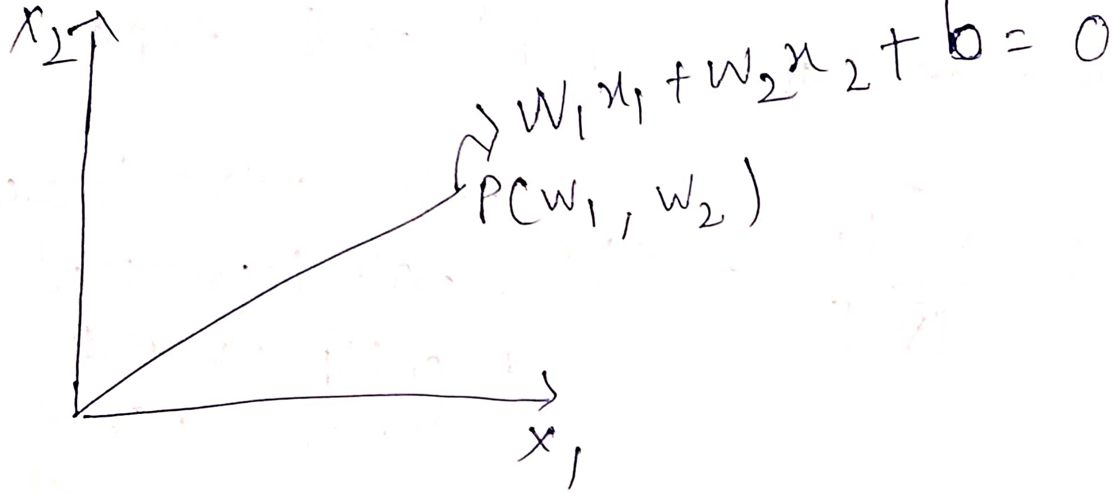
Support Vector Machine (SVM)

① Classfication → SVC (Support vector classfier)

② Regression → SVR (Support vector Regressor)

basics

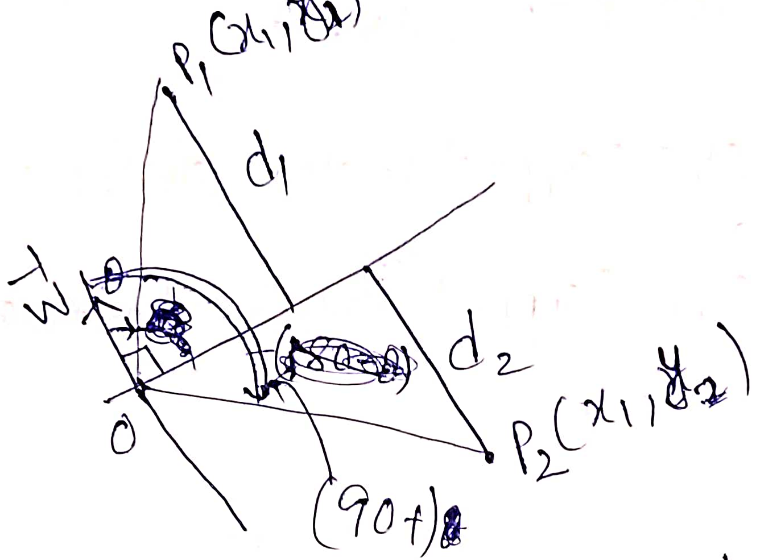


$w_1x_1 + w_2x_2 + b = 0$

$\Rightarrow w^T x + b = 0 \quad \left[\because \; \overbrace{\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \begin{bmatrix} x_1 & x_2 \end{bmatrix}}^{w_1x_1 + w_2x_2} \right]$

$\Rightarrow w^T x = 0 \quad [\because \text{ intercept } b = 0]$

equation of a line passing through origin.



Distance of a point to the plane

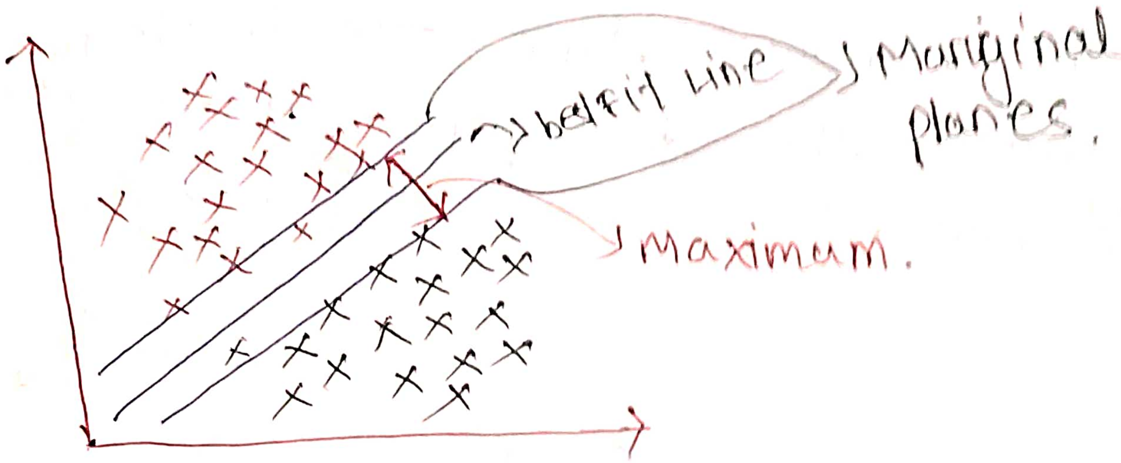
is $d_1 = \dfrac{w^T P_1}{\|\vec{w}\|} = +ve$ [always +ve because angle between $\vec{w}$ and $\vec{OP}$ always less then 90°]

$d_2 = \dfrac{w^T P_2}{\|\vec{w}\|} = -ve$ [angle between $\vec{w}$ and $\vec{OP_2}$ is always greater then 90°]

Geometric Intution Behind SVM

→ best fit line as per Logistic regression.

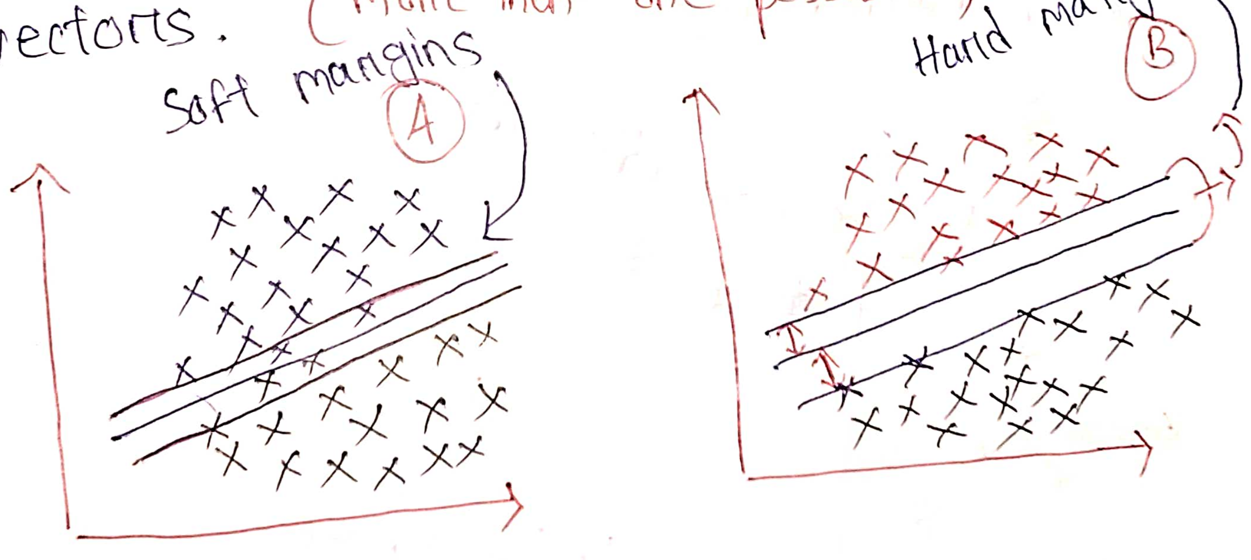
with Supperr vector classifier



→ Along with the best fit Line we ~~try to~~ make two marginals planes such that distance between two planes is maximum.

Support vectors

The lines on which data points touches the marginal plane is called support vectors. (more then one possibles)

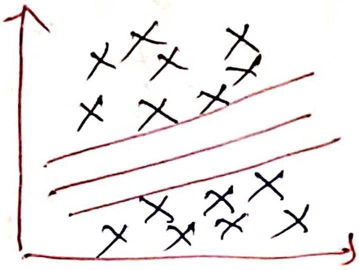


Q which is better marginal planes (A or B)

definitely B is better because it has maximum marginal distance.

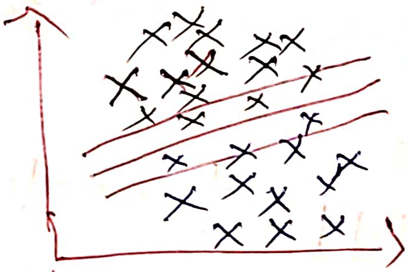
Hard Margins (without errors)

If the marginal plane able to clearly separate the data points it called hard margins



Soft Margins (with error)

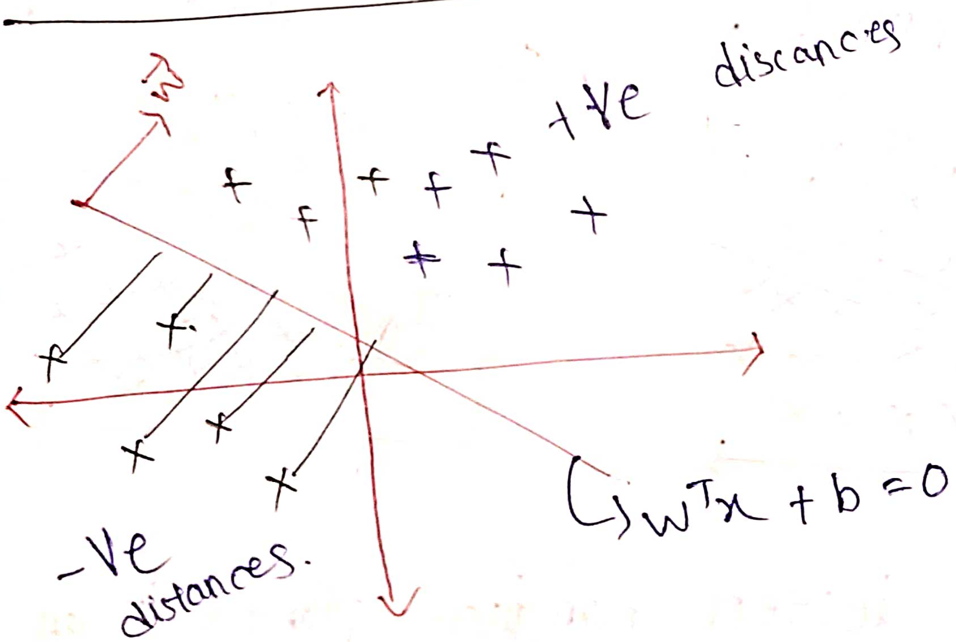
If the plane not able to clearly separate the data points then it's called soft margin



→ In real life we get soft margins

→ marginal planes are equidistant ($d_1 = d_2$)

SVM mathematical Intution



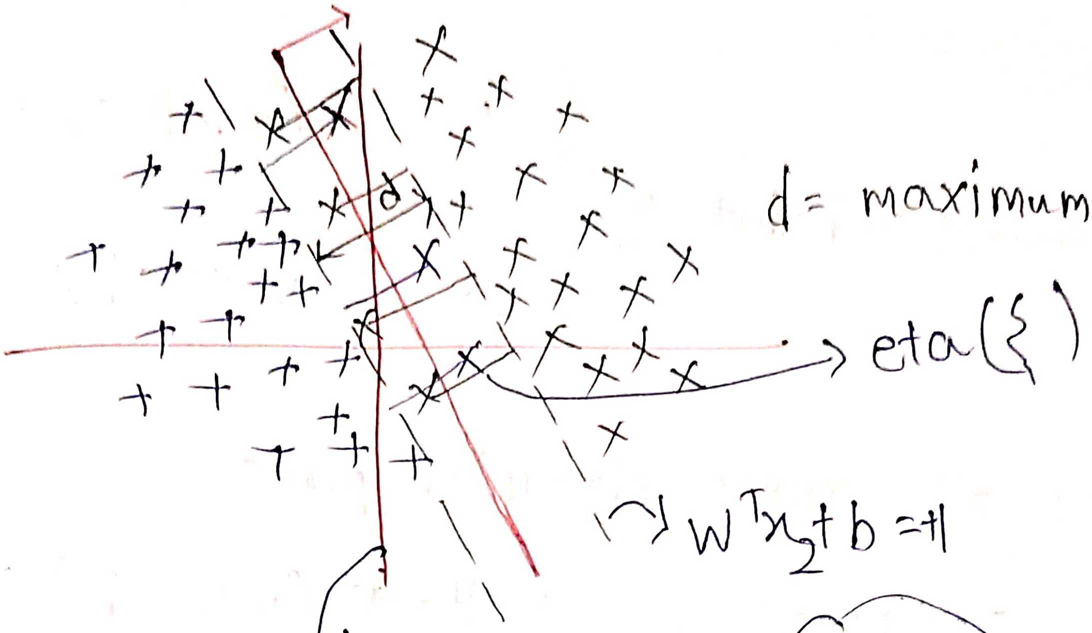


Fig-1

$$w^T x_1 + b = +1$$
$$w^T x_2 + b = -1$$
$$(-) \quad (-) \quad (+)$$

$$\Rightarrow w^T(x_1 - x_2) = 2$$

$$\Rightarrow \frac{w^T(x_1 - x_2)}{\|w\|} = \frac{2}{\|w\|}$$ (for converting ~~it~~ to unit vector we devide $\|w\|$)

Cost function

Maximize $\frac{2}{\|w\|}$ by changing w,b

⇓

distance between marginal planes.

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

For all correct classified points

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

Maximize $\frac{2}{||w||}$ with changing w,b

$\Downarrow$

minimize $\frac{||w||}{2}$ by changing w,b

So

<u>cost function</u>

Soft Margins eta

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

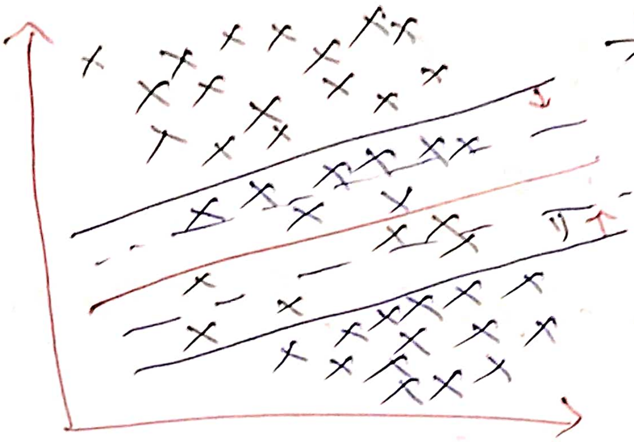
$\Downarrow$

Hinge loss

C_i = How many points we can ignore for misclassification

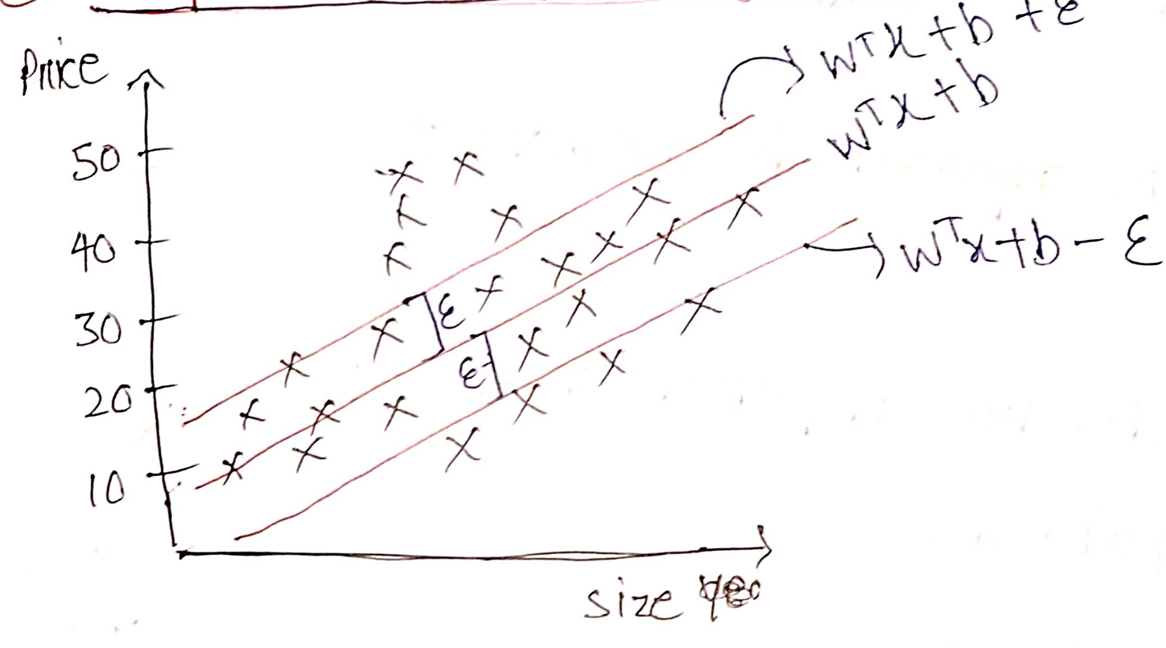
For fig-1 its $C_i = 7$

ξ_i = summation of the distance of incorrect data points from the marginal points.



→ As both C_i & eta is high for this plot we can reduce this by closing marginal planes.

② Support Vector Regressor (SVR)



∴ ε = Marginal error from best fit line.

Cost function

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

→ Hinge Loss

Constraint

$$|y_i - w_i x_i| \le \varepsilon + \xi_i$$

for point lies inside ε

$$|y_i - w_i x_i| \le \varepsilon$$

for point lies outside ε, ξ_i (eta) added

$$|y_i - w_i x_i| \le \varepsilon + \xi_i$$

$\varepsilon \to$ marginal error
$\xi \to$ Error outside margin.

Q. SVM impacted by outliers ?

→ yes

Q. Standardize required or not ?

→ Required.

SVM kernels

In this figure our line not able to classify data points (accuracy 50% or less)

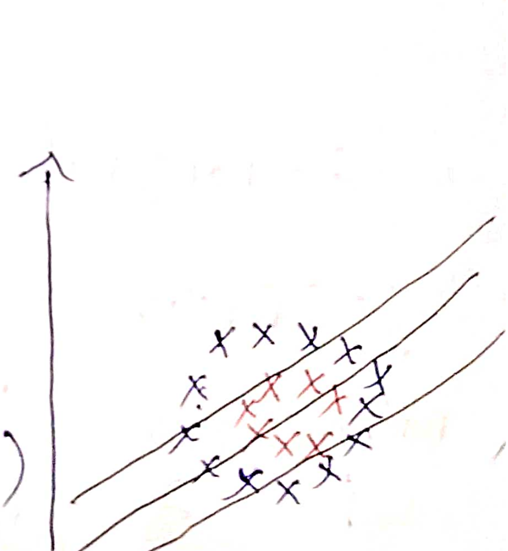
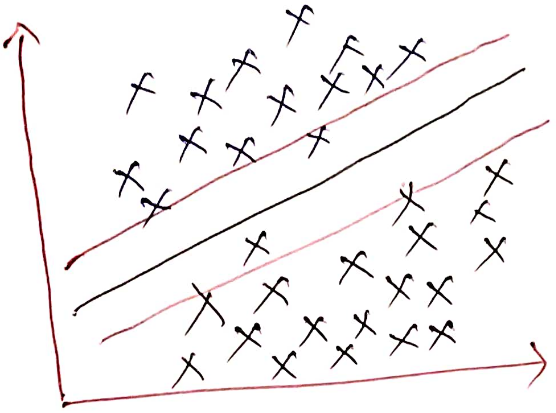


fig-2

for that we use different transformation convert to 3D by using SVM kernels.

examples

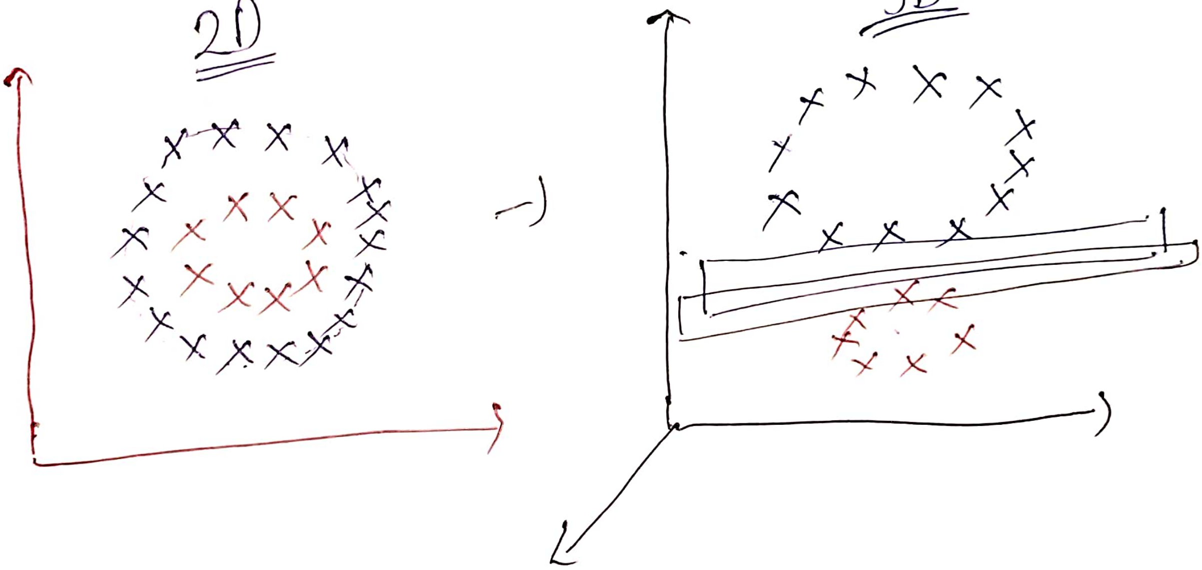


Kernel
↓
Linear SVC

kernel = linear.

2D

3D



Types

1. polynomial kernel
2. RBF Kernel (default)
3. Sigmoid kernel.
4. Linear (~~default~~)