Mo Tu We Th Fr Sa Su

Memo No.

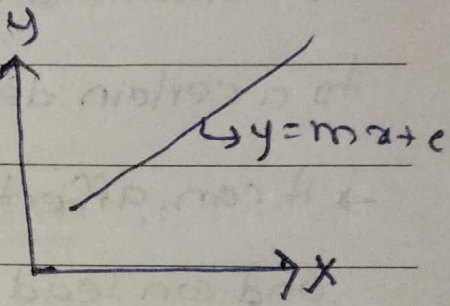
Date / /

Support Vector Machine

→ classification → SVC (support Vector C.)

→ Regression → SVR (support vector regression)

Basic equation of line bye:



$$y = mx + c$$

≅

$$y = \beta_0 + \beta_1 x$$

≅ ↓ can also be written as

$$ax + by + c = 0$$

⇓ → coefficient

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

Let us have a one equation be

$$a_1 x_1 + b x_2 + c = 0 \quad \text{—— (A)}$$

Eq^n (A) can also be represented as

$$w_1x_1 + w_2x_2 + b = 0 \quad \text{——— (B)}$$

$$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \begin{bmatrix} x_1 & x_2 \end{bmatrix} = w_1x_1 + w_2x_2$$

$\therefore \; w = [w_1 \quad w_2]$

$\Downarrow$ transpose

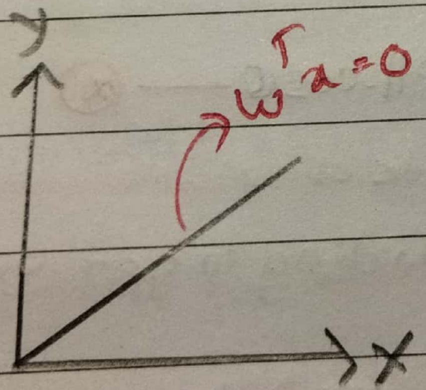
$$w^T = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

$\therefore$ Eq^n (B) can also be represented as

$$w^Tx + b = 0 \quad \text{——— } eq^n \text{ (C)}$$

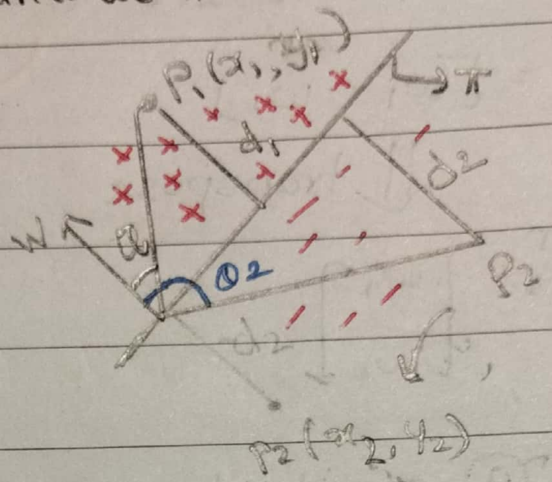
If Eq^n c passes through the origin then,

$b = 0$.



Memo No.
Date / /

let say, we have a line a 2D line. and let denotes it by π. and we have a vector W, which is $\perp$ to line π.



let $P_1(x_1, y_1)$ and $P_2(x_2, y_2)$ be two points.

Now, we have to find distance between ~~to~~ P_1 and line π and P_2 and line π.

So, the distance between P_1 and line is.

$$d_1 = \frac{w^T \cdot P_1}{\|w\|} \Rightarrow \|W\| \cdot \|P\| \cdot \cos\theta \text{ --- } (*)$$

$\|w\|$ ← magnitude of w
↳ we are dividing to get unit vector.

The angle θ_1 is less than 90.
↳ so, value of d_1 is always +ve.

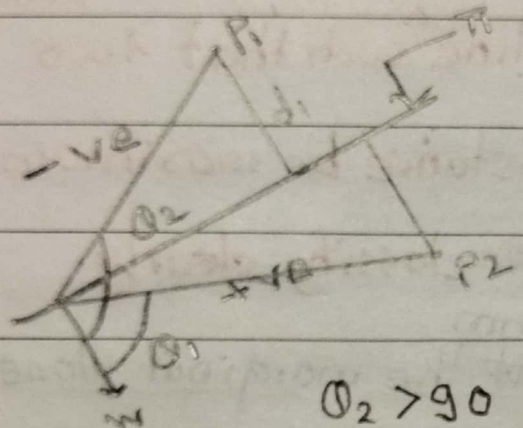
So, we can conclude, the distance above, the plane is always positive since θ is always less than 90°.

The angle θ_2 is always greater than $90°$

↳ So, the value of d_2 is always -ve.

So, we can conclude, the distance of the any point lying below the plane is negative.

Another case:



-ve if vector is in oposite direction.

+ve if vector is pointing in same direction.

$\theta_2 > 90$

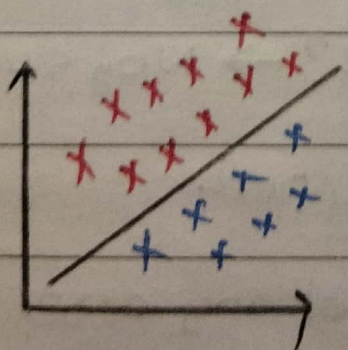
$\theta_1 < 90$

Important Note

We are just separating data. Above plane +1, below plane -1

Geometric Intution Behind Suppor Vector machine

Logistic → Basically used for binary classification



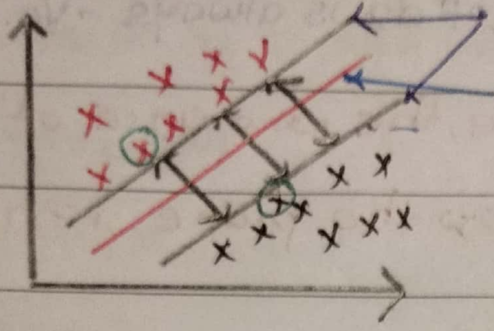
⇒ logistic regression

Mo Tu We Th Fr Sa Su

Memo No.

Date / /

Support Vector classifier



marginal plane such that distance between 2 marginalplane is maximum.

best fitline

Aim: To create a best fit line along with 2 marginal plane such that two marginal plane's distance be maximum so it can seperate or classify clearly.

The nearest datapoint ~~that~~ from the marginal plane pass as known as support vector. We can have more than 1 support vector

Question:

In a real world, do you think we will get clearly seperable line? → No, so many dataset will get overlapped.

Hardmargin: We won't get any error and marginal plane clearly classify the datapoints.

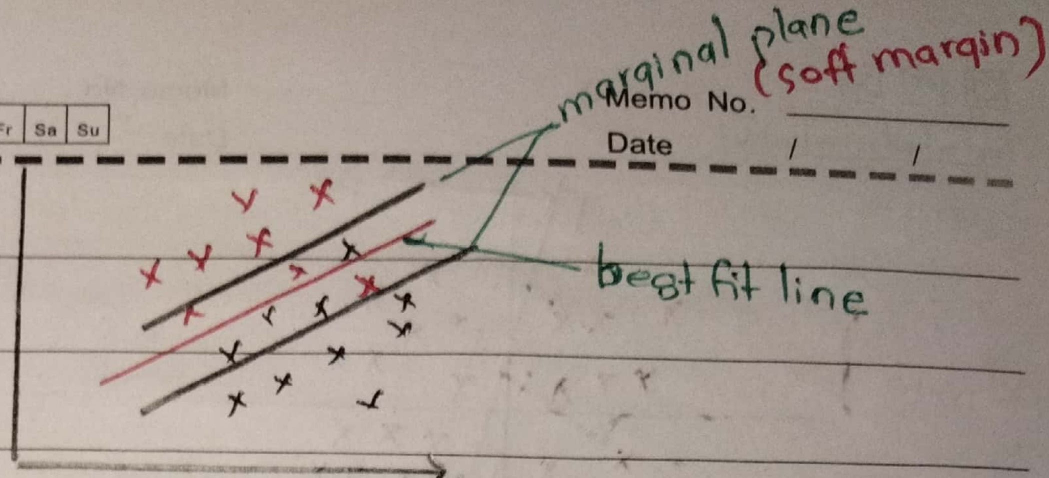


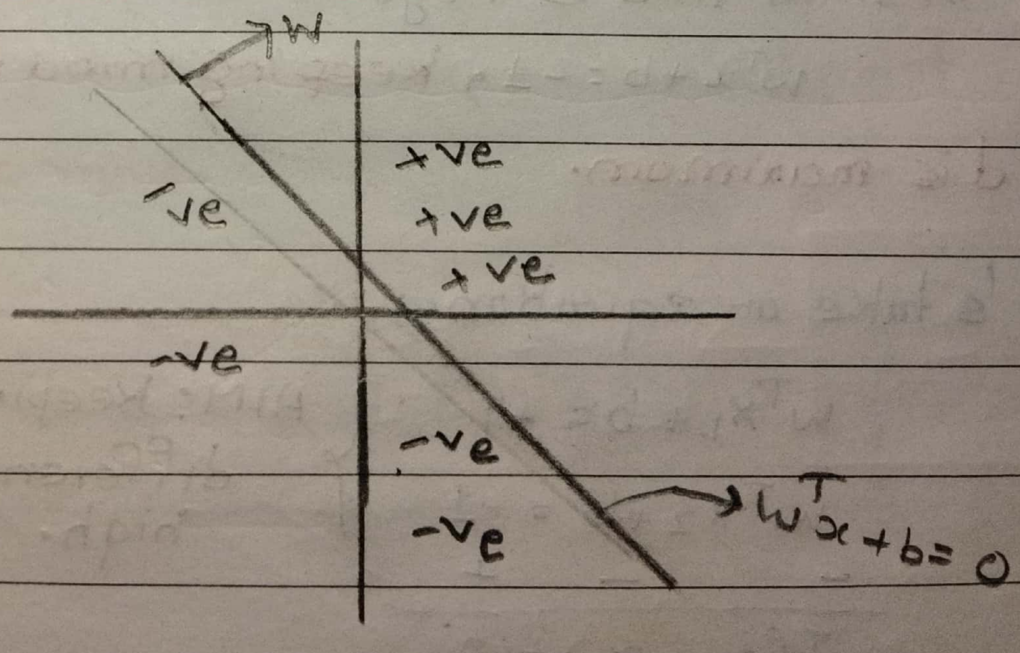
Fig: Real world dataset & SVC.

Soft Margin
↳ marginal plane with error.

* Model with higher 'D' is better model where D is distance between two marginal plane

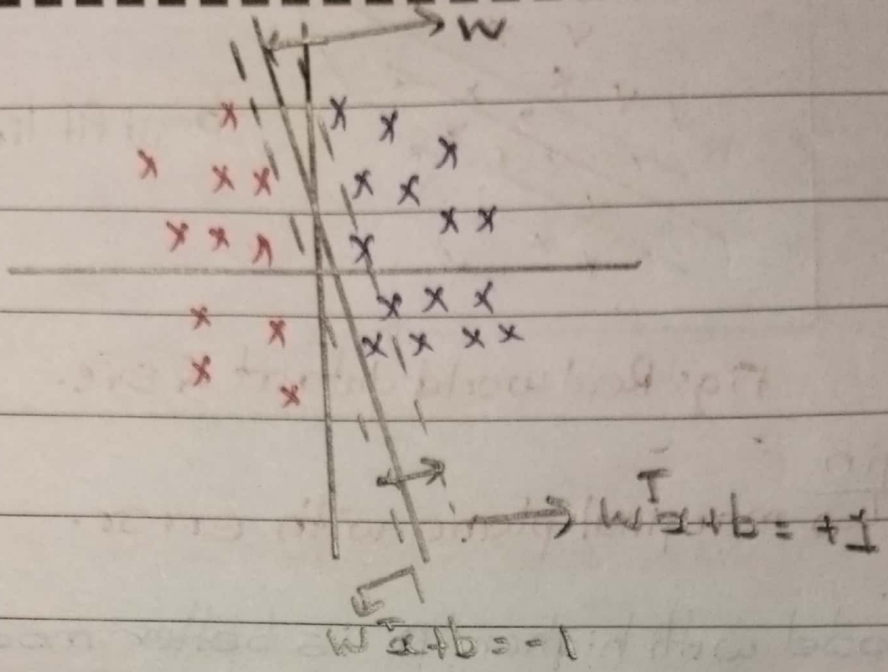
* The marginal plane should be in equidistance from the best fit line.

SVM mathematical Intuition



Memo No.

Date / /



Explanation:

Since, at the direction w is point, it is +ve.

So, $w^Tx+b=+1$ for the marginal plane with is above the best fit line

and, another one, which is below the best fit line is negative so.

$w^Tx+b=-1$, keeping mind the d is maximum.

let's take an equations:

$$
\begin{array}{r}
w^Tx_1+b=+1 \\
w^Tx_2+b=-1 \\
- \quad\quad - \quad\quad + \\
\hline
w^T(x_1-x_2)=2
\end{array}
$$

AIM: Keeping the difference very high.

Memo No. ______

Date / /

For calculating the unit vector,

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

cost function

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

changing w, b

Explanation: we have to change the value of w, and b to get the marginal plane such that we will get max^m distance.

Constraint such that

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

actual data

→ It is for all correct classified point

So, simply

actual data

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

Mo Tu We Th Fr Sa Su

Cost function

Memo No.

Date / /

Again

Maximize $\longrightarrow \frac{2}{||w||} \equiv$ minimize $\rightarrow \frac{||w||}{2}$

changing w, b changing w, b

This is for hard margin

$\therefore$ Cost function $\Rightarrow$

minimize $\longrightarrow \frac{||w||}{2} + C_i \sum_{i=1}^{n} \xi_i$

changing w, b

Eta

This is for soft margin.

C_i = How many point we can ignore for misclassification

we ignore some overlapped datapoints

hyper parameter

$\boxed{C_i \sum_{i=1}^{n} \xi_i}$ $\longrightarrow$ It gives the summation of the distance of incorrect data point from the marginal plane.

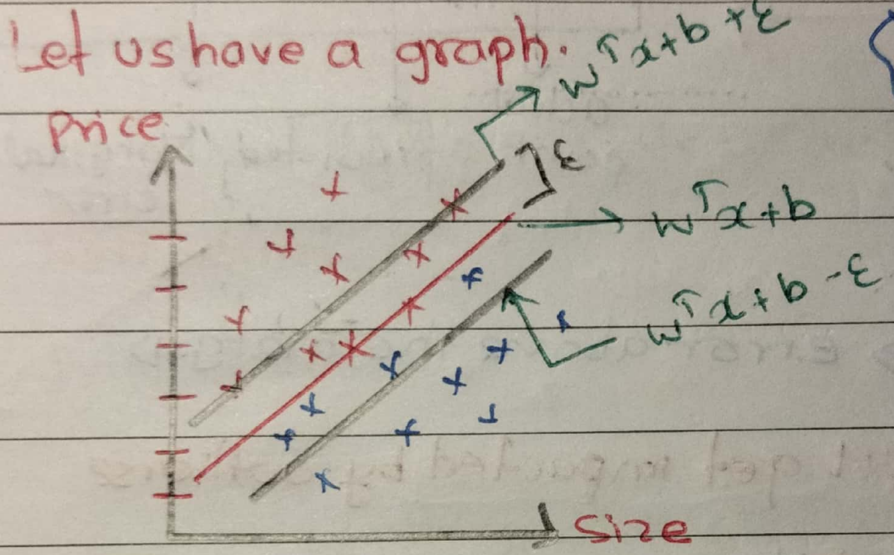
hinge loss

Mo Tu We Th Fr Sa Su   Memo No.   Date / /

Restricting ξ_i value will give the best marginal line.

Support Vector Regressor

Let us have a graph.



$\epsilon \Rightarrow$ marginal error ~~calculated from the overlapped dataset from the best fit line~~

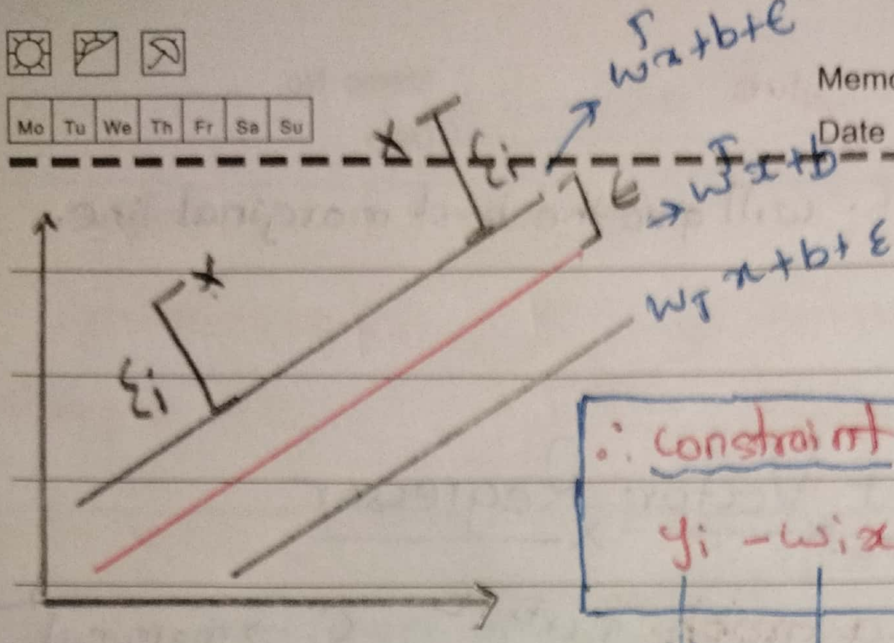
Cost function

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

$\xrightarrow{}$ hinge loss

Constraint:

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



Epsilon
Etai

$\therefore$ constraint

$$y_i - w_i x_i \le \epsilon + \xi_i$$

actual point — predicted — Marginal error

ξ Etai → error above the margin

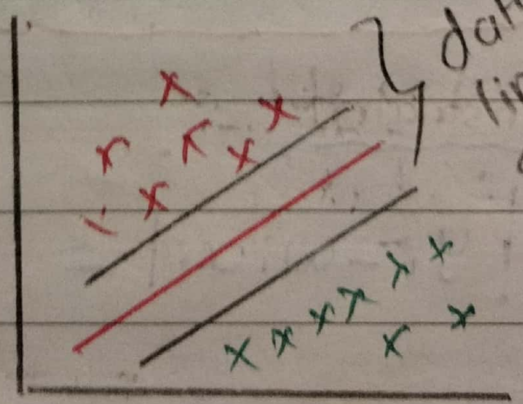
Q. Will SVM get impacted by outliers?

→ Yes.

Q. Will SVM needs data be standadize?

→ Yes

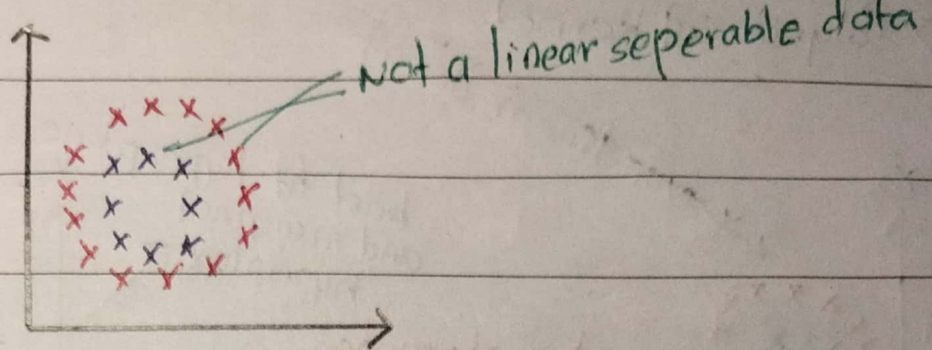
SVM Kernel



} dataset is linearly seperable.

linear SVC → we call kernel

But, when our datasets are not linearly seperable,



So, for solving these sorts of problem, we have SVM kernel.

What it does?

It perform some transformation technique [nothing but any mathematical formula] on data

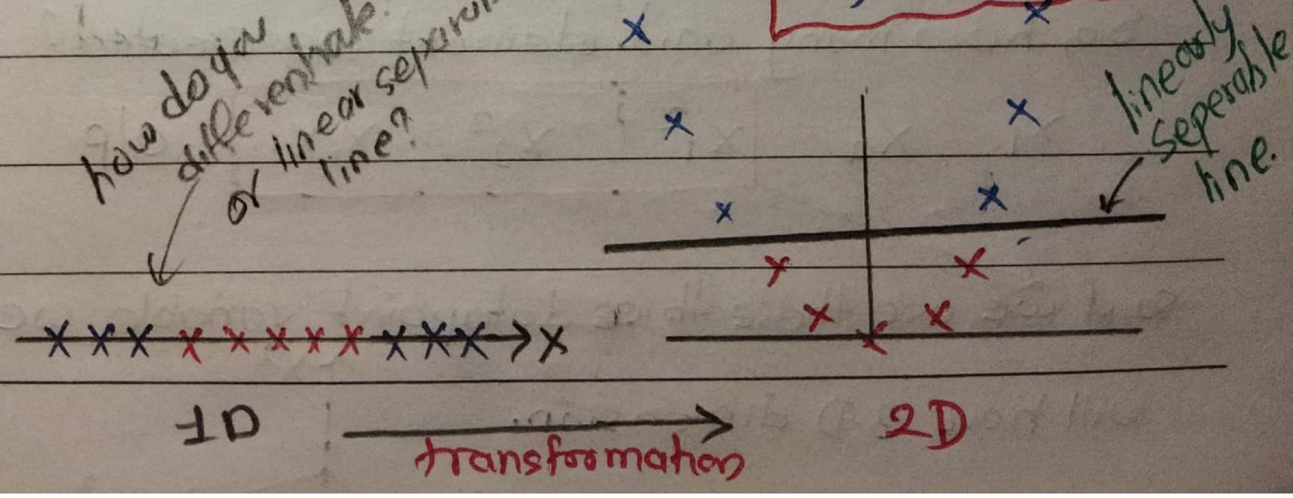
↓

Increasing the dimension of data.

Practical:

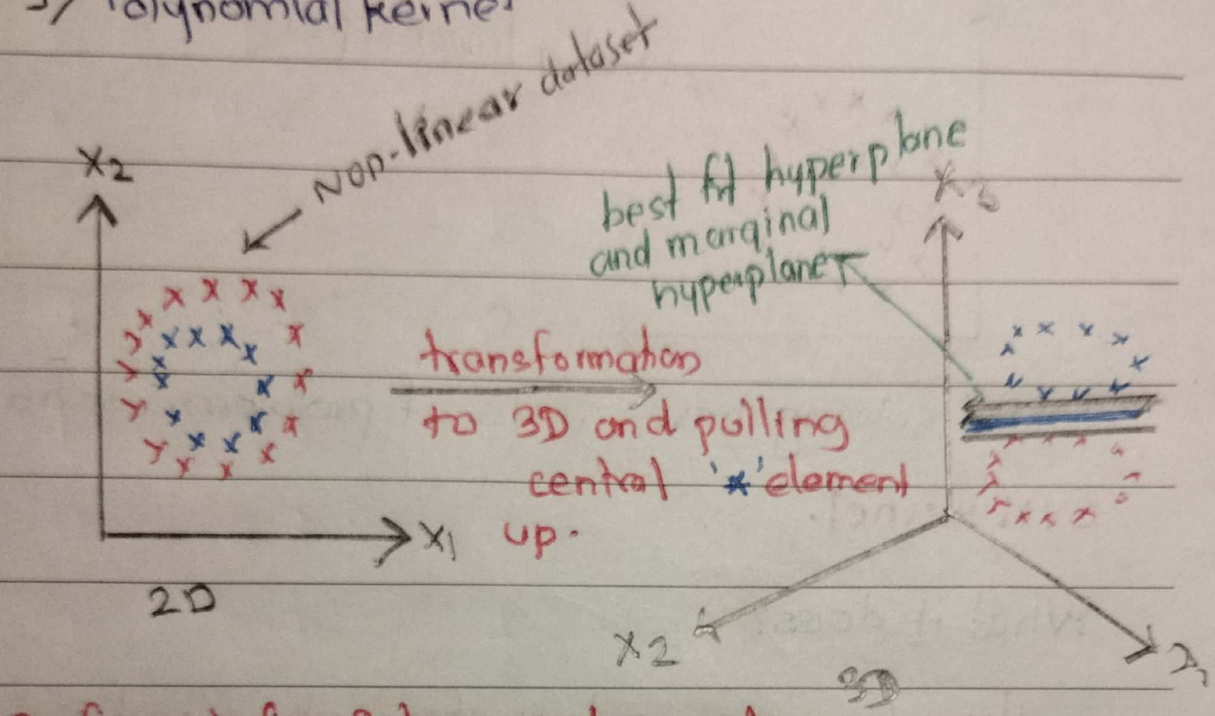
We have following datapoints:

Two steps here
1) 2D transformation
2) $y = x^2$



SVM Kernel types:-

1) Polynomial kernel



So, formula for Polynomial kernel

$$f(x_1, x_2) = (x_1^T \cdot x_2 + 1)^d$$

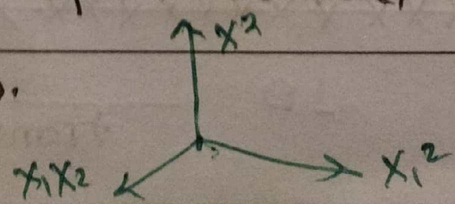
$d \leftarrow$ dimension
Here $d=3$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \cdot \begin{bmatrix} x_1 & x_2 \end{bmatrix} = \begin{bmatrix} x_1^2 & x_1 . x_2 \\ x_1 . x_2 & x_2^2 \end{bmatrix}$$

So, how many unit element we created?

x_1 x_2 $\boxed{x_1^2 \quad x_2^2 \quad x_1 x_2}$ o/p

So, if we use these three datapoint variable we will have 3-D dimension.



Memo No.

Date / /

2> Radial Basis function (RBf)

-> used for classification of non linear

$$K(x_1, x_2) = \exp\left(-\frac{\|x_1 - x_2\|}{2\sigma^2}\right)$$

where x_1, x_2 are vector point

Thus $\|x_1 - x_2\| \longrightarrow$ difference of equilidian distance.