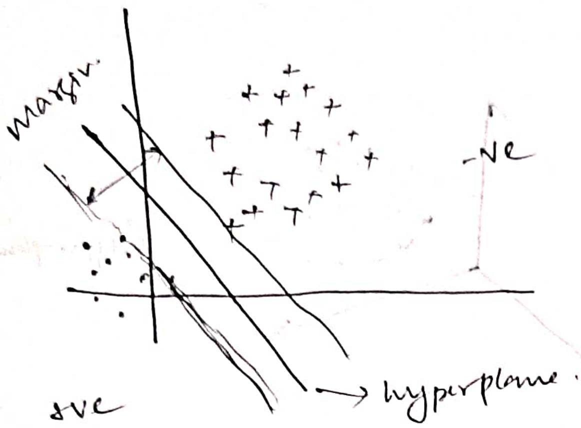
SVM

→ Basic steps:

1. Select 2 hyperplanes (2D) which separates data with no points b/w them.
2. Maximise distance.
3. Average line will be decision boundary.



$y = w^T x + b$

let $b = 0$,

$y = w^T x$.

In this case,

$w^T x + b \geq 1$ for $y_i = +1$

i.e. point in +ve region

& vice-versa.

$\therefore \; y_i * (w^T x_i + b_i) \geq 1$

$\therefore$ Optimization function $= (w^*, b^*)$ for $\max \frac{2}{\|w\|}$

$\hookrightarrow \min$ of $\frac{\|w\|}{2}$

(grad. descent)

For this case

$\therefore$ For this case:

$$(w^* \; b^*) = \min \frac{\|w\|}{2} + c \sum_{i=1}^{n} \xi_i$$

→ Regularization (hyperparameter)

How many errors can model can consider

Summation of values of error

→ SVM Kernels for Non-Linear cases.

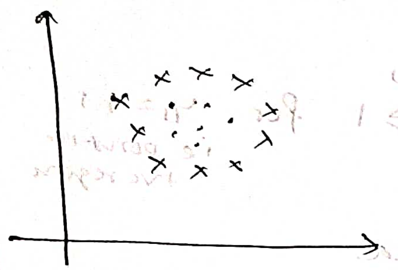
Types of Kernels:-

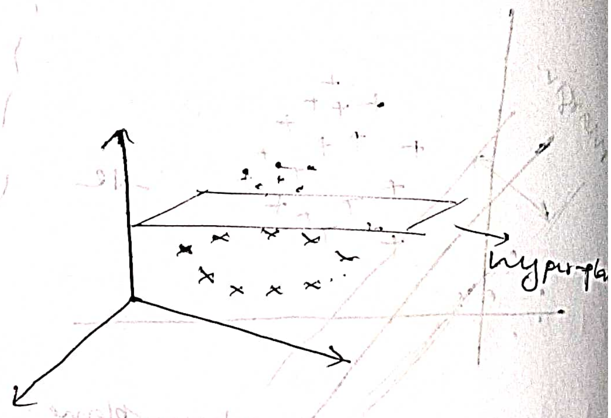
i. Polynomial Kernels
ii. RBF Kernels
iii. SIGMOID Kernels.

SOME TERMS:

1. Hard Margin - No error
2. Soft Margin - Some errors

→ How does kernel work.



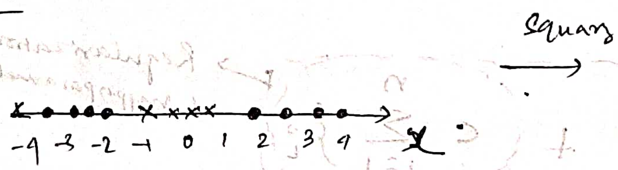


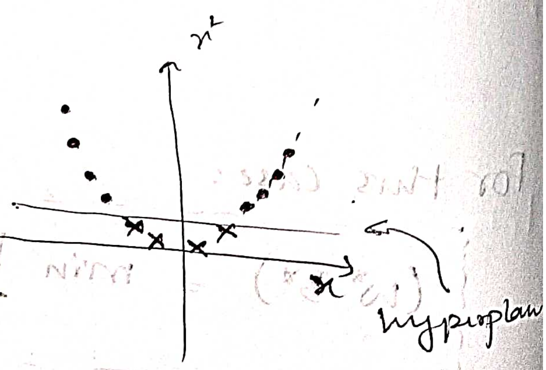
In this example 2d to 3d.

→ SVM kernel does some kind of transformation from lower to higher dimension.

Polynomial Kernel

1-D

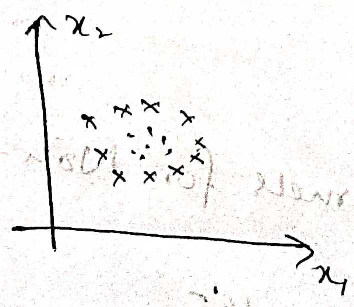




Convert 1D to 2D

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

Vector operation



$y = f(x_1, x_2)$,

{ (x_1, x_2) together is data.

$\therefore \quad f(x_1, x_2) = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \cdot [x_1 \ x_2] + 1$

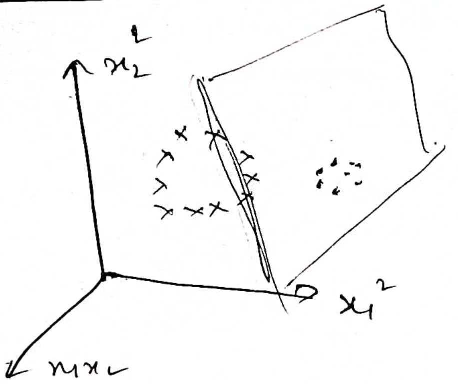
[~~dot~~ vector product so x_1x_2 only one]

$= \begin{bmatrix} x_1^2 & x_1x_2 \\ x_1x_2 & x_2^2 \end{bmatrix}$

$\therefore \quad x_1^2 \quad x_2^2 \quad x_1x_2$

$\therefore \quad x_1 \quad x_2 \quad y \quad x_1^2 \quad x_2^2 \quad x_1x_2$ } Use all 5 to create hyperplane. 5 dimensions.

Using only x_1^2, x_2^2, x_1x_2



Generalized

$$\boxed{K(x_i \, x_j) = (x_i x_j + 1)^d}$$

RBF Kernel

$$\boxed{K(x_i \, x_j) = e^{-\gamma (\|x_i - x_j\|)^2}}$$

Almost like ... seeing neighbourhood elements. Different from K-means.

Sigmoid Kernel

$$\boxed{K(x, y) = \tanh(\alpha x^T y + c}$$

→ <u>Hyper-Parameters to tune:</u>

(i) Regularization (C & errors)

(ii) Gamma (γ) (Distance factor from neighbours etc)

(iii) Margin

(iv) Kernel type