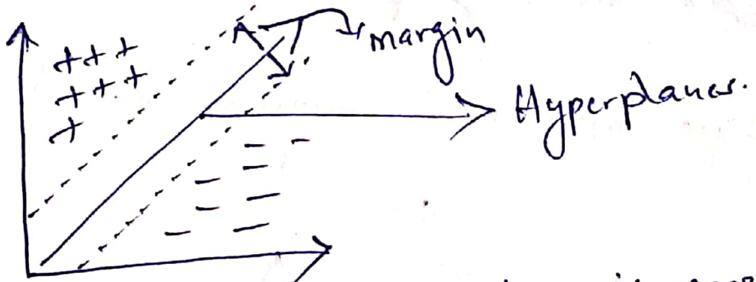
Support vector machines

problem statement:- Supervised Classification and Regression problems

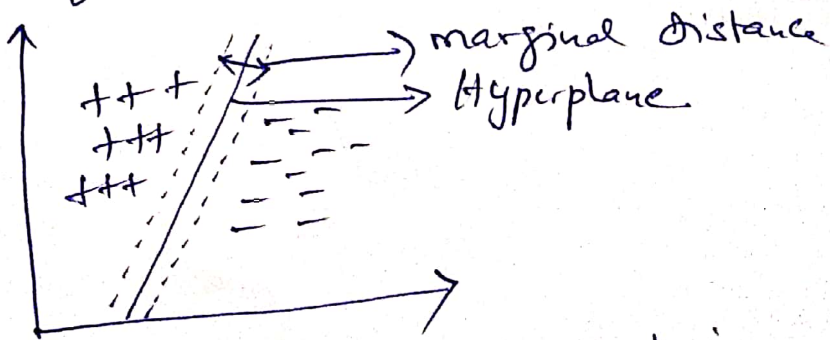
Support vectors
Hyperplanes
Marginal Distance
Linear Separable points
Non-Linear Separable

} What ?

Aim:- Linearly separates the classes using hyperplanes.



→ Distance b/w the two Dashed line is marginal distance
→ Distance is very important
→ Multiple hyperplanes is also possible.
→ Maximal marginal Distance should be obtained.

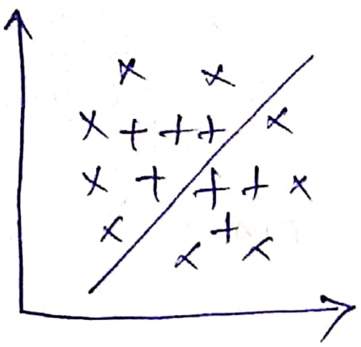


→ Hyperplane should be selected which is having maximal marginal distance.

→ The term 'hyperplanes' is mostly used in 2D and 3D ...

Non-Linear Separable:-

2-D

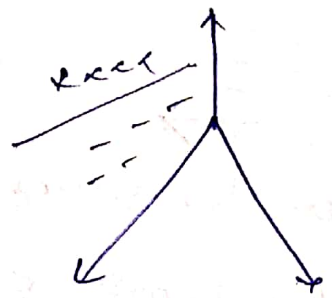


? ≤50%. Accuracy.

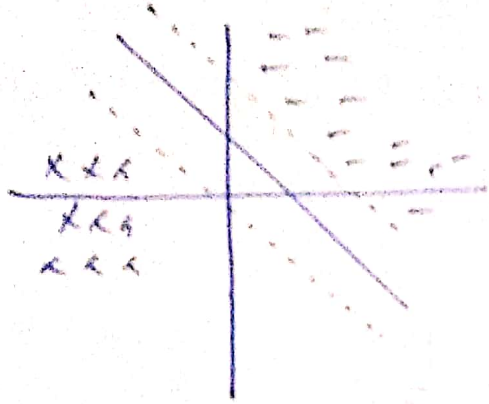
How to solve this problem? → SVM kernels.

Support vector → points that lies on the marginal line.

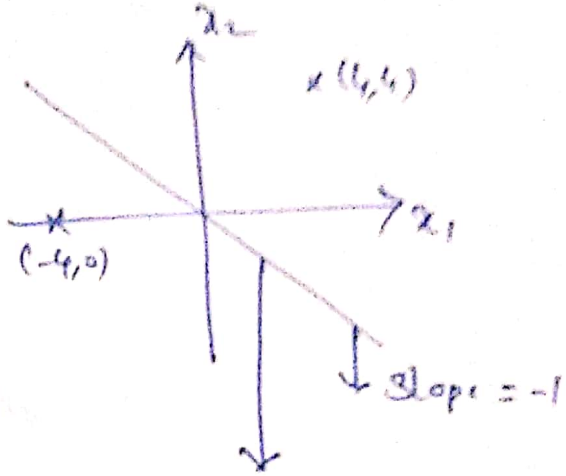
SVM Kernels - Converts low dimension to high dimension



Mathematical Concepts in SVM:



Derive the marginal plane.



$$y = w^T x + b$$

$$y = w^T x + 0$$
$$= w^T x$$

$w^T x + b = 0$ (eqn of hyperplane)

(b is 0 $\Rightarrow$ it passes thru origin)
intercept

$$= \begin{bmatrix} -1 \\ 0 \end{bmatrix} \begin{bmatrix} -4 & 0 \end{bmatrix} \quad w = \text{Slope}$$

$$= 4 \ (+ve)$$

Plotting any points below the hyperplane is +ve

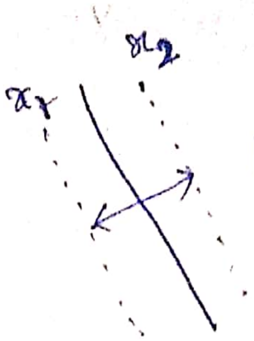
For (4,4)

$$y = w^T x + b$$
$$= w^T x$$
$$= \begin{bmatrix} -1 \\ 0 \end{bmatrix} \begin{bmatrix} 4 & 4 \end{bmatrix}$$
$$= -4 \ (-ve)$$

Plotting any points above the hyperplane is -ve.

$W^T x + b = -1$

$W^T x + b = +1$



how to Compute the distance $x_2 - x_1$

$$
\begin{aligned}
W^T x_1 + b &= +1 \\
- W^T x_2 \underset{(-)}{+} b &= \underset{(-)}{-}1 \\
\hline
W^T (x_1 - x_2) + 0 &= 2
\end{aligned}
$$

Normalize the magnitude. w to get rid of W^T

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

↓ maximum distance

<u>Objective:</u> Maximize the margin but still make the Correct predictions.

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

This can also be written as

$$y_i * W^T x_i + b_i \geq 1$$

If this does not hold true then there is misclassification

Loosen our model

$$\text{Max } \frac{2}{||W||} \rightarrow \min \frac{||W||}{2} + c_i * \sum_{i=1}^{n} \xi_i \quad \text{(Zeta)}$$

c_i - How many errors?

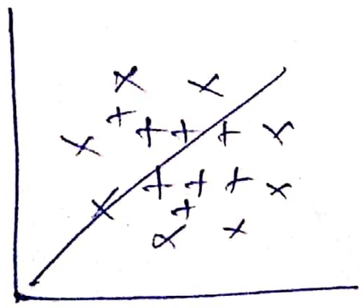
(Zeta) ξ_i - value of the error (Distance from the ~~hyperplane~~ marginal plane and the point).

C_i - Regularization parameter.

ξ_i - Slack variable.

SVM Kernels:

Margin
- Soft - Tolerate misclassification
- Hard - very rigid in classification, cause overfitting.

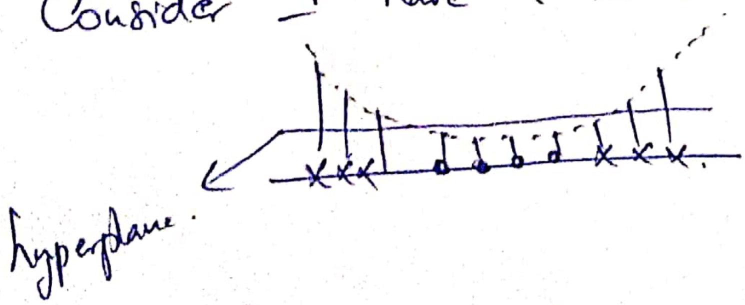


① polynomial kernel

② Radial Basis Function kernel

③. Sigmoid kernel.

Consider I have 1 dim.



$y = f(x) = x^2$

polynomial Kernel

x_1, x_2 - features.

$y = f(x_1, x_2)$

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

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

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

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