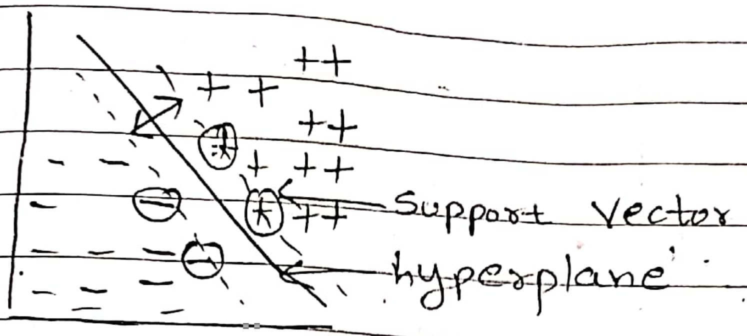
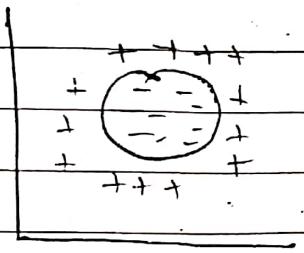
15/22

SUPPORT VECTOR MACHINE (SVM)

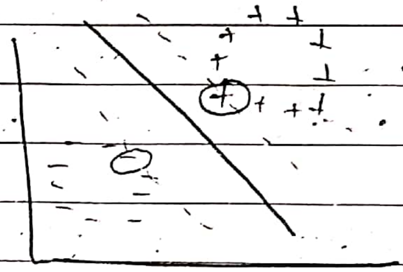


Linearly separable data

It is a supervised machine learning algori used for binary as well as multiclass classification. It is derived in 1992 It has a lower error rates



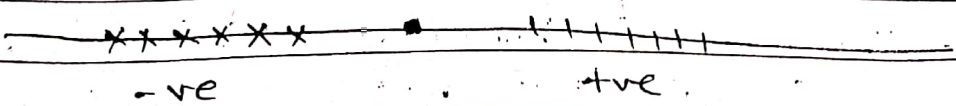
Non linearly separable



Linearly separabl

Kernel function

SVM maps input space the into a output space using a non linear mapping function such tha the data points becomes linearly separab in output space. Once the points become linearly separable then SVM discovers optimal separating hyperplane



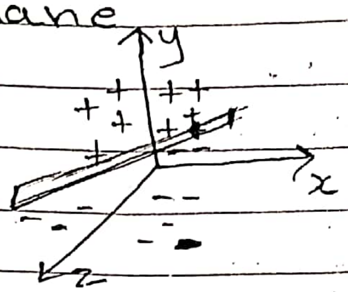
For one dimensional data the hyperpla is a point.

For two dimensional data the hyperplan

Page No.
Date

7/22

is a line.
~~Fore~~ For 3 dimensional data the hyperplane is a plane



For more than 3-D data it is called as hyper-plane.
The objective of SVM is to find optimal separating hyperplane because it correctly classifies the training data & also generalize better with unseen data

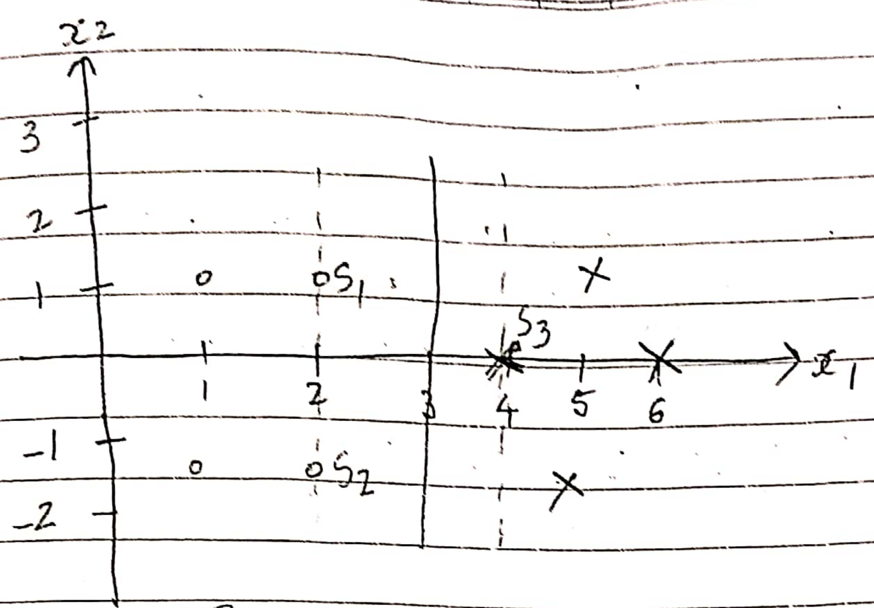
★ Relationship between margin & optimal hyper-plane

If a hyperplane is very close to the data points its margin will be very small. If the hyperplane is sufficiently away from the datapoints its margin will be large. The optimal hyperplane would be 1 with the biggest margin. Therefore, the objective of SVM is to find the optimal separating hyperplane which maximizes the margin of the training data.

Q] Find the optimal hyperplane for the set of data points {(1,1), (2,1), (1,-1), (2,-1); (4,0), (5,1), (5,-1), (6,0)}.

A] Class 1 - {(1,1), (2,1) (1,-1), (2,-1)}
Class 2 - {(4,0), (5,1), (5,-1), (6,0)}

8/8/22



$S_1 = \left\{\begin{bmatrix} 2 \\ 1 \end{bmatrix}\right\}$, $S_2 = \left\{\begin{bmatrix} 2 \\ -1 \end{bmatrix}\right\}$, $S_3 = \left\{\begin{bmatrix} 4 \\ 0 \end{bmatrix}\right\}$

Equation of a plane

$$y = ax + b$$

Equation of plane in SVM is

$$w^T \cdot x = 0 \qquad (w^T x + b = 0 \leftarrow \text{bias})$$

$\hookrightarrow y - ax - b = 0$

$$w = \begin{bmatrix} -b \\ -a \\ 1 \end{bmatrix} \qquad x = \begin{bmatrix} 1 \\ x \\ y \end{bmatrix}$$

$$w^T x = y - ax - b$$

Augment the support vector with 1 has a biased input i.e. $\tilde{S}_1 = \left\{\begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}\right\}$, $\tilde{S}_2 = \left\{\begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}\right\}$, $\tilde{S}_3 = \left\{\begin{bmatrix} 4 \\ 0 \\ 1 \end{bmatrix}\right\}$

* Find out 3 parameters α_1, α_2 & α_3 whic are lagregin multipliers for each support vectors.

The equation of the hyperplane that discriminates positive class from negative class is given as

18/10/22

$$\boxed{\tilde{W} = \sum_{i=1}^{n} \alpha_i S_i}$$

where α_i is lagreain multipliers
S_i - augmented support vector
n - no of support vectors

$$\boxed{\tilde{W} = \alpha_1 \tilde{S_1} + \alpha_2 \tilde{S_2} + \alpha_1 \tilde{S_3}}$$

Scanned with CamScanner

8/10/22

A) We need to find out values of α_1, α_2 & α_3 which are lagraine multipliers by creating 3 simultaneous equations for 3 support vector.

$$\alpha_1 \tilde{S}_1 \tilde{S}_1 + \alpha_2 \tilde{S}_1 \tilde{S}_2 + \alpha_3 \tilde{S}_1 \tilde{S}_3 = -1 \quad (1)$$
$$\alpha_1 \tilde{S}_1 \tilde{S}_2 + \alpha_2 \tilde{S}_2 \tilde{S}_2 + \alpha_3 \tilde{S}_2 \tilde{S}_3 = -1 \quad (2)$$
$$\alpha_1 \tilde{S}_1 \tilde{S}_3 + \alpha_2 \tilde{S}_2 \tilde{S}_3 + \alpha_3 \tilde{S}_3 \tilde{S}_3 = +1 \quad (3)$$

$$6\alpha_1 + 4\alpha_2 + 9\alpha_3 = -1$$
$$4\alpha_1 + 6\alpha_2 + 9\alpha_3 = -1$$
$$9\alpha_1 + 9\alpha_2 + 17\alpha_3 = +1$$

$\tilde{S}_1 = \begin{Bmatrix} 2 \\ 1 \\ 1 \end{Bmatrix}$, $\tilde{S} = \begin{Bmatrix} x_1 \\ x_2 \\ b \end{Bmatrix}$

$\tilde{S}_2 = \begin{Bmatrix} 2 \\ -1 \\ 1 \end{Bmatrix}$

$\tilde{S}_3 = \begin{Bmatrix} 4 \\ 0 \\ 1 \end{Bmatrix}$

$\alpha_1 = -3.25$

$\alpha_2 = -3.25$

$\alpha_3 = 3.5$

$$\tilde{W} = -3.25 \begin{Bmatrix} 2 \\ 1 \\ 1 \end{Bmatrix} + (-3.25) \begin{Bmatrix} 2 \\ -1 \\ 1 \end{Bmatrix} + 3.5 \begin{Bmatrix} 4 \\ 0 \\ 1 \end{Bmatrix}$$

$$= \begin{Bmatrix} -6.5 \\ -3.25 \\ -3.25 \end{Bmatrix} + \begin{Bmatrix} -6.5 \\ 3.25 \\ -3.25 \end{Bmatrix} + \begin{Bmatrix} 14 \\ 0 \\ 3.5 \end{Bmatrix}$$

$$= \begin{Bmatrix} 1 \\ 0 \\ -3 \end{Bmatrix}$$

$W = \begin{Bmatrix} 1 \\ 0 \end{Bmatrix}$ $\downarrow$ hyperplane $b = \{-3\}$ $\downarrow$ bias

W gives orientation of hyperplane & b

2/8/22

is offset ~~for~~ for the origin from the hyperplane, Thus, W & b together define the hyperplane
$\boxed{y = wx+b}$ is the equation of hyperplane where x is input vector if $y<0$ then x is classified as negative class belongs to negative class if $y>0$ then x is positive class belonging to positive class.
If $y=0$ the x is on hyper plane.
Using given SVM model classify the following input data

$x = \begin{Bmatrix} 2 \\ 2 \end{Bmatrix}$ $x = \begin{Bmatrix} 5 \\ 5 \end{Bmatrix}$

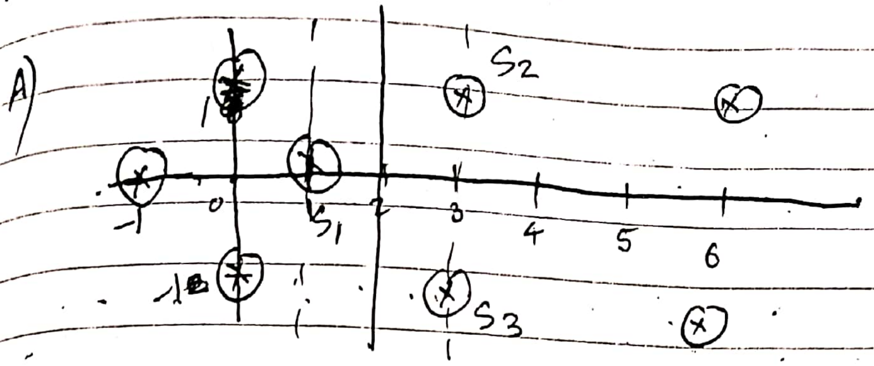
$$y = w^T x + b$$
$$= \begin{Bmatrix} 1 \\ 0 \end{Bmatrix} \begin{Bmatrix} 2 \\ 2 \end{Bmatrix} - 3$$
$$= 2 - 3$$
$$= -1$$

$$y = w^T x + b$$
$$= \begin{Bmatrix} 1 \\ 0 \end{Bmatrix} \begin{Bmatrix} 5 \\ 5 \end{Bmatrix} - 3$$
$$= 5 - 3$$
$$= 2$$

Q) Following are data samples given in 2D space $+ve = \left\{ \begin{pmatrix} 3 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ -1 \end{pmatrix}, \begin{pmatrix} 6 \\ 1 \end{pmatrix}, \begin{pmatrix} 6 \\ -1 \end{pmatrix} \right\}$.
$-ve = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \end{pmatrix} \right\}$
Using this sample data Construct a hyperplane which discriminates +ve & -ve class

22/8/22

A)



$$\tilde{s}_1 = \begin{pmatrix}1\\0\\1\end{pmatrix}, \quad \tilde{s}_2 = \begin{pmatrix}3\\1\\1\end{pmatrix}, \quad \tilde{s}_3 = \begin{pmatrix}3\\-1\\1\end{pmatrix}$$

$$\cancel{s_1 = \begin{pmatrix}1\\0\end{pmatrix}, \ s_2 = \begin{pmatrix}3\\1\end{pmatrix}, \ s_3 = \begin{pmatrix}3\\-1\end{pmatrix}, \ b = 1}$$

$$\tilde{s}_1 \cdot \tilde{s}_1 = \begin{pmatrix}1\\0\\1\end{pmatrix}\begin{pmatrix}1\\0\\1\end{pmatrix} = 1+0+1 = 2$$

$$\tilde{s}_1 \cdot \tilde{s}_2 = \begin{pmatrix}1\\0\\1\end{pmatrix}\begin{pmatrix}3\\1\\1\end{pmatrix} = 3+0+1 = 4$$

$$\tilde{s}_1 \cdot \tilde{s}_3 = \begin{pmatrix}1\\0\\1\end{pmatrix}\begin{pmatrix}3\\-1\\1\end{pmatrix} = 3+0+1 = 4$$

$$\tilde{s}_2 \cdot \tilde{s}_2 = \begin{pmatrix}3\\1\\1\end{pmatrix}\begin{pmatrix}3\\1\\1\end{pmatrix} = 9+1+1 = 11$$

$$\tilde{s}_2 \cdot \tilde{s}_3 = \begin{pmatrix}3\\1\\1\end{pmatrix}\begin{pmatrix}3\\-1\\+1\end{pmatrix} = 9-1+1 = 9$$

$$\tilde{s}_3 \cdot \tilde{s}_3 = \begin{pmatrix}3\\-1\\1\end{pmatrix}\begin{pmatrix}3\\-1\\1\end{pmatrix} = 9+1+1 = 11$$

$$\alpha_1 2 + \alpha_2 4 + 4\alpha_3 = -1 \quad \text{—} (1)$$
$$\alpha_1 4 + \alpha_2 11 + \alpha_3 9 = +1 \quad \text{—} (2)$$

29/8/22

$\alpha_1 = -2.5$

$\alpha_2 = 0.75$

$\alpha_3 = 0.75$

$$\tilde{W} = -2.5\begin{Bmatrix}1\\0\\1\end{Bmatrix} + 0.75\begin{Bmatrix}3\\1\\1\end{Bmatrix} + 0.75\begin{Bmatrix}3\\-1\\1\end{Bmatrix}$$

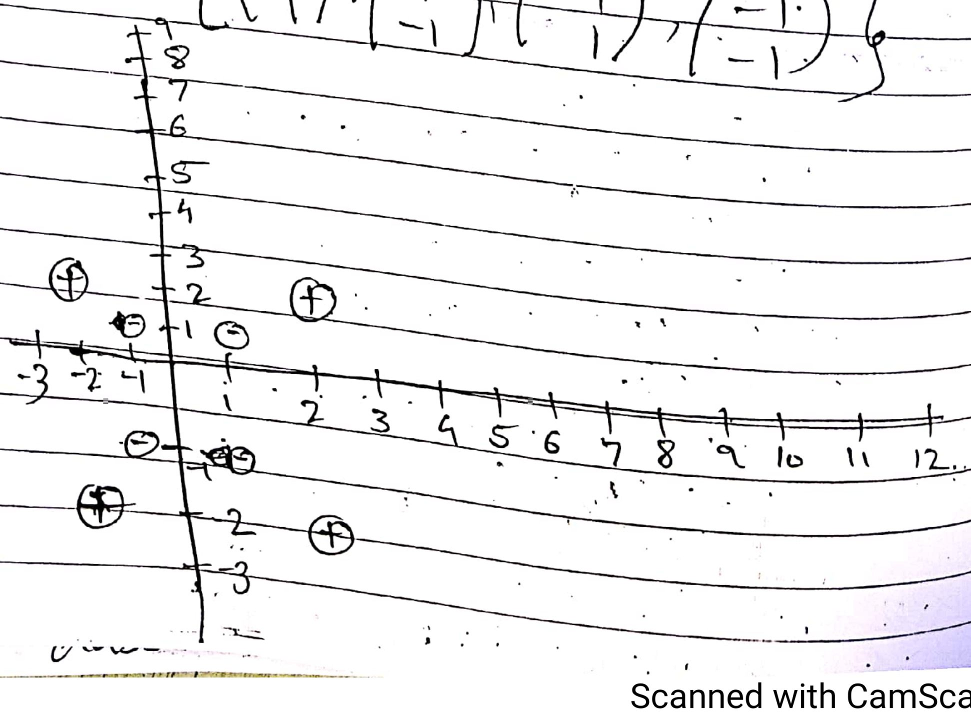
$$= \begin{Bmatrix}-3.5\\0\\-3.5\end{Bmatrix} + \begin{Bmatrix}2.25\\0.75\\0.75\end{Bmatrix} + \begin{Bmatrix}2.25\\-0.75\\0.75\end{Bmatrix}$$

$$= \begin{Bmatrix}1\\0\\-2\end{Bmatrix}$$

$$W = \begin{Bmatrix}1\\0\end{Bmatrix} \quad b = -2$$

Q3) $+ve = \left\{ \begin{pmatrix}2\\2\end{pmatrix}, \begin{pmatrix}2\\-2\end{pmatrix}, \begin{pmatrix}-2\\2\end{pmatrix}, \begin{pmatrix}-2\\-2\end{pmatrix} \right\}$

$-ve = \left\{ \begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}1\\-1\end{pmatrix}, \begin{pmatrix}-1\\1\end{pmatrix}, \begin{pmatrix}-1\\-1\end{pmatrix} \right\}$



[illegible]/22

This an eg of non linear dataset. To solve this using SVM we need to convert the given data set to a linear dataset using a mapping function $\phi(x)$ which can transform the data to a feature's space where a separating hyperplane can be found. The mapping function or the transformation function is

$$\boxed{\phi(x) = \phi\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}}$$

$$\phi\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{cases} \begin{bmatrix} (4-x_2) + |x_1 - x_2| \\ (4-x_1) + |x_1 - x_2| \end{bmatrix} & \text{if } \left(\sqrt{x_1^2 + x_2^2}\right) \geq 2 \\ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} & \text{otherwise} \end{cases}$$

All the positive data samples satisfy the condition of $\sqrt{x_1^2 + x^2} \geq 2$. ∴ All the positive labelled data are required to be transformed or mapped using mapping function $\phi(x)$

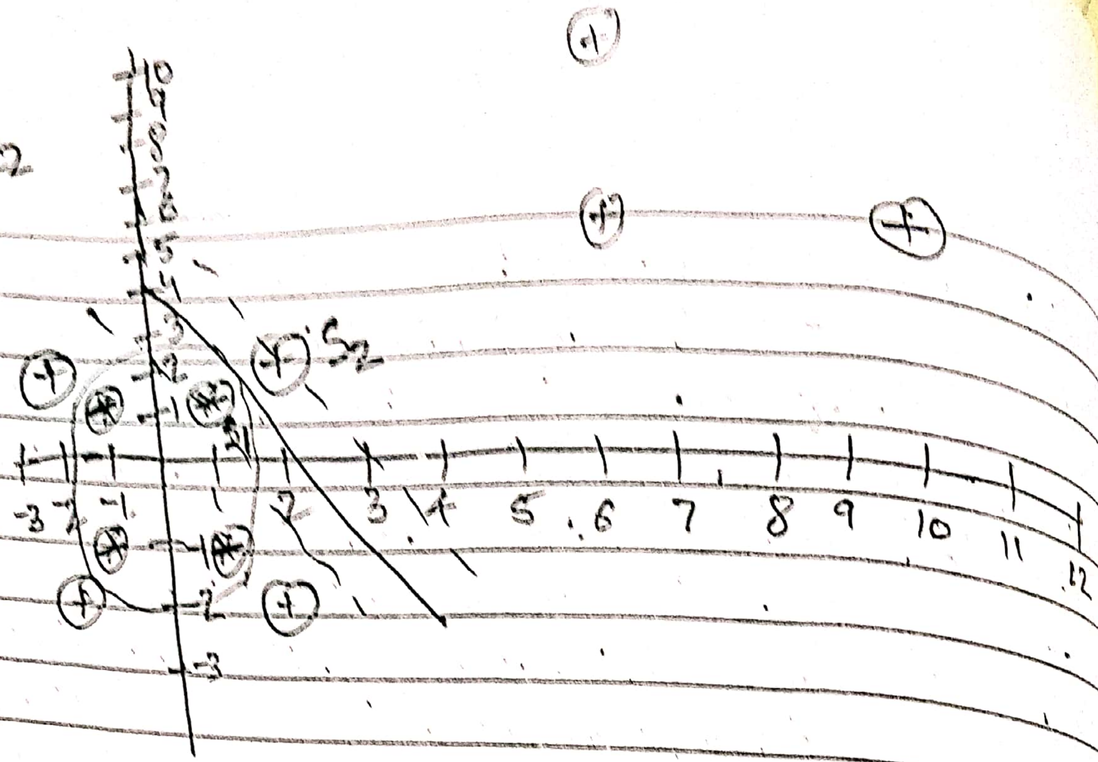
$$\phi\begin{pmatrix} 2 \\ 2 \end{pmatrix} = \begin{bmatrix} (4-2) + (2-2) \\ (4-2) + (2-2) \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

$$\phi\begin{pmatrix} 2 \\ -2 \end{pmatrix} = \begin{bmatrix} 4+2 + |2+2| \\ 4-2 + |2+2| \end{bmatrix} = \begin{bmatrix} 10 \\ 6 \end{bmatrix}$$

$$\phi\begin{pmatrix} -2 \\ 2 \end{pmatrix} = \begin{bmatrix} 4-2 + |-2-2| \\ 4+2 + |-2-2| \end{bmatrix} = \begin{bmatrix} 6 \\ 10 \end{bmatrix}$$

$$\phi\begin{pmatrix} -2 \\ -2 \end{pmatrix} = \begin{bmatrix} (4+2) + |-2+2| \\ (4+2) + |-2+2| \end{bmatrix} = \begin{bmatrix} 6 \\ 6 \end{bmatrix}$$

27/8/22



$\tilde{S}_1 = \begin{bmatrix}1\\1\\1\end{bmatrix}$ $\quad \tilde{S}_2 = \begin{bmatrix}2\\2\\1\end{bmatrix}$

$\alpha_1 \tilde{S}_1 \tilde{S}_1 + \alpha_2 \tilde{S}_1 \tilde{S}_2 = -1$ — (1)

$\alpha_1 \tilde{S}_1 \tilde{S}_2 + \alpha_2 \tilde{S}_2 \tilde{S}_2 = +1$ — (2)

$\tilde{S}_1 \tilde{S}_1 = \begin{bmatrix}1\\1\\1\end{bmatrix} \begin{bmatrix}1\\1\\1\end{bmatrix} = 1 + 1 + 1 = 3$

$\tilde{S}_2 \tilde{S}_2 = \begin{bmatrix}2\\2\\1\end{bmatrix} \begin{bmatrix}2\\2\\1\end{bmatrix} = 4 + 4 + 1 = 9$

$\tilde{S}_1 \tilde{S}_2 = \begin{bmatrix}1\\1\\1\end{bmatrix} \begin{bmatrix}2\\2\\1\end{bmatrix} = 2 + 2 + 1 = 5$

$\alpha_1 3 + 5\alpha_2 = -1$

$\alpha_1 5 + 9\alpha_2 = 1$

$\alpha_1 = -7$

$\alpha_2 = 4$

22/8/22

$$\tilde{W} = -7\begin{Bmatrix}1\\1\\1\end{Bmatrix} + 4\begin{Bmatrix}2\\2\\1\end{Bmatrix}$$

$$= \begin{Bmatrix}-7\\-7\\-7\end{Bmatrix} + \begin{Bmatrix}8\\8\\4\end{Bmatrix}$$

$$= \begin{Bmatrix}1\\1\\-3\end{Bmatrix}$$

$$W = \begin{Bmatrix}1\\1\end{Bmatrix} \qquad b = -3$$

For non linear data the classification function of input data x is given as

$$\boxed{f(x) = \sum_{i=1}^{n} \alpha_i \tilde{S}_i \, \phi(\tilde{x})}$$

Classify the input data $x = \begin{pmatrix}2\\3\end{pmatrix}$

$$\phi(x) = \begin{bmatrix}(4-2) + |2-3| \\ (4-\cancel{8}) + |2-3|\end{bmatrix}$$

$$= \begin{bmatrix}1\cancel{2}+1\\2\cancel{1}+1\end{bmatrix}$$

$$= \begin{bmatrix}2\cancel{8}\\3\cancel{2}\end{bmatrix}$$

$$\phi(\tilde{x}) = \begin{bmatrix}2\\3\\1\end{bmatrix}$$

$$f(x) = \alpha_1 \tilde{S}_1 \phi(\tilde{x}) + \alpha_2 \tilde{S}_2 \phi(\tilde{x})$$

$$= -7\begin{Bmatrix}1\\1\\1\end{Bmatrix}\begin{Bmatrix}2\\3\\1\end{Bmatrix} + 4\begin{Bmatrix}2\\2\\1\end{Bmatrix}\begin{Bmatrix}2\\3\\1\end{Bmatrix}$$

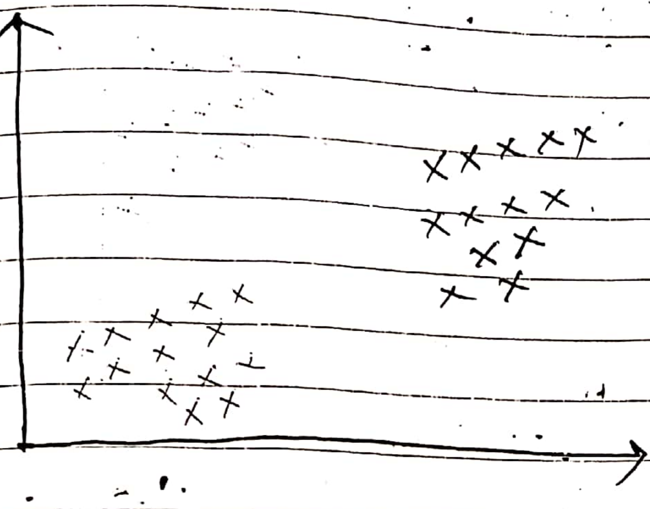
2/8/22

$= -42 + 44$

$= 2$

/8/22

* Multiclassification using SVM –

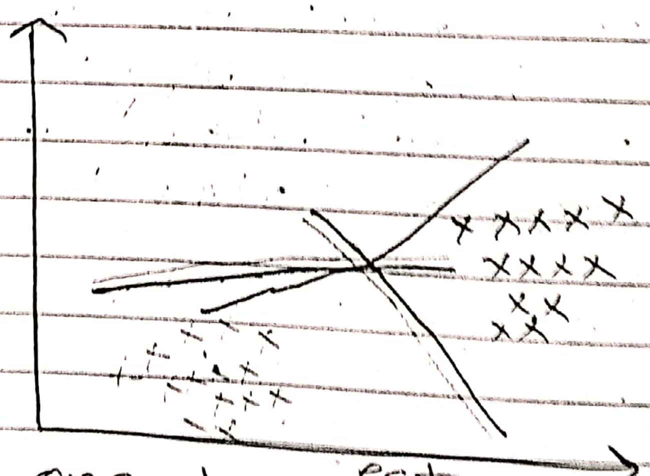


In general SVM doesn't support multiclass classification, it supports binar classification & separating datapoints into 2 classes. For multiclass classificatio the same principle is utilized by breaking down multiclassification proble into multiple binary classification problem. The idea is to map datapoint to high dimensional space to gain mutua linear separation between every two classes. This is called a one to one approach which break downs a multi class problem into multiple binary classification problem. If binary classifi ~~per~~ for each pair of classes. Another approach one can use is one to rest approach.
• In this approach the break down is said to a binary classifier: ~~per~~ for each class. A single SVM does binary classification. So, for datapoints with w
... to rest approach the

24/8/22

classifier can use $\boxed{\text{'m' SVMs}}$. Each would predict membership in one of the 'm' classes. In one to one approach the classifier can use $\boxed{\frac{m(m-1)}{2} \text{ SVMs}}$.

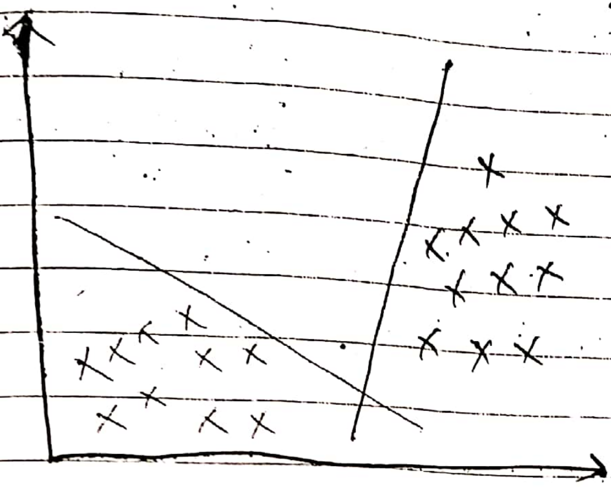
In the given diagram apply the approaches we get the following observations. In one to one approach we need a hyperplane to separate between every 2 classes neglecting the points of the 3rd class. This means the separation text into accounts only the points into 2 classes in the current splits.



In one to ~~raised~~ rest approach we need a hyperplane to separate between a class & all other at once. This means the separation takes all points into account dividing them into 2 groups. The group for a class points and a group for all other points.

24/8/22

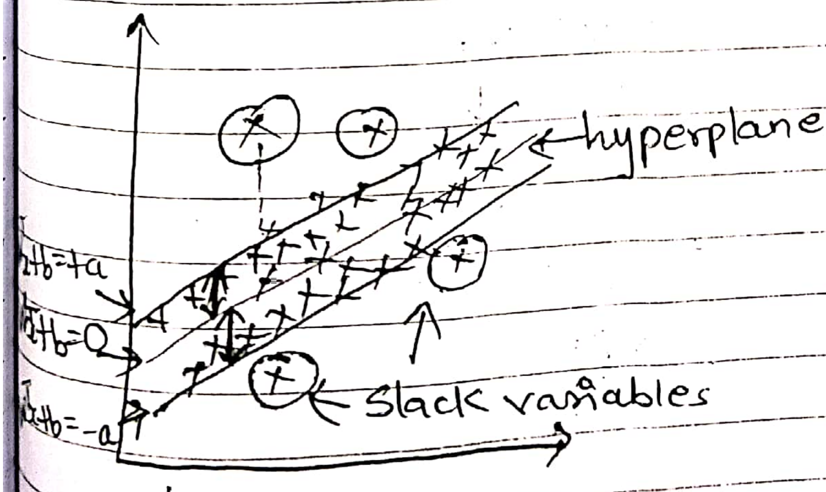
eg) A plane tries to maximize separation between green points & all other points at once



One-to-rest

One of the most common real-world problem for multiclass problem using SVM is text classification. For eg) Classifying new articles, tweets or scientific paper

★ (SVR) Support Vector Regression -



Datapoints outside the insensitive tube are called as slack variables. The data points following outside or on the boundary

1/8/22

of error margin are called as sup vectors.

$$y = w^T x + b \qquad y_i - W^T x - b - a =$$

$$y - W^T x - b = 0 \qquad y_i - W^T x - b + a = 0$$

The constraint to satisfy SVR is:

$$\boxed{-a \leq y - w^T x - b \leq a}$$

Assignment 2

?/8/22

8) Write & explain various kernel function used in SVM

A) ~~Kernel function is a method~~ SVM Algorithms use a group of mathematical functions that are known as kernels. Kernel function is a method used to take data as input and transform it into the required form of processing data. So, Kernel function generally transforms the training set of data so that a non linear decision surface is able to transform to a linear equation in a higher number of dimension spaces

Types of kernel function used in SVM –

1) Linear Kernel – It is the most basic type of kernel, usually one dimensional in nature. It proves to be the best function when there are lots of features. Linear kernel functions are faster than other functions

2) Gaussian Radial Basis function (RBF) – It is one of the most preferred and used kernel functions in SVM It is usually choosen for non linear data. It helps to make proper separation when there is no prior knowledge of data

3) Polynomial Kernel – It is a more generalized representation of the linear kernel. It is not as preferred as other kernel functions as it is less efficient and accurate

25/8/22

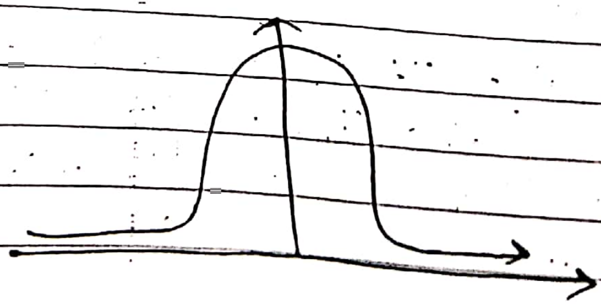
4) Sigmoid Kernel - It is mostly preferred for neural networks. This kernel function is similar to a two-layer perceptron model of the neural network, which works as an activation function for neurons.

5) Bessel function kernel - It is mainly used for removing the cross term in mathematical functions.

6) Anova Kernel - It is also known as a radial basis function kernel. It usually performs well in multidimensional regression problems.

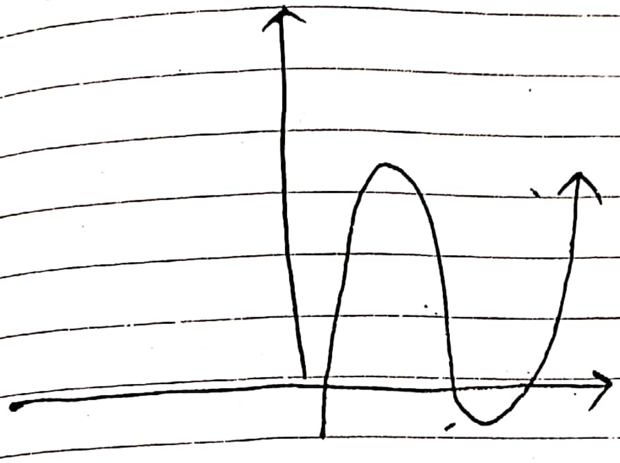
1) Linear Kernel -

2) Gaussian Kernel Radial Basis function (RBF) -

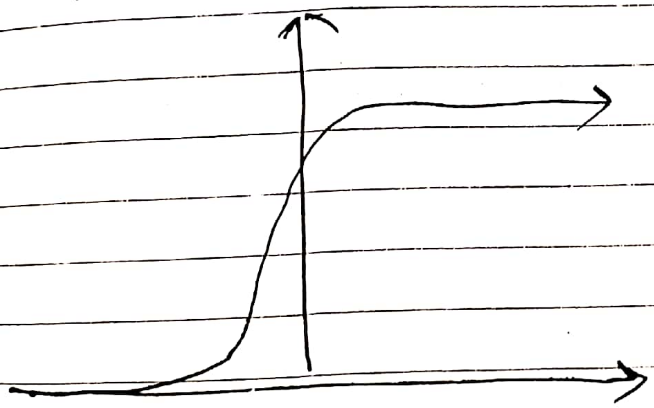


25/8/22

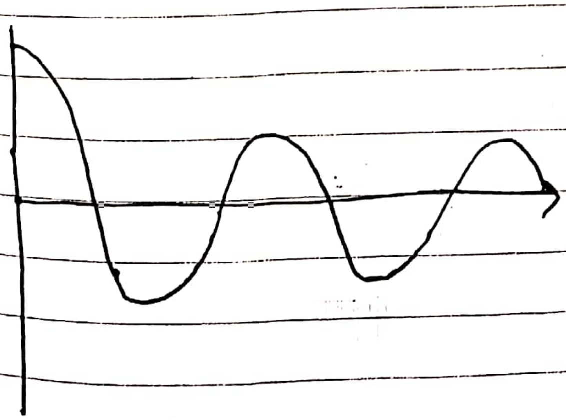
3) Polynomial Kernel -



4) Sigmoid Kernel -



5) Bessel function Kernel:-



25/8/22

3) <u>Anova Kernel</u> -