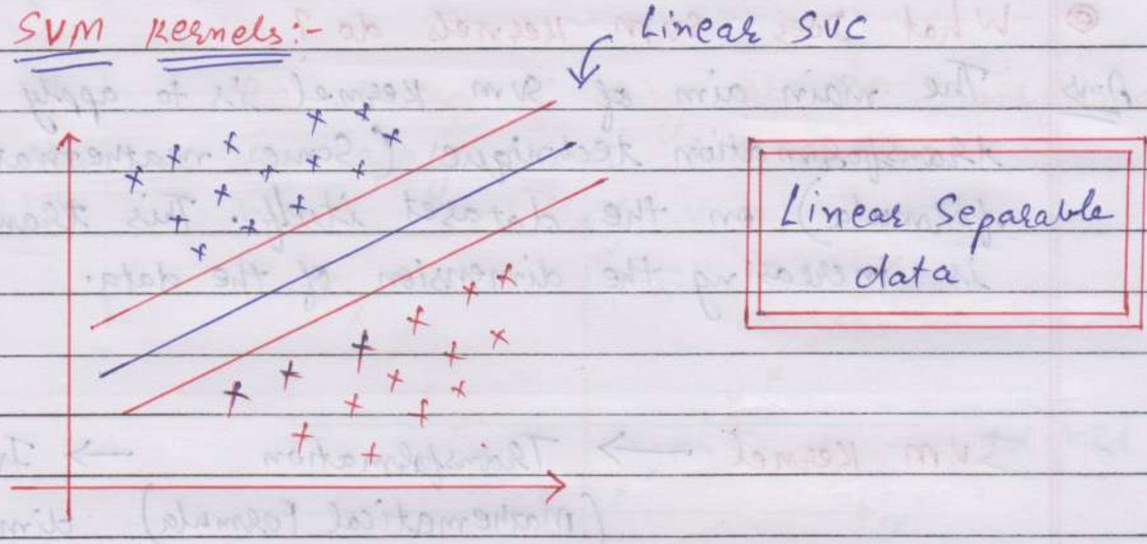
Machine Learning DAY 6 30/10/2022

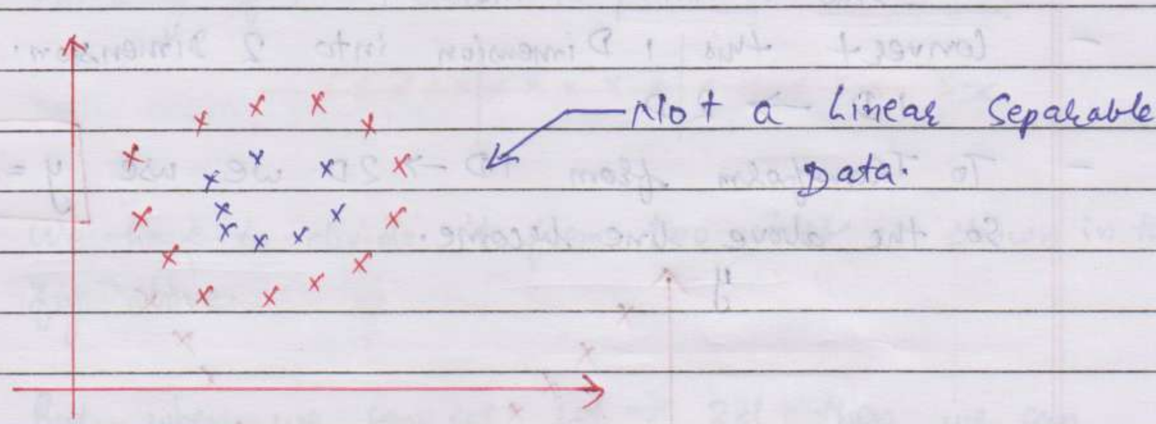
- SVM Kernels
- ROC AUC Curve

★ SVM Kernels:-



- When we create this type of Best Fit Line and marginal plane, we are actually ~~seperating~~ solving the Linear Seperable data

- This type of SVC is basically called as Linear SVC



- If data is not a Linear Separable data, you will not be able to create best Fit Line and not able to create a marginal plane and even though we create it, the accuracy will be very Low.

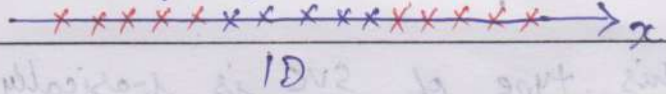
- For this type of problems, we have some more SVM Kernels.

What Does SVM Kernels do ?

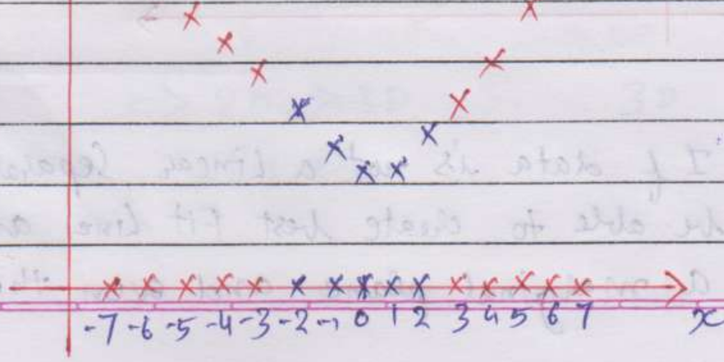
Ans The main aim of SVM Kernel is to apply some transformation technique (Some mathematical formula) on the dataset itself. This transformation is increasing the dimension of the data.

SVM Kernel ⟶ Transformation (Mathematical Formula) ⟶ Increasing the dimension of the data.

Suppose all points are falling on the same Line.



- This Line is in 1 Dimension.
- Convert this 1 Dimension into 2 Dimension.
  1D ⟶ 2D
- To Transform from 1D ⟶ 2D we use $\boxed{y = x^2}$
  So the above line become.



$y = x^2$

So if $x = -7$, $y = 49$

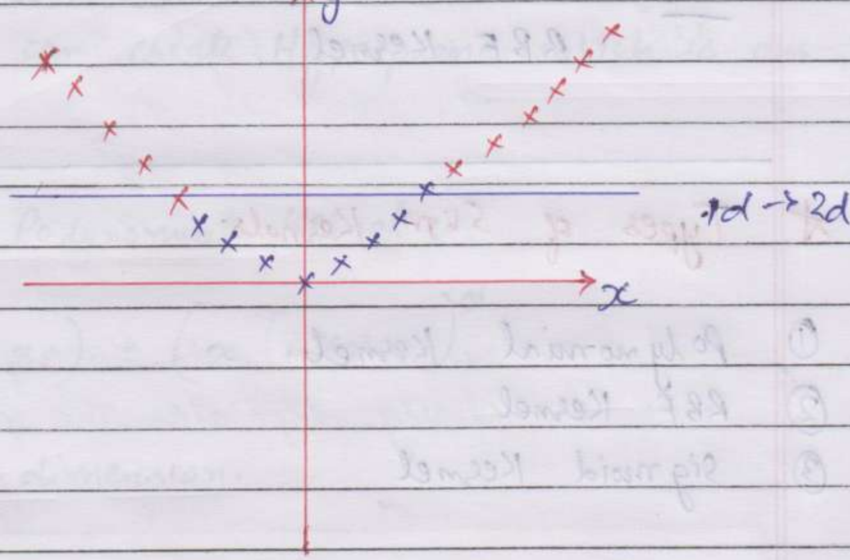
Similarly $x = -3$, $y = 9$

$x = 2$, $y = 4$

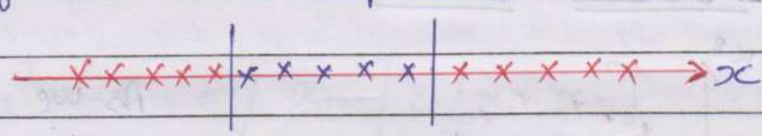
$x = 6$, $y = 36$ and so on.

What is the advantage of doing this 2D transformation?

- Now we can apply Linear SVM or SVC



Initially if there were the points in 1d:



We have to divide it from two sides as shown in the fig above.

- But when we convert 1d → 2d then we can divide all the points using single Line which is called Linear SVC

Assigment :- What Kind of transformation you will try to apply to convert all points of fig 1 to points shown in fig 2.

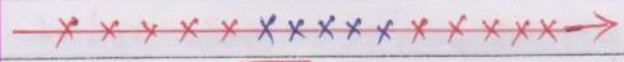


Fig 1

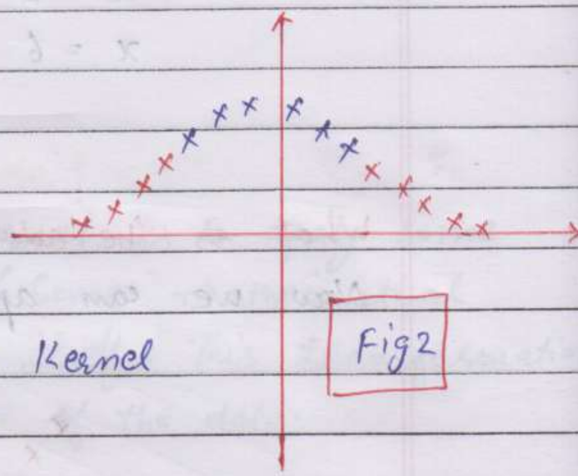


Fig2

Hint :- Radial Basis Function Kernel
RBF Kernel

★ Types of SVM Kernels:

① Polynomial Kernel
② RBF Kernel
③ Sigmoid Kernel

① Polynomial Kernel

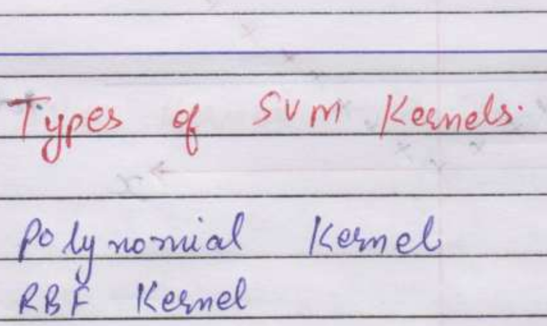


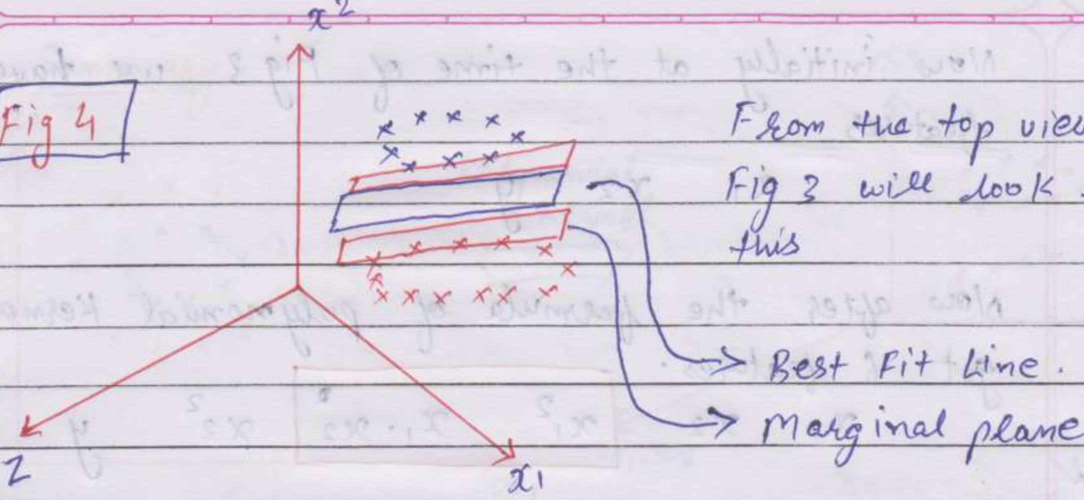
Fig 3

As we know that the data shown in fig 3 is not a Linear Separable data, so we cannot create a best fit line.

This is a 2d Graph, if we convert it in 3d, then we will get

Fig 4

From the top view Fig 3 will look like this



our main aim is to increase dimension from 2d → 3d so that we can create Hyperplane which is our SVM.

Formula of Polynomial Kernel :-

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

where $d \Rightarrow$ dimension.

- If we are converting from 2d → 3d the value of d will be 3

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

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

Now we have 3 unique values in the above formula.

x_1^2 $\quad x_1 \cdot x_2 \quad$ x_2^2

Now initially at the time of Fig 3 we have 3 features

$x_1 \quad x_2 \quad y$

Now after the formula of polynomial Kernel we have got 6 features.

$x_1 \quad x_2 \quad \boxed{x_1^2 \quad x_1 \cdot x_2} \quad x_2^2 \quad y$

These $x_1^2 \quad x_1 \cdot x_2 \quad x_2^2$ can be plotted as the 3rd Dimension in Fig 4. This means that in Fig 4

x_1 will be x_1^2

x_2 will be x_2^2

z will be $x_1 \cdot x_2$

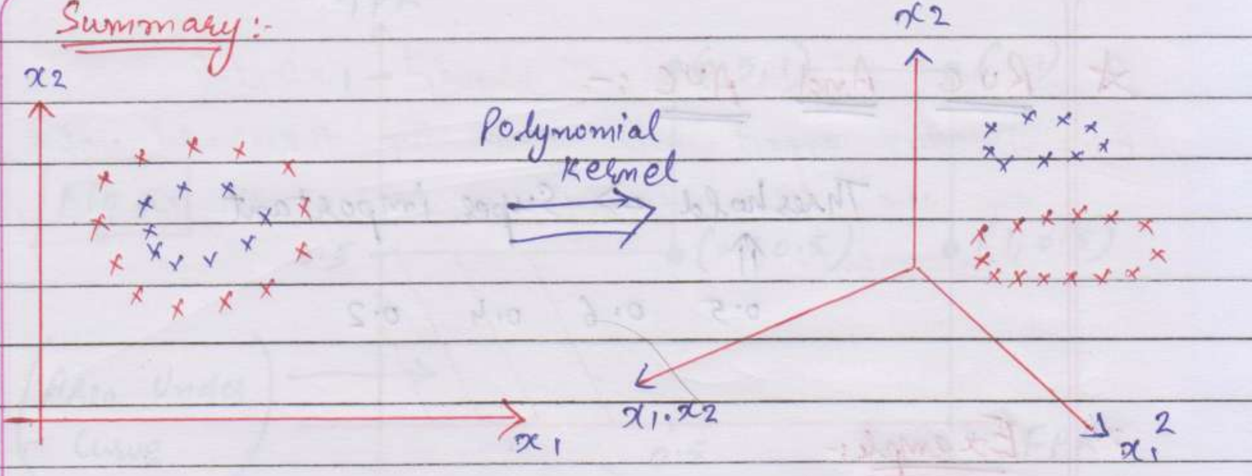
and once we have all these points, we will be able to clearly separate the points

⇒ Now Suppose initially initially in Fig 3 we get Linear SVC accuracy of 50%.

⇒ Now if we ~~can~~ change the Kernel to polynomial Kernel, then it will create all the additional features internally

⇒ And by using these features, it will try to create 3d plot and from that it will try to create SVM

Summary:-



Dataset

$x_1 \quad x_2 \quad y$

Dataset

$x_1 \quad x_2 \quad x_1^2 \quad x_1 \cdot x_2 \quad x_2^2 \quad y$

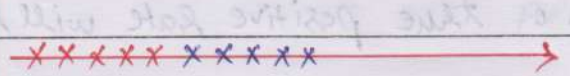
Polynomial Kernel Formula :-

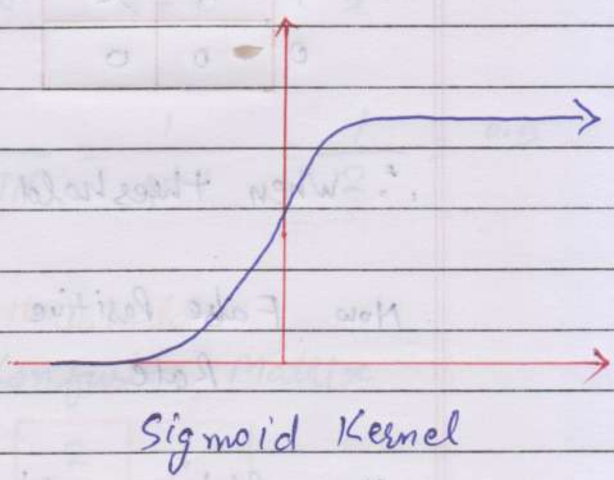
$$f(x_1, x_2) = \left(x_1^T \cdot x_2 + 1\right)^d$$

③ RBF Kernel :-

$$K(\vec{x}, \vec{l_i}) = e^{-\frac{\|\vec{x} - \vec{l_i}\|^2}{2\sigma^2}}$$

③ Sigmoid Kernel :-





Sigmoid Kernel

Sigmoid Kernel $\Rightarrow \frac{1}{1+e^{-x}}$

★ ROC And AUC :-

Threshold ⇒ Super important

⇑

0.5 0.6 0.4 0.2

Example:-

Actual o/p y	Probability $\hat{y}$	$y(0)$
1	0.8	1
0	0.96	1
1	0.4	1
1	0.3	1
0	0.2	1
1	0.7	1

← If we put threshold = 0

Confusion Matrix will be.

	1	0
1	4	2
0	0	0

True Positive Rate $= \frac{TP}{TP+FN} = \frac{4}{4+0} = 1$

∴ When threshold is 0 true positive rate will be 1

Now False Positive Rate $= \frac{FP}{FP+TN} = \frac{2}{2+0} = 1$

Now Plot a diagram.

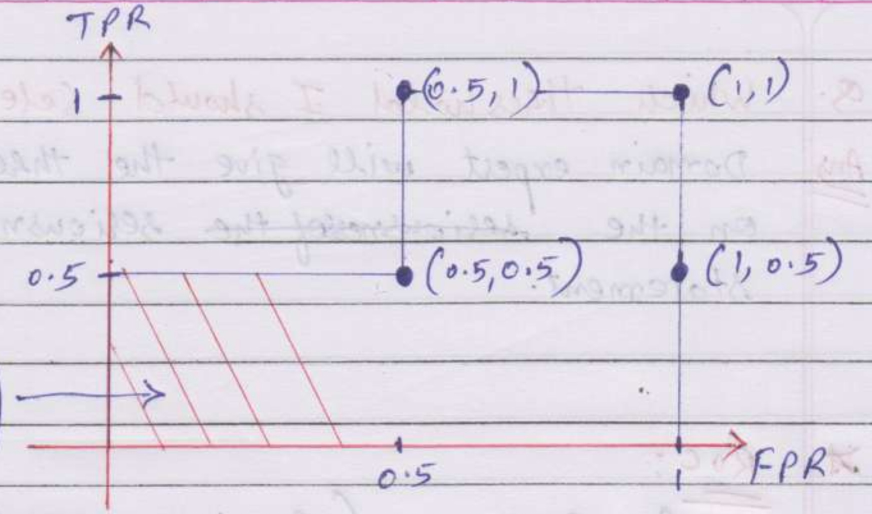


Fig 5

Now if we take 0.2 as threshold

y	$\hat{y}$	$y(0.2)$
1	0.8	1
0	0.96	1
1	0.4	1
1	0.3	1
0	0.2	0
1	0.7	1

Confusion Matrix

Predicted \ [Actual]	1	0
1	4	1
0	0	1

predicted

$$\text{True Positive Rate} = \frac{TP}{TP+FN} = \frac{4}{4+0} = \frac{4}{4} = \cancel{0.8}\ 1$$

$$\text{False Positive Rate} = \frac{FP}{FP+TN} = \frac{1}{1+1} = \frac{1}{2} = 0.5$$

Now if we take 0.4 as threshold.

y	$\hat{y}$	$y(0.4)$
1	0.8	1
0	0.96	1
1	0.4	0
1	0.3	0
0	0.2	0
1	0.7	1

Confusion Matrix

2	1
2	1

TPR = 0.5
FPR = 0.5

Q. Which threshold I should select?

Ans Domain expert will give the threshold value depending on the ~~seriousness of the~~ seriousness of the problem statement.

★ ROC:

An ROC curve (Receiver operating characteristic curve) is a graph showing the performance of a classification model at all classification thresholds.