Total Marks of Question no.		Examiner	
		Moderator	
		Re-Assessor	

Space for Marks	Question No.	START WRITING HERE

Q1.

a) Gaussian Kernel :-

Gaussian kernel is defined as.

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

Applications where prior knowledge is not available, then gaussian kernel is used.

b) Hierarchical clustering :-

It is unsupervised machine learning algorithm, which is used to group the unlabelled data set in to a cluster known as heirarchical clustering.

In this algorithm we develop the hierarchy of clusters in the form of tree, & this tree-shaped structure is known as dendrogram:

It has two approches :-

1. Agglomerative :- It is a bottom up approach, in which the algorithm starts with taking all data-point as a single cluster & merging them untill one cluster is left.

2. Division :- It is the reverse of agglomeration approach as it is top-down approach.

c) Optimization methods in Machine Learning :-

- Gradient descent
- Stochastic Gradient descent
- Adaptive learning rate method.

Total Marks of Question no.		Examiner	
		Moderator	
		Re-Assessor	

Space for Marks	Question No.	START WRITING HERE
		- Conjugate learning sub method
		- Derivation free optimisation
		- Zeroth order optimisation
		- For meta learning.
	d)	Two application of expectation maximization clustering :-
		- Use to calculate the Gaussian density of a function.
		- It helps to fill in the missing data during a sample.
	e)	Differential between DBSCAN & k-means algo:-
		- DBSCAN algorithm can not efficiently handel high dimensional data set.
		- k-means does not work well with outliers & noisy data set.
		- DBSCAN algorithm is capable of creating arbitrary shaped clusters.
	g)	Lazy learner / Eager learner
		i) It stores data set without learning from it / i) When it recives data set it start classifying
		ii) Start classifying data when it recives test data. / ii) It does not wait for test-data to learn

Scanned with CamScanner

Total Marks of Question no.		Examiner	
		Moderator	
		Re-Assessor	

Space for Marks	Question No.	START WRITING HERE

iii) It takes less time for learning & more time for classifying data

iii) It takes long time for learning & less time for classifying data.

g) Lift as a performance metric in clustering :-

- Lift is a defined as a strength of any rule.

$$\text{Lift} = \frac{\text{Support}(P,Q)}{\text{Support}(P) * \text{Support}(Q)}$$

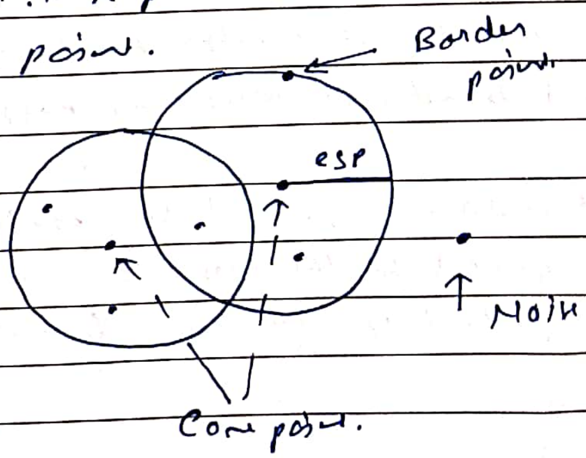
It is the ratio of the observed support measure of expected support if P & Q are independent of each other.

If lift = 1 → the probability of appearance of antecedent & consequent.

h) DBSCAN:-

i) Border point:- A point which has fewer points than MinPts within esp but it is in the neighbourhood of a core point.

ii) Noise:- A point which not a core point or border point.



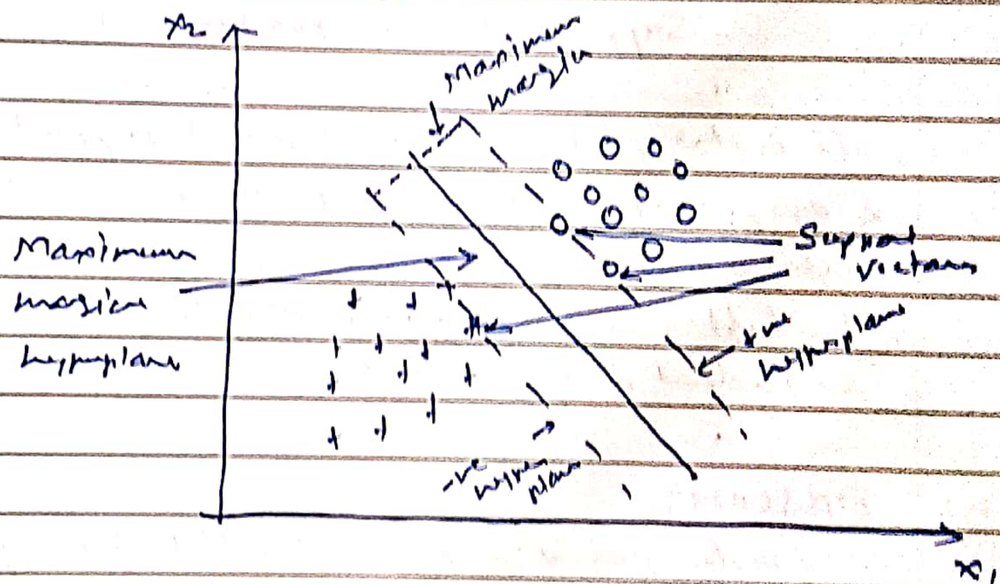
Total Marks of Question no.		Examiner	
		Moderator	
		Re-Assessor	

Space for Marks	Question No.	START WRITING HERE

Q2

7) SVM (Support Vector Machine)

- SVM is a supervised learning algorithm.
- The goal of SVM is to create a best fit line or decision boundary that can segregate n-dimensional space into classes.
- SVM chooses the extreme points / vectors that helps in creating the hyperplane.



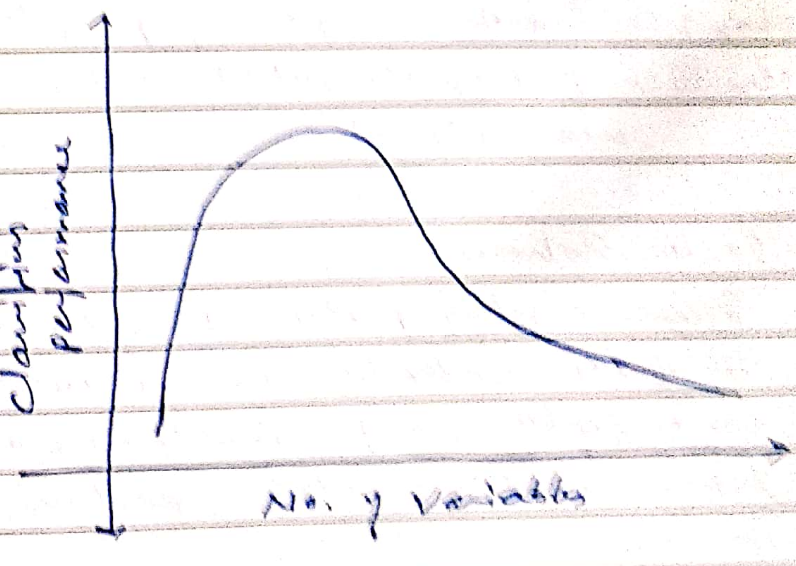
Hyperplane :- There can be multiple lines / decision boundary to segregate the classes in n-dimensional space, but we need to find out the best decision boundary that helps to classify the data points. This best boundary is known as hyperplane.

Support vectors :- The data points or vectors that are the closest to the hyperplane & which affect the position of the hyperplane are termed as support vectors.

Total Marks of Question no.		Examiner	
		Moderator	
		Re-Assessor	

Space for Marks	Question No.	START WRITING HERE
	b)	Dimension reduction importance in Machine learning:

- It assist in information packing of diminishing the storage room required.
- It decreases the time which is required for doing same calculations. If the dimensions are less then processing will be less, added advantage of having less dimensions is permission to use calculation suited for condition.
- It handles with multi-colinearity that is used to enhance the execution of the model. It evaluates excess highlights.
- It helpful for noise elimination.
- The classifier performance normally will degrade for a large no. of features.



Total Marks of Question no.		Examiner	
		Moderator	
		Re-Assessor	

Space for Marks	Question No.	START WRITING HERE

c) Dimensionality reduction techniques:-

a) Feature selection:-

- Given a set of features $F = \{x_1, \dots, x_n\}$

The feature selection problem is to find a subset $F' \subseteq F$ that maximizes the learner's ability to classify the patterns.

Finally F' should maximize scaling function.

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \xrightarrow{\text{feature selection}} \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{im} \end{bmatrix}$$

Feature selection steps:-

Feature selection is an optimization problem.

Step I:- Search the space of possible feature subset

Step II:- Pick up the subset that is optimal or near optimal w.r.t. some objective function.

b) Feature Extraction:-

- Suppose a set of features $F = \{x_1, \dots, x_n\}$ is given
- The feature extraction task is to map F to some feature set F'' that will maximize the learner's ability to classify patterns.

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \xrightarrow{\text{feature extraction}} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = f\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Total Marks of Question no.		Examiner	
		Moderator	
		Re-Assessor	

Space for Marks	Question No.	START WRITING HERE

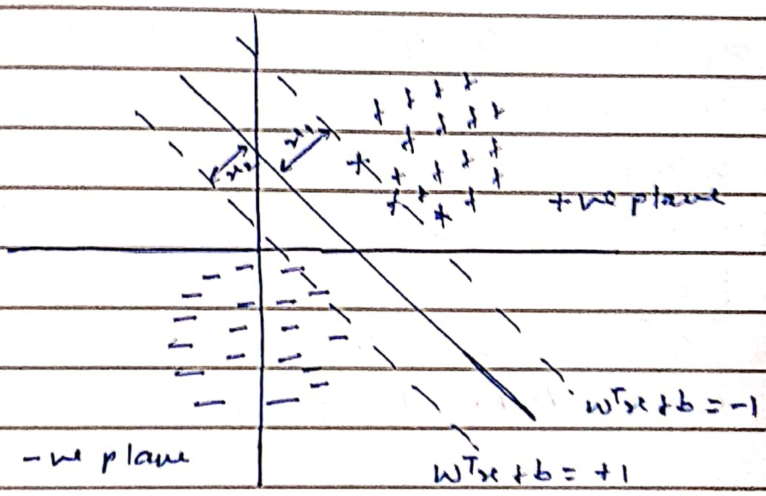
- A projection matrix w is computed from N-Dimensional to M-dimensional vectors to achieve low errors.

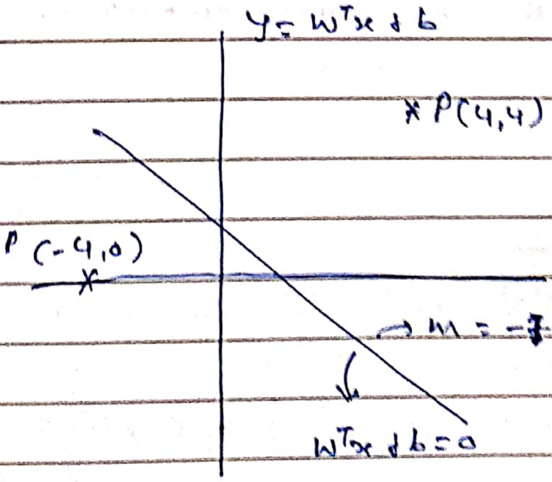
$$z = w^T x$$

- Principle Component analysis & Independent Component analysis are the feature extraction methods.

Q3

a)





We know that, $m = -1$

let c as $b = 0$

when $(x_1, y_1) = (-4, 0)$

Total Marks of Question no.		Examiner	
		Moderator	
		Re-Assessor	

Space for Marks	Question No.	START WRITING HERE

$y = w^T x + 0$

$= \begin{bmatrix} -1 \\ 0 \end{bmatrix} \begin{bmatrix} -4 & 0 \end{bmatrix}$

$= 4$ +ve value

↳ pt. will be always +ve

When $(x_2, y_2) = (4, 4)$

$y = w^T x + 0$

$= \begin{bmatrix} -1 \\ 0 \end{bmatrix} \begin{bmatrix} 4 & 4 \end{bmatrix}$

$= -4 \rightarrow$ -ve value

↳ pt will be always -ve

Compute $\rightarrow x_2 - x_1$

Consider +ve & -ve plane equation:

$w^T x_1 + b = -1$

$w^T x_2 + b = 1$

$\overline{- \quad - \quad -}$

$w^T (x_2 - x_1) = 2$

$\therefore \frac{w^T}{||w||} (x_2 - x_1) = \boxed{\frac{2}{||w||}} \rightarrow$ Optimization function.

Total Marks of Question no.		Examiner	
		Moderator	
		Re-Assessor	

Space for Marks	Question No.	START WRITING HERE

We need to maximize optimization function update (w^*, b^*) to max $\frac{2}{\|w\|}$

Such that;

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

We can also write above equation as,

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

Further:

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

where, C_i = How many errors are there?
ϵ_i = value of the errors.

Q3 b) Steps of developing Machine Learning applications:-

i) Collection of Data:-
- Collecting the samples from a website & extracting data.
  - From RSS feed or an API
  - From devices to collect wind speed measurement.
  - Publicity available data.

ii) Preparation of the i/p data:-
- Check i/p data & its useable format.

Total Marks of Question no.		Examiner	
		Moderator	
		Re-Assessor	

Space for Marks	Question No.	START WRITING HERE

iii) Analyse the i/p data :-
- I/p data is analysed & garbage value is checked if any.

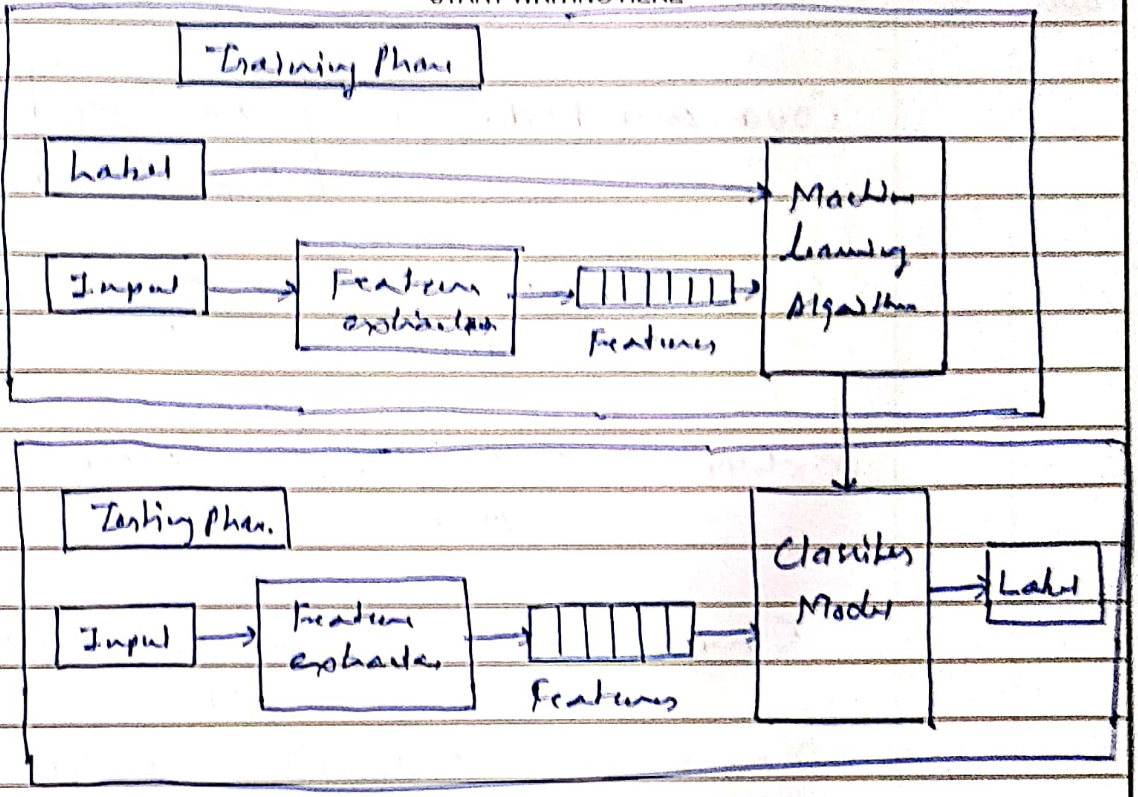
iv) Train the algorithm :-
- Good clean data from the first two steps is given as i/p to the algorithm. The algorithm extracts information as knowledge.
- In case of unsupervised learning, training steps is not there because target value is not present.

v) Test the algorithm:
- In this step, the information learned in earlier step, is used.
- In supervised learning, known values can be used to evaluate the algorithm.
- In unsupervised learning, matrix are used to evaluate the success.

vi) Use it :-
- Real program is developed to do some task & once again it is checked if all the steps are correct.

Total Marks of Question no.		Examiner	
		Moderator	
		Re-Assessor	

Space for Marks	Question No.	START WRITING HERE



Q4
(a) PCA :-

X	2.3	0.7	2.1	1.2	3.2	2.2	2.6	2.0	1.6	1.1
y	2.2	0.9	2.4	2.4	3.0	3.1	1.8	1.2	1.5	0.8

Principal Component :-

$\Sigma x = 19$

$\Sigma y = 19.3$

$$X_m = \frac{\Sigma x}{N} = \frac{19}{10} = 1.9$$

$$Y_m = \frac{\Sigma y}{N} = \frac{19.3}{10} = 1.93$$

Total Marks of Question no.		Examiner	
		Moderator	
		Re-Assessor	

Space for Marks	Question No.	START WRITING HERE

Covariance Matrix, $C = \begin{bmatrix} C_{xx} & C_{xy} \\ C_{yx} & C_{yy} \end{bmatrix}$

$\therefore C_{xx} = \frac{\Sigma(x - x_m)^2}{N-1} = \frac{5.14}{9} = 0.571$

$\therefore C_{yy} = \frac{\Sigma(y - Y_m)^2}{N-1} = \frac{6.092}{9} = 0.676$

$\therefore C_{xy} = C_{yx} = \frac{\Sigma(x - x_m)(y - y_m)}{N-1}$

$= \frac{3.714}{9} = 0.412$

$C = \begin{bmatrix} C_{xx} & C_{xy} \\ C_{yx} & C_{yy} \end{bmatrix} = \begin{bmatrix} 0.571 & 0.412 \\ 0.412 & 0.676 \end{bmatrix}$

$\det(S - \lambda I) = 0$

$\therefore \det\left(\begin{bmatrix} 0.571-\lambda & 0.412 \\ 0.412 & 0.676-\lambda \end{bmatrix}\right) = 0$

$\therefore (0.571 - \lambda)(0.676 - \lambda) - (0.412)(0.412) = 0$

$\therefore \lambda^2 - 1.247\lambda + 0.216 = 0$

$\therefore \boxed{\lambda_1 = 0.414}$ & $\boxed{\lambda_2 = 0.776}$

Total Marks of Question no.		Examiner	
		Moderator	
		Re-Assessor	

Space for Marks	Question No.	START WRITING HERE

For λ_1

$$CV = \lambda V$$

$$\therefore \begin{bmatrix} 0.571 & 0.412 \\ 0.412 & 0.676 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = 0.414 \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$$

$$\therefore 0.571 x_1 + 0.412 y_1 = 0.414 x_1$$
$$0.412 x_1 + 0.676 y_1 = 0.414 x_1$$

$$\therefore 0.157 x_1 = -0.412 y_1$$
$$0.412 x_1 = 0.262 y_1$$

$$\therefore x_1 = -2.624 y_1 \quad \text{OR} \quad x_1 = -0.636 y_1$$

Assume $y_1 = 1$

$$\Rightarrow \begin{bmatrix} -2.624 \\ 1 \end{bmatrix} \Rightarrow \sqrt{(-2.624)^2 + (1)^2} = 2.808$$

$$\begin{bmatrix} -2.624/2.808 \\ 1/2.808 \end{bmatrix} = \begin{bmatrix} -0.9344 \\ 0.356 \end{bmatrix} = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} -0.636 \\ 1 \end{bmatrix} \Rightarrow \sqrt{(-0.636)^2 + (1)^2} = 1.184$$

$$\begin{bmatrix} -0.636/1.184 \\ 1/1.184 \end{bmatrix} = \begin{bmatrix} -0.537 \\ 0.844 \end{bmatrix} = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$$

Total Marks of Question no.		Examiner	
		Moderator	
		Re-Assessor	

Space for Marks	Question No.	START WRITING HERE

for x_2

$$\begin{bmatrix} 0.571 & 0.412 \\ 0.412 & 0.676 \end{bmatrix} \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = 0.776 \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}$$

$$0.571x_2 + 0.412y_2 = 0.776x_2$$

$$0.412\,x_2 + 0.676y_2 = 0.776x_2$$

$$0.205x_2 = 0.412y_2$$

$$0.412y_2 = 0.1y_2$$

$$x_2 = 2.009y_2 \quad \text{OR} \quad x_2 = 0.242y_2$$

Assume $y_2 = 1$

$$\Rightarrow \begin{bmatrix} 2.009 \\ 1 \end{bmatrix} \Rightarrow \sqrt{(2.009)^2 + (1)^2} = 2.244$$

$$\begin{bmatrix} 2.009/2.244 \\ 1/2.244 \end{bmatrix} = \begin{bmatrix} 0.895 \\ 0.445 \end{bmatrix} = \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} 0.242 \\ 1 \end{bmatrix} \Rightarrow \sqrt{(0.242)^2 + (1)^2} = 1.028$$

$$\begin{bmatrix} 0.242/1.028 \\ 1/1.028 \end{bmatrix} = \begin{bmatrix} 0.235 \\ 0.972 \end{bmatrix} = \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}$$

Total Marks of Question no.		Examiner	
		Moderator	
		Re-Assessor	

Space for Marks	Question No.	START WRITING HERE
	Q4	
	b)	Kernel tricks in Machine Learning :-

- Kernel trick is to find the plane which can separate, classify or split the data with maximum margin is also called street width.
- The distance from the point (x, y) to a line $Ax + By + C = 0$ is $\frac{|Ax + By + C|}{\sqrt{A^2 + B^2}}$
- In order to maximize the margin, the distance betn H_0 & H_1 is then

$$\frac{|w \cdot x + b|}{||w||} = \frac{1}{||w||},$$ so the total distance between H_1 & H_2 is $\frac{2}{||w||}$

- In order to maximize the margin, minimize $||w||$

The major Kernel tricks are :-

1) Gaussian Kernel
2) Gaussian Kernel Radial Basis function
3) Sigmoid Kernel
4) Polinomial Kernel

Total Marks of Question no.		Examiner	
		Moderator	
		Re-Assessor	

Space for Marks	Question No.	START WRITING HERE

1) Gaussian Kernel :- It is used to perform transformation when there is no prior knowledge data

$$K(x,y) = e^{-\left(\frac{\|x-y\|^2}{2\sigma^2}\right)}$$

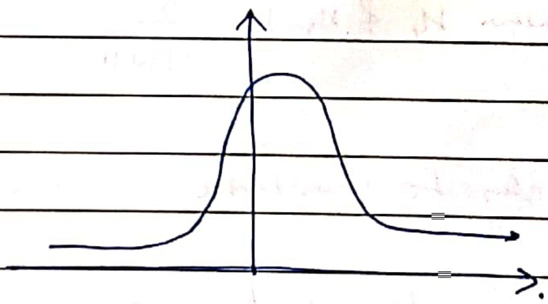
2) Gaussian Kernel Radial Basis function:- RBF is same as above kernel function, adding radial basis method to improve the transformation.

$$K(x,y) = e^{-(\gamma \|x-y\|^2)}$$

$K(x,x_1) + K(x,x_2)$ Simplified formula

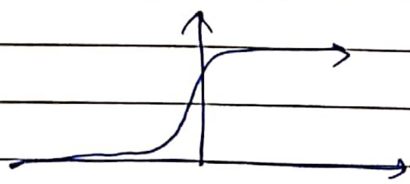
$K(x,x_1) + K(x,x_2) > 0$ (+ve class)

$K(x,x_1) + K(x,x_2) = 0$ (-ve class)



3) Sigmoid Kernel : This function is equivalent to the two layer perceptron model of NN, which is used as a activation function for artificial neurons.

$$K(x,y) = \tanh(\gamma . x^T y + r)$$

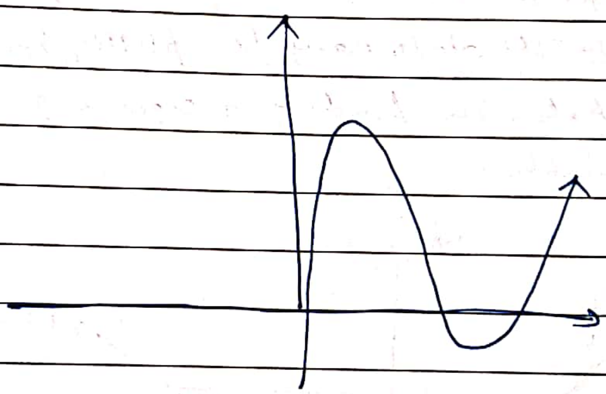


Total Marks of Question no.		Examiner	
		Moderator	
		Re-Assessor	

Space for Marks	Question No.	START WRITING HERE

4) Polynomial Kernel :- It represents the similarity of vectors in the training set of data in a feature space over polynomials of the original variables used in the kernel.

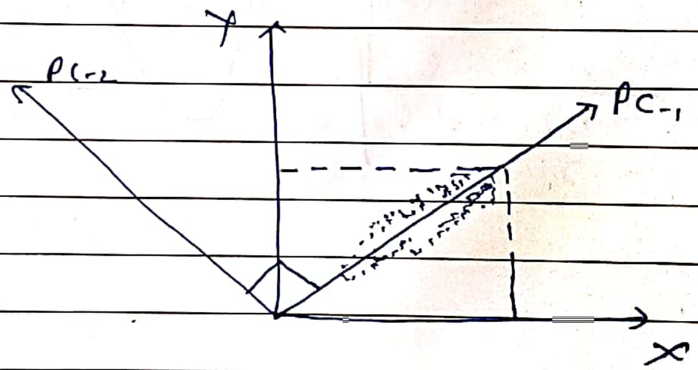
$$K(x,y) = \tanh(\gamma . x^T y + r)^d, \gamma > 0$$



Total Marks of Question no.		Examiner	
		Moderator	
		Re-Assessor	

Space for Marks	Question No.	START WRITING HERE
	Q5(a)	Principal Component Analysis :- (PCA)

- PCA is an unsupervised learning technique for reducing the dimensionality of the data.
- It increases interpretability of data yet at the same time it minimizes information loss.
- It helps to find most significant features in the data set & makes the data easy for plotting in 2D & 3D.
- PCA helps in finding a sequence of linear combinations of variables.



In this system, variables are changed into another arrangement of variables, which are straight blend of unique variables. These new arrangements of variables are known as principal components.

They are calculated so that first principal component represents a large portion of the conceivable variety of unique information after which each succeeding component has the most conceivable variance.

- The second principal component should be symmetrical to the primary principal component.
- The principle components are sensitive to the size of estimation.

		Examiner	
Total Marks of Question no.		Moderator	
		Re-Assessor	

Space for Marks	Question No.	START WRITING HERE

The PCA algorithm is based on mathematical concepts such as:
- Variance & covariance
- Eigenvalues & Eigen vectors

Steps in PCA algorithm:
1) Getting data set
2) Representing data into a structure
3) Standardizing the data
4) Calculating the covariance
5) Calculating the Eigen values & Eigen Vectors
6) Sorting the Eigen vectors
7) Calculating the new features of Principal Component
8) Remove less or unimportant features from the new data set.

Applications of PCA :-
- It is used as the dimensionality reduction techniques in various AI applications such as Computer Vision, Image compression etc.

- It can also be used as finding hidden patterns if data has high dimensions.

- Some fields where PCA is used are Finance, data mining, Psychology etc.

Total Marks of Question no.		Examiner	
		Moderator	
		Re-Assessor	

Space for Marks	Question No.	START WRITING HERE

Q5(b) Average linkage :-

	A	B
P_1	1	1
P_2	1.5	1.5
P_3	5	5
P_4	3	4
P_5	4	4
P_6	3	3.5

Step 1 :-

Distance Matrix :-

P_1	0					
P_2	0.707	0				
P_3	5.656	4.949	0			
P_4	3.605	2.915	2.236	0		
P_5	4.242	3.535	1.414	1	0	
P_6	5.201	2.5	1.802	$\boxed{0.5}$	1.118	0
	P_1	P_2	P_3	P_4	P_5	P_6

0.5 is the smallest, so P_4 & P_6 have smallest distance.

Step II

P1	0				
P2	$\boxed{0.707}$	0			
P3	5.656	4.949	0		
P4,P6	4.403	2.707	2.019	0	
P5	4.242	3.535	1.414	1.059	0
	P1	P2	P3	P4,P6	P5

Total Marks of Question no.		Examiner	
		Moderator	
		Re-Assessor	

Space for Marks	Question No.	START WRITING HERE

0.707 is the smallest i.e. P1 & P2 have smallest distance.

Step III :-

P1, P2	0			
P3	5.302	0		
P4, P6	3.55	2.019	0	
P5	3.888	1.414	$\boxed{1.059}$	0
	P1, P2	P3	P4, P6	P5

1.059 is the smallest, P4, P6 & P5 are combined together.

Step IV :-

P1, P2	0		
P3	5.302	0	
P4, P5, P6	3.66	$\boxed{1.817}$	0
	P1, P2	P3	P4, P5, P6

1.817 is smallest P4, P5, P6 & P3 are combined together.

Step V :-

P1, P2	0	
P3, P4, P5, P6	4.07	0
	P1, P2	P3, P4, P5, P6

Total Marks of Question no.		Examiner	
		Moderator	
		Re-Assessor	

Space for Marks	Question No.	START WRITING HERE

Step VI :-

Dendrogram :-