VIT Vidyalankar Institute of Technology
Accredited A+ by NAAC
(Autonomous College Affiliated to University of Mumbai)

End Semester Examination (CBSGS-C scheme) -(2022-23)

Date: 15/11/2022	Branch: Computer Engineering	Time: 2 Hr.
Semester: VII	Subject: Machine Learning	Marks: 50

N B. :- All Questions are Compulsory

		CO	BL
Q. 1)	Attempt any Five (2 Marks Each)		
a)	Define Gaussian Kernel	CO4	L1
b)	Define Hierarchical Clustering	CO5	L1
c)	List different optimization methods in machine learning.	CO4	L1
d)	State any two applications of Expectation Maximization Clustering	CO5	L1
e)	Differentiate between DBSCAN and K-means algorithm	CO5	L1
f)	State the difference between lazy learner and eager learner	CO5	L1
g)	Define Lift as a performance metric in clustering	CO5	L1
h)	How are border and Noise points calculated in DBSCAN?	CO6	L2
Q. 2)	Attempt any two. (5 Marks Each)		
a)	What is SVM? Explain the following terms: hyperplane, separating hyperplane, margin and support vectors with example.	CO4	L1
b)	Why Dimensionality Reduction is very important in Machine Learning Algorithm?	CO6	L1
c)	Explain Dimensionality Reduction Techniques in Machine Learning: (a) Feature Selection (b) Feature Extraction	CO6	L1
Q 3)	Attempt any One (10 Marks Each)		
a)	What is the goal of Support Vector Machine (SVM)? How to compute the margin?	CO4	L1
b)	Explain the steps of developing Machine Learning applications.	CO5	L1
Q 4)	Attempt any One (10 Marks Each)		
a)	Apply PCA on the following data and find the principle component.	CO6	L2

X	2.3	0.7	2.1	1.2	3.2	2.2	2.6	2	1.6	1.1
Y	2.2	0.9	2.4	2.4	3	3.1	1.8	1.2	1.5	0.8

		CO	BL
b)	Explain kernel tricks in machine learning mathematically.	CO4	L1
Q 5)	Attempt any One (10 Marks Each)		
a)	Write a short note on: PCA	CO6	L1
b)	For the given set of points identify clusters using average linkage	CO5	L2

	A	B
P1	1	1
P2	1.5	1.5
P3	5	5
P4	3	4
P5	4	4
P6	3	3.5

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 algo, which is used to group the unlabeled datasets into a cluster known as hierarchical clustering.

(c) Optimization methods in Machine Learning:-

- Gradient descent
- Stochastic Gradient descent
- Adaptive learning rate method
- Conjugate gradient method.
- Derivative-free optimisation
- Zerothe order optimisation
- For meta learning.

(d) Two applications of Expectation Maximization Clustering :-

- Use to calculate the Gaussian density of a function.

Re-Assessor

START WRITING HERE

Space for Marks	Question No.	

- It helps to fill in the missing data during a sample.

(e) <u>Differentiate between DBSCAN & k-means algorithm:-</u>
- DBSCAN clustering cannot efficiently handle high dimensional datasets.
- kmeans clustering does not work well with outliers & noisy datasets.
- DBSCAN clustering efficiently handles outliers & noisy dataset.

(f)

Lazy learner	Eager learner
(i) It stores dataset without learning from it.	(i) When it receives dataset, it starts classifying.
(ii) Start classifying data when it receives test data	(ii) It does not wait for test-data to learn.
(iii) It takes less time for learning & more time for classifying data.	(iii) It takes long time for learning & less time classifying data.

Question no.		Moderator	
		Re-Assessor	

Space for Marks	Question No.	START WRITING HERE

(g) <u>Lift as a performance metric in clustering:-</u>

- Lift is defined as the strength of any rule.

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

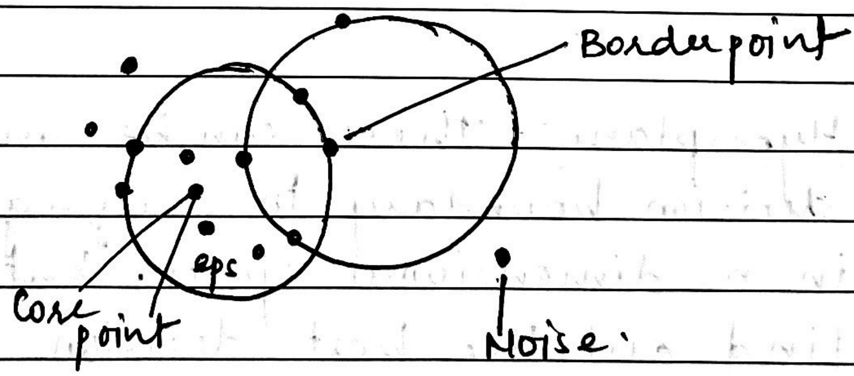
It is the ratio of the observed support measure & expected support if P & Q are independent of each other.

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

(h) <u>DBSCAN:-</u>

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

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

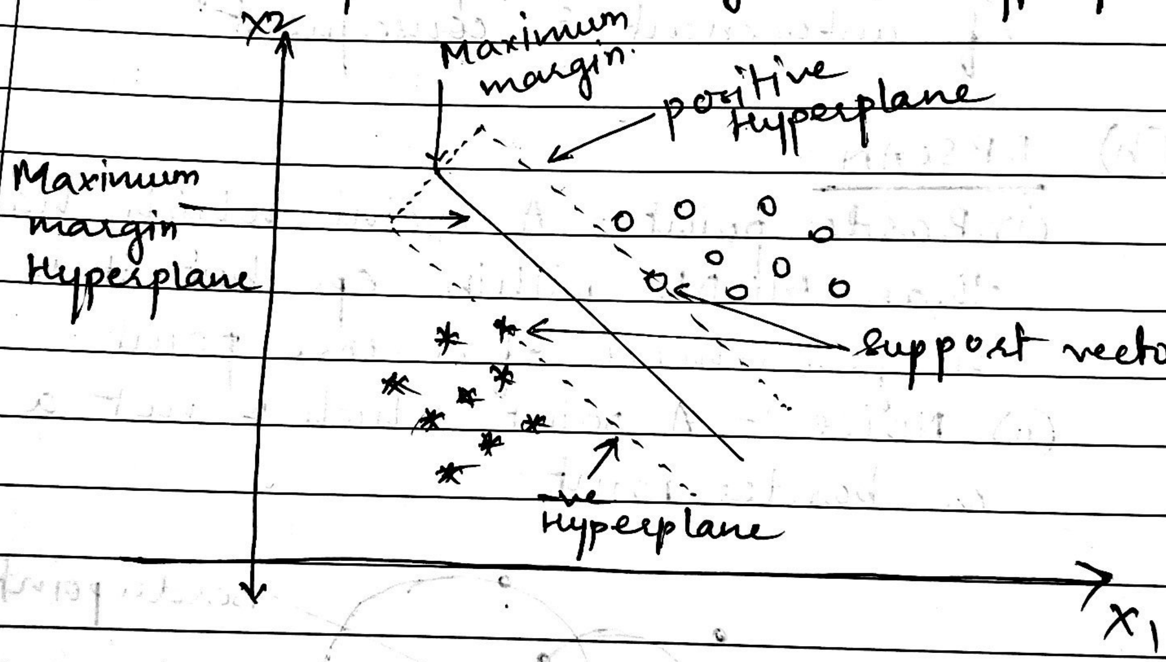


Total Marks of Question no.		Moderator	
		Re-Assessor	

Space for Marks	Question No.	START WRITING HERE

(Q2)
(a) SVM (Support Vector Machine) :-

- SVM is a supervised learning algorithm
- The goal of the SVM algorithm is to create the best 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.

Scanned with CamScanner

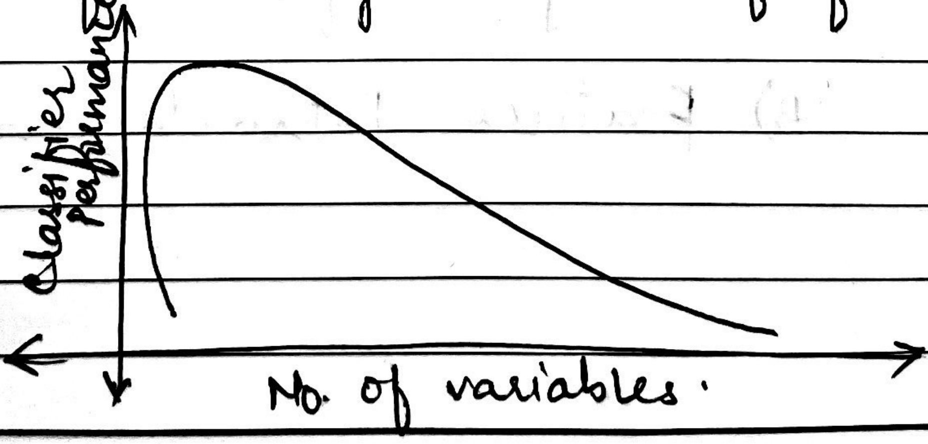
Total Marks of Question no.		Examiner	
		Moderator	
		Re-Assessor	

Space for Marks	Question No.	START WRITING HERE

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 Vector.

(b) Dimension reduction importance in Machine Learning:-

- It assists in information packing & diminishing the storage room required.
- It decreases the time which is required for doing ~~new~~ same calculations. If the dimensions are less then processing will be less, added advantage of having less dimensions is permission to use calculations unfit for countless.
- It handles with multi-collinearity that is used to enhance the execution of the model. It evacuates excess highlights.
- It is helpful in noise evacuation.
- The classifier's performance usually will degrade for a large number of features.



Re-Assessor

Space for Marks	Question No.	START WRITING HERE

(c) <u>Dimensionality reduction technique:-</u>

(A) <u>Feature Selection:-</u>

- Given a set of features $F = \{x_1, \ldots, x_n\}$. 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 scoring function:

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \xrightarrow{\text{feature selection}} \begin{bmatrix} x_{i_1} \\ x_{i_2} \\ \vdots \\ x_{i_M} \end{bmatrix}$$

<u>Feature selection steps:-</u>

feature selection is an optimization problem.

<u>Step-I</u>:- Search the space of possible feature subsets.

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

~~(B) Feature Extraction:-~~

		Re-Assessor	

Space for Marks	Question No.	START WRITING HERE

(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.

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

- A projection matrix w is computed from N-dimensional to M-dimensional vectors to achieve low error.
- $z = w^T x$.
- Principle component analysis & independent component analysis are the feature extraction methods.

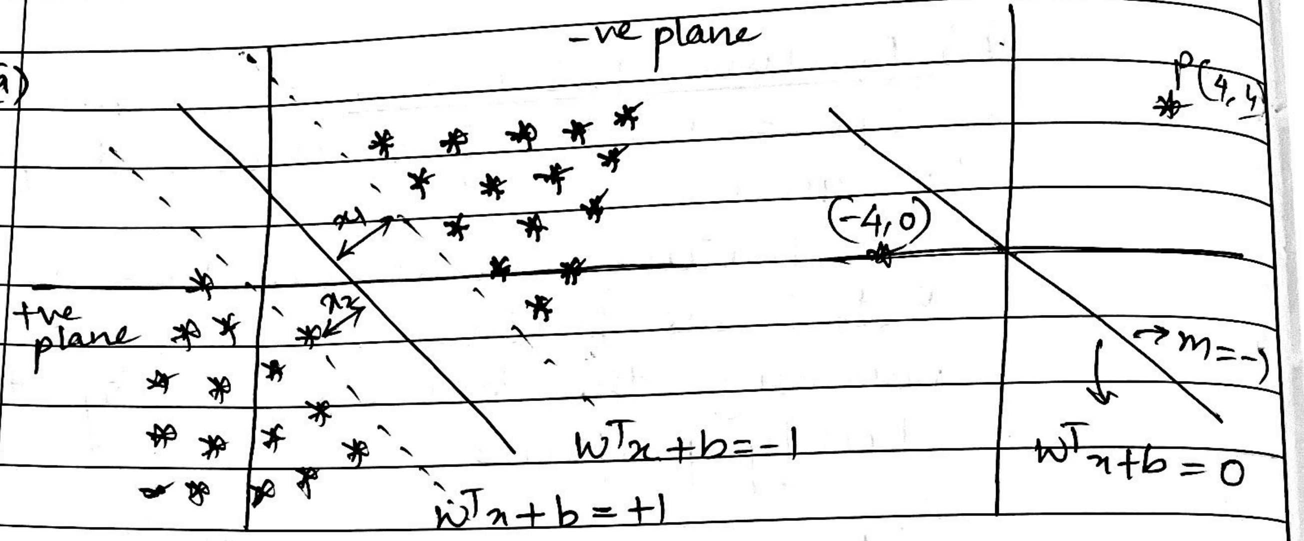
Total Marks of Question no.		Moderator	
		Re-Assessor	

START WRITING HERE

Space for Marks	Question No.

(Q3)(a)

$y = w^T x + b$



we know that, $m = -1$

let c or $b = 0$

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

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

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

$= 4 \Rightarrow$ +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 $\rightarrow x_2 - x_1$

Question no.		Moderator	
		Re-Assessor	

Space for Marks	Question No.	START WRITING HERE

consider +ve & -ve plane equation :-

$$w^T x_1 + b = -1$$
$$w^T x_2 + b = 1$$
$$(-) \quad (-) \quad (-)$$

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

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

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,

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

Question no. | | Re-Assessor

Space for Marks	Question No.	START WRITING HERE

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 error.

(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.
  - Publicly available data.

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

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

(iv) **Train the algorithm :-**
- Good clean data from the first two

pace for Marks	Question No.	START WRITING HERE

steps is given as i/p to the algorithm. The algorithm extracts information or knowledge.

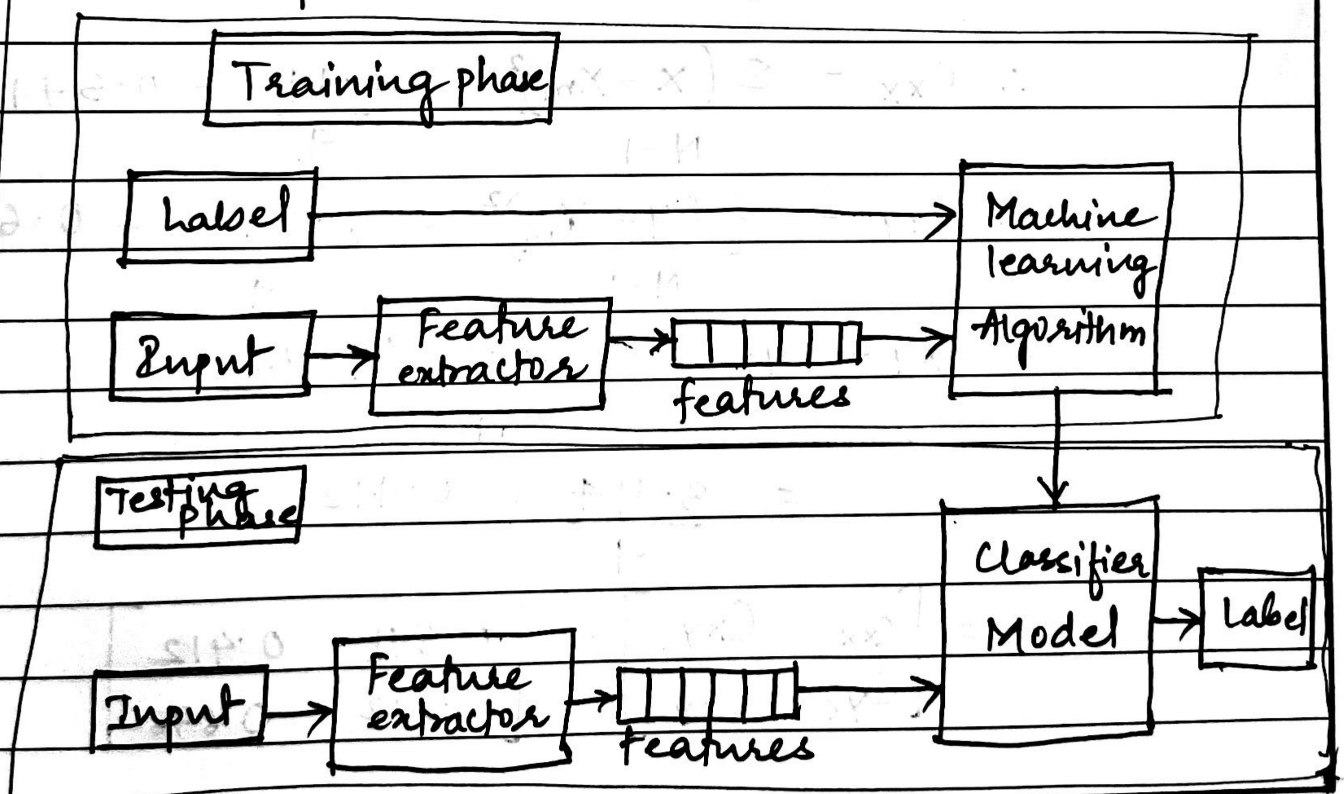
- In case of unsupervised learning, training step is not there because target value is not present.

(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, matrices 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.



Space for Marks	Question No.
	(Q4)
	(a)

PCA :-

X	2.3	0.7	2.1	1.2	3.2	2.2	2.6	2	1.6	1.1
Y	2.2	0.9	2.4	2.4	3	3.1	1.8	1.2	1.5	0.8

Principal component :-

$$\Sigma X = 19$$

$$\Sigma Y = 19.3$$

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

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

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

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

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

$$\therefore C_{xy} = C_{yx} = \frac{\Sigma (X - X_m)(Y - Y_m)}{N-1}$$

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

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

Total Marks of Question no.		Moderator	
		Re-Assessor	

Space for Marks	Question No.	START WRITING HERE

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

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

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

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

$\therefore \boxed{\lambda 1 = 0.414}$

$\therefore \boxed{\lambda 2 = 0.776}$

For $\lambda 1$,

$Cv = \lambda v$

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

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

$0.412 x_1 + 0.676 y_1 = 0.414 x_1$

$\therefore 0.157 x_1 = -0.412 y_1$

$0.412 x_1 = -0.262 y_1$

$\therefore x_1 = -2.624 y_1$ OR $x_1 = -0.636 y_1$

for ks	Question No.

Assume $Y_1 = 1$

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

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

Assume $Y_1 = 1$

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

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

For λ_2 :-

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

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

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

$$\therefore 0.205 x_2 = 0.412 y_2$$

$$0.412 x_2 = 0.1 y_2$$

$\therefore x_2 = 2.009 y_2$ OR $x_2 = 0.242 y_2$

Assume $Y_2 = 1$

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

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

Assume $Y_2 = 1$

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

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

Space for Marks	Question No.	START WRITING HERE

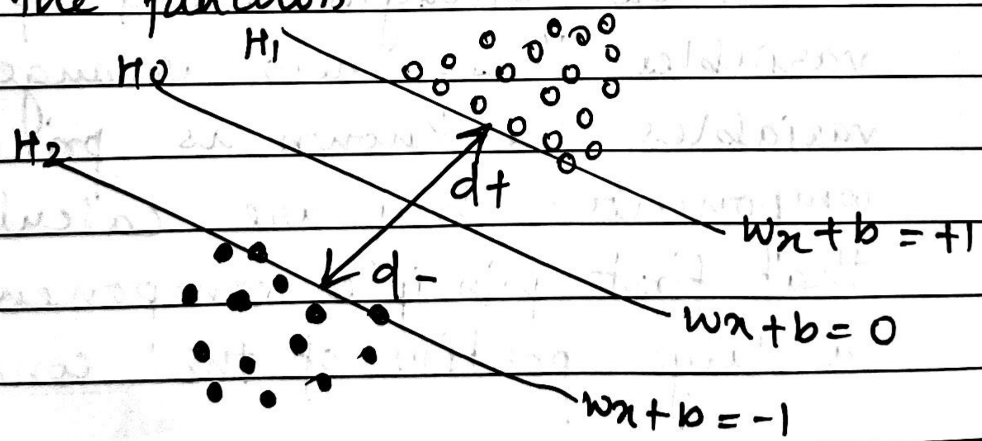
(Q4.b)

Kernel tricks in machine learning mathematically:-

- 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 $\dfrac{|Ax + By + C|}{\sqrt{A^2 + B^2}}$
- In order to maximise the margin, the distance between H_0 & H_1 is then $\dfrac{|w \cdot x + b|}{||w||} = \dfrac{1}{||w||}$, so the total distance

between H_1 and H_2 is $\dfrac{2}{||w||}$

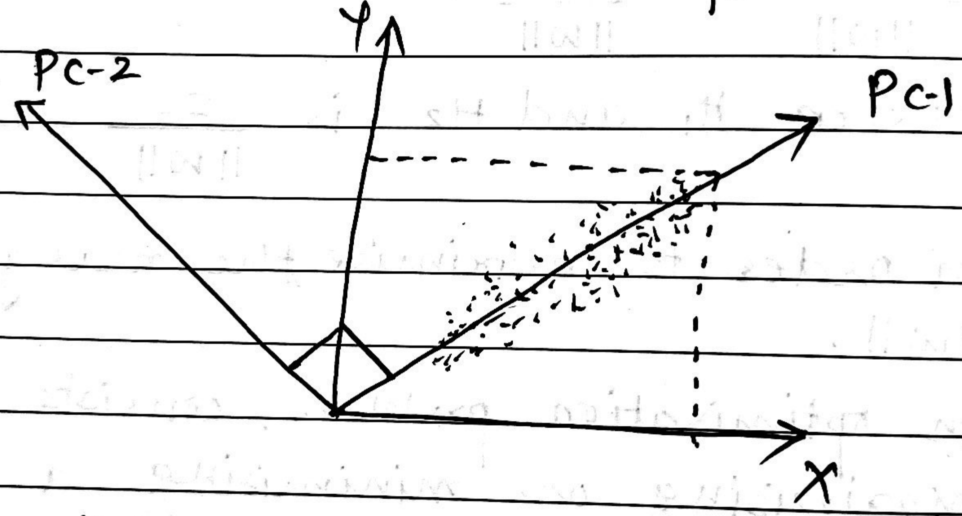
- In order to maximise the margin, minimise $||w||$.
- An optimisation problem, consists of either maximising or minimising a real valued function by choosing i/p values from within an allowed set & computing the value of the function.



Space for Marks | Question No.

(Q5a) **PCA :-**

- Principle Component Analysis is an unsupervised learning technique for reducing the dimensionality of data.
- It increases interpretability yet, at the same time, it minimizes information loss.
- It helps to find the most significant features in a dataset & 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 principle components. They are calculated so that first principle component & represents a large portion of the conceivable

Space for Marks	Question No.	START WRITING HERE

variety of unique information after which each succeeding component has the most conceivable variance.

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

(Q5b) Average linkage :-

	A	B
P1	1	1
P2	1.5	1.5
P3	5	5
P4	3	4
P5	4	4
P6	3	3.5

Step-I :-

Distance Matrix :-

	P1	P2	P3	P4	P5	P6
P1	0					
P2	0.707	0				
P3	5.656	4.949	0			
P4	3.605	2.915	2.236	0		
P5	4.242	3.535	1.414	1	0	
P6	5.201	2.5	1.802	0.5	1.118	0

START WRITING HERE

Space for Marks	Question No.

∴ 0.5 is smallest. P4 & P6 have smallest distance

Step II :-

P1	0				
P2	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

0.707 is smallest. P1 & P2 have smallest distance.

Step III :-

P1, P2	0			
P3	~~5~~ 5.302	0		
P4,P6	3.55	2.019	0	
P5	3.888	1.414	1.059	0
	P1,P2	P3	P4,P6	P5

1.059 is smallest. P4,P6 and P5 are combined together.

for s	Question No.	START WRITING HERE

Step-IV :-

P1, P2	0		
P3	5.302	0	
P4, P5, P6	3.66	1.817	0
	P1, P2	P3	P4, P5, P6

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

Step-V :-

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

Step-VI :-

Dendogram :-