

11/03/23

UNIT-3

Clustering and Regression

⇒ Clustering:

- Clustering (or) cluster analysis is a machine learning techniques, which groups the unlabelled dataset. It can be defined as "A way of grouping the data points into different clusters, consisting of similar data points."
- * The objects with the possible similarities remain in a group that has less (or) no similarities with another group.
 - * It does it by finding some similar patterns in the unlabelled dataset such as shape, size, color behaviour etc.... & divides them as per the presence & absence of the similar patterns.
 - * It is an Unsupervised learning method, hence no supervision is provided to the algorithm & it deals with the unlabeled dataset.
 - * After applying this clustering technique, each cluster (or) group is provided with a cluster id. ml system can use this id to simplify the processing of large & complex datasets.
 - * The clustering technique is commonly used for statistical data analysis. Some most common users of this technique are,
 1. Market Segmentation.
 2. Statistical data Analytics.
 3. Social network Analysis.

→ R

X

O

✓

X

✓

15/03/

→ K

20

Ans

1.

2.

3.

Data

a₁(

a₂(

a₃(

a₄(

a₅(

a₆(

chine learning
dataset.

g the data
similar

remain
es with
patterns
size, color
the presence

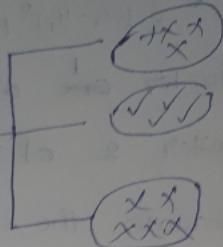
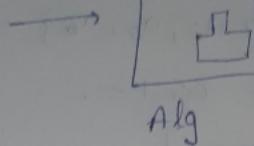
service no
o & it
each
ks id.
the

sed for
non

4. Image Segmentation
5. Anomaly detection etc...

→ Raw data:

$\begin{matrix} x & x & v \\ o & o & o \\ \checkmark & x & o \\ x & x & \\ \checkmark & \checkmark & \checkmark \end{matrix}$



15/03/23

→ K means clustering Alg to divide following data into
2 cluster set

x_1	1	2	2	3	4	5
x_2	1	1	3	2	3	5

Ans: Choosing randomly 2 cluster sets

1. choosing randomly 2 cluster sets

$$v_1 = (2, 1)$$

$$v_2 = (2, 3)$$

2. finding the distance b/w the cluster center &
each data points.

Distance Table:-

Data point	Distance from $v_1 (2, 1)$	Distance from $v_2 (2, 3)$	Assigned Centre
$a_1 (1, 1)$	1	$\sqrt{5} = 2.236$	v_1
$a_2 (2, 1)$	0	2	v_1
$a_3 (2, 3)$	2	0	v_2
$a_4 (3, 2)$	$\sqrt{2} = 1.414$	$\sqrt{2} = 1.414$	v_1
$a_5 (4, 3)$	$2\sqrt{2} = 2.828$	2	v_2
$a_6 (5, 5)$	5	$\sqrt{13} = 3.605$	v_2

→ Euclidean distance

$$(x_1, x_2) \quad (y_1, y_2)$$

$$= \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}.$$

→ 3. cluster ~~one~~ of $v_1 = \{a_1, a_2, a_4\}$

cluster 2 of $v_2 = \{a_3, a_5, a_6\}$

4. Recalculate the cluster centres.

$$v_1 = \frac{1}{3} [(1, 1) + (2, 1) + (3, 2)]$$

$$= \frac{1}{3} [(6, 4)] = (2, 1.33)$$

$$v_2 = \frac{1}{3} [(2, 3) + (4, 3) + (5, 5)]$$

$$= \frac{1}{3} [(11, 11)] = (3.67, 3.67)$$

5. Repeat from step 2, until we get same cluster centre (v_1) same cluster elements as in the previous iteration.

Data point	$v_1 (2, 1.33)$	$v_2 (3.67, 3.67)$	Assigned Centre
$a_1 (1, 1)$	1.05	3.78	v_1
$a_2 (2, 1)$	0.33	3.15	v_1
$a_3 (2, 3)$	1.67	1.8	v_1
$a_4 (3, 2)$	1.204	1.8	v_1
$a_5 (4, 3)$	2.605	0.75	v_2
$a_6 (5, 5)$	4.74	1.88	v_2

cluster 1 of $v_1 = \{a_1, a_2, a_3, a_4\}$

cluster 2 of $v_2 = \{a_5, a_6\}$

$$v_1 = \frac{1}{4} [(1,1) + (2,1) + (3,2) + (2,3)]$$

$$= \frac{1}{4} (8,7) = (2,1.75)$$

$$v_2 = \frac{1}{2} [(4,3) + (5,5)]$$

$$= \frac{1}{2} (9,8) = (4.5, 4)$$

→ so, cluster elements are not same as previous

iteration Elements of v_1 are $(1,1), (2,1), (3,2), (2,3)$

Data point	$v_1 (2,1.75)$	$v_2 (4.5, 4)$	Assigned Centre
$a_1 (1,1)$	1.25	4.61	v_1
$a_2 (2,1)$	0.75	3.9	v_1
$a_3 (2,3)$	1.25	2.69	v_1
$a_4 (3,2)$	1.03	2.5	v_2
$a_5 (4,3)$	2.36	1.12	v_2
$a_6 (5,5)$	4.42	1.12	v_2

→ cluster 1 of $v_1 = \{a_1, a_2, a_3, a_4\}$

cluster 2 of $v_2 = \{a_5, a_6\}$

~~$\frac{1}{4}$~~ cluster elements are same as
in the previous iteration. so, our clusters

cluster 1 = $\{(1,1), (2,1), (2,3), (3,2)\}$

cluster 2 = $\{(4,3), (5,5)\}$

16/03/23

→ HAC:- Hierarchical Agglomerative clustering.

* HAC starts with one cluster, individual item in its own cluster & iteratively merge clusters until all the items belong to one cluster.

* Bottom up approach is followed to merge the clusters together.

* Dendrogram (or) pictorially used to represent the HAC.

* HAC represents 3 techniques:

1. single, nearest (or) single linkage.

2. Complete-farthest distance (or) complete linkage.

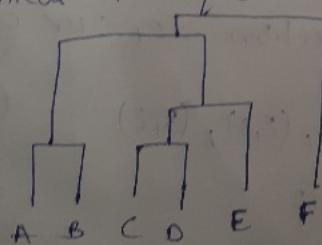
3. Average distance (or) Average linkage.

1. S.N:- This is the diff b/w the closest members of the two clusters.

2. C.L:- This is the diff b/w the members that are farthest apart.

3. A.L:- This Method involves looking at a distance b/w all pairs & averages all of these distance. this is also called unweighted pair group mean average

* Dendrogram:- A tree like structure which represent hierarchical technique.



Problem:-

* find the clusters using single link technique & draw the denrogram diagram.

	α	γ
P ₁	0.40	0.53
P ₂	0.22	0.38
P ₃	0.35	0.32
P ₄	0.26	0.19
P ₅	0.08	0.41
P ₆	0.45	0.30

linkage.

Sol:-

1. Euclidean distance

$$(\alpha_1, \gamma_1) (\alpha_2, \gamma_2) = \sqrt{(\alpha_1 - \alpha_2)^2 + (\gamma_1 - \gamma_2)^2}$$

	P ₁	P ₂	P ₃	P ₄	P ₅	P ₆
P ₁	0					
P ₂	0.23	0				
P ₃	0.22	0.15	0			
P ₄	0.37	0.20	0.15	0		
P ₅	0.34	0.34	0.26	0.29	0	
P ₆	0.23	0.25	0.11	0.22	0.39	0

Distance Table

→ Smallest element = 0.11

P₃ P₆

3 6

Recalculate the distance Matrix

(2)	P ₁	P ₂	P ₃ P ₆	P ₄	P ₅
P ₁	0				
P ₂	0.23	0			
P ₃ P ₆	0.22	0.15	0		
P ₄	0.37	0.20	0.15	0	
P ₅	0.34	0.34	0.26	0.29	0

Small value = 0.15

$\min(\text{dist}(x, y))$

$$\min(\text{dist}(P_3, P_6), P_1) = \min(\text{dist}((P_3, P_1), (P_6, P_1))).$$

$$= \min(\text{dist}(0.22, 0.23))$$

$$= 0.22$$

(3) updated dist Matrix.

	P ₁	P ₂ P ₅	P ₃ P ₆	P ₄
P ₁	0			
P ₂ P ₅	0.23	0		
P ₃ P ₆	0.22	0.15	0	
P ₄	0.37	0.20	0.15	0

small element = 0.15

\Rightarrow New cluster $[(P_2, P_5), (P_3, P_6)]$

	P_1	$P_2 P_5 P_3 P_6$	P_4
P_1	0	0.22	0.37
$P_2 P_5$	0.22	0	0.15
$P_3 P_6$	0.37	0.15	(0.22) b merge $(P_2 P_5)$ b
P_4			<small>small.</small>

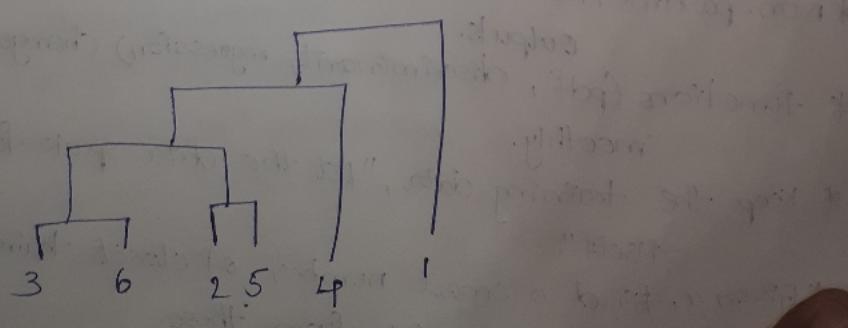
$\rightarrow \min(\text{dist}[(C P_2 P_5 P_3 P_6), P_1])$ and disjoint (a)
 $= \min(\text{dist}[(C P_2 P_5), P_1] [P_3 P_6, P_1])$ b
 $= \min(\text{dist}(0.22) (0.22))$ disjoint - overlapping
 $= 0.22$. [distance between 2 elements]

$\rightarrow \min(\text{dist}[(C P_2 P_5 P_3 P_6), P_4])$ overlap & bellow
 $= 0.15$. overlap & below

\Rightarrow updated

	P_1	$P_2 P_5 P_3 P_6 P_4$
P_1	0	0.22
$P_2 P_3$	0.22	0
$P_6 P_4$		

Diagram :-



17/03/23

→ Agglomerative clustering:

Distance b/w two groups G_i & G_j

1) single link

$$d(G_i, G_j) = \min d(x^i, x^j)$$

2) Complete-link

$$d(G_i, G_j) = \max d(x^i, x^j)$$

Dendrogram :- Decompose data objects into several levels of nested partitioning (tree of clusters) called a dendrogram.

⇒ Density Estimation:

* Given the training set $\{x\}$ drawn iid from $p(x)$

* Divide data into bins of size h

* Histogram :-

$$p(x) = \frac{\#\{x-h < x \leq x+h\}}{2Nh}$$

⇒ Non-parametric methods:

* parametric (single global model), semi-parametric (small number of local models).

* Non-parametric : similar inputs have similar outputs.

* functions (pdf, discriminant, regression) change smoothly.

* Keep the training data, "let the data speak for itself".

* Given x , find a small number of closest training instances & interpolate from these.

* Aka Lazy/memory based / case-based / instance-based learning.

⇒ Kernal Estimator:

kernal function, eg:- Gaussian

Kernal:-

$$K(u) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{u^2}{2}\right]$$

Kernal Estimator: (Parzen window)

$$\hat{p}(x) = \frac{1}{Nh} \sum_{t=1}^n K\left[\frac{x-x^t}{h}\right]$$

→ K= Nearest Neighbour Estimator

$$\hat{p}(x) = \frac{1}{2N} d_k(x)$$

Generalisation of multivariate data:

1. Kernal density Estimator

2. multivariate Gaussian kernal density

↓
1. Spherical

2. Ellipsoidal

→ Hamming distance:

$$H_0(x, x^t) = \sum_{j=1}^d I(x_j \neq x_j^t)$$

→ Non-parametric classification:

1. kernal estimator

2. KNN Estimator

→ Non-parametric Regression:

1. Running mean Smoother

2. Kernal smoothers.

$$1. \hat{g}(\alpha) = \frac{\sum_{t=1}^N w_t (\alpha - \alpha_t)^+ \cdot x_t^+}{\sum_{t=1}^N w_t (\alpha - \alpha_t)^+}$$

$$2. \hat{g}(\alpha) = \frac{\sum_{t=1}^N k_t (\alpha - \alpha_t)^+ \cdot x_t^+}{\sum_{t=1}^N k_t (\alpha - \alpha_t)^+}$$

\Rightarrow Linear Discrimination:

$$g_i(\alpha_i / w_i) = w^T x + w = \sum w_j x_j + w.$$

* likelihood vs discriminant based classification.

1. Assume model for $P(c_i | \alpha)$, use Baye's rule

to calculate $p(c_i | \alpha) g_i(\alpha) = \log p(c_i | \alpha)$.

2. D.B.C = Assume model for $p(\alpha | d_i)$ NO
density estimation. Estimating the boundary
is enough. No need to accurately estimate the
densities inside the boundaries.

18/03/23

\rightarrow KNN Algorithm:

* KNN is a Non-parametric classification Method.

* It is used for classification & regression.

* It is very simple & easy to implement supervised
Machine learning algorithm.

Advantages:-

* more effective if training data is large.

→ problem :-

S.No	maths	CS	Result
1	4	3	Fail
2	6	7	Pass
3	7	8	Pass
4	5	5	Fail
5	8	8	Pass
6	6	8	Pass

$$x = (\text{math} - 6, \text{CS} - 8)$$

$$(1) \text{ Let } x = 3 \Rightarrow \text{prob} =$$

(a)

x_2, y_2 - Observed values

x_1, y_1 - Actual values.

→ Euclidean Distance :-

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

$$1. = \sqrt{(6-4)^2 + (8-3)^2}$$

$$= \sqrt{4 + 25} = \sqrt{29} = 5.38$$

$$2. \sqrt{(6-6)^2 + (8-7)^2} = 1$$

$$3. \sqrt{(6-7)^2 + (8-8)^2} = 1$$

$$4. \sqrt{(6-5)^2 + (8-5)^2} = \sqrt{10} = 3.16$$

$$5. \sqrt{(6-8)^2 + (8-8)^2} = \sqrt{4} = 2$$

→ As per result $K=3$ nearest

2-pass

3-pass

5-pass

so, 3 greater than 0, probability of pass
is high.

Hence, $x = (\text{math} = 6, \text{CS} = 8)$ is pass.

seeks

→ Decision tree:

- * Decision tree is a tree like a predictive model.
- used for classification & regression.
- * it is a supervised learning method.

⇒ Learning Algorithm used:

① Cost → Regression → Gini index.

② ID3 Algorithm - Classification

↓ Entropy, Information Gain.

$$\rightarrow \text{Entropy} = \sum_{(D)}^C -P_i \log_2(P_i)$$

$$\rightarrow \text{Gain}(D, A) = \text{Entropy}(D) - \sum_{\text{Value}(A)} \frac{|D_v|}{|D|} \times \text{Entropy}(D_v)$$

problem:

It is used to generate Decision Tree from a Dataset.

→ Classify the given training set using ID3 Algorithm

Construct the decision tree.

<u>size</u>	<u>color</u>	<u>shape</u>	<u>class</u>
Small	Yellow	Round	A
Big	Yellow	Round	A
Big	Red	Round	A
Small	Red	Round	A
Small	Black	(Round) Round	B
Big	Black	Cube	B
Big	Yellow	Cube	B
Big	Black	Round	B
Small	Yellow	Cube	B

mode)

Sol:-

$$A=4 \quad \checkmark \quad \left(\frac{4}{9} \log_2 \left(\frac{4}{9} \right) + \frac{5}{9} \log_2 \left(\frac{5}{9} \right) \right) H = 0.992$$
$$B=5 \quad \checkmark \quad \left(\frac{2}{5} \log_2 \left(\frac{2}{5} \right) + \frac{3}{5} \log_2 \left(\frac{3}{5} \right) \right) H$$
$$m=2$$

Calculate the Entropy (O).

$$\text{Entropy}(O) = - \sum_{i=1}^m p_i \log_2(p_i)$$
$$= - \sum_{i=1}^2 4/9 \log_2(4/9)$$

$$= - \frac{4/9 \log_2(4/9)}{2} - \frac{5/9 \log_2(5/9)}{2}$$

$$= (-0.444 * -1.171) - (0.555 * -0.850)$$

$$= 0.520 + 0.472 H = 0.992$$

Identify the splitting attribute, calculate the gain.

* Attribute - size:

	A	B	
Small	2	2	-4
Big	2	3	-5

$$\text{Gain}(O, \text{size}) = 0.992 - \left[\frac{4}{9} \left(-\frac{2}{4} \log_2 \left(\frac{2}{4} \right) - \frac{2}{4} \log_2 \left(\frac{2}{4} \right) \right) + \frac{5}{9} \left(-\frac{3}{5} \log_2 \left(\frac{3}{5} \right) - \frac{2}{5} \log_2 \left(\frac{2}{5} \right) \right) \right]$$

$$= 0.992 - [0.444 + 0.539]$$

$$= 0.992 - 0.983$$

$$= 0.009$$

* Attribute - color:

	A	B	
yellow	2	2	-4
red	2	0	-2
black	3	0	-3

$$\begin{aligned} \text{Gain}(D, \text{color}) &= 0.992 - \left[\frac{4}{9} \left(-\frac{2}{4} \log_2 \left(\frac{2}{4} \right) - \frac{2}{4} \log_2 \left(\frac{2}{4} \right) \right) + \right. \\ &\quad \left. \frac{2}{9} \left(-\frac{2}{2} \log_2 \left(\frac{2}{2} \right) \right) + \right. \\ &\quad \left. \frac{3}{9} \left(-\frac{2}{3} \log_2 \left(\frac{2}{3} \right) - \frac{3}{3} \log_2 \left(\frac{3}{3} \right) \right) \right] \\ &= 0.992 - [0.4444 + 0.0] \\ &= 0.55. \end{aligned}$$

size	1
Small	
Big	
Big	
Small	

* Attribute - shape

$$\begin{array}{ccc} A & B & C \\ \hline \text{Round} & 4 & 2 \\ \text{cube} & 3 & 0 \end{array}$$

$$\begin{aligned} \text{Gain}(D, \text{shape}) &= 0.992 - \left[\frac{6}{9} \left(-\frac{4}{6} \log_2 \left(\frac{4}{6} \right) - \frac{2}{6} \log_2 \left(\frac{2}{6} \right) \right) + \right. \\ &\quad \left. \frac{3}{9} \left(-\frac{3}{3} \log_2 \left(\frac{3}{3} \right) \right) \right] \\ &= 0.992 - 0.612 \\ &= 0.38. \end{aligned}$$

→ Information Gain

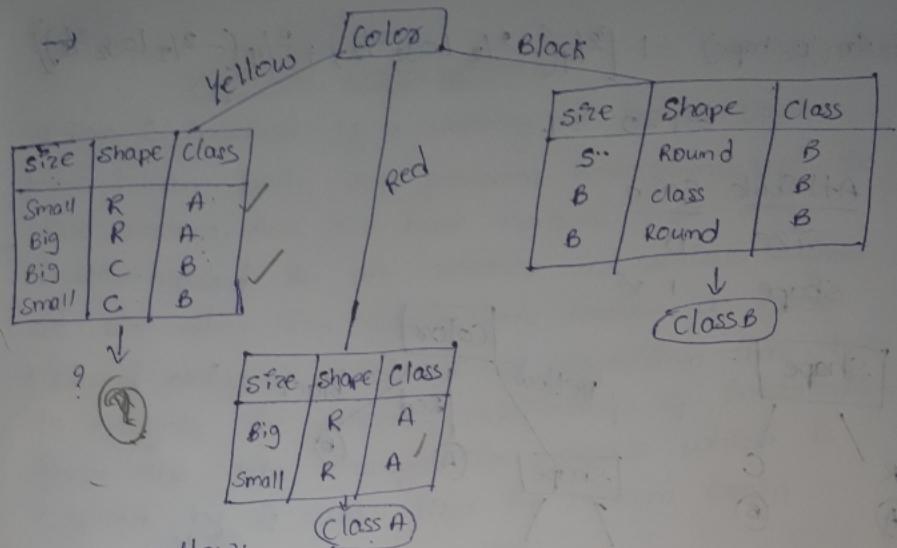
Attribute	Gain
size	0.009
color	0.55
shape	0.379

* color has the largest gain, so it becomes the root node.

$$\frac{2}{4} \log_2(2/4) +$$

+

$$-\frac{3}{3} \log_2(3/3)]$$



yellow:

$$\text{Entropy}(D) = -\frac{2}{4} \log_2 \frac{2}{4} - \frac{2}{4} \log_2 \frac{2}{4}$$

$$= 1.$$

→ Attribute = size:

	A	B	
small	1	1	2
Big	1	1	2

$$\text{Gain}(D, \text{size}) = 1 - \left[\frac{2}{4} \left(-\frac{1}{2} \log_2 \frac{1}{2} - \frac{1}{2} \log_2 \frac{1}{2} \right) + \frac{2}{4} \left(-\frac{1}{2} \log_2 \frac{1}{2} - \frac{1}{2} \log_2 \frac{1}{2} \right) \right]$$

$$= 1 - [0.5 + 0.5]$$

$$= 0.$$

→ Attribute : shape:

	A	B	
Round	2	0	2
Cube	0	2	2

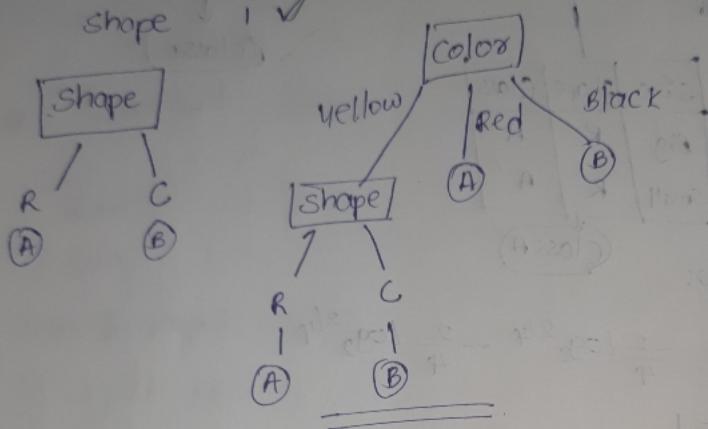
$$\text{Gain}(D, \text{shape}) = 1 - [2/4(-2/2 \log_2 2/2) + 2/4(-2/2 \log_2 2/2)]$$

25/03/

$$= 1 - 0 = 1.$$

Attribute Gain

size	Gain
shape	1 ✓



* IPS:-

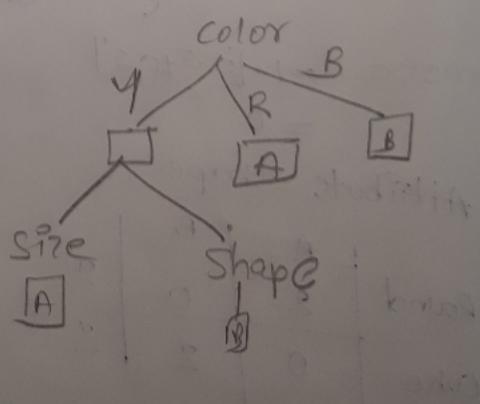
$$* \text{Entropy} = -\frac{P}{P+m} \log_2 \left(\frac{P}{P+m} \right) - \frac{m}{P+m} \log_2 \left(\frac{m}{P+m} \right)$$

* Average Info.:

$$I(\text{Attribute}) = \sum \frac{P_i + m_i}{P+m} \text{Entropy}(s_i)$$

$$\text{Gain} = \text{Entropy}(s) - I(\text{Attribute}).$$

→ ~~Entropy~~.



1) Find the clusters using single linkage technique

Sindhuja. V

40111233

'E1' Section
CSE

	x	y
P ₁	0.40	0.53
P ₂	0.22	0.38
P ₃	0.35	0.32
P ₄	0.26	0.19
P ₅	0.08	0.41
P ₆	0.45	0.30

Soln:

Step 1: Euclidean distance

$$\text{Distance} [(x, y) (a, b)] = \sqrt{(x-a)^2 + (y-b)^2}$$

Step 2:

Get the distance Matrix

Table:

	P ₁	P ₂	P ₃	P ₄	P ₅	P ₆
P ₁	0					
P ₂	0.234	0				
P ₃	0.22	0.16	0			
P ₄	0.37	0.20	0.16	0		
P ₅	0.34	0.14	0.28	0.28	0	
P ₆	0.24	0.26	0.11	0.22	0.39	0

$$d(P_2, P_1)$$

$$\begin{aligned} &= \sqrt{(0.22 - 0.40)^2 + (0.38 - 0.53)^2} \\ &= \sqrt{0.08 + 0.02} \\ &= 0.234 \end{aligned}$$

$$d(P_3, P_1)$$

$$\begin{aligned} &= \sqrt{(0.35 - 0.40)^2 + (0.32 - 0.53)^2} \\ &= \sqrt{0.0025 + 0.04} \\ &= 0.22 \end{aligned}$$

$$d(P_3, P_2)$$

$$\begin{aligned} &= \sqrt{(0.35 - 0.22)^2 + (0.32 - 0.38)^2} \\ &= \sqrt{0.0169 + 0.0036} \\ &= \sqrt{0.0205} = 0.15 \end{aligned}$$

$$d(P_4, P_1)$$

$$\begin{aligned} &= \sqrt{(0.26 - 0.40)^2 + (0.19 - 0.53)^2} \\ &= \sqrt{0.0196 + 0.1156} \\ &= \sqrt{0.1352} = 0.37 \end{aligned}$$

$$d(P_4, P_2)$$

$$\begin{aligned} &= \sqrt{(0.26 - 0.22)^2 + (0.19 - 0.38)^2} \\ &= \sqrt{0.0016 + 0.0361} \\ &= \sqrt{0.0377} = 0.20 \end{aligned}$$

(3)

$$d(P_4, P_3)$$

$$\begin{aligned} &= \sqrt{(0.08 - 0.35)^2 + (0.41 - 0.32)^2} \\ &= \sqrt{0.0081 + 0.0169} \\ &= 0.15 \end{aligned}$$

$$d(P_5, P_1)$$

$$\begin{aligned} &= \sqrt{(0.08 - 0.40)^2 + (0.41 - 0.53)^2} \\ &= \sqrt{0.1024 + 0.0144} \\ &= 0.34 \end{aligned}$$

$$d(P_5, P_2)$$

$$\begin{aligned} &= \sqrt{(0.08 - 0.22)^2 + (0.41 - 0.38)^2} \\ &= \sqrt{0.0196 + 0.0009} \\ &= \sqrt{0.0205} = 0.14 \end{aligned}$$

$$d(P_5, P_3)$$

$$\begin{aligned} &= \sqrt{(0.08 - 0.26)^2 + (0.41 - 0.19)^2} \\ &= \sqrt{0.0324 + 0.0484} \\ &= 0.28 \end{aligned}$$

$$d(P_6, P_1)$$

$$\begin{aligned} &= \sqrt{(0.45 - 0.40)^2 + (0.30 - 0.53)^2} \\ &= \sqrt{0.0025 + 0.0529} \\ &= 0.24 \end{aligned}$$

$d(P_6, P_2)$

$$= \sqrt{(0.45 - 0.22)^2 + (0.30 - 0.38)^2}$$

$$= \sqrt{0.0529 + 0.0064}$$

$$= 0.25$$

$d(P_6, P_3)$

$$= \sqrt{(0.45 - 0.35)^2 + (0.30 - 0.32)^2}$$

$$= \sqrt{0.01 + 0.0004}$$

$$= 0.11$$

$d(P_6, P_4)$

$$= \sqrt{(0.45 - 0.26)^2 + (0.30 - 0.19)^2}$$

$$= \sqrt{0.0361 + 0.0121}$$

$$= 0.22$$

$d(P_6, P_5)$

$$= \sqrt{(0.45 - 0.08)^2 + (0.30 - 0.41)^2}$$

$$= \sqrt{0.1369 + 0.0121}$$

$$= \sqrt{0.149}$$

$$= 0.39$$

Step 3:

Find the minimum element from the distance matrix
∴ 0.11 is the smallest value (P_3, P_6)

(i) Recalculate the distance matrix,
 $\min(\text{dist}(x,y))$

Table Q:

	P_1	P_2	P_3, P_6	P_4	P_5
P_1	0				
P_2	0.23	0			
P_3, P_6	0.22	0.15	0		
P_4	0.37	0.20	0.15	0	
P_5	0.34	0.14	0.28	0.29	0

$$\min(\text{dist}((P_3, P_6), P_1))$$

$$\min(\text{dist}((P_3, P_1), (P_6, P_1)))$$

$$\min(\text{dist}(0.22, 0.23))$$

$$= 0.22$$

$$\min(\text{dist}((P_3, P_6), P_2))$$

$$\min(\text{dist}((P_3, P_2), (P_6, P_2)))$$

$$\min(\text{dist}(0.15, 0.22))$$

$$= 0.15$$

$$\min(\text{dist}((P_3, P_6), P_4))$$

$$\min(\text{dist}((P_3, P_4), (P_6, P_4)))$$

$$\min(\text{dist}(0.15, 0.22))$$

$$= 0.15$$

$$\min(\text{dist}((P_3, P_6), P_5))$$

$$\min(\text{dist}((P_3, P_5), (P_6, P_5)))$$

$$\min(\text{dist}(0.28, 0.39))$$

$$= 0.28$$

\therefore Smallest element is $0.15 (P_2, P_5)$

Table 3:

	P_1	$P_2 \ P_6$	$P_3 \ P_4$	P_5
P_1	0			
$P_2 \ P_5$	0.23	0		
$P_3 \ P_6$	0.22	0.15	0	
P_4	0.37	0.20	0.15	0

$$\min(\text{dist}((P_2, P_5), P_1))$$

$$\min(\text{dist}(P_2, P_1), (P_5, P_1))$$

$$\min(\text{dist}(0.23, 0.37))$$

$$= 0.23$$

$$\min(\text{dist}(P_3, P_6), (P_2, P_6))$$

$$\min(\text{dist}((P_3, P_2), (P_3, P_4), (P_6, P_2), (P_6, P_5)))$$

$$\min(0.15, 0.28, 0.26, 0.39)$$

$$= 0.15$$

$$\min(\text{dist}(P_2, P_6), P_4))$$

$$\min(\text{dist}((P_2, P_4), (P_6, P_4)))$$

$$\min(\text{dist}(0.20, 0.28))$$

$$= 0.20$$

\therefore Smallest element = 0.15

$$[(P_3, P_6), (P_2, P_5)]$$

Table : 4

	P_1	$P_2 P_5 P_3 P_6$	P_4
P_1	0		
$P_2 P_5 P_3 P_6$	0.23	0	
P_4	0.39	0.16	0

$$\min(\text{dist}((P_2 P_5 P_3 P_6), P_1))$$

$$\min(\text{dist}((P_2, P_1), (P_5, P_1), (P_3, P_1), (P_6, P_1)))$$

$$= \min(\text{dist}(0.23, 0.34, 0.22, 0.23))$$

$$= 0.23$$

$$\min(\text{dist}((P_2 P_5 P_3 P_6), P_4))$$

$$\min(\text{dist}((P_2, P_4), (P_5, P_4), (P_3, P_4), (P_6, P_4)))$$

$$= \min(\text{dist}(0.20, 0.28, 0.15, 0.22))$$

$$= 0.15$$

$$\therefore \text{Smallest element} = 0.15 [P_4, P_2 P_5 P_3 P_6]$$

Table 5:

	P_1	$P_2 P_5 P_3 P_6 P_4$
P_1	0	
$P_2 P_5 P_3 P_6 P_4$	0.22	0

$$\min(\text{dist}((P_2, P_5), (P_3, P_6), (P_4, P_1)))$$

$$\min(\text{dist}((P_2, P_1), (P_5, P_1), (P_3, P_1), (P_6, P_1), (P_4, P_1)))$$

$$\min(\text{dist}(0.23, 0.34, 0.22, 0.26, 0.34))$$

$$\approx 0.22$$

Dendrogram:

