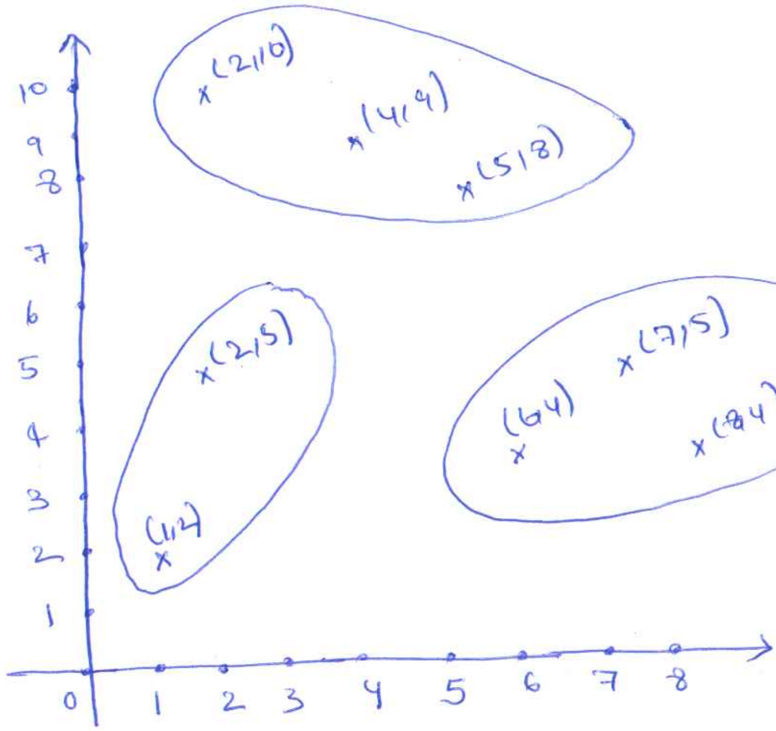
S. Ramesh kumar
2021205868

1). (i) A1 (2,10), A2 (~~8~~ 2,5), A3 (8,4), A4 (5,8), A5 (7,5), A6 (6,4), A7 (1,2), A8 (4,9)



(ii)

	A1	A2	A3	A4	A5	A6	A7	A8
A1	0	$\sqrt{25}$	$\sqrt{36}$	$\sqrt{13}$	$\sqrt{50}$	$\sqrt{52}$	$\sqrt{65}$	$\sqrt{5}$
A2		0	$\sqrt{37}$	$\sqrt{18}$	$\sqrt{25}$	$\sqrt{17}$	$\sqrt{10}$	$\sqrt{20}$
A3			0	$\sqrt{25}$	$\sqrt{2}$	$\sqrt{2}$	$\sqrt{53}$	$\sqrt{41}$
A4				0	$\sqrt{13}$	$\sqrt{17}$	$\sqrt{52}$	$\sqrt{2}$
A5					0	$\sqrt{2}$	$\sqrt{45}$	$\sqrt{25}$
A6						0	$\sqrt{29}$	$\sqrt{29}$
A7							0	$\sqrt{58}$
A8								0

Initial center points c1 – (2,10), c2 (5,8), c3(1,2)

and let C1, C2, C3 be the 3 clusters.

A1:

$d(A_1, C_1) = 0$

$d(A_1, C_2) = \sqrt{13} > 0$

$d(A1, C3) = \sqrt{65} > 0$

min = 0, $A1 \in C1$

A2:

$d(A2, C1) = 5$

$d(A2, C2) = 4.24$

$d(A2, C3) = 3.16$

min = 3.16, $A2 \in C3$

A3:

$d(A3, C1) = 6$

$d(A3, C2) = 5$

$d(A3, C3) = 7.28$

min = 5, $A3 \in C2$

A4:

$d(A4, C1) = \sqrt{13}$

$d(A4, C2) = 0$

$d(A4, C3) = \sqrt{52}$

min = 0, $A4 \in C2$

A5:

$d(A5, C1) = 7.07$

$d(A5, C2) = 3.6$

$d(A5, C3) = 6.70$

min = 3.6, $A5 \in C2$

A6:

$d(A6, C1) = 7.21$

$d(A6, C2) = 4.12$

$d(A6, C3) = 5.38$

min = 4.12,
$A6 \in C2$

A7:

$d(A7, C1) = \sqrt{65}$

$d(A7, C2) = \sqrt{52}$

$d(A7, C3) = 0$

$A7 \in C3$

A8:

$d(A8, C1) = \sqrt{5}$

$d(A8, C2) = \sqrt{2}$

$d(A3, C3) = \sqrt{58}$

$\sqrt{2}$ is min, $A8 \in C2$

clusters formed after iteration are $\{A1\}$, $\{A3, A4, A5, A6, A8\}$, $\{A2, A7\}$

Centers formed are (2,5), (6,6), (1.5, 3.5).

(iii) After the 2nd iteration, the results would be

c1 {A1, A8}, c2 {A3, A4, A5, A6}, c3 {A2, A7}

and centers are (3, 9.5), (6.5, 5.25), (1.5, 3.5)

After the 3rd iteration, the results would be

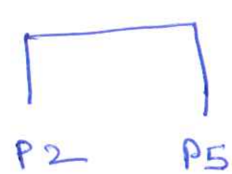
c1 {A1, A4, A8}, c2 {A3, A5, A6}, c3 {A2, A7}

and centers are (3.66, 9), (7, 4.33), (1.5, 3.5).

16) Similarity Matrix — ~~Single~~ Complete Link

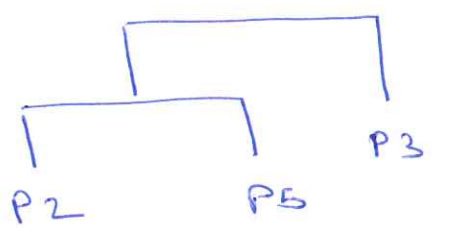
	P1	P2	P3	P4	P5
P1	1.0	0.10	0.40	0.55	0.35
P2	0.10	1.0	0.64	0.47	0.98
P3	0.41	0.64	1.0	0.44	0.85
P4	0.55	0.47	0.44	1.0	0.76
P5	0.35	0.98	0.85	0.76	1.0

Max = 0.98.



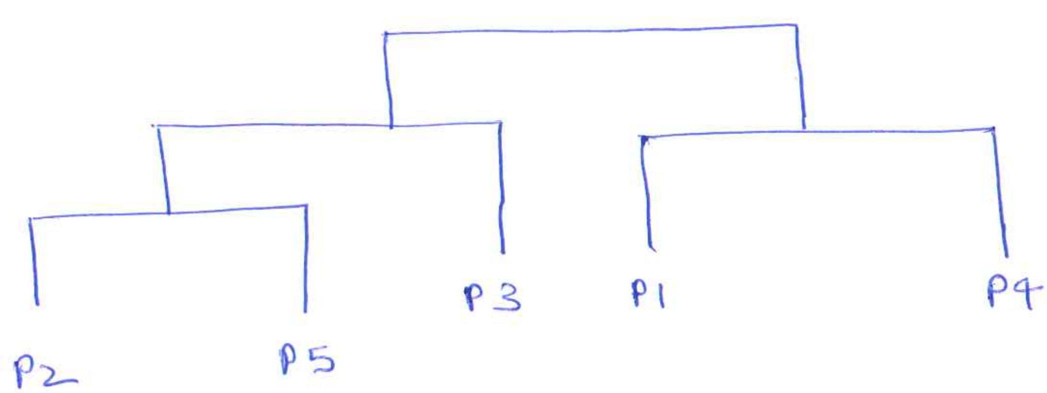
	P1	{P2, P5}	P3	P4
P1	1.0	0.10	0.40	0.55
{P2, P5}	0.10	1.0	0.64	0.47
P3	0.41	0.64	1.0	0.44
P4	0.55	0.47	0.44	1.0

Max = 0.64.



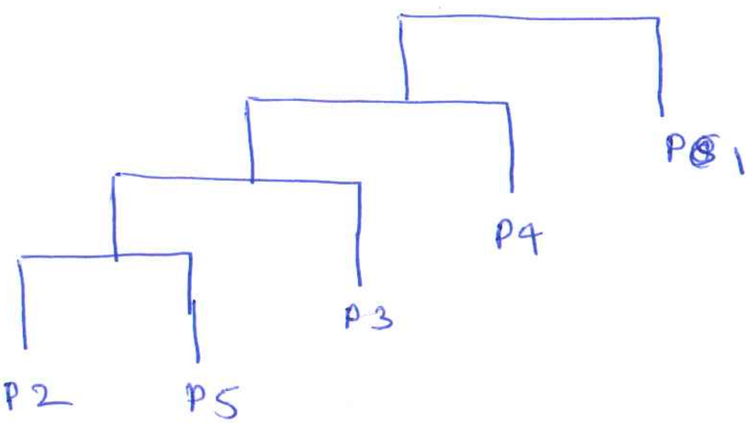
	P1	{P2, P3, P5}	P4
P1	1.0	0.10	0.55
{P2, P3, P5}	0.10	1.0	0.44
P4	0.55	0.44	1.0

Max = 0.55



(b) In the same way, for the ~~complete~~ Single link.

the dendogram is



17). $\{6, 12, 18, 24, 30, 42, 48\}$.

(a)

(i) Centroid $\{18, 45\}$. Two clusters are

c1 $\{6, 12, 18, 24, 30\}$, and

c2 $\{42, 48\}$

Squared Error for C1 $= (18-6)^2 + (18-12)^2 + (18-18)^2 + (18-24)^2 + (18-30)^2$

$= 144 + 36 + 0 + 36 + 144 = 360$

Squared Error for C2 $= (45-42)^2 + (45-48)^2$

$= 9 + 9 = 18$

Total Squared Error $= 360 + 18 = 378$

(ii) {15, 40}

Two clusters are c1 {6, 12, 18, 24} and c2 {30, 42, 48}

Squared Error for c1 = $(15-6)^2 + (15-12)^2 + (15-18)^2 + (15-24)^2$

$= 81 + 9 + 9 + 81 = 180$

Squared Error for c2 = $(30-40)^2 + (42-40)^2 + (48-40)^2$

$= 100 + 4 + 64$

$= 168$

Total Squared Error = 348

(b) Yes, both the sets of centroids represent stable solutions.

(c) The two clusters produced by single link are

{6, 12, 18, 24, 30} and {42, 48}

(d) K-means seems to produce the most natural clustering in this situation.

(e) It produces contiguous clusters, However density is also acceptable and center based is also acceptable because one set of centers gives the desired clusters.

(f) K-means is not good at finding clusters of different sizes, atleast in cases when there is no proper separation. The reasons for this is the objective of minimizing squared error causes it to break the larger cluster. Thus for the given problem, the low error clustering solution is unnatural one.

23) Similarity Matrix

	P1	P2	P3	P4
P1	1	0.8	0.65	0.55
P2	0.8	1	0.7	0.6
P3	0.65	0.7	1	0.9
P4	0.55	0.6	0.9	1

The two clusters are {P1, P2}, {P3, P4}

The dissimilarity matrix = 1 − Similarity Matrix

	P1	P2	P3	P4
P1	0.0	0.2	0.35	0.45
P2	0.20	0.0	0.30	0.40
P3	0.35	0.30	0.0	0.10
P4	0.45	0.40	0.10	0.0

Let x be the average distance of a point to other points in its cluster.

Let y indicate the minimum of average distance of a point to points in another cluster.

For point P1,

Silhoutte Coefficient $= 1 - \frac{x}{y}$

$$= 1 - \frac{0.2}{\frac{0.35+0.45}{2}} = 1 - \frac{0.2}{0.4} = 0.5$$

For point P2,

Silhouette Coefficient $= 1 - x/y$

$$= 1 - \frac{0.2}{\frac{0.3+0.4}{2}} = 1 - \frac{0.\cancel{2}4}{0.\cancel{35}7} = \frac{3}{7} = 0.42$$

For point P3,

Silhouette Coefficient $= 1 - x/y$

$$= 1 - \frac{0.1}{\frac{0.35+0.30}{2}} = 1 - \frac{0.\cancel{2}04}{0.\cancel{65}13} = \frac{9}{13} = 0.6923$$

For point P4,

Silhouette Coefficient $= 1 - x/y$

$$= \frac{1 - 0.1}{\frac{0.45 + 0.40}{2}} = 1 - \frac{0.\cancel{2}04}{0.\cancel{85}\,17} = \frac{13}{17}$$

$$= 0.7647.$$

For cluster C1, Silhouette Coefficient $= \frac{0.50 + 0.42}{2}$

$$= \frac{0.92}{2} = 0.46.$$

For cluster C2, Silhouette Coefficient $= \frac{9}{13} + \frac{13}{17} = 0.7285$

For overall cluster, Silhouette Coefficient $= \frac{0.46 + 0.7285}{2}$

$$= 0.5942$$