Usama Sarwar
Fall-BSCS-090-B

Question No : 1
Step # 01

	x	y
z_1	0.4	0.53
z_2	0.22	0.38
z_3	0.35	0.32
z_4	0.26	0.19
z_5	0.08	0.41
z_6	0.45	0.31

Step 2:
find the distance matrix by using:
Ecliclean distance

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

Distance $(z_1, z_2) = 0.33$
Distance $(z_1, z_3) = 0.26$
Distance $(z_1, z_4) = 0.35$
Distance $(z_1, z_5) = 0.44$

Similarly we will calculate for all data points and distance
matrix is given by:

	z_1	z_2	z_3	z_4	z_5	z_6
z_1	0	0.33	0.26	0.35	0.44	0.28
z_2	0.33	0	0.19	0.23	0.01	0.31
z_3	0.26	0.19	0	0.22	0.36	0.12
z_4	0.35	0.23	0.22	0	0.4	0.3
z_5	0.44	0.01	0.36	0.4	0	0.48
z_6	0.28	0.31	0.12	0.3	0.48	0

→ The minimum distance between the point $Z_2 + Z_5$ and they both from the first cluster. The dendrogram is drawn to represent the first cluster.

$$\downarrow \quad \downarrow$$
$$Z_5 \quad Z_2$$

Step # 3:

Update the distance matrix between $(Z_2, Z_5) Z_1$
using single link $\min((Z_2, Z_1), (Z_5, Z_1))$
$$\min(0.33, 0.44)$$
$$0.33$$

Distance matrix between (Z_2, Z_5) and Z_1 using complex Link:
$$\text{Max}((Z_2, Z_1), (Z_5, Z_1))$$
$$0.44$$

distance $(Z_2, Z_5) Z_1$) using average linkage is given by:
$$\frac{1}{2}[0.33 + 0.44] = 0.55$$

Update Matrix is given by:

	Z_1	Z_2, Z_5	Z_3	Z_4	Z_6
Z_1	0	0.33	0.26	0.35	0.28
Z_2, Z_5	0.33	0	0.19	0.23	0.31
Z_3	0.26	0.19	0	0.22	0.12
Z_4	0.35	0.23	0.22	0	0.3
Z_6	0.28	0.31	0.12	0.3	0

distance between $(Z_2, Z_5) + (Z_3) = 0.19$
$$\text{dist}((Z_2, Z_5) + (Z_4)) = 0.23$$
$$\text{dist}((Z_2, Z_5) + (Z_6)) = 0.31$$
$$= \frac{1}{2}[0.33 + 0.44]$$
$$= 0.55$$

Update matrix is given by:

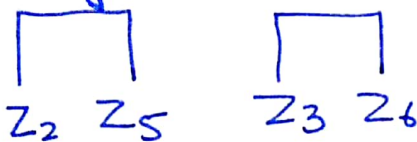|	z_1	z_2, z_5	z_3	z_4	z_6
z_1	0	0.33	0.26	0.35	0.28
z_2, z_5	0.33	0	0.19	0.23	0.31
z_3	0.26	0.19	0	0.22	0.12
z_4	0.35	0.23	0.22	0	0.3
z_6	0.28	0.31	0.12	0.3	0

dist between $(z_2, z_5) + z_3) = 0.19$

dist $(z_2, z_5) + z_4)) = 0.23$

dist $(z_2, z_5) + (z_6) = 0.31$

The minimum distance is b/w $z_3 + z_6$ and they form the 2nd cluster.
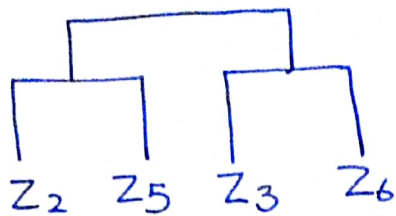
dendogram:

z_2 z_5 z_3 z_6

Step 4:
Again update distance matrix using the newly formed cluster

	z_1	z_2, z_5	z_3, z_6	z_4
z_1	0	0.33	0.26	0.35
z_2, z_5	0.33	0	0.19	0.23
z_3, z_6	0.26	0.19	0	0.22
z_4	0.35	0.23	0.22	0

Dendogram now look:

Z_2 Z_5 Z_3 Z_6

Step#05:

Update Distance Matrix:

	Z_1	Z_2,Z_5	Z_3,Z_6	Z_4
Z_1	0	0.33	0.26	0.35
Z_2,Z_5	0.33	0	0.19	0.23
Z_3,Z_6	0.26	0.19	0	0.22
Z_4	0.35	0.23	0.22	0

Dendogram new look:

	Z_1	Z_2,Z_5,Z_3,Z_6	Z_4
Z_1	0	0.26	0.35
Z_2,Z_5,Z_3,Z_6	0.26	0	0.22
Z_4	0.35	0.22	0

Dist $(Z_1, Z_3, Z_5, Z_4, Z_2)$

Min $(0.33, 0.26)$

Dist (Z_4) $Z_2, Z_5, Z_3, Z_4)$

Min $= 0.23, 0.22$

Min $= 0.22$

Z_2 Z_5 Z_3 Z_6 Z_4
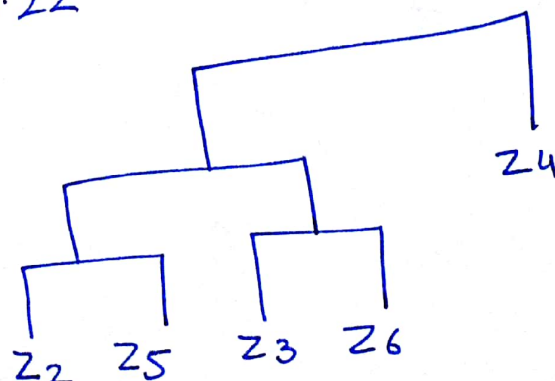
Step 6:
Update again:

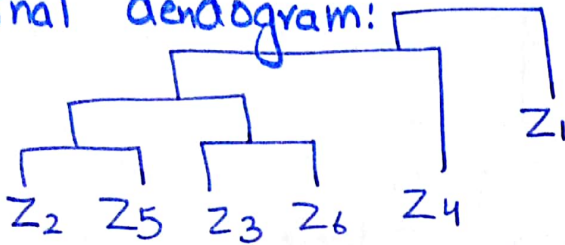|	Z_1	$Z_2 Z_5 Z_6 Z_3 Z_4$
Z_1	0	$\boxed{0.33}$
Z_2, Z_5, Z_6, Z_3, Z_4	$\boxed{0.33}$	0

distance b/w $(Z_1 + (Z_2 Z_5 Z_4 Z_3 Z_6))$

$\min(0.33, 0.35)$

0.33

Final dendogram:

$$Z_2 \quad Z_5 \quad Z_3 \quad Z_6 \quad Z_4 \qquad Z_1$$

Question #2:

→ DBSCAN clustering
   Given data set

	x	y
A_1	2	10
A_2	2	5
A_3	8	4
A_4	5	8
A_5	7	5
A_6	6	4
A_7	1	2
A_8	4	9

$\varepsilon = 2$, midpoint $= 2$

	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8
A_1	0	5	8.49	3.61	7.07	7.21	8.06	2.24
A_2	5	0	6.08	4.24	5	4.12	3.61	4.47
A_3	8.49	6.08	0	5	1.41	2	7.28	6-4
A_4	3.61	4.24	5	0	3.61	4.12	7.21	1.41
A_5	7.07	5	1.41	3.61	0	1.41	6.71	5
A_6	7.21	4.12	2	4.12	1.41	0	5.39	5.39
A_7	8.06	3.16	7.28	7.21	6.71	5.39	0	7.62
A_8	2.24	4.47	6.4	1.41	5	5.39	7.62	0

Number of element $\varepsilon = 2$ distance from
$A_1 = 1 (A_1$ itself$)$ which less than midpoint so
A_1 is an outlier.

A_2 and A_7, A_1 is outlier

A_4 and A_8 is core point

A_3, A_5, A_6 is also an core point

We have two cluster

$C_1 = \{A_4, A_8\}$

$C_2 = \{A_3, A_5, A_6\}$

Core point