

Data Mining (CS4038)

Quiz No.5

Roll No: _____

Section: _____

Date: 06-05-2024

Question No.1

(10 marks)

Consider the following data points and apply Complete Linkage clustering. Draw dendrogram and cut the dendrogram at a suitable height to find appropriate no. of clusters.

A(2,8), B(9,15), C(12,11), D(19,3), E(15,11), F(8,4), G(5,10), H(1,6)

	A	B	C	D	E	F	G	H
A	0							
B	14	0						
C	13	7	0					
D	22	22	15	0				
E	16	10	③	10	0			
F	8	12	11	12	14	0		
G	11	9	8	21	11	9	0	
H	5	17	16	21	19	9	8	0

$$\text{dist}((C,E),A) = \frac{13+16}{2} = 14.5$$

$$\text{dist}((C,E),B) = \frac{7+10}{2} = 8.5$$

$$\text{dist}((C,E),D) = \frac{15+10}{2} = 12.5$$

$$\text{dist}((C,E),F) = \frac{11+14}{2} = 12.5$$

$$\text{dist}((C,E),G) = \frac{8+11}{2} = 9.5$$

$$\text{dist}((C,E),H) = \frac{16+19}{2} = 17.5$$

	A	B	C,E	D	F	G	H
A	0						
B	14	0					
C,E	14.5	8.5	0	12.5	12.5	9.5	17.5
D	22	22	12.5	0			
F	8	12	12.5	12	0		
G	11	9	9.5	21	9	0	
H	5	17	17.5	21	9	8	0

$$\text{dist}((A,H),B) = \frac{14+17}{2} = 15.5$$

$$\text{dist}((A,H),C,E) = \frac{14.5+17.5}{2} = 16$$

$$\text{dist}((A,H),D) = \frac{22+21}{2} = 21.5$$

$$\text{dist}((A,H),F) = \frac{8+9}{2} = 8.5$$

$$\text{dist}((A,H),G) = \frac{11+8}{2} = 9.5$$

	A,H	B	C,E	D	F	G	H
A,H	0						
B	15.5	0	16	21.5	8.5	9.5	
C,E	16	8.5	0				
D	21.5	22	13.5	0			
F	8.5	12	12.5	12	0		
G	9.5	9	9.5	21	9	0	
H							0

(A,H)F	B	(C,E)	D	G
B	21.5	0		
(C,E)	14.2	(8.5)	0	
D	16.75	22	13.5	0
G	9.25	9	9.5	21
				0

$$\text{dist}((A,H)F, B) =$$

$$\text{dist}((A,H)F, (C,E)) = \frac{16 + 12.5}{2} = 14.2$$

$$\text{dist}((A,H)F, D) = \frac{21.5 + 12}{2} = \frac{33.5}{2} = 16.75$$

$$\text{dist}((A,H)F, G) = \frac{9.5 + 9}{2} = 9.25$$

(A,H)F	(C,E)B	D	G
(A,H)F	0		
(C,E)B	17.85	0	
D	16.75	17.75	0
G	(9.25)	9.25	9.25 21

$$\text{dist}((C,E)B, (A,H)F) = \frac{14.2 + 21.5}{2} = 17.85$$

$$\text{dist}((C,E)B, D) = \frac{13.5 + 22}{2} = 17.75$$

$$\text{dist}((C,E)B, G) = \frac{9.5 + 9}{2} = 9.25$$

(A,H)G	CEB	D
(A,H)G	0	
CEB	(13.5)	0
D	18.88	17.75

$$\text{dist}((A,H)G, CEB) = \frac{17.8 + 9.25}{2} = 13.5$$

$$\text{dist}((A,H)G, D) = \frac{16.75 + 21}{2} = \frac{18.88}{2}$$

(AHFG)(CEB)	D
(AHFG)(CEB)	0
D	18.3

$$\text{dist}((AHFG)(CEB), D) = \frac{18.88 + 17.75}{2} = 18.3$$

Data Mining (CS4038)

Quiz No.5

Roll No: _____

Section: _____

Date: 06-04-2024

Question No.1

(10 marks)

Consider the following data points and apply Group Average clustering. Draw dendrogram and cut the dendrogram at a suitable height to find appropriate no. of clusters.

A(12,6), B(4,7), C(10,11), D(9,3), E(15,9), F(7,2), G(15,10), H(13,16)

using Manhattan

	A	B	C	D	E	F	G	H
A	0	9	7	6	6	9	7	11
B	9	0	10	9	13	8	14	18
C	7	10	0	4	7	12	6	8
D	6	9	4	0	12	3	13	17
E	6	13	7	12	0	15	1	9
F	9	8	12	3	15	0	16	20
G	7	14	6	13	1	16	0	8
H	11	18	8	17	9	20	8	0

	A	B	C	D	(E,G)	F	H
A	0						
B	9	0					
C	7	10	0				
D	6	9	4	0			
(E,G)	6.5	13.5	6.5	12.5	0	15.5	8.5
F	9	8	12	3	15.5	0	
H	11	18	8	17	8.5	20	0

$$\text{dist}((E,G), A) = \text{Avg}(\text{dist}(E,A), \text{dist}(G,A)) \\ = (6, 7) / 2 = 6.5$$

$$\text{dist}((E,G), B) = \text{Avg}(\text{dist}(E,B), \text{dist}(G,B)) \\ = (13, 14) / 2 = 13.5$$

$$(E,G), C = 7 + 6 / 2 = 6.5$$

$$(E,G), D = 12 + 13 / 2 = 12.5$$

$$(E,G), F = 15 + 16 / 2 = 15.5$$

$$(E,G), H = 9 + 8 / 2 = 8.5$$

	A	B	C	(D,F)	(E,G)	H
A	0					
B	9	0				
C	7	10	0			
(D,F)	7.5	8.5	8	0	14	18.5
(E,G)	6.5	13.5	6.5	14	0	
H	11	18	8	18.5	8.5	0

$$\text{dist}((D,F), A) = (6 + 9) / 2 = 7.5$$

$$\text{dist}((D,F), B) = (9 + 8) / 2 = 8.5$$

$$\text{dist}((D,F), (E,G)) = (12.5 + 15.5) / 2 = 14$$

$$\text{dist}((D,F), H) = (17 + 20) / 2 = 18.5$$

$$\text{dist}((D,F), C) = 4 + 12 / 2 = 8$$

Data Mining (CS4038)

Quiz No.5

Roll No: _____

Section: _____

Date: 06-05-2024

Question No.1

(10 marks)

Consider the following data points and apply Group Average clustering. Draw dendrogram and cut the dendrogram at a suitable height to find appropriate no. of clusters.

A(12,6), B(4,7), C(10,11), D(9,3), E(15,9), F(7,2), G(15,10), H(13,16)

	(E,G) A	B	C	(D,F)	H
(E,G) A	0	11.25	6.75	10.25	9.75
B	11.25	0			
C	6.75	10	0		
(D,F)	10.25	8.5	8	0	
H	9.75	18	8	18.5	0

$$\begin{aligned} ((E,G)A, B) &= (13.5 + 9)/2 = 11.25 \\ ((E,G)A, C) &= (6.5 + 7)/2 = 6.75 \\ ((E,G)A, (D,F)) &= (14 + 6.5)/2 = 10.25 \\ ((E,G)A, H) &= (8.5 + 11)/2 = 9.75 \end{aligned}$$

	(E,G) A	B	(D,F)	H
(E,G) A	0			
B	10.625	0		
(D,F)	9.125	8.5	0	
H	8.875	18	18.5	0

$$\begin{aligned} ((E,G)A) B &= (11.25 + 10)/2 = 10.625 \\ ((E,G)A) (D,F) &= (10.25 + 8)/2 = 9.125 \\ ((E,G)A) H &= (9.75 + 8)/2 = 8.875 \end{aligned}$$

	(E,G) A	(D,F) B	H
(E,G) A	0		
(D,F) B	9.875	0	
H	8.875	18.25	0

$$\begin{aligned} ((D,F)B) H &= (18.5 + 18)/2 = 18.25 \\ ((E,G)A) ((D,F)B) &= (9.125 + 10.625)/2 = 9.875 \end{aligned}$$

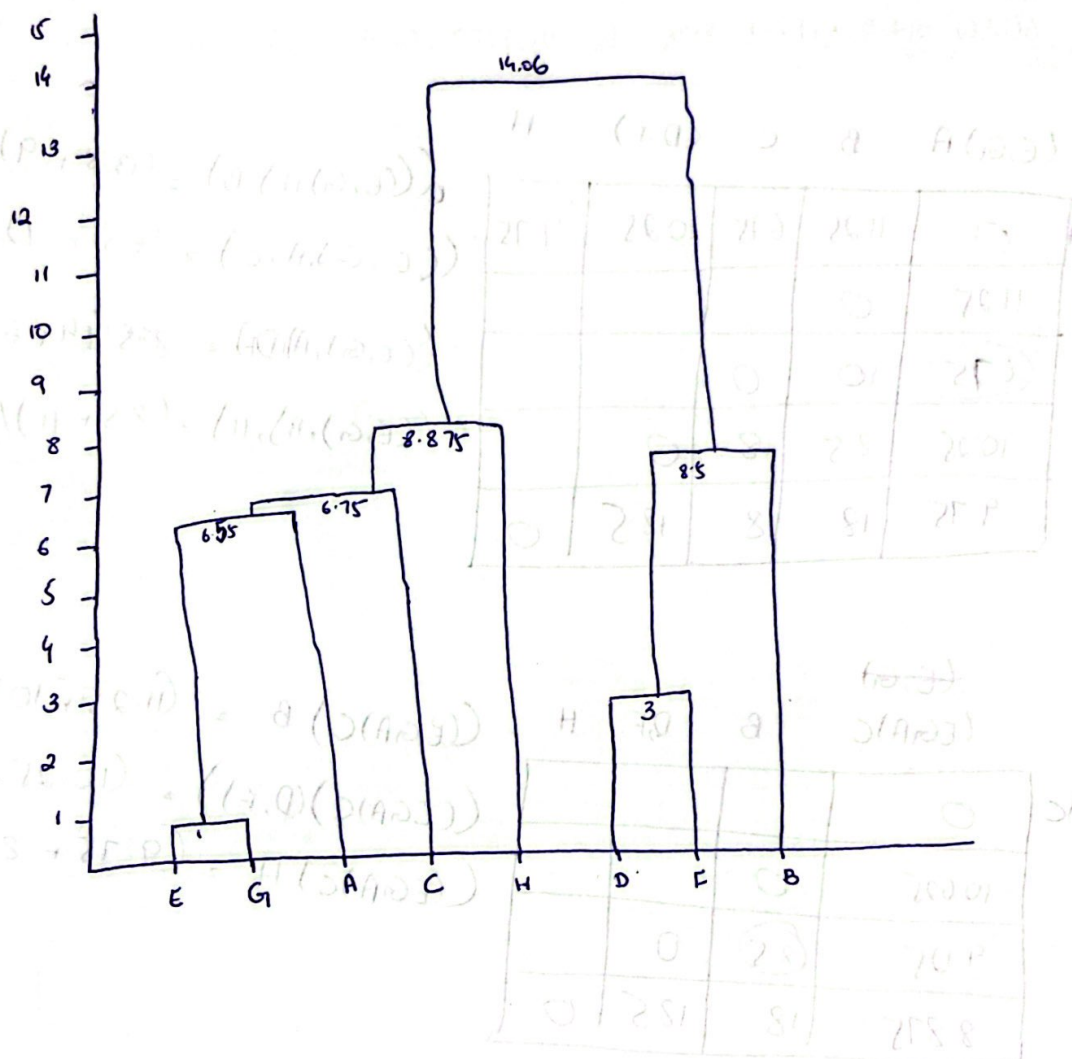
(EGAC) H (DF) B

(EGAC) H

0	14.06
14.06	0

(DF) B

$$((EGAC)H)(DF)B = (9.875 + 18.25) / 2 = 14.06$$



Data Mining (CS4038)

Quiz No.5

Roll No: _____

Section: _____

Date: 06-05-2024

Question No.1

(10 marks)

Consider the following data points and apply Group Average clustering. Draw dendrogram and cut the dendrogram at a suitable height to find appropriate no. of clusters.

A(12,6), B(4,7), C(10,11), D(9,3), E(15,9), F(7,2), G(15,10), H(13,16)

using Euclidean.

	A	B	C	D	E	F	G	H
A	0	8.06	5.38	4.24	4.24	6.4	5	10.04
B	8.06	0	7.2	6.4	11.1	5.8	11.4	12.7
C	5.38	7.2	0	8.06	5.39	9.4	5.09	5.83
D	4.24	6.4	8.06	0	8.4	2.2	9.2	13.60
E	4.24	11.1	5.39	8.4	0	10.6	7	7
F	6.4	5.8	9.4	2.2	10.6	0	11.3	6.32
G	5	11.4	5.09	9.2	1	11.3	0	6.32
H	10.04	12.7	5.83	13.60	7	6.32	6.32	0

	A	B	C	D	(E, G)	F	H
A	0	8.06	5.38	4.24	4.62	6.4	10.04
B		0	7.2	6.4	11.25	5.8	12.7
C			0	8.06	5.24	9.4	5.83
D				0	8.8	2.2	13.60
(E, G)	4.62	11.25	5.24	8.8	0	10.95	6.66
F						0	6.32
H							0

$$\text{dist}((E, G), A) = \text{Avg}(E, A), (G, A) \\ = (4.24 + 5)/2 = 4.62$$

$$\text{dist}((E, G), B) = (11.1 + 11.4)/2 = 11.25$$

$$\text{dist}((E, G), C) = (5.39 + 5.09)/2 = 5.24$$

$$\text{dist}((E, G), D) = (8.4 + 9.2)/2 = 8.8$$

$$\text{dist}((E, G), F) = (10.6 + 1)/2 = 10.95$$

$$\text{dist}((E, G), H) = (7 + 6.32)/2 = 6.66$$

$$\text{dist}((D, F), A) = (4.24 + 6.4)/2 = 5.32$$

$$\text{dist}((D, F), B) = (6.4 + 5.8)/2 = 6.1$$

$$\text{dist}((D, F), C) = (8.06 + 9.4)/2 = 8.73$$

$$\text{dist}((D, F), (E, G)) = (8.8 + 10.95)/2 = 9.87$$

$$\text{dist}((D, F), H) = (13.60 + 6.32)/2 = 9.96$$

	A	B	C	(D,F)	(E,G)	H
A	0	8.06	5.38	34.12	<u>4.62</u>	10.04
B		0	7.2	6.1	11.25	12.7
C			0	8.73	5.24	5.83
D,F	34.12	6.1	8.73	0	9.87	9.96
E,G					0	6.66
H						0

(AEGC)

	A,(EG)	B	C	(D,F)	H
A,(EG)	0	89.65	<u>5.31</u>	21.995	8.35
B		0	7.2	6.1	12.7
C			0	8.73	5.83
D,F				0	9.96
H					0

$$(A,(EG),B) = (8.06 + 11.25)/2 = 9.655$$

$$(A,(EG),C) = (5.38 + 5.24)/2 = 5.31$$

$$(A,(EG),D,F) = (34.12 + 9.87)/2 = 21.995$$

$$(A,(EG),H) = (10.04 + 6.66)/2 = 8.35$$

(AEG)C B (D,F) H

$$(AEG),C,B = (9.655 + 7.2)/2 = 8.42$$

(AEG)C	0	842	15.36	7.09
B		0	<u>6.1</u>	12.7
D,F			0	9.96
H				0

$$((AEG),C),D,F = (21.995 + 8.73)/2 = 15.36$$

$$((AEG),C),H = (8.35 + 5.83)/2 = 7.09$$

(AEG)C B(D,F) H

(AEG)C	0	11.89	<u>7.09</u>
B(D,F)		0	11.33
H			0

$$((B,D,F),AEG),C = (842 + 15.36)/2 = 11.89$$

$$(B,D,F),H = (12.7 + 9.96)/2 = 11.33$$

(AEGC)H

BDF

$$((AEGC).H) BDF = (11.89 + 11.33)/2 = 11.61$$

(AEGC)H

0	11.61
11.61	0

BDF

