Khantil & Aman
HW4

Problem 2 Hierarchical Clustering

Euclidean Distance Matrix:

	A	B	C	D	E	F	G	H	I	J
A	0.00	7.00	4.12	5.39	3.16	3.61	6.40	6.08	8.06	10.63
B		0.00	3.16	2.83	8.54	5.83	5.83	8.49	11.31	8.00
C			0.00	1.41	5.39	2.83	4.00	5.83	8.60	7.62
D				0.00	6.08	3.16	3.16	5.66	8.49	6.32
E					0.00	3.00	5.39	3.61	5.00	9.43
F						0.00	2.83	3.16	5.83	7.07
G							0.00	3.16	5.83	4.24
H								0.00	2.83	6.32
I									0.00	8.00
J										0.00

(a)

Single Link Hierarchical Clustering:

(Similarity of two clusters is based on the two most similar closest points in different clusters)

Step 1: {C} will merge with {D} at 1.41

	A	B	C,D	E	F	G	H	I	J
A	0.00	7.00	4.12	3.16	3.61	6.40	6.08	8.06	10.63
B		0.00	2.83	8.54	5.83	5.83	8.49	11.31	8.00
C,D			**0.00**	**5.39**	**2.83**	**3.16**	**5.66**	**8.49**	**6.32**
E				0.00	3.00	5.39	3.61	5.00	9.43
F					0.00	2.83	3.16	5.83	7.07
G						0.00	3.16	5.83	4.24
H							0.00	2.83	6.32
I								0.00	8.00
J									0.00

Step 2: {C,D} will merge with {B} at 2.83

Here 2.83 distance is same for {C,D} and {B} & {H} and {I}, so I have chosen the first one.

	A	C,D,B	E	F	G	H	I	J
A	0.00	4.12	3.16	3.61	6.40	6.08	8.06	10.63
C,D,B		**0.00**	**5.39**	**2.83**	**3.16**	**5.66**	**8.49**	**6.32**
E			0.00	3.00	5.39	3.61	5.00	9.43
F				0.00	2.83	3.16	5.83	7.07
G					0.00	3.16	5.83	4.24
H						0.00	2.83	6.32
I							0.00	8.00
J								0.00

Step 3: {C,D,B} will merge with {F} at 2.83

	A	C,D,B,F	E	G	H	I	J
A	0.00	3.61	3.16	6.40	6.08	8.06	10.63
C,D,B,F		**0.00**	**3.00**	**2.83**	**3.16**	**5.83**	**6.32**
E			0.00	5.39	3.61	5.00	9.43
G				0.00	3.16	5.83	4.24
H					0.00	2.83	6.32
I						0.00	8.00
J							0.00

Step 4: {C,D,B,F} will merge with {G} 2.83

	A	C,D,B,F,G	E	H	I	J
A	0.00	3.61	3.16	6.08	8.06	10.63
C,D,B,F,G		**0.00**	**3.00**	**3.16**	**5.83**	**4.24**
E			0.00	3.61	5.00	9.43
H				0.00	2.83	6.32
I					0.00	8.00
J						0.00

Step 5: {H} will merge with {I} 2.83

	A	C,D,B,F,G	E	**H,I**	J
A	0.00	3.61	3.16	**6.08**	10.63
C,D,B,F,G		0.00	3.00	**3.16**	4.24
E			0.00	**3.61**	9.43
H,I				**0.00**	**6.32**
J					0.00

Step 5: {C,D,B,F,G} will merge with {E} 3.00

	A	C,D,B,F,G,E	**H,I**	J
A	0.00	3.16	**6.08**	10.63
C,D,B,F,G,E		0.00	**3.16**	4.24
H,I			**0.00**	**6.32**
J				0.00

Step 6: {C,D,B,F,G,E} will merge with {H,I} at height 3.16

	A	C,D,B,F,G,E,H,I	J
A	0.00	3.16	10.63
C,D,B,F,G,E,H,I		**0.00**	**4.24**
J			0.00

Step 7: {C,D,B,F,G,E,H,I} will merge with {A} at height 3.16

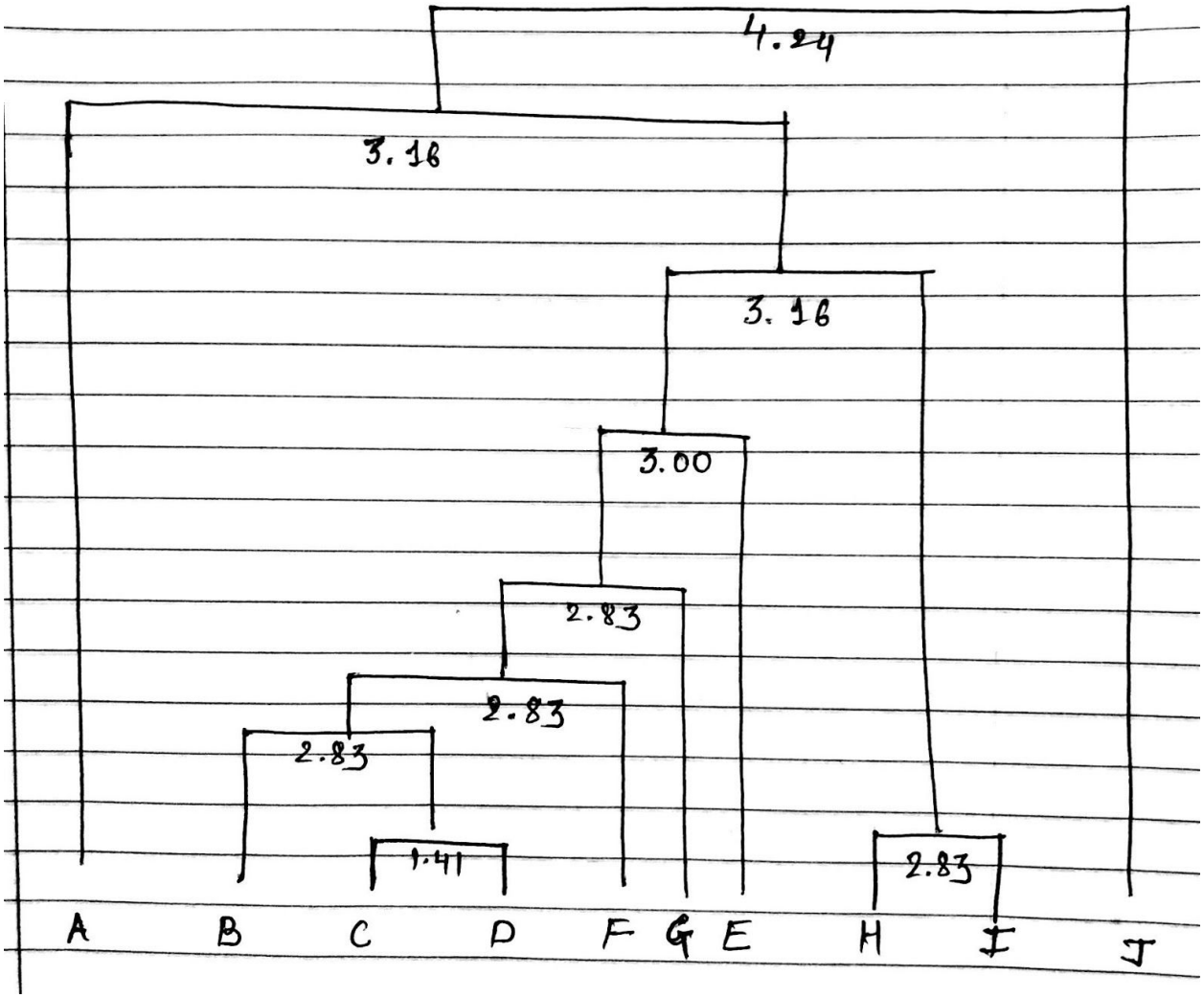
	C,D,B,F,G,E,H,I,A	J
C,D,B,F,G,E,H,I,A	**0.00**	**4.24**
J		0.00

Step 8: {C,D,B,F,G,H,I,A} will merge with {J} at height 4.24

Dendrogram:

Khantil, Aman

Single Link



Complete Link Hierarchical Clustering:
(Similarity of two clusters is based on the two **least** similar (most distant) points in different clusters)

Step 1: {C} will merge with {D} at 1.41

	A	B	**C,D**	E	F	G	H	I	J
A	0.00	7.00	**5.39**	3.16	3.61	6.40	6.08	8.06	10.63
B		0.00	**3.16**	8.54	5.83	5.83	8.49	11.31	8.00
C,D			**0.00**	**6.08**	**3.16**	**4.00**	**5.83**	**8.60**	**7.62**
E				0.00	3.00	5.39	3.61	5.00	9.43
F					0.00	2.83	3.16	5.83	7.07
G						0.00	3.16	5.83	4.24
H							0.00	2.83	6.32
I								0.00	8.00
J									0.00

Step 2: {G} will merge with {F} at 2.83

	A	B	C,D	E	**G,F**	H	I	J
A	0.00	7.00	5.39	3.16	6.40	6.08	8.06	10.63
B		0.00	3.16	8.54	**5.83**	8.49	11.31	8.00
C,D			0.00	6.08	4.00	5.83	8.60	7.62
E				0.00	5.39	3.61	5.00	9.43
G,F					**0.00**	**3.16**	**5.83**	**7.07**
H						0.00	2.83	6.32
I							0.00	8.00
J								0.00

Step 4: {H} will merge with {I} at 2.83

	A	B	C,D	E	G,F	**H,I**	J
A	0.00	7.00	5.39	3.16	6.40	**8.06**	10.63
B		0.00	3.16	8.54	5.83	**11.31**	8.00
C,D			0.00	6.08	4.00	**8.60**	7.62
E				0.00	5.39	**5.00**	9.43
G,F					0.00	**5.83**	7.07
H,I						**0.00**	**8.00**
J							0.00

Step 5: {A} will merge with {E} at 3.16

	A,E	B	C,D	G,F	H,I	J
A,E	**0.00**	**8.54**	**5.39**	**6.40**	**8.06**	**10.63**
B		0.00	3.16	5.83	11.31	8.00
C,D			0.00	4.00	8.60	7.62
G,F				0.00	5.83	7.07
H,I					0.00	8.00
J						0.00

Step 6: {B} will merge with {C,D} at 3.16

	A,E	**B,C,D**	G,F	H,I	J
A,E	0.00	**8.54**	6.40	8.06	10.63
B,C,D		**0.00**	**5.83**	**11.31**	**8.00**
G,F			0.00	5.83	7.07
H,I				0.00	8.00
J					0.00

Step 7: {B,C,D} will merge with {G,F} at 5.83

	A,E	**B,C,D,F,G**	H,I	J
A,E	0.00	**8.54**	8.06	10.63
B,C,D,F,G		**0.00**	**11.31**	**8.00**
H,I			0.00	8.00
J				0.00

Step 7: {B,C,D,F,G} will merge with {J} at 8.00

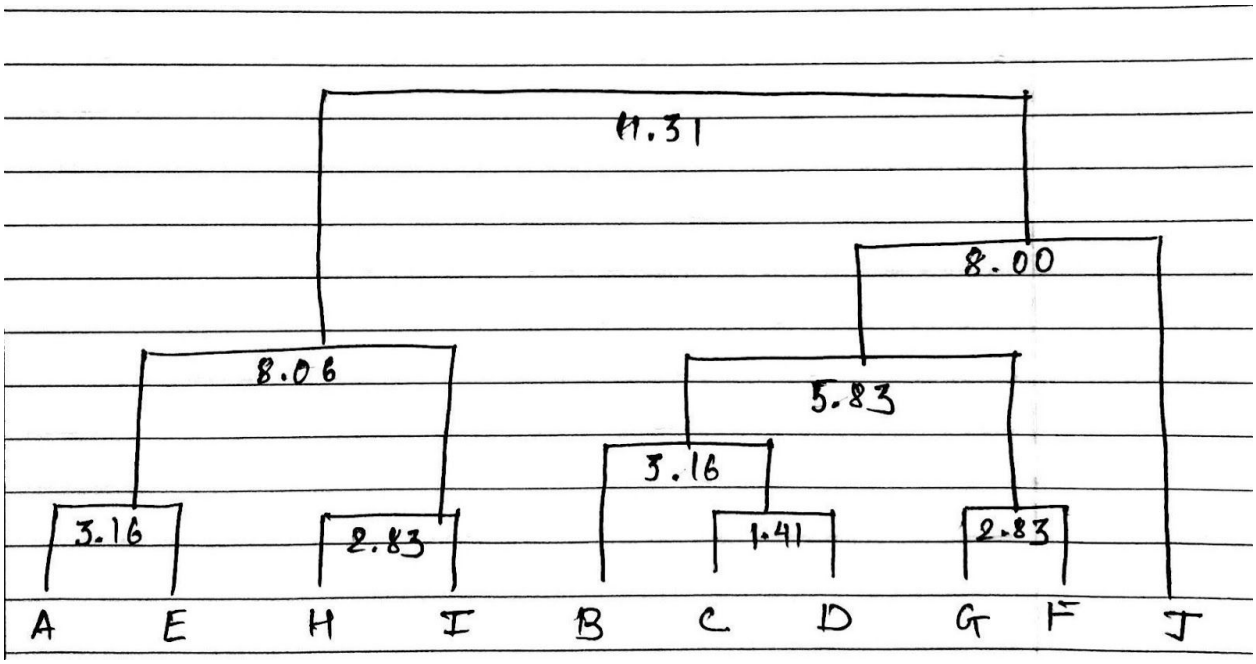
	A,E	**B,C,D,F,G,J**	H,I
A,E	0.00	10.63	8.06
B,C,D,F,G,J		**0.00**	**11.31**
H,I			0.00

Step 8: {A,E} will merge with {H,I} at 8.06

	A,E,H,I	**B,C,D,F,G,J**
A,E,H,I	0.00	**11.31**
B,C,D,F,G,J		**0.00**

Step 9: {A,E,H,I} will merge with {B,C,D,F,G,J} at 11.31

Dendrogram:



(b) If we assume there are three clusters, which of the single and complete link hierarchical clustering will give better resulted clusters?

Single Link Hierarchical:
3 Clusters: {B,C,D,E,G,H,I} , {J} and {A}
Single Link SSE: 107.875 ((Calculated in Q2.pynb)

Complete Link Hierarchical:
3 Clusters: {B,C,D,G,F} , {J} and {A,E,H,I}
SSE: 72.0 (Calculated in Q2.pynb)

As we can visualize from dendrogram and cluster formation, Single link tends to provide elongated clusters whereas complete link short and circular clusters.
Conclusion: If there are three clusters, based on Sum of Squared Error calculated and dendrogram, Complete link gives better result.

(c) Compare your resulted clusters from 2(b) with the resulted clusters using K-means in Question 1 by calculating their corresponding Sum of Squared Error (SSE). Based on their SSE results, which resulted clusters, 1(b) or 2(b), are better?

From Q1, SSE for K-Means Clustering is : 60.83333333333333 (Calculated in Q1.pynb)

Comparing results with Q1(b) and Q2(b),
K-Means provide better results, as it provides least Sum of Squared Errors.