

COMPSCI 762 Tutorial 10

Solution for Week 10 Review Question 2 Part 2 – Complete Linkage Agglomerative Clustering

Luke Chang

May 2021

The University of Auckland

Review Question 2: Agglomerative Clustering

Use complete-linkage (MAX) agglomerative clustering to group the data described in Exercise 1. Show the dendrogram.

	A1	A2	A3	A4	A5	A6	A7	A8
A1	0	$\sqrt{45}$	$\sqrt{63}$	$\sqrt{57}$	$\sqrt{41}$	$\sqrt{28}$	$\sqrt{95}$	$\sqrt{6}$
A2		0	$\sqrt{55}$	$\sqrt{49}$	$\sqrt{35}$	$\sqrt{11}$	$\sqrt{5}$	$\sqrt{25}$
A3			0	$\sqrt{11}$	$\sqrt{23}$	$\sqrt{54}$	$\sqrt{47}$	$\sqrt{65}$
A4				0	$\sqrt{2}$	$\sqrt{7}$	$\sqrt{26}$	$\sqrt{5}$
A5					0	$\sqrt{5}$	$\sqrt{21}$	$\sqrt{35}$
A6						0	$\sqrt{13}$	$\sqrt{27}$
A7							0	$\sqrt{53}$
A8								0

Review Question 2: Agglomerative Clustering

Use complete-linkage (MAX) agglomerative clustering to group the data described in Exercise 1. Show the dendrogram.

	A1	A2	A3	A4	A5	A6	A7	A8
A1	0	$\sqrt{45}$	$\sqrt{63}$	$\sqrt{57}$	$\sqrt{41}$	$\sqrt{28}$	$\sqrt{95}$	$\sqrt{6}$
A2		0	$\sqrt{55}$	$\sqrt{49}$	$\sqrt{35}$	$\sqrt{11}$	$\sqrt{5}$	$\sqrt{25}$
A3			0	$\sqrt{11}$	$\sqrt{23}$	$\sqrt{54}$	$\sqrt{47}$	$\sqrt{65}$
A4				0	$\sqrt{2}$	$\sqrt{7}$	$\sqrt{26}$	$\sqrt{5}$
A5					0	$\sqrt{5}$	$\sqrt{21}$	$\sqrt{35}$
A6						0	$\sqrt{13}$	$\sqrt{27}$
A7							0	$\sqrt{53}$
A8								0

- Agglomerative clustering starts with considering each data point as a cluster.
- Complete-linkage uses the maximum distances between all observations of the two sets.

Level	# Clusters	Clusters
0	8	$\{A1\}, \{A2\}, \{A3\}, \{A4\}, \{A5\}, \{A6\}, \{A7\}, \{A8\}$

Review Question 2: Agglomerative Clustering

- Complete-linkage uses the maximum distances between all observations of the two sets.
- Since each point is a cluster, we simply merge the two cluster with minimum distance.

	A1	A2	A3	A4	A5	A6	A7	A8
A1	0	$\sqrt{45}$	$\sqrt{63}$	$\sqrt{57}$	$\sqrt{41}$	$\sqrt{28}$	$\sqrt{95}$	$\sqrt{6}$
A2		0	$\sqrt{55}$	$\sqrt{49}$	$\sqrt{35}$	$\sqrt{11}$	$\sqrt{5}$	$\sqrt{25}$
A3			0	$\sqrt{11}$	$\sqrt{23}$	$\sqrt{54}$	$\sqrt{47}$	$\sqrt{65}$
A4				0	$\sqrt{2}$	$\sqrt{7}$	$\sqrt{26}$	$\sqrt{5}$
A5					0	$\sqrt{5}$	$\sqrt{21}$	$\sqrt{35}$
A6						0	$\sqrt{13}$	$\sqrt{27}$
A7							0	$\sqrt{53}$
A8								0

Level	# Clusters	Clusters
0	8	$\{A1\}, \{A2\}, \{A3\}, \{A4\}, \{A5\}, \{A6\}, \{A7\}, \{A8\}$
1	7	$\{A1\}, \{A2\}, \{A3\}, \{A4, A5\}, \{A6\}, \{A7\}, \{A8\}$

Review Question 2: Agglomerative Clustering

Complete-linkage Proximity Matrix

	A1	A2	A3	A4	A5	A6	A7	A8
A1	0	$\sqrt{45}$	$\sqrt{63}$	$\sqrt{57}$	$\sqrt{41}$	$\sqrt{28}$	$\sqrt{95}$	$\sqrt{6}$
A2		0	$\sqrt{55}$	$\sqrt{49}$	$\sqrt{35}$	$\sqrt{11}$	$\sqrt{5}$	$\sqrt{25}$
A3			0	$\sqrt{11}$	$\sqrt{23}$	$\sqrt{54}$	$\sqrt{47}$	$\sqrt{65}$
A4				0	$\sqrt{2}$	$\sqrt{7}$	$\sqrt{26}$	$\sqrt{5}$
A5					0	$\sqrt{5}$	$\sqrt{21}$	$\sqrt{35}$
A6						0	$\sqrt{13}$	$\sqrt{27}$
A7							0	$\sqrt{53}$
A8								0

- A4 and A5 are in one cluster now, therefore we want to update our proximity matrix.
- We merge the columns and rows which contain A4, A5 by only keeping the maximum value (complete-linkage).
- The distance within the same cluster is always 0 (trivial).

Review Question 2: Agglomerative Clustering

Complete-linkage Proximity Matrix at Level 1

	A1	A2	A3	A4, A5	A6	A7	A8
A1	0	$\sqrt{45}$	$\sqrt{63}$	$\sqrt{57}$	$\sqrt{28}$	$\sqrt{95}$	$\sqrt{6}$
A2		0	$\sqrt{55}$	$\sqrt{49}$	$\sqrt{11}$	$\sqrt{5}$	$\sqrt{25}$
A3			0	$\sqrt{23}$	$\sqrt{54}$	$\sqrt{47}$	$\sqrt{65}$
A4, A5				0	$\sqrt{7}$	$\sqrt{26}$	$\sqrt{35}$
A6					0	$\sqrt{13}$	$\sqrt{27}$
A7						0	$\sqrt{53}$
A8							0

Level	# Clusters	Clusters
0	8	$\{A1\}, \{A2\}, \{A3\}, \{A4\}, \{A5\}, \{A6\}, \{A7\}, \{A8\}$
1	7	$\{A1\}, \{A2\}, \{A3\}, \{A4, A5\}, \{A6\}, \{A7\}, \{A8\}$

- $\sqrt{5}$ is the next minimum distance between clusters when using complete-linkage, so we merge A2 and A7.

Review Question 2: Agglomerative Clustering

Complete-linkage Proximity Matrix at Level 1 – Merge A2 and A7

	A1	A2	A3	A4, A5	A6	A7	A8
A1	0	$\sqrt{45}$	$\sqrt{63}$	$\sqrt{57}$	$\sqrt{28}$	$\sqrt{95}$	$\sqrt{6}$
A2	$\sqrt{45}$	0	$\sqrt{55}$	$\sqrt{49}$	$\sqrt{11}$	$\sqrt{5}$	$\sqrt{25}$
A3		$\sqrt{55}$	0	$\sqrt{23}$	$\sqrt{54}$	$\sqrt{47}$	$\sqrt{65}$
A4, A5		$\sqrt{49}$		0	$\sqrt{7}$	$\sqrt{26}$	$\sqrt{35}$
A6		$\sqrt{11}$			0	$\sqrt{13}$	$\sqrt{27}$
A7	$\sqrt{95}$	$\sqrt{5}$	$\sqrt{47}$	$\sqrt{26}$	$\sqrt{13}$	0	$\sqrt{53}$
A8		$\sqrt{25}$				$\sqrt{53}$	0

Level	# Clusters	Clusters
0	8	{A1}, {A2}, {A3}, {A4}, {A5}, {A6}, {A7}, {A8}
1	7	{A1}, {A2}, {A3}, {A4, A5}, {A6}, {A7}, {A8}
2	6	{A1}, {A2, A7}, {A3}, {A4, A5}, {A6}, {A8}

- Complete the rows and column which contain A2 and A7
- Merge A2 and A7 into one cluster to replace A2, and then remove A7

Review Question 2: Agglomerative Clustering

Complete-linkage Proximity Matrix at Level 2

	A1	A2, A7	A3	A4, A5	A6	A8
A1	0	$\sqrt{95}$	$\sqrt{63}$	$\sqrt{57}$	$\sqrt{28}$	$\sqrt{6}$
A2, A7		0	$\sqrt{55}$	$\sqrt{49}$	$\sqrt{13}$	$\sqrt{53}$
A3			0	$\sqrt{23}$	$\sqrt{54}$	$\sqrt{65}$
A4, A5				0	$\sqrt{7}$	$\sqrt{35}$
A6					0	$\sqrt{27}$
A8						0

- $\sqrt{6}$ is the next minimum distance between clusters when using complete-linkage, so we merge A1 and A8.

Level	# Clusters	Clusters
0	8	$\{A1\}, \{A2\}, \{A3\}, \{A4\}, \{A5\}, \{A6\}, \{A7\}, \{A8\}$
1	7	$\{A1\}, \{A2\}, \{A3\}, \{A4, A5\}, \{A6\}, \{A7\}, \{A8\}$
2	6	$\{A1\}, \{A2, A7\}, \{A3\}, \{A4, A5\}, \{A6\}, \{A8\}$
3	5	$\{A1, A8\}, \{A2, A7\}, \{A3\}, \{A4, A5\}, \{A6\}$

Review Question 2: Agglomerative Clustering

Complete-linkage Proximity Matrix at Level 2 – Merge A1 and A8

	A1	A2, A7	A3	A4, A5	A6	A8
A1	0	$\sqrt{95}$	$\sqrt{63}$	$\sqrt{57}$	$\sqrt{28}$	$\sqrt{6}$
A2, A7	$\sqrt{95}$	0	$\sqrt{55}$	$\sqrt{49}$	$\sqrt{13}$	$\sqrt{53}$
A3	$\sqrt{63}$		0	$\sqrt{23}$	$\sqrt{54}$	$\sqrt{65}$
A4, A5	$\sqrt{57}$			0	$\sqrt{7}$	$\sqrt{35}$
A6	$\sqrt{28}$				0	$\sqrt{27}$
A8	$\sqrt{6}$	$\sqrt{53}$	$\sqrt{65}$	$\sqrt{35}$	$\sqrt{27}$	0

Level	# Clusters	Clusters
0	8	$\{A1\}, \{A2\}, \{A3\}, \{A4\}, \{A5\}, \{A6\}, \{A7\}, \{A8\}$
1	7	$\{A1\}, \{A2\}, \{A3\}, \{A4, A5\}, \{A6\}, \{A7\}, \{A8\}$
2	6	$\{A1\}, \{A2, A7\}, \{A3\}, \{A4, A5\}, \{A6\}, \{A8\}$
3	5	$\{A1, A8\}, \{A2, A7\}, \{A3\}, \{A4, A5\}, \{A6\}$

Review Question 2: Agglomerative Clustering

Complete-linkage Proximity Matrix at Level 3

	A1, A8	A2, A7	A3	A4, A5	A6
A1, A8	0	$\sqrt{95}$	$\sqrt{65}$	$\sqrt{57}$	$\sqrt{28}$
A2, A7		0	$\sqrt{55}$	$\sqrt{49}$	$\sqrt{13}$
A3			0	$\sqrt{23}$	$\sqrt{54}$
A4, A5				0	$\sqrt{7}$
A6					0

- $\sqrt{7}$ is the next minimum distance between clusters when using complete-linkage, so we merge {A4, A5} and {A6}.

Level	# Clusters	Clusters
0	8	{A1}, {A2}, {A3}, {A4}, {A5}, {A6}, {A7}, {A8}
1	7	{A1}, {A2}, {A3}, {A4, A5}, {A6}, {A7}, {A8}
2	6	{A1}, {A2, A7}, {A3}, {A4, A5}, {A6}, {A8}
3	5	{A1, A8}, {A2, A7}, {A3}, {A4, A5}, {A6}
4	4	{A1, A8}, {A2, A7}, {A3}, {A4, A5, A6}

Review Question 2: Agglomerative Clustering

Complete-linkage Proximity Matrix at Level 3 – Merge {A4, A5} and {A6}

	A1, A8	A2, A7	A3	A4, A5	A6
A1, A8	0	$\sqrt{95}$	$\sqrt{65}$	$\sqrt{57}$	$\sqrt{28}$
A2, A7		0	$\sqrt{55}$	$\sqrt{49}$	$\sqrt{13}$
A3			0	$\sqrt{23}$	$\sqrt{54}$
A4, A5				0	$\sqrt{7}$
A6					0

Level	# Clusters	Clusters
0	8	{A1}, {A2}, {A3}, {A4}, {A5}, {A6}, {A7}, {A8}
1	7	{A1}, {A2}, {A3}, {A4, A5}, {A6}, {A7}, {A8}
2	6	{A1}, {A2, A7}, {A3}, {A4, A5}, {A6}, {A8}
3	5	{A1, A8}, {A2, A7}, {A3}, {A4, A5}, {A6}
4	4	{A1, A8}, {A2, A7}, {A3}, {A4, A5, A6}

Review Question 2: Agglomerative Clustering

Complete-linkage Proximity Matrix at Level 4

	A1, A8	A2, A7	A3	A4, A5, A6
A1, A8	0	$\sqrt{95}$	$\sqrt{65}$	$\sqrt{57}$
A2, A7		0	$\sqrt{55}$	$\sqrt{49}$
A3			0	$\sqrt{54}$
A4, A5, A6				0

- $\sqrt{7}$ is the next minimum distance between clusters when using complete-linkage, so we merge {A2, A7} and {A4, A5, A6}.

Level	# Clusters	Clusters
0	8	{A1}, {A2}, {A3}, {A4}, {A5}, {A6}, {A7}, {A8}
1	7	{A1}, {A2}, {A3}, {A4, A5}, {A6}, {A7}, {A8}
2	6	{A1}, {A2, A7}, {A3}, {A4, A5}, {A6}, {A8}
3	5	{A1, A8}, {A2, A7}, {A3}, {A4, A5}, {A6}
4	4	{A1, A8}, {A2, A7}, {A3}, {A4, A5, A6}
5	3	{A1, A8}, {A2, A7, A4, A5, A6}, {A3}

Review Question 2: Agglomerative Clustering

Complete-linkage Proximity Matrix at Level 5

	A1, A8	A2, A7, A4, A5, A6	A3
A1, A8	0	$\sqrt{95}$	$\sqrt{65}$
A2, A7, A4, A5, A6		0	$\sqrt{55}$
A3			0

Level	# Clusters	Clusters
0	8	$\{A1\}, \{A2\}, \{A3\}, \{A4\}, \{A5\}, \{A6\}, \{A7\}, \{A8\}$
1	7	$\{A1\}, \{A2\}, \{A3\}, \{A4, A5\}, \{A6\}, \{A7\}, \{A8\}$
2	6	$\{A1\}, \{A2, A7\}, \{A3\}, \{A4, A5\}, \{A6\}, \{A8\}$
3	5	$\{A1, A8\}, \{A2, A7\}, \{A3\}, \{A4, A5\}, \{A6\}$
4	4	$\{A1, A8\}, \{A2, A7\}, \{A3\}, \{A4, A5, A6\}$
5	3	$\{A1, A8\}, \{A2, A7, A4, A5, A6\}, \{A3\}$
6	2	$\{A1, A8\}, \{A2, A7, A4, A5, A6, A3\}$
7	1	$\{A1, A8, A2, A7, A4, A5, A6, A3\}$

Note: Use the sequence from the 2nd last level to draw the dendrogram. (I assume you do not keep the sequence at the last level, since it is one cluster which contains all data points.)