COMP 20008
SEMESTER 2, 2018

Agglomerative Clustering Example

Using the 1-dimensional dataset (1,2.7,3.3,4.6,6.0,7.5,9.1,10.8,12.6,16.5), agglomerative hierarchical clustering and Euclidean distance, show the dendrogram for each of single linkage and complete linkage criteria.

Dissimilarity Matrix

	1	2.7	3.3	4.6	6	7.5	9.1	10.8	12.6	16.5
1	0									
2.7	1.7	0								
3.3			0							
4.6				0						
6					0					
7.5						0				
9.1							0			
10.8								0		
12.6									0	
16.5										0

Initially, how many clusters do we have?

$(2.7 - 1)^2$

Inter-point distance Matrix

Step1: Calculate Distances between every pair of observation: Euclidean Distance

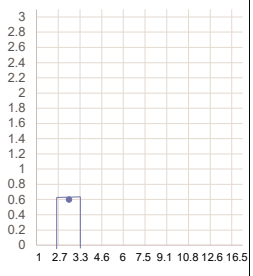
Using the 1-dimensional dataset (1,2.7,3.3,4.6,6.0,7.5,9.1,10.8,12.6,16.5), agglomerative hierarchical clustering and Euclidean distance, show the dendrogram for each of single linkage and complete linkage criteria.

Dissimilarity Matrix

	1	2.7	3.3	4.6	6	7.5	9.1	10.8	12.6	16.5
1	0									
2.7	1.7	0								
3.3	2.3	0.6	0							
4.6	3.6	1.9	1.3	0						
6	5	3.3	2.7	1.4	0					
7.5	6.5	4.8	4.2	2.9	1.5	0				
9.1	8.1	6.4	5.8	4.5	3.1	1.6	0			
10.8	9.8	8.1	7.5	6.2	4.8	3.3	1.7	0		
12.6	11.6	9.9	9.3	8	6.6	5.1	3.5	1.8	0	
16.5	15.5	13.8	13.2	11.9	10.5	9	7.4	5.7	3.9	0

2.7 and 3.3

Inter-point distance Matrix



Dendrogram Plot
X-axis → observations , Y-axis → distances

Step 2: Choose the most similar two observations to merge (i.e. Closest) (i.e. pair with the minimum distance in Dissimilarity Matrix)

Using the 1-dimensional dataset (1,2.7,3.3,4.6,6.0,7.5,9.1,10.8,12.6,16.5), agglomerative hierarchical clustering and Euclidean distance, show the dendrogram for each of single linkage and complete linkage criteria.

Dissimilarity Matrix

	1	2.7-3.3	4.6	6	7.5	9.1	10.8	12.6	16.5
1	0								
2.7-3.3	1.7	0							
4.6	3.6	?	0						
6	5	?	1.4	0					
7.5	6.5	?	2.9	1.5	0				
9.1	8.1	?	4.5	3.1	1.6	0			
10.8	9.8	?	6.2	4.8	3.3	1.7	0		
12.6	11.6	?	8	6.6	5.1	3.5	1.8	0	
16.5	15.5	?	11.9	10.5	9	7.4	5.7	3.9	0

D(4.6,2.7)
D(4.6,3.3)

	1	2.7	3.3	4.6	6
1	0				
2.7	1.7	0			
3.3	2.3	0.6	0		
4.6	3.6	1.9	1.3	0	
6	5	3.3	2.7	1.4	0
7.5	6.5	4.8	4.2	2.9	1.5
9.1	8.1	6.4	5.8	4.5	3.1
10.8	9.8	8.1	7.5	6.2	4.8
12.6	11.6	9.9	9.3	8	6.6
16.5	15.5	13.8	13.2	11.9	10.5

Inter-point distance Matrix

Step 3: Update Dissimilarity Matrix: Calculate the distance between Cluster12 and all other observations (calculate linkage using min)

Iteration # 1 — Single Linkage – Hierarchical Clustering

Using the 1-dimensional dataset (1,2.7,3.3,4.6,6.0,7.5,9.1,10.8,12.6,16.5), agglomerative hierarchical clustering and Euclidean distance, show the dendrogram for each of single linkage and complete linkage criteria.

Dissimilarity Matrix

	1	2.7-3.3	4.6	6	7.5	9.1	10.8	12.6	16.5
1	0								
2.7-3.3	1.7	0							
4.6	3.6	1.3	0						
6	5	?	1.4	0					
7.5	6.5	?	2.9	1.5	0				
9.1	8.1	?	4.5	3.1	1.6	0			
10.8	9.8	?	6.2	4.8	3.3	1.7	0		
12.6	11.6	?	8	6.6	5.1	3.5	1.8	0	
16.5	15.5	?	11.9	10.5	9	7.4	5.7	3.9	0

Inter-point distance Matrix

	1	2.7	3.3	4.6	6
1	0				
2.7	1.7	0			
3.3	2.3	0.6	0		
4.6	3.6	1.9	1.3	0	
6	5	3.3	2.7	1.4	0
7.5	6.5	4.8	4.2	2.9	1.5
9.1	8.1	6.4	5.8	4.5	3.1
10.8	9.8	8.1	7.5	6.2	4.8
12.6	11.6	9.9	9.3	8	6.6
16.5	15.5	13.8	13.2	11.9	10.5

Step 3: Update Dissimilarity Matrix: Calculate the distance between Cluster12 and all other observations (calculate linkage using min)

Iteration # 1 — Single Linkage – Hierarchical Clustering

Using the 1-dimensional dataset (1,2.7,3.3,4.6,6.0,7.5,9.1,10.8,12.6,16.5), agglomerative hierarchical clustering and Euclidean distance, show the dendrogram for each of single linkage and complete linkage criteria.

Dissimilarity Matrix

	1	2.7-3.3	4.6	6	7.5	9.1	10.8	12.6	16.5
1	0								
2.7-3.3	1.7	0							
4.6	3.6	1.3	0						
6	5	2.7	1.4	0					
7.5	6.5	?	2.9	1.5	0				
9.1	8.1	?	4.5	3.1	1.6	0			
10.8	9.8	?	6.2	4.8	3.3	1.7	0		
12.6	11.6	?	8	6.6	5.1	3.5	1.8	0	
16.5	15.5	?	11.9	10.5	9	7.4	5.7	3.9	0

Inter-point distance Matrix

	1	2.7	3.3	4.6	6
1	0				
2.7	1.7	0			
3.3	2.3	0.6	0		
4.6	3.6	1.9	1.3	0	
6	5	3.3	2.7	1.4	0
7.5	6.5	4.8	4.2	2.9	1.5
9.1	8.1	6.4	5.8	4.5	3.1
10.8	9.8	8.1	7.5	6.2	4.8
12.6	11.6	9.9	9.3	8	6.6
16.5	15.5	13.8	13.2	11.9	10.5

Step 3: Update Dissimilarity Matrix: Calculate the distance between Cluster12 and all other observations (calculate linkage using min)

Iteration # 1 — Single Linkage – Hierarchical Clustering

Using the 1-dimensional dataset (1,2.7,3.3,4.6,6.0,7.5,9.1,10.8,12.6,16.5), agglomerative hierarchical clustering and Euclidean distance, show the dendrogram for each of single linkage and complete linkage criteria.

Dissimilarity Matrix

	1	2.7-3.3	4.6	6	7.5	9.1	10.8	12.6	16.5
1	0								
2.7-3.3	1.7	0							
4.6	3.6	1.3	0						
6	5	2.7	1.4	0	*Fill in the ?*				
7.5	6.5	4.2	2.9	1.5	0				
9.1	8.1	?	4.5	3.1	1.6	0			
10.8	9.8	?	6.2	4.8	3.3	1.7	0		
12.6	11.6	?	8	6.6	5.1	3.5	1.8	0	
16.5	15.5	?	11.9	10.5	9	7.4	5.7	3.9	0

Inter-point distance Matrix

	1	2.7	3.3	4.6	6
1	0				
2.7	1.7	0			
3.3	2.3	0.6	0		
4.6	3.6	1.9	1.3	0	
6	5	3.3	2.7	1.4	0
7.5	6.5	4.8	4.2	2.9	1.5
9.1	8.1	6.4	5.8	4.5	3.1
10.8	9.8	8.1	7.5	6.2	4.8
12.6	11.6	9.9	9.3	8	6.6
16.5	15.5	13.8	13.2	11.9	10.5

Step 3: Update Dissimilarity Matrix: Calculate the distance between Cluster12 and all other observations (calculate linkage using min)

Iteration # 1 — Single Linkage – Hierarchical Clustering

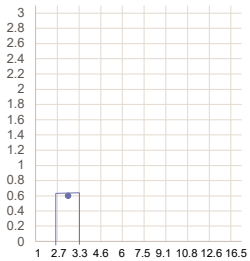
Using the 1-dimensional dataset (1,2.7,3.3,4.6,6.0,7.5,9.1,10.8,12.6,16.5), agglomerative hierarchical clustering and Euclidean distance, show the dendrogram for each of single linkage and complete linkage criteria.

Dissimilarity Matrix

	1	2.7-3.3	4.6	6	7.5	9.1	10.8	12.6	16.5
1	0								
2.7-3.3	1.7	0							
4.6	3.6	1.3	0						
6	5	2.7	1.4	0					
7.5	6.5	4.2	2.9	1.5	0				
9.1	8.1	5.8	4.5	3.1	1.6	0			
10.8	9.8	7.5	6.2	4.8	3.3	1.7	0		
12.6	11.6	9.3	8	6.6	5.1	3.5	1.8	0	
16.5	15.5	13.2	11.9	10.5	9	7.4	5.7	3.9	0

Inter-point distance Matrix

This is one iteration COMPLETED



Dendrogram Plot
X-axis → observations , Y-axis → distances

Iteration # 2 — Single Linkage – Hierarchical Clustering

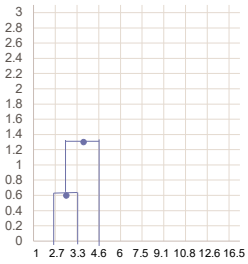
Using the 1-dimensional dataset (1,2.7,3.3,4.6,6.0,7.5,9.1,10.8,12.6,16.5), agglomerative hierarchical clustering and Euclidean distance, show the dendrogram for each of single linkage and complete linkage criteria.

Dissimilarity Matrix

	1	2.7-3.3	4.6	6	7.5	9.1	10.8	12.6	16.5
1	0								
2.7-3.3	1.7	0							
4.6	3.6	1.3	0						
6	5	2.7	1.4	0					
7.5	6.5	4.2	2.9	1.5	0				
9.1	8.1	5.8	4.5	3.1	1.6	0			
10.8	9.8	7.5	6.2	4.8	3.3	1.7	0		
12.6	11.6	9.3	8	6.6	5.1	3.5	1.8	0	
16.5	15.5	13.2	11.9	10.5	9	7.4	5.7	3.9	0

2.7-3.3 and 4.6

Inter-point distance Matrix



Dendrogram Plot
X-axis → observations , Y-axis → distances

Step 2: Choose the most similar two observations to merge (i.e. Closest)
(i.e. pair with the minimum distance in Dissimilarity Matrix)

Iteration # 2 — Single Linkage – Hierarchical Clustering

Using the 1-dimensional dataset (1,2.7,3.3,4.6,6.0,7.5,9.1,10.8,12.6,16.5), agglomerative hierarchical clustering and Euclidean distance, show the dendrogram for each of single linkage and complete linkage criteria.

Dissimilarity Matrix

	1	2.7-3.3-4.6	6	7.5	9.1	10.8	12.6	16.5
1	0	1.7						
2.7-3.3-4..6	1.7	0						
6	5	?	0					
7.5	6.5	?	1.5	0				
9.1	8.1	?	3.1	1.6	0			
10.8	9.8	?	4.8	3.3	1.7	0		
12.6	11.6	?	6.6	5.1	3.5	1.8	0	
16.5	15.5	?	10.5	9	7.4	5.7	3.9	0

Inter-point distance Matrix

	1	2.7-3.3	4.6	6
1	0	1.7	3.6	
2.7-3.3	1.7	0		
4.6	3.6	1.3	0	
6	5	2.7	1.4	0
7.5	6.5	4.2	2.9	1.5
9.1	8.1	5.8	4.5	3.1
10.8	9.8	7.5	6.2	4.8
12.6	11.6	9.3	8	6.6
16.5	15.5	13.2	11.9	10.5

Step 3: Update Dissimilarity Matrix: Calculate the distance between Cluster12 and all other observations (calculate linkage using min)

Iteration # 2 — Single Linkage – Hierarchical Clustering

Using the 1-dimensional dataset (1,2.7,3.3,4.6,6.0,7.5,9.1,10.8,12.6,16.5), agglomerative hierarchical clustering and Euclidean distance, show the dendrogram for each of single linkage and complete linkage criteria.

Dissimilarity Matrix

	1	2.7-3.3-4.6	6	7.5	9.1	10.8	12.6	16.5
1	0	1.7						
2.7-3.3-4..6	1.7	0						
6	5	?	0					
7.5	6.5	?	1.5	0				
9.1	8.1	?	3.1	1.6	0			
10.8	9.8	?	4.8	3.3	1.7	0		
12.6	11.6	?	6.6	5.1	3.5	1.8	0	
16.5	15.5	?	10.5	9	7.4	5.7	3.9	0

D(4.6,2.7-3.3)
D(4.6,2.73.3)

	1	2.7-3.3	4.6	6
1	0			
2.7-3.3	1.7	0		
4.6	3.6	1.3	0	
6	5	2.7	1.4	0
7.5	6.5	4.2	2.9	1.5
9.1	8.1	5.8	4.5	3.1
10.8	9.8	7.5	6.2	4.8
12.6	11.6	9.3	8	6.6
16.5	15.5	13.2	11.9	10.5

Inter-point distance Matrix

Step 3: Update Dissimilarity Matrix: Calculate the distance between Cluster12 and all other observations (calculate linkage using min)

Iteration # 2 — Single Linkage – Hierarchical Clustering

Using the 1-dimensional dataset (1,2.7,3.3,4.6,6.0,7.5,9.1,10.8,12.6,16.5), agglomerative hierarchical clustering and Euclidean distance, show the dendrogram for each of single linkage and complete linkage criteria.

Dissimilarity Matrix

	1	2.7-3.3-4.6	6	7.5	9.1	10.8	12.6	16.5
1	0							
2.7-3.3-4..6	1.7	0						
6	5	1.4	0					
7.5	6.5	?	1.5	0				
9.1	8.1	?	3.1	1.6	0			
10.8	9.8	?	4.8	3.3	1.7	0		
12.6	11.6	?	6.6	5.1	3.5	1.8	0	
16.5	15.5	?	10.5	9	7.4	5.7	3.9	0

Fill in the ?

Inter-point distance Matrix

	1	2.7-3.3	4.6	6
1	0			
2.7-3.3	1.7	0		
4.6	3.6	1.3	0	
6	5	2.7	1.4	0
7.5	6.5	4.2	2.9	1.5
9.1	8.1	5.8	4.5	3.1
10.8	9.8	7.5	6.2	4.8
12.6	11.6	9.3	8	6.6
16.5	15.5	13.2	11.9	10.5

Step 3: Update Dissimilarity Matrix: Calculate the distance between Cluster12 and all other observations (calculate linkage using min)

Iteration # 2 — Single Linkage – Hierarchical Clustering

Using the 1-dimensional dataset (1,2.7,3.3,4.6,6.0,7.5,9.1,10.8,12.6,16.5), agglomerative hierarchical clustering and Euclidean distance, show the dendrogram for each of single linkage and complete linkage criteria.

Dissimilarity Matrix

	1	2.7-3.3-4.6	6	7.5	9.1	10.8	12.6	16.5
1	0							
2.7-3.3-4..6	1.7	0						
6	5	1.4	0					
7.5	6.5	2.9	1.5	0				
9.1	8.1	4.5	3.1	1.6	0			
10.8	9.8	6.2	4.8	3.3	1.7	0		
12.6	11.6	8	6.6	5.1	3.5	1.8	0	
16.5	15.5	11.9	10.5	9	7.4	5.7	3.9	0

Inter-point distance Matrix

	1	2.7-3.3	4.6	6
1	0			
2.7-3.3	1.7	0		
4.6	3.6	1.3	0	
6	5	2.7	1.4	0
7.5	6.5	4.2	2.9	1.5
9.1	8.1	5.8	4.5	3.1
10.8	9.8	7.5	6.2	4.8
12.6	11.6	9.3	8	6.6
16.5	15.5	13.2	11.9	10.5

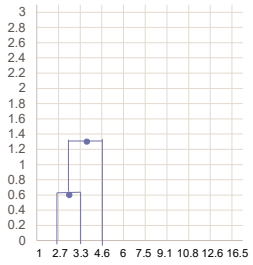
Step 3: Update Dissimilarity Matrix: Calculate the distance between Cluster12 and all other observations (calculate linkage using min)

Iteration # 2 — Single Linkage – Hierarchical Clustering

Using the 1-dimensional dataset (1,2.7,3.3,4.6,6.0,7.5,9.1,10.8,12.6,16.5), agglomerative hierarchical clustering and Euclidean distance, show the dendrogram for each of single linkage and complete linkage criteria.

Dissimilarity Matrix

	1	2.7-3.3-4.6	6	7.5	9.1	10.8	12.6	16.5
1	0							
2.7-3.3-4..6	1.7	0						
6	5	1.4	0					
7.5	6.5	2.9	1.5	0				
9.1	8.1	4.5	3.1	1.6	0			
10.8	9.8	6.2	4.8	3.3	1.7	0		
12.6	11.6	8	6.6	5.1	3.5	1.8	0	
16.5	15.5	11.9	10.5	9	7.4	5.7	3.9	0



Dendrogram Plot
X-axis → observations , Y-axis → distances

Inter-point distance Matrix

This is one more iteration COMPLETED

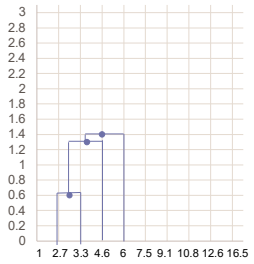
Iteration # 3 — Single Linkage – Hierarchical Clustering

Using the 1-dimensional dataset (1,2.7,3.3,4.6,6.0,7.5,9.1,10.8,12.6,16.5), agglomerative hierarchical clustering and Euclidean distance, show the dendrogram for each of single linkage and complete linkage criteria.

Dissimilarity Matrix

	1	2.7-3.3-4.6	6	7.5	9.1	10.8	12.6	16.5
1	0							
2.7-3.3-4..6	1.7	0						
6	5	1.4	0					
7.5	6.5	2.9	1.5	0				
9.1	8.1	4.5	3.1	1.6	0			
10.8	9.8	6.2	4.8	3.3	1.7	0		
12.6	11.6	8	6.6	5.1	3.5	1.8	0	
16.5	15.5	11.9	10.5	9	7.4	5.7	3.9	0

2.7-3.3-4.6 and 6
Let's call it cluster A



Dendrogram Plot
X-axis → observations , Y-axis → distances

Inter-point distance Matrix

This is one iteration COMPLETED

Step 2: Choose the most similar two observations to merge (i.e. Closest)
(i.e. pair with the minimum distance in Dissimilarity Matrix)

Iteration # 3 — Single Linkage – Hierarchical Clustering

Using the 1-dimensional dataset (1,2.7,3.3,4.6,6.0,7.5,9.1,10.8,12.6,16.5), agglomerative hierarchical clustering and Euclidean distance, show the dendrogram for each of single linkage and complete linkage criteria.

Dissimilarity Matrix

	1	A	7.5	9.1	10.8	12.6	16.5
1	0	1.7					
A	1.7	0					
7.5	6.5	?	0				
9.1	8.1	?	1.6	0			
10.8	9.8	?	3.3	1.7	0		
12.6	11.6	?	5.1	3.5	1.8	0	
16.5	15.5	?	9	7.4	5.7	3.9	0

Fill in the ?

Inter-point distance Matrix

	1	2.7-3.3-4.6	6	7.5
1	0			
2.7-3.3-4..6	1.7	0		
6	5	1.4	0	
7.5	6.5	2.9	1.5	0
9.1	8.1	4.5	3.1	1.6
10.8	9.8	6.2	4.8	3.3
12.6	11.6	8	6.6	5.1
16.5	15.5	11.9	10.5	9

Step 3: Update Dissimilarity Matrix: Calculate the distance between Cluster12 and all other observations (calculate linkage using min)

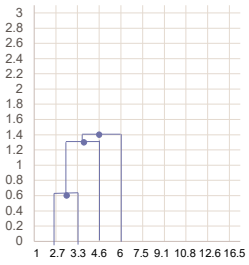
Iteration # 3 — Single Linkage – Hierarchical Clustering

Using the 1-dimensional dataset (1,2.7,3.3,4.6,6.0,7.5,9.1,10.8,12.6,16.5), agglomerative hierarchical clustering and Euclidean distance, show the dendrogram for each of single linkage and complete linkage criteria.

Dissimilarity Matrix

	1	A	7.5	9.1	10.8	12.6	16.5
1	0						
A	1.7	0					
7.5	6.5	1.5	0				
9.1	8.1	3.1	1.6	0			
10.8	9.8	4.8	3.3	1.7	0		
12.6	11.6	6.6	5.1	3.5	1.8	0	
16.5	15.5	10.5	9	7.4	5.7	3.9	0

Inter-point distance Matrix



Dendrogram Plot
X-axis → observations , Y-axis → distances

This is another iteration COMPLETED

How many Iterations will be running the algorithm ?

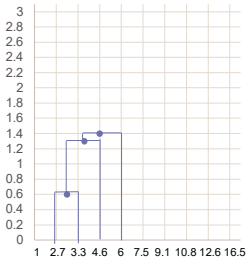
Single Linkage – Hierarchical Clustering

Using the 1-dimensional dataset (1,2.7,3.3,4.6,6.0,7.5,9.1,10.8,12.6,16.5), agglomerative hierarchical clustering and Euclidean distance, show the dendrogram for each of single linkage and complete linkage criteria.

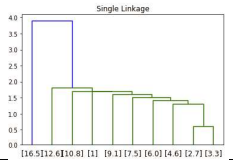
Dissimilarity Matrix

	1	A	7.5	9.1	10.8	12.6	16.5
1	0						
A	1.7	0					
7.5	6.5	1.5	0				
9.1	8.1	3.1	1.6	0			
10.8	9.8	4.8	3.3	1.7	0		
12.6	11.6	6.6	5.1	3.5	1.8	0	
16.5	15.5	10.5	9	7.4	5.7	3.9	0

Inter-point distance Matrix



The one generated by Python



How many Iterations will be running the algorithm ?

Linkage Methods

Other cluster merging methods:

Dissimilarity Matrix

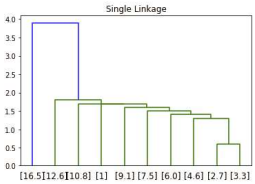
	1	A	7.5	9.1	10.8	12.6	16.5
1	0						
A	1.7	0					
7.5	6.5	?	0				
9.1	8.1	?	1.6	0			
10.8	9.8	?	3.3	1.7	0		
12.6	11.6	?	5.1	3.5	1.8	0	
16.5	15.5	?	9	7.4	5.7	3.9	0

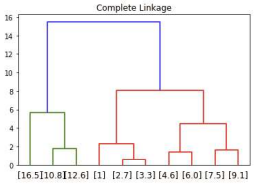
Inter-point distance Matrix

	1	2.7-3.3-4..6	6	7.5
1	0			
2.7-3.3-4..6	1.7	0		
6	5	1.4	0	
7.5	6.5	2.9	1.5	0
9.1	8.1	4.5	3.1	1.6
10.8	9.8	6.2	4.8	3.3
12.6	11.6	8	6.6	5.1
16.5	15.5	11.9	10.5	9

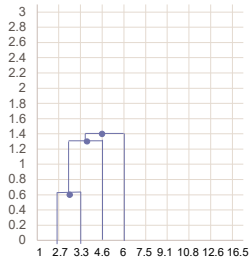
- Single linkage: D({2.7, 3.3, 4.6}, 6) = min {2.9,1.5} = 1.5
- Complete linkage: D({2.7, 3.3, 4.6}, 6) = max {2.9,1.5} = 2.9

Linkage Methods





Dendrogram through Python packages



3-iterations dendrogram