CZ4041/CE4041: Machine Learning

Week 10: Clustering

Nerd Joke Time

- Hollywood Executive: Give me an idea for a movie.
- Me: It's a movie about high-school girls trying to figure out which clique they belong to. They move from one clique to the next until they minimize their differences. It's called K-means Girls.



Question 1

• MIN or Single Link: Distance of two clusters is based on the two most closest points in the different clusters

	P1	P2	P3	P4	P5
P1	0.00	0.90	0.10	0.65	0.20
P2	0.90	0.00	0.70	0.60	0.50
P3	0.10	0.70	0.00	0.40	0.30
P4	0.65	0.60	0.40	0.00	0.80
P5	0.20	0.50	0.30	0.80	0.00

Agglomerative Clustering Algorithm

Basic algorithm:

Compute the proximity matrix
Let each data point be a cluster
Repeat
Merge the two closest clusters
Update the proximity matrix
Until only a single cluster remains

Step 1: Merge the two closest clusters (smallest distance)

	P1	P2	P3	P4	P5
P1	0.00	0.90	0.10	0.65	0.20
P2	0.90	0.00	0.70	0.60	0.50
P3	0.10	0.70	0.00	0.40	0.30
P4	0.65	0.60	0.40	0.00	0.80
P5	0.20	0.50	0.30	0.80	0.00

Distance matrix

3 2 4

• Step 2: Update proximity matrix based on MIN: proximity of two clusters is based on the two closest points in different clusters (smallest distance)

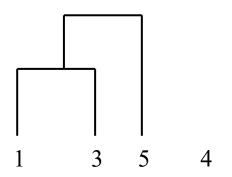
	P1∪P3		P2	P4	P5
		0.10	0.00	0.65	000
P1UP3	0.0	00	0.70	0.40	0.20
	0.10	0.00	0.70	0.40	0.30
P2	0.	70	0.00	0.60	0.50
P4	0.4	40	0.60	0.00	0.80
P5	0.2	20	0.50	0.80	0.00

Distance matrix

3 2 4

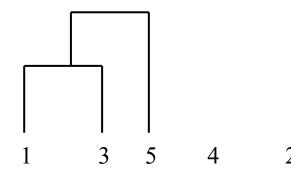
Step 1: Merge the two closest clusters (smallest distance)

	P1UP3	P2	P4	P5
P1∪P3	0.00	0.70	0.40	0.20
P2	0.70	0.00	0.60	0.50
P4	0.40	0.60	0.00	0.80
P5	0.20	0.50	0.80	0.00



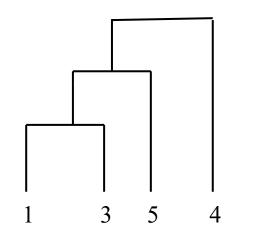
• Step 2: Update proximity matrix based on MIN: proximity of two clusters is based on the two closest points in different clusters (smallest distance)

	P1UP3UP5		P2	P4
	0.00	0.00	0.70	0.40
P1UP3UP5	0.00		0.50	0.40
	0.20	0.00	0.30	0.80
1				
P2	0	50	0.00	0.60



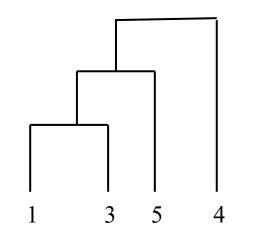
Step 1: Merge the two closest clusters (smallest distance)

	P1UP3UP5	P2	P4
P1UP3UP5	0.00	0.50	0.40
P2	0.50	0.00	0.60
P4	0.40	0.60	0.00



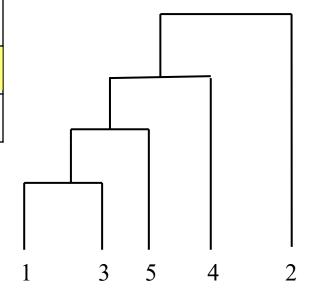
• Step 2: Update proximity matrix based on MIN: proximity of two clusters is based on the two closest points in different clusters (smallest distance)

	P1UP3U	P2	
	0.00	0.40	0.50
P1UP3UP5UP4	0.0	0.50	
	0. 1 0	0.00	0.00
P2	0.50		0.00



Step 1: Merge the two closest clusters (smallest distance)

	P1UP3UP5UP4	P2
P1UP3UP5UP4	0.00	0.50
P2	0.50	0.00



Question 2

• MAX or Complete Link: Similarity of two clusters is based on the two least similar points in the different clusters

	P1	P2	P3	P4	P5
P1	1.00	0.90	0.10	0.65	0.20
P2	0.90	1.00	0.70	0.60	0.50
P3	0.10	0.70	1.00	0.40	0.30
P4	0.65	0.60	0.40	1.00	0.80
P5	0.20	0.50	0.30	0.80	1.00

1 2 3 4 5

Similarity matrix



Step 1: Merge the two closest clusters (largest similarity)

	P1	P2	P3	P4	P5
P1	1.00	0.90	0.10	0.65	0.20
P2	0.90	1.00	0.70	0.60	0.50
P3	0.10	0.70	1.00	0.40	0.30
P4	0.65	0.60	0.40	1.00	0.80
P5	0.20	0.50	0.30	0.80	1.00

Similarity matrix

4

5

• Step 2: Update proximity matrix based on MAX: proximity of two clusters is based on the two farthest points in different clusters (smallest similarity)

	P1∪P2	P3	P4	P5
	1 00 0 00	0.10	0.65	0.20
P1∪P2	1.00	0.10	0.60	0.20
	0.90 1.00	0.70	0.00	0.50
P3	0.10	1.00	0.40	0.30
P4	0.60	0.40	1.00	0.80
P5	0.20	0.30	0.80	1.00

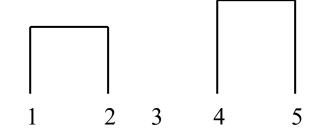
Similarity matrix

2 3

3

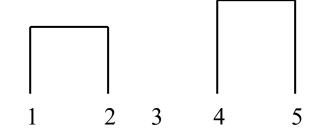
Step 1: Merge the two closest clusters (largest similarity)

	P1∪P2	P3	P4	P5
P1∪P2	1.00	0.10	0.60	0.20
P3	0.10	1.00	0.40	0.30
P4	0.60	0.40	1.00	0.80
P5	0.20	0.30	0.80	1.00



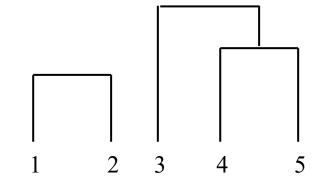
• Step 2: Update proximity matrix based on MAX: proximity of two clusters is based on the two farthest points in different clusters (smallest similarity)

	P1∪P2	P3	P4∪P5	
P1UP2	1.00	0.10	0.20	
P3	0.10	1.00	0.30	
	0.60	0.40	1.00	0.00
P4UP5	0.20	0.30	1.00	
	0.20	0.30	0.80	1.00



Step 1: Merge the two closest clusters (largest similarity)

	P1∪P2	P3	P4∪P5
P1∪P2	1.00	0.10	0.20
P3	0.10	1.00	0.30
P4∪P5	0.20	0.30	1.00



• Step 2: Update proximity matrix based on MAX: proximity of two clusters is based on the two farthest points in different clusters (smallest similarity)

	P1∪P2	P3 UP4UP5				
P1∪P2	1.00	0.10				
		1 00 00	<u>-</u>		ı——	
P3UP4UP5	0.10	1.00				
	0.20	0.30 1.00	-			
C:.						
Sir	nilarity matrix					
			1	2	3	

Step 1: Merge the two closest clusters (largest similarity)

	P1∪P2	P3UP4UP5
P1∪P2	1.00	0.10
P3UP4UP5	0.10	1.00

