task_10.4

Machine Learning (WiSe 2025/2026)

Author: Suvansh Shukla
Matriculation No. 256245

Assignment 10 Task 4

Part A Single Link

To do this task we require a distance matrix.

The distance matrix is a tabular representation of the distance from every point to every other point.

The distance measure used here is euclidean distance.

For Hierarchical agglomerative clustering via Single Link as cluster distance we need a distance matrix:

	A	B	C	D	E	F
A	0.00					
B	1.41	0.00				
C	6.00	5.10	0.00			
D	6.26	6.80	11.88	0.00		
E	5.39	5.00	9.43	4.03	0.00	
F	4.12	3.61	8.06	4.61	1.41	0.00

Looking at the distance matrix we see that we can merge A & B and E & F together.

Both having distanace value = 1.41.

We start off by merging A & B

This would update the matrix like so:

	{A,B}	C	D	E	F
{A,B}	0.00				
C	5.10	0.00			
D	6.26	11.88	0.00		
E	5.00	9.43	4.03	0.00	

	{A,B}	C	D	E	F
F	3.61	8.06	4.61	1.41	0.00

Now we merge E & F, updating the matrix to this:

	{A,B}	C	D	{E,F}
{A,B}	0.00			
C	5.10	0.00		
D	6.26	11.88	0.00	
{E,F}	3.61	8.06	4.03	0.00

Now we see that the smallest distance is 3.61 between {A,B} & {E,F}.

This means we merge them next and update the matrix:

	{A,B,E,F}	C	D
{A,B,E,F}	0.00		
C	5.10	0.00	
D	4.03	11.88	0.00

Now we see that the smallest distance is 4.03 between {A,B,E,F} and D.

This means we merget them and update the matrix again:

	{A,B,E,F,C}	D
{A,B,E,F,C}	0.00	
D	4.03	0.00

Finally we merge D with {A,B,E,F,C} resulting in {A,B,E,F,C,D}.

Part B Centroid Link

In Centroid Linkage HAC, we have to calculate a new centroid after merging points.

Then we need to re-calculate distances of each point to this new centroid and update the distance matrix.

So our initially distance matrix is the same:

	A	B	C	D	E	F
A	0.00					
B	1.41	0.00				
C	6.00	5.10	0.00			

	A	B	C	D	E	F
D	6.26	6.80	11.88	0.00		
E	5.39	5.00	9.43	4.03	0.00	
F	4.12	3.61	8.06	4.61	1.41	0.00

We start off again by merging A & B.

The new centroid for this is {A,B}: (8.50, 8.50).

Now we re-calculate distance of all points to each other and to this new centroid, then update the matrix.

	{A,B}	C	D	E	F
{A,B}	0.00				
C	5.5227	0.00			
D	6.5000	11.88	0.00		
E	5.1478	9.43	4.03	0.00	
F	3.8079	8.06	4.61	1.41	0.00

Now we see that the smallest distance is between E & F, so we merge and calculate new centroid.

New Centroid for {E,F}: (6.50, 4.50)

Now we re-calculate distance of all points to each other and to this new centroid, then update the matrix.

	{A,B}	C	D	{E,F}
{A,B}	0.00			
C	5.10	0.00		
D	6.26	11.88	0.00	
{E,F}	4.4721	8.7464	4.2720	0.00

Now the smallest distance is 4.2720 between D & {E,F}. So we merge them and calculate the new centroid.

New Centroid for {D,E,F}: (4.50, 5.25)

	{A,B}	C	{E,F,D}
{A,B}	0.00		
C	10.2133	0.00	
{E,F,D}	5.1539	11.88	0.00

Now the smallest distance is 5.1539 between {A,B} and {E,F,D}.

So we merge them and calculate the new centroid and update the matrix.

New Centroid for {AB,EFD}: (6.50, 6.88)

	{A,B,E,F,D}	C
{A,B,E,F,D}	0.00	
C	7.7952	0.00

Finally we just merge C into {A,B,E,F,D} giving us: {A, B, E, F, D, C}

Part C Cluster Distance

In cluster distance we calculate the average distance of the cluster to all other clusters/points.

This means if we merge A & B, we calculate the average of $\frac{d(A,C)+d(B,C)}{2}$.

This is repeated for all points C, D, E, F.

Starting with the normal distance matrix:

	A	B	C	D	E	F
A	0.00					
B	1.41	0.00				
C	6.00	5.10	0.00			
D	6.26	6.80	11.88	0.00		
E	5.39	5.00	9.43	4.03	0.00	
F	4.12	3.61	8.06	4.61	1.41	0.00

We merge A & B and update the matrix using the average method.

So the new matrix is:

	{C}	{D}	{E}	{F}	{A,B}
{C}	0.0000				
{D}	11.8849	0.0000			
{E}	9.4340	4.0311	0.0000		
{F}	8.0623	4.6098	1.4142	0.0000	
{A,B}	5.5495	6.5329	5.1926	3.8643	0.0000

Now we see that the smallest distance is 1.4142 between E & F.

We merge them and update the matrix by calculating the average distances between points and clusters.

	{C}	{D}	{A,B}	{E,F}
{C}	0.0000			
{D}	11.8849	0.0000		
{A,B}	5.5495	6.5329	0.0000	
{E,F}	8.7481	4.3205	4.5285	0.0000

The smallest distance now is 4.3205 between D & {E,F}. So we merge them.

Then calculate average distances and update the matrix.

	{C}	{A,B}	{D,E,F}
{C}	0.0000		
{A,B}	5.5495	0.0000	
{D,E,F}	9.7937	5.1966	0.0000

This Finally gives us the smallest distance of 5.1966 between {A,B} and {D,E,F}.

So we merge them and update the matrix.

	{C}	{A,B,D,E,F}
{C}	0.0000	
{A,B,D,E,F}	8.0960	0.0000

And finally we merge C & {A, B, D, E, F}.