

Hierarchical Clustering Exercise (Single Linkage)

This exercise introduces students to hierarchical clustering using a realistic toy dataset. The dataset represents casino customer behavior with two features per customer. Students will calculate distances, identify the merging sequence, and draw a dendrogram.

Adjusted Toy Dataset (Casino Customers)

Customer	Avg Bet per Visit (USD)	Visits per Month
A	3	4
B	2	4
C	3	3
D	3	5
E	9	7
F	8	7
G	9	6
H	9	8
I	4	4
J	10	7
T	5	4
U	7	7

DBSCAN Hyperparameters:

eps = 1

min_samples = 2

	A (3,4)	B (2,4)	C (3,3)	D (3,5)	E (9,7)	F (8,7)	G (9,6)	H (9,8)	I (4,4)	J (10,7)	T (5,4)	U (7,7)
A(3,4)	0	1	1	1	6.7	9.8	6.3	7.2	1	7.6		
B(2,4)	1	0	1.4	1.4	7.6	6.7	7.2	8	2	8.5		
C(3,3)	1	1.4	0	2	7.2	6.4	6.7	7.8	1.4	8		
D(3,5)	1	1.4	2	0	6.3	5.3	6	6.7	1.4	7.2		
E(9,7)	6.7	7.6	7.2	6.3	0	1	1	1	5.8	1		
F(8,7)	9.8	6.7	6.4	5.3	1	0	1.4	1.4	5	2		
G(9,6)	6.3	7.2	6.7	6	1	1.4	0	2	5.3	1.4		
H(9,8)	7.2	8	7.8	6.7	1	1.4	2	0	6.4	1.4		
I(4,4)	1	2	1.4	1.4	5.8	5	5.3	6.4	0	6.7		
J(10,7)	7.6	8.5	8	7.2	1	2	1.4	1.4	6.7	0		
T(5,4)												
U(7,7)												

Tasks

1. Create a scatter plot
2. Compute all pairwise Euclidean distances between customers.
3. Create a full distance matrix.
4. For each point A-U, list all of its neighbours based on **eps**.
5. Create a table showing each point and the **number of neighbors** in its neighborhood.
6. For each point A-U, determine if it is a core point, non-core point and noise.
7. Simulate DBSCAN
 - 7.1. Select the a core point and begin a new cluster
 - 7.2. Expand the cluster using its neighbours
 - 7.3. Add newly discovered core points to the expansion queue.
 - 7.4. Continue until no more points can be reached
 - 7.5. Repeat until all points are assigned or labeled noise.
8. Create a table showing each point and its cluster.

