

Clustering Assignment

1. Implement Hierarchical clustering algorithm and apply it on following distance matrix. Show the resulting dendograms after applying average linkage approach.

a.

	A	B	C	D	E	F
A	0	0.71	5.66	3.61	4.24	3.20
B	0.71	0	4.95	2.92	3.54	2.50
C	5.66	4.95	0	2.24	1.41	2.50
D	3.61	2.92	2.24	0	1.0	0.5
E	4.24	3.54	1.41	1.0	0	1.12
F	3.20	2.50	2.50	0.5	1.12	0

b. Apply same algorithm on distance matrix given in matrix.xlsx file.

2. Implement BIRCH algorithm and apply it on following datasets (T, L, B parameters can be user given if not specified):

a. $D1 = \{0.5, 0.25, 0, 0.65, 1, 1.4, 1.1\}$

given $T=0.15, L=2, B=2$

b. $D2 = \{(6,2), (7,2), (3,4), (7,4), (8,4), (2,6), (4,5), (4,7), (3,8)\}$

Apply the same on data given in files data1.xlsx and data2.xlsx as well.

3. Implement DBSCAN algorithm and apply it on datasets given in files data1.xlsx and data2.xlsx (epsilon, MinPts parameters can be user given if not specified):

Let $\epsilon = 1$ or 0.5

$\text{MinPts} = 30$ or 10

Apply the same on USCensusData (download from <https://archive.ics.uci.edu/ml/datasets/US+Census+Data+%281990%29>)

4. Implement OPTICS algorithm and apply it on datasets given in files data1.xlsx, data2.xlsx and opticsdata.xlsx (for this $\epsilon = 0.02$, $\text{minPts} = 500$) and output each point's reachability distance, core distance and order of points in the reachability graph.

5. Implement COBWEB clustering algorithm using following dataset:

INSTANCE LABEL	COLOUR	NUCLEI	TAILS
a	White	1	1
b	White	2	2
c	Black	2	2
d	Black	3	1