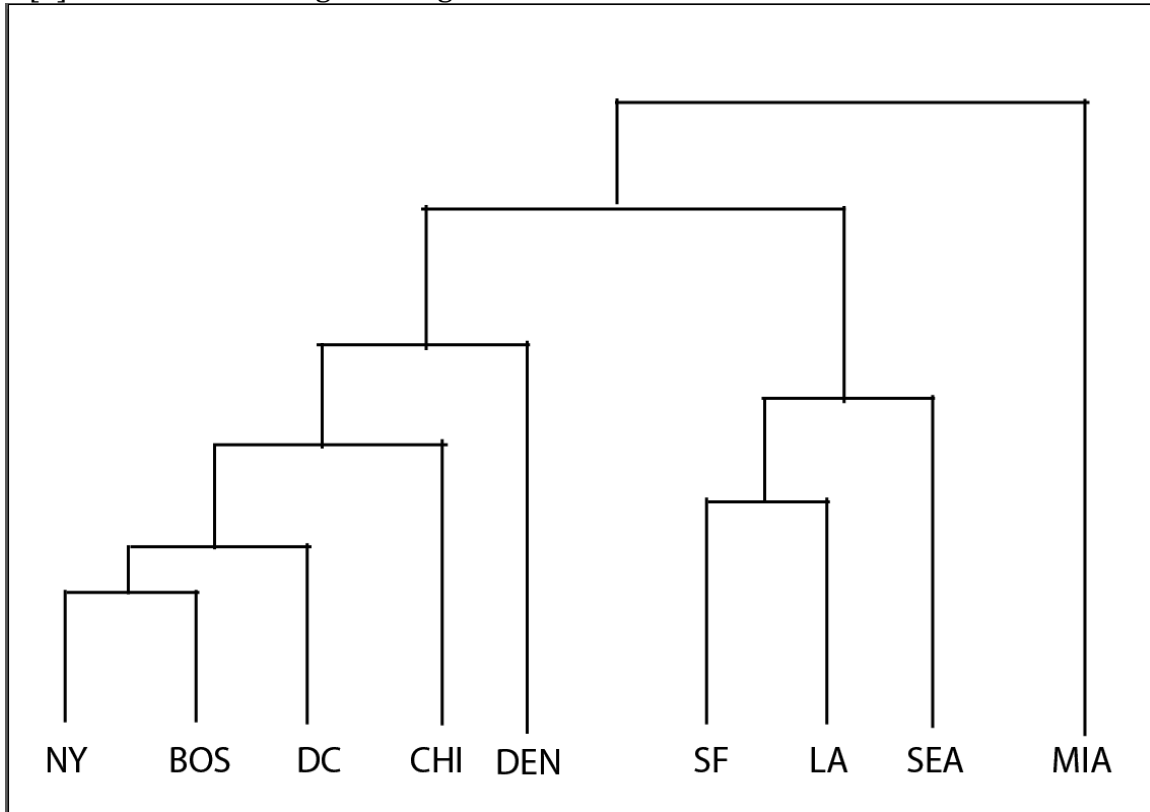


Problem 1. [20] Unsupervised Learning by Clustering

Consider the following information about distances between pairs of 9 U.S. cities:

{a} [10] Perform {manually} hierarchical agglomerative clustering using single-linkage and the above data.

i. [8] Show the resulting dendrogram:



ii. [2] What clusters of cities are created if you want 3 clusters?

Answer: {BOS-NY-DC-CHI-DEN} {MIA} {SF-LA-SEA}

{b} [10] Show the results of one iteration of k-means clustering assuming  $k = 2$  and the initial cluster centers are defined as  $c1 = \{38.0, 103.0\}$  and  $c2 = \{30.0, 78.0\}$ .

(TABLE BEFORE ALG, RUN)

CITY	X	Y	C1 Euclidian	C2 Euclidian	Dist from C1 (abs val)	Dist from C2 (abs val)	Cluster:
BOS	42.4	71.1	71.054	55.926	20.870	9.748	c2
NY	41.7	74	72.583	57.856	19.341	8.219	c2
DC	38.9	77	75.343	61.283	16.581	5.459	c2
MIA	25.8	80.2	84.761	72.549	7.163	3.958	c2
CHI	41.9	87.7	79.515	66.345	12.409	1.287	c2
SEA	47.6	122.3	99.021	88.792	7.097	18.218	c1
SF	37.8	122.4	106.687	97.268	14.763	25.885	c1
LA	34.1	118.2	106.291	96.834	14.367	25.489	c1
DEN	39.7	105	92.136	81.044	0.212	11.334	c1
c1	38	103	91.924	80.802	0.000	11.122	c1
c2	30	78	80.802	67.882	11.122	0.000	c2

(TABLE AFTER ALG, RUN)

CITY	X	Y	C1 Euclidian	C2 Euclidian	Dist from C1 (abs val)	Dist from C2 (abs val)	Closest cluster:
BOS	42.4	71.1	82.175	50.047	26.720	5.061	c2
NY	41.7	74	83.500	52.195	25.394	3.735	c2
DC	38.9	77	85.910	55.970	22.984	1.325	c2
MIA	25.8	80.2	94.278	68.120	14.616	7.043	c2
CHI	41.9	87.7	89.592	61.471	19.303	2.356	c2
SEA	47.6	122.3	107.280	85.212	1.614	20.045	c1
SF	37.8	122.4	114.395	94.011	5.500	27.159	c1
LA	34.1	118.2	114.025	93.562	5.131	26.790	c1
DEN	39.7	105	100.961	77.104	7.934	13.726	c1
c1	39	116	108.894	87.235	0.000	21.659	c1
c2	37	78	87.235	57.983	21.659	0.000	c2

i. [3] Give the list of cities in each of the initial 2 clusters.

BOS	c2
NY	c2
DC	c2
MIA	c2
CHI	c2
SEA	c1
SF	c1
LA	c1
DEN	c1

ii. [4] Give the coordinates of the new cluster centers. {Use Euclidean distance between {latitude, longitude} coordinates.}

Centroid 1 = (39, 116)

Centroid 2 = (37, 78)

iii. [3] Give the list of cities in the 2 clusters based on the new cluster centers computed in {ii}.

(Centroid1)	Cluster 1:
SEA	(47.6, 122.3)
SF	(37.8, 122.4)
LA	(34.1, 118.2)
DEN	(39.7, 105.0)

(Centroid2)	Cluster 2:
BOS	(42.4, 71.1)
NY	(41.7, 74.0)
DC	(38.9, 77.0)
MIA	(25.8, 80.2)
CHI	(41.9, 87.7)