CS 540: Introduction to Artificial Intelligence

Homework Assignment #3 Solutions

Problem 1. [20] Unsupervised Learning by Clustering

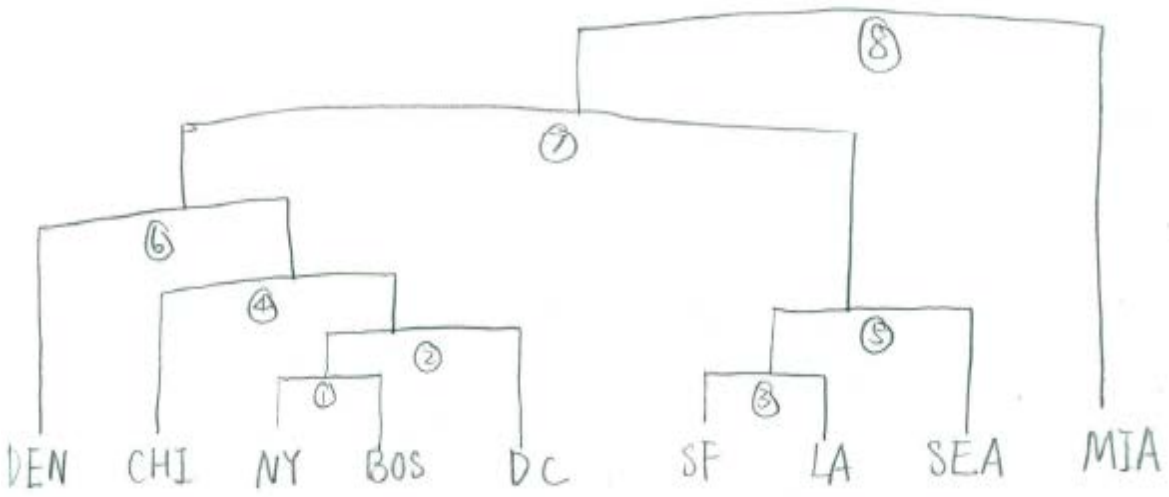
Consider the following information about distances between pairs of 9 U.S. cities:

	BOS	NY	DC	MIA	CHI	SEA	SF	LA	DEN
BOS	0	206	429	1504	963	2976	3095	2979	1949
NY	206	0	233	1308	802	2815	2934	2786	1771
DC	429	233	0	1075	671	2684	2799	2631	1616
MIA	1504	1308	1075	0	1329	3273	3053	2687	2037
CHI	963	802	671	1329	0	2013	2142	2054	996
SEA	2976	2815	2684	3273	2013	0	808	1131	1307
SF	3095	2934	2799	3053	2142	808	0	379	1235
LA	2979	2786	2631	2687	2054	1131	379	0	1059
DEN	1949	1771	1616	2037	996	1307	1235	1059	0

The (latitude, longitude) locations of these cities are: 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), SEA (47.6, 122.3), SF (37.8, 122.4), LA (34.1, 118.2), and DEN (39.7, 105.0).

(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?

```
Cluster 1: DEN, CHI, NY, BOS, DC
Cluster 2: LA, SF, SEA
Cluster 3: MIA
```

(b) [10] Show the results of one iteration of ***k*-means clustering** assuming $k = 2$ and the initial cluster centers are defined as c_1 = (38.0, 103.0) and c_2 = (30.0, 78.0)

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

```
Cluster 1: SEA, SF, LA, DEN
Cluster 2: BOS, NY, DC, MIA, CHI
```

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

```
Center of Cluster 1: (39.8, 116.98)
Center of Cluster 2: (38.14,78)
```

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

```
Cluster 1: SEA, SF, LA, DEN
Cluster 2: BOS, NY, DC, MIA, CHI
```