Gustavo Padilha Polleti
9345193

Homework 3

PCS 5024 April 29, 2019

Implement *k*-means manually

1. Center of the first cluster after one interaction. **Answer: [5.171, 3.171]**
2. Center of the second cluster after two interactions. **Answer: [5.3, 4.0]**
3. Center of the third cluster when it converges. **Answer: [6.2, 3.025]**
4. Number of interactions until it converges. **Answer: 3**

X_i	Clusters		
	red	green	blue
0	0.300	0.860	0.632
1	1.628	2.154	1.903
2	0.400	0.985	0.361
3	1.500	1.965	1.811
4	1.221	1.208	1.562
5	1.217	1.746	1.500
6	1.304	1.803	1.603
7	0.510	0.608	0.224
8	1.253	1.503	1.612
9	0.283	0.922	0.500

X_i	cluster
0	red
1	red
2	blue
3	red
4	green
5	red
6	red
7	blue
8	red
9	red

Clusters	Coordinates	
	x-axis	y-axis
Red	5.171	3.171
Green	5.5	4.2
Blue	6.45	2.95

Table 1: Distances, Clusters and Centroids, respectively, for first Iteration

X_i	Clusters		
	red	green	blue
0	0.729	1.077	0.604
1	0.633	1.581	1.851
2	1.094	1.565	0.292
3	0.472	1.281	1.768
4	1.080	0.000	1.570
5	0.242	1.300	1.451
6	0.281	1.253	1.557
7	1.530	1.628	0.292
8	0.633	0.566	1.595
9	0.846	1.300	0.453

X_i	cluster
0	blue
1	red
2	blue
3	red
4	green
5	red
6	red
7	blue
8	green
9	blue

Clusters	Coordinates	
	x-axis	y-axis
Red	4.8	3.05
Green	5.3	4.0
Blue	6.2	3.025

Table 2: Distances, Clusters and Centroids, respectively, for second iteration

X_i	Clusters		
	red	green	blue
0	1.110	1.000	0.347
1	0.250	1.304	1.605
2	1.422	1.500	0.225
3	0.180	1.000	1.510
4	1.346	0.283	1.368
5	0.206	1.044	1.200
6	0.112	0.985	1.302
7	1.901	1.664	0.506
8	0.808	0.283	1.346
9	1.201	1.221	0.202

X_i	cluster
0	blue
1	red
2	blue
3	red
4	green
5	red
6	red
7	blue
8	green
9	blue

Clusters	Coordinates	
	x-axis	y-axis
Red	4.8	3.05
Green	5.3	4.0
Blue	6.2	3.025

Table 3: Distances, Clusters and Centroids, respectively, for third and last iteration

Application of *k*-means

1. Dataset A. **A2**
2. Dataset B. **B2**
3. Dataset C. **C2**
4. Dataset D. **D1**
5. Dataset E. **E2**
6. Dataset F. **F2**

Hierarchical Clustering

1. Complete link. **2.1095**
2. Single link. **0.9220**
3. Average link. **1.4129**
4. Most robust to noise. **Average link**