Homework 4

**Task 1**

1. After the first interaction the centroid of the clusters are (2.0, 8.0) and (5.875, 3.625).

|  |  |
| --- | --- |
| Cluster 1 | Cluster 2 |
| (2, 8)  (2, 8) | (2, 3)  (2, 3)  (7, 3)  (7, 3)  (7, 4)  (7, 4)  (8, 5)  (7, 4) |

Graphical user interface

Description automatically generated with low confidence

1. After the first interaction the centroid of the clusters are (2.0, 8.0) and (5.875, 3.625).

|  |  |
| --- | --- |
| Cluster 1 | Cluster 2 |
| (2, 8)  (2, 8) | (2, 3)  (2, 3)  (7, 3)  (7, 3)  (7, 4)  (7, 4)  (8, 5)  (7, 4) |

Graphical user interface

Description automatically generated with low confidence

1. After the first interaction the centroid of the clusters are (2.0, 5.5) and (7.166666666666667, 3.8333333333333335).

|  |  |
| --- | --- |
| Cluster 1 | Cluster 2 |
| (2, 3)  (2, 3)  (2, 8)  (2, 8) | (7, 3)  (7, 3)  (7, 4)  (7, 4)  (8, 5)  (7, 4) |

A picture containing scatter chart

Description automatically generated

1. After the first interaction the centroid of the clusters are (5.571428571428571, 3.4285714285714284) and (4.0, 7.0).

|  |  |
| --- | --- |
| Cluster 1 | Cluster 2 |
| (2, 3)  (2, 3)  (7, 3)  (7, 3)  (7, 4)  (7, 4)  (8, 5)  (7, 4) | (2, 8)  (2, 8) |

A picture containing scatter chart

Description automatically generated

**Task 2**

1. The SSE
2. 2
3. I set up the
4. 5
5. It takes a long time to compete this analysis.

**Task 3**

1. The distance between the two farthest members, which are (4.6,2.9) and (6.7,3.1), is 2.1095.
2. The distance between the two closest members, which are (5.0,3.0) and (5.9,3.2) is 0.9220.
3. The average distance between all pairs is 1.4129. This can be done by adding all of the distances which equals 22.6061 and then divide it by 16, which is the total number of different distances

Work

A: (4.7, 3.2) E: (5.9, 3.2)

B: (4.9, 3.1) F: (6.7, 3.1)

C: (5.0, 3.0) G: (6.0, 3.0)

D: (4.6, 2.9) H: (6.2, 2.8)

AE:

AF:

AG: *.*3153

AH:

BE:

BF:

BG:

BH:

CE:

CF:

CG:

CH:

DE:

DF:

DG:

DH:

**Additional Questions**

1. I spend around 10 hours completing this assignment.
2. One of the aspects that I found the most challenging was creating the k means algorithm from scratch.
3. I really enjoyed the assignment since it helped me polish my understanding on the k means machine learning algorithm and my coding skills. There is nothing that I would have changed.

Github:

Please go to the folder named Homework 4 to see the code for task 1 and task 2.