Name: CHAN King Yeung

SID: 1155119394

SEEM2460 Assignment 3

Question 1

1. We constructed data matrix as follow,

Then we calculated the distances based on the formula of Euclidean distance

for

We used K-means algorithm and obtained the following distance matrices

with

Customer A, B, and F are assigned into cluster 1; Customer C, D, E, G, and I are assigned into cluster 2; Customer H is assigned into cluster 3;

*# note here: customer C and G have same distance between cluster 2 and 3, but we assigned them into cluster 2*

We computed new centroids and repeated the above steps.

with

Customer G no longer belongs to cluster 2 and being assigned into cluster 3

with

Customer C no longer belongs to cluster 2 and being assigned into cluster 3

with

We stopped here since there is no more new assignment to cluster.

Cluster 1: customer A, B, and F; Cluster 2: customer D, E, and I; Cluster 3: customer H, G, and C

1. Partial source codes are pasted here for your convenience

|  |
| --- |
| # Step-1: Assign 3 initial centroids.  centroids = {  # Please specify three centroids below  1: [1, 2], 2: [4, 4], 3: [6, 4],  }  # Step-2: Continue until all assigned categories don't change any more  df = assignment(df, centroids)  while True:  closest\_centroids = df['closest'].copy(deep=True)  # Please determine the order of update() and assignment() below, Hint: 2 lines of codes  df = assignment(df, centroids)  centroids = update(centroids)  if closest\_centroids.equals(df['closest']):  break |

My source code for your reference

<https://colab.research.google.com/drive/1Wr3nHm4VQkbkvjWil3yIPEL-8v5yJIhx>