**502571-3 2nd Trimester 2022/2023 HW#5**

# Topics: K-Means Algo Sections: 2233 & 3827

# Due Date: Monday 13/02/2023 – 2:00 pm

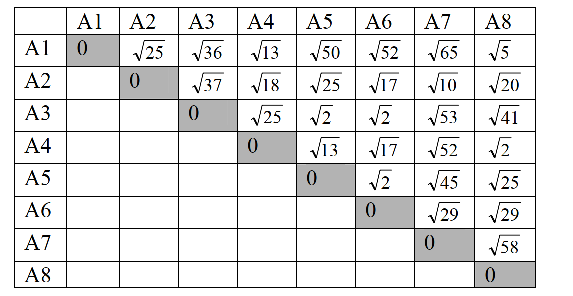
# \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Use the K-Means algorithm and Euclidean distance to cluster the following 8 instances into 3 clusters: A1 = (2, 10); A2 = (2, 5); A3 = (8, 4); A4 = (5, 8); A5 = (7, 5); A6 = (6, 4); A7 = (1, 2); A8 = (4, 9).

Suppose that the initial seeds (center of each cluster or centroid) are: A1, A4, and A7. Run the K-Means algorithm until the convergence (the algorithm is stopped).

Solution:

﻿The distance matrix based on the Euclidean distance is given below:



﻿d(a,b) denotes the Eucledian distance between a and b. It is obtained directly from the distance matrix or

calculated as follows: d(a,b)=sqrt((xb-xa)2+(yb-ya)2))

seed1=A1=(2,10), seed2=A4=(5,8), seed3=A7=(1,2)

Iteration 1 – start:

A1:

d(A1, seed1)=0 as A1 is seed1

d(A1, seed2)= 13 >0

d(A1, seed3)= 65 >0

* A1 ∈ cluster1

A2:

d(A2,seed1)= 25 = 5

d(A2, seed2)= 18 = 4.24

d(A2, seed3)= 10 = 3.16 smaller

* A2 ∈ cluster3

A3:

d(A3, seed1)= 36 = 6

d(A3, seed2)= 25 = 5 smaller

d(A3, seed3)= 53 = 7.28

* A3 ∈ cluster2

A4:

d(A4, seed1)= 13

d(A4, seed2)=0 as A4 is seed2

d(A4, seed3)= 52 >0

* A4 ∈ cluster2

A5:

d(A5, seed1)= 50 = 7.07

A6:

d(A6, seed1)= 52 = 7.21d(A5, seed2)= 13 = 3.60 smaller

d(A5, seed3)= 45 = 6.70

* A5 ∈ cluster2

d(A6, seed2)= 17 = 4.12 smaller

d(A6, seed3)= 29 = 5.38

* A6 ∈ cluster2

A7:

d(A7, seed1)= 65 >0

d(A7, seed2)= 52 >0

d(A7, seed3)=0 as A7 is seed3

* A7 ∈ cluster3

A8:

d(A8, seed1)= 5

d(A8, seed2)= 2 smaller

d(A8, seed3)= 58

A8 ∈ cluster2

new clusters: 1: {A1}, 2: {A3, A4, A5, A6, A8}, 3: {A2, A7}

End of Iteration 1

=============================

Iteration 2

centers of the new clusters:

C1= (2, 10), C2= ((8+5+7+6+4)/5, (4+8+5+4+9)/5) = (6, 6), C3= ((2+1)/2, (5+2)/2) = (1.5, 3.5)

We would need two more iterations. After the 2nd iteration the results would be:

1: {A1, A8}, 2: {A3, A4, A5, A6}, 3: {A2, A7}

with centers C1=(3, 9.5), C2=(6.5, 5.25) and C3=(1.5, 3.5).

End of Iteration 1

=============================

Iteration 3

After the 3rd epoch, the results would be:

1: {A1, A4, A8}, 2: {A3, A5, A6}, 3: {A2, A7}

with centers C1=(3.66, 9), C2=(7, 4.33) and C3=(1.5, 3.5).

When we perform iteration 4, we will obtain the same clusters with the same seeds (centroids) as in iteration 3. So, we stop the algorithm.