**ELG5255 Applied Machine Learning**

REPORT of: Group Assignment 3 (Group-18)

Part 1: Calculations

Use the k-means algorithm and Euclidean distance to cluster the following 5 data points into 2 clusters: A1=(2,5), A2=(5,8), A3=(7,5), A4=(1,2), A5=(4,9). Suppose that the initial centroids (centers of each cluster) are A2 and A4. Using k-means, cluster the 5 points and show the followings for one iteration only:

1. Show step-by-step the performed calculations to cluster the 5 points.
2. Draw a 10 by 10 space with all the clustered 5 points and the coordinates of the new centroids
3. Calculate the silhouette score and WSS score.

**Solution:**

1. Suppose d (a, b) denotes the Euclidean distance between a and b.

It is obtained directly from the distance matrix or calculated as follows:

* d (a, b) = sqrt (( - )2 + ( - )2)

One iteration-start:

|  |  |  |  |
| --- | --- | --- | --- |
| A | Data | Centroid A2(5, 8) | Centroid A4(1, 2) |
| 1 | (2, 5) | d(A1,A2) = = **4.24** | d(A1,A4) = = **3.16** |
| 2 | (5, 8) | d(A2,A2) = = **0** | d(A2,A4) = = **7.21** |
| 3 | (7, 5) | d(A3,A2) = = **3.16** | d(A3,A4) = = **6.71** |
| 4 | (1, 2) | d(A4,A2) = = **7.21** | d(A4,A4) = = **0** |
| 5 | (4, 9) | d(A5,A2) = = **1.41** | d(A5,A4) = = **7.62** |

One iteration-end

In Centroid A2 (5, 8) € Cluster1 (C1), A4 (1, 2) € Cluster2 (C2):

|  |  |
| --- | --- |
| d (A1,A2) = 4.24, d (A1,A4) = 3.16  d (A1,A4) is less than d (A1,A2)  A1 € C2 | d (A4,A2) = 7.21, d (A4,A4) = 0  d (A4,A4) is less than d (A4,A2)  A4 € C2 |
| d (A2,A2) = 0, d (A2,A4) = 7.21  d (A2,A2) is less than d (A2,A4)  A2 € C1 | d (A5,A2) = 1.41, d (A5,A4) = 7.62  d (A5,A2) is less than d (A5,A4)  A5 € C1 |
| d (A3,A2) = 3.16, d (A3,A4) = 6.71  d (A3,A2) is less than d (A3,A4)  A3 € C1 |  |

C1 🡪 {A2, A3, A5}, C2 🡪 {A1, A4}

b)

|  |  |
| --- | --- |
|  |  |
|  |  |

c) C1 🡪 {A2, A3, A5}, C2 🡪 {A1, A4}

In C1 (A2)

**Cohesion**: d (A3, A2) = **3.16**, d (A5, A2) = **1.41**

**a (A2) = = 2.51**

**Separation:** d (A1, A2) = **4.24** d (A4, A2) = **7.21**

**b (A2) = = 5.73**

In C1 (A3)

**Cohesion**: d (A2, A3) = **3.61**, d (A5, A3) = = **5**

**a (A3) = = 4.31**

**Separation:** d (A1, A3) = = **5**, d (A4, A3) = = **6.71**

**b (A3) = = 5.86**

In C1 (A5)

**Cohesion**: d (A2, A5) = **1.41**, d (A3, A5) = = **5**

**a (A3) = = 3.21**

**Separation:** d (A1, A5) = = **4.47**, d (A4, A5) = 7.**62**

**b (A3) = = 6.05**

In C2 (A1)

**Cohesion**: d (A4, A1) = 3.16

**a (A3) = = 3.16**

**Separation:** d (A2, A1) = **4.24**, d (A3, A1) = = **5**

d (A5, A1) = = **4.47**

**b (A1) = = 4.57**

In C2 (A4)

**Cohesion**: d (A1, A4) = 3.16

**a (A4) = = 3.16**

**Separation:** d (A2, A4) = **7.21**, d (A3, A4) = **6.71**

d (A5, A4) = **7.62**

**b (A4) = = 7.18**

a(A1) = 3.16, b(A1) = 4.57

a(A2) = 2.51, b(A2) = 5.73

a(A3) = 4.31, b(A3) = 5.86

a(A4) = 3.16, b(A4) = 7.18

a(A5) = 3.21, b(A5) = 6.05

S (i) =

S (A1) =  **= = 0.31**

S (A2) =  **= = 0.56**

S (A3) =  **= = 0.26**

S (A4) =  **= = 0.56**

S (A5) =  **= = 0.47**

**The silhouette score = = 0.43**

C1 🡪 {A2, A3, A5} {(5, 8), (7, 5), (4, 9)}

Xc1 =  **= 5.33, Yc1 = = 7.33**

New Centroid (Cc1)= (5.33, 7.33)

C2 🡪 {A1, A4} {(2, 5), (1, 2)}

Xc2 =  **= 1.5, Yc2 = = 3.5**

New Centroid (Cc2)= (1.5, 3.5)

WSS =

WSS = + + +

+ + + +

+ + = 18.33.

**WSS score = 18.33**