# COMP 472: Artificial Intelligence

## k-means

## *Solutions*

**Question 1** Consider the following data set with two attributes for each of seven individuals.

| | a1 | a2 |
|---|---|---|
| Data1 | 1.0 | 1.0 |
| Data2 | 1.5 | 2.0 |
| Data3 | 3.0 | 4.0 |
| Data4 | 5.0 | 7.0 |
| Data5 | 3.5 | 5.0 |
| Data6 | 4.5 | 5.0 |
| Data7 | 3.5 | 4.5 |

(a) The data set is to be grouped into two clusters. Initialize the clusters using Data1 and Data4 as initial centroids to the two clusters. That is, allocate the remaining individual to one of the two clusters using the Euclidean distance.

| | Individuals | Centroid |
|---|---|---|
| Group 1 | 1 | (1.0, 1.0) |
| Group 2 | 4 | (5.0, 7.0) |

*For Data2:*
*Distance to Centroid 1:* $\sqrt{(1.5-1.0)^2+(2.0-1.0)^2} \approx 1.1$
*Distance to Centroid 2:* $\sqrt{(1.5-5.0)^2+(2.0-7.0)^2} \approx 6.1$

| | Distance to Centroid 1 | Distance to Centroid 2 |
|---|---|---|
| Data1 | 0.0 | 7.2 |
| Data2 | 1.1 | 6.1 |
| Data3 | 3.6 | 3.6 |
| Data4 | 7.2 | 0.0 |
| Data5 | 4.7 | 2.5 |
| Data6 | 5.3 | 2.1 |
| Data7 | 4.3 | 2.9 |

(b) Recalculate the centroids based on the current partition, reassign the individuals based on the new centroids. Which individuals (if any) changed clusters as a result?

*For Group 1:*

$$\frac{1.0 + 1.5 + 3.0}{3} = 1.83$$

$$\frac{1.0 + 2.0 + 4.0}{3} \approx 2.33$$

| | Individuals | Centroid |
|---|---|---|
| Group 1 | 1, 2, 3 | (1.83, 2.33) |
| Group 2 | 4, 5, 6, 7 | (4.125, 5.375) |

| | Distance to Centroid 1 | Distance to Centroid 2 |
|---|---|---|
| Data1 | 1.6 | 5.4 |
| Data2 | 0.5 | 4.3 |
| Data3 | 2.0 | 1.8 |
| Data4 | 5.6 | 1.9 |
| Data5 | 3.1 | 0.7 |
| Data6 | 3.8 | 0.5 |
| Data7 | 2.7 | 1.1 |

*Data3 changed from Group 1 to Group 2. All the other individuals remained in the same cluster.*