# CHA2555 Artificial Intelligence

# Practical 18: Clustering 1 - Solution

**Activity 1:**

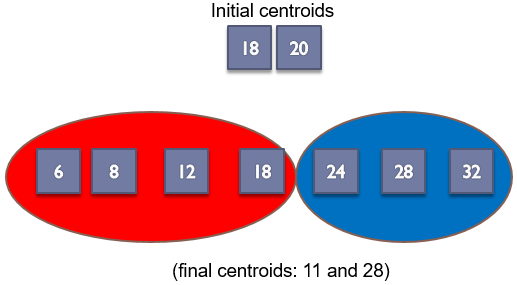
Given the following 1D data: {6, 8, 18, 28, 12, 32, 24}, choose your own initial centroids and perform *k*-means. E.g., you may start with a pair of initial centroids (18, 20), perform the k-means procedures until it converges. The following table can help you find the final centroids following the k-means algorithm, and you can create more when you work on your initial centroids. Try other centroids of your own choice (Stay within the range 6 to 32, round up/down averages if needed) and compare if the final results are consistent.

Iterate until converged:

1. Compute distance from all data points to all *k* centroids
2. For each **data point**, assign it to the cluster whose current centroid it is nearest
3. For each **centroid**, compute the average (mean) of all points assigned to it
4. Replace the *k* centroids with the new averages

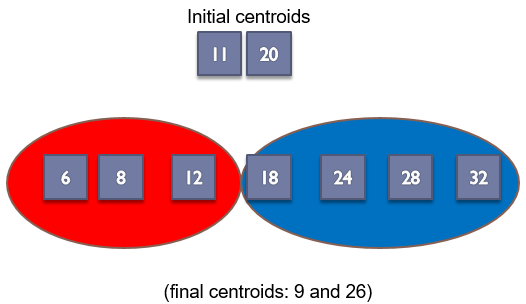
With initial centroids (18, 20)

|  |  |  |
| --- | --- | --- |
| Centroids | Number of iterations | Cluster assignment |
| 18, 20 | 1 | cluster 1 (18): 6, 8, 12, 18  cluster 2 (20): 24, 28, 32 |
| 11, 28 | 2 | cluster 1 (11): 6, 8, 12, 18  cluster 2 (28): 24, 28, 32 |
| 11, 28 (converged, as the centroids do not move any more) | 3 | cluster 1 (11): 6, 8, 12, 18  cluster 2 (28): 24, 28, 32 |



With initial centroids (11, 20)

|  |  |  |
| --- | --- | --- |
| Centroids | Number of iterations | Cluster assignment |
| 11, 20 | 1 | cluster 1 (11): 6, 8, 12  cluster 2 (20): 18, 24, 28, 32 |
| 8.6, 25.5 | 2 | cluster 1 (8.6): 6, 8, 12  cluster 2 (25.5): 18, 24, 28, 32 |
| 8.6, 25.5 (converged) | 3 | cluster 1 (8.6): 6, 8, 12  cluster 2 (25.5): 18, 24, 28, 32 |



As we saw, the final results can vary based on choice of initial centroids, which may be controlled by setting seeds for the program. Some seeds can result in poor convergence rate, or convergence to sub-optimal clustering. So, we need a way of choosing them, .e.g, trying out multiple starting points, selecting good seeds with another method.