Navigatelabsai Assignment 1

Part-3 K-Means Clustering Algorithm

November 9, 2024

Step 1: Dataset Creation

We'll use six 2D points:

$$\text{Data points} = \{(1,2), (1,4), (3,2), (5,8), (8,8), (8,10)\}$$

Let's initialize the centroids randomly:

$$C_1 = (1,2), \quad C_2 = (5,8)$$

Step 2: Iteration 1

Step 2.1: Calculate the Distance of Each Point to Both Centroids

We'll use the Euclidean distance formula:

$$\text{distance} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

Distance from each point to $C_1 = (1,2)$:

$$d((1,2), C_1) = 0$$

$$d((1,4), C_1) = \sqrt{(1-1)^2 + (4-2)^2} = 2$$

$$d((3,2), C_1) = \sqrt{(3-1)^2 + (2-2)^2} = 2$$

$$d((5,8), C_1) = \sqrt{(5-1)^2 + (8-2)^2} = \sqrt{52} \approx 7.21$$

$$d((8,8), C_1) = \sqrt{(8-1)^2 + (8-2)^2} = \sqrt{85} \approx 9.22$$

$$d((8,10), C_1) = \sqrt{(8-1)^2 + (10-2)^2} = \sqrt{113} \approx 10.63$$

Distance from each point to $C_2 = (5,8)$:

$$d((1,2), C_2) = \sqrt{(1-5)^2 + (2-8)^2} = \sqrt{52} \approx 7.21$$

$$d((1,4), C_2) = \sqrt{(1-5)^2 + (4-8)^2} = \sqrt{32} \approx 5.66$$

$$d((3,2), C_2) = \sqrt{(3-5)^2 + (2-8)^2} = \sqrt{40} \approx 6.32$$

$$d((5,8), C_2) = 0$$

$$d((8,8), C_2) = \sqrt{(8-5)^2 + (8-8)^2} = 3$$

$$d((8,10), C_2) = \sqrt{(8-5)^2 + (10-8)^2} = \sqrt{13} \approx 3.61$$

Step 2.2: Assign Points to the Nearest Cluster

Based on the distances calculated above, assign each point to the cluster with the nearest centroid.

Point	$d(C_1)$	$d(C_2)$	Cluster
(1,2)	0.00	7.21	C_1
(1,4)	2.00	5.66	C_1
(3,2)	2.00	6.32	C_1
(5,8)	7.21	0.00	C_2
(8,8)	9.22	3.00	C_2
(8,10)	10.63	3.61	C_2

So, after the first iteration, the clusters are:

$$\text{Cluster 1: } (1,2), (1,4), (3,2)$$

$$\text{Cluster 2: } (5,8), (8,8), (8,10)$$

Step 2.3: Update Centroids

Compute the new centroid for each cluster by taking the average of all points in that cluster.

New centroid C_1:

$$C_1 = \left(\frac{1+1+3}{3}, \frac{2+4+2}{3} \right) = (1.67, 2.67)$$

New centroid C_2:

$$C_2 = \left(\frac{5+8+8}{3}, \frac{8+8+10}{3} \right) = (7.00, 8.67)$$

Step 3: Iteration 2

Step 3.1: Recalculate Distances and Reassign Points

Using the updated centroids $C_1 = (1.67, 2.67)$ and $C_2 = (7.00, 8.67)$, repeat the distance calculations and assignments as in the previous iteration.

After two or more iterations, the centroids will stabilize, and the assignments will no longer change, meaning the algorithm has converged.