k-Means Homework

k-Means Clustering Using Manhattan Distance

Data Set:

Pattern	x	y
a	0.9	0.8
b	0.2	0.2
c	0.7	0.6
d	-0.1	-0.6
e	0.5	0.5

Parameters:

- Number of clusters (k): 2
- Initial centroids: First two instances
  - **Centroid 1:** a $(0.9, 0.8)$
  - **Centroid 2:** b $(0.2, 0.2)$

Iteration 1

Assignment Step

Calculate Manhattan distance from each instance to each centroid.

Manhattan Distance Formula:

$$d((x_1, y_1), (x_2, y_2)) = |x_1 - x_2| + |y_1 - y_2|$$

Distances to Centroid 1 (a):

- **a:** $d(a, a) = |0.9 - 0.9| + |0.8 - 0.8| = 0 + 0 = 0$
- **b:** $d(b, a) = |0.2 - 0.9| + |0.2 - 0.8| = 0.7 + 0.6 = 1.3$

- **c:** $d(c, a) = |0.7 - 0.9| + |0.6 - 0.8| = 0.2 + 0.2 = 0.4$
- **d:** $d(d, a) = |-0.1 - 0.9| + |-0.6 - 0.8| = 1.0 + 1.4 = 2.4$
- **e:** $d(e, a) = |0.5 - 0.9| + |0.5 - 0.8| = 0.4 + 0.3 = 0.7$

Distances to Centroid 2 (b):

- **a:** $d(a, b) = |0.9 - 0.2| + |0.8 - 0.2| = 0.7 + 0.6 = 1.3$
- **b:** $d(b, b) = |0.2 - 0.2| + |0.2 - 0.2| = 0 + 0 = 0$
- **c:** $d(c, b) = |0.7 - 0.2| + |0.6 - 0.2| = 0.5 + 0.4 = 0.9$
- **d:** $d(d, b) = |-0.1 - 0.2| + |-0.6 - 0.2| = 0.3 + 0.8 = 1.1$
- **e:** $d(e, b) = |0.5 - 0.2| + |0.5 - 0.2| = 0.3 + 0.3 = 0.6$

Cluster Assignments:

Pattern	Distance to Centroid 1	Distance to Centroid 2	Assigned Cluster
a	0	1.3	Cluster 1
b	1.3	0	Cluster 2
c	0.4	0.9	Cluster 1
d	2.4	1.1	Cluster 2
e	0.7	0.6	Cluster 2

Update Step

Recalculate centroids based on current cluster assignments.

Cluster 1: {a, c}

- **Centroid 1:**
  - $x = \frac{0.9+0.7}{2} = 0.8$
  - $y = \frac{0.8+0.6}{2} = 0.7$
  - **New Centroid 1:** $(0.8, 0.7)$

Cluster 2: {b, d, e}

- **Centroid 2:**
  - $x = \frac{0.2+(-0.1)+0.5}{3} = \frac{0.6}{3} = 0.2$
  - $y = \frac{0.2+(-0.6)+0.5}{3} = \frac{0.1}{3} \approx 0.0333$

- **New Centroid 2:** $(0.2, 0.0333)$

Iteration 2

Assignment Step

Calculate Manhattan distance from each instance to updated centroids.

Distances to Centroid 1 $(0.8, 0.7)$**:**

- **a:** $d(a, \text{Centroid 1}) = |0.9 - 0.8| + |0.8 - 0.7| = 0.1 + 0.1 = 0.2$
- **b:** $d(b, \text{Centroid 1}) = |0.2 - 0.8| + |0.2 - 0.7| = 0.6 + 0.5 = 1.1$
- **c:** $d(c, \text{Centroid 1}) = |0.7 - 0.8| + |0.6 - 0.7| = 0.1 + 0.1 = 0.2$
- **d:** $d(d, \text{Centroid 1}) = |-0.1 - 0.8| + |-0.6 - 0.7| = 0.9 + 1.3 = 2.2$
- **e:** $d(e, \text{Centroid 1}) = |0.5 - 0.8| + |0.5 - 0.7| = 0.3 + 0.2 = 0.5$

Distances to Centroid 2 $(0.2, 0.0333)$**:**

- **a:** $d(a, \text{Centroid 2}) = |0.9 - 0.2| + |0.8 - 0.0333| = 0.7 + 0.7667 = 1.4667$
- **b:** $d(b, \text{Centroid 2}) = |0.2 - 0.2| + |0.2 - 0.0333| = 0 + 0.1667 = 0.1667$
- **c:** $d(c, \text{Centroid 2}) = |0.7 - 0.2| + |0.6 - 0.0333| = 0.5 + 0.5667 = 1.0667$
- **d:** $d(d, \text{Centroid 2}) = |-0.1 - 0.2| + |-0.6 - 0.0333| = 0.3 + 0.6333 = 0.9333$
- **e:** $d(e, \text{Centroid 2}) = |0.5 - 0.2| + |0.5 - 0.0333| = 0.3 + 0.4667 = 0.7667$

Cluster Assignments:

Pattern	Distance to Centroid 1	Distance to Centroid 2	Assigned Cluster
a	0.2	1.4667	Cluster 1
b	1.1	0.1667	Cluster 2
c	0.2	1.0667	Cluster 1
d	2.2	0.9333	Cluster 2
e	0.5	0.7667	Cluster 1

Update Step

Recalculate centroids based on current cluster assignments.

Cluster 1: {a, c, e}

- **Centroid 1:**
  - $x = \frac{0.9+0.7+0.5}{3} = \frac{2.1}{3} = 0.7$
  - $y = \frac{0.8+0.6+0.5}{3} = \frac{1.9}{3} \approx 0.6333$
  - **New Centroid 1:** $(0.7, 0.6333)$

Cluster 2: {b, d}

- **Centroid 2:**
  - $x = \frac{0.2+(-0.1)}{2} = \frac{0.1}{2} = 0.05$
  - $y = \frac{0.2+(-0.6)}{2} = \frac{-0.4}{2} = -0.2$
  - **New Centroid 2:** $(0.05, -0.2)$

Summary After Two Iterations

Final Centroids:

- **Centroid 1:** $(0.7, 0.6333)$
- **Centroid 2:** $(0.05, -0.2)$

Cluster Assignments:

- **Cluster 1:** {a, c, e}
- **Cluster 2:** {b, d}

Closest Instances to Each Centroid:

- **Centroid 1:** Instances **a** and **c** are equally closest (distance = 0.2), and **e** is also close (distance = 0.5).
- **Centroid 2:** Instance **b** is the closest (distance = 0.1667), followed by **d** (distance = 0.9333).
