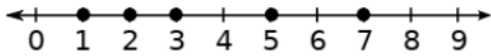


Doris Chen

Lesson 22 (K-Means Algorithm) Apply K-Means clustering to the five points shown below with the number of clusters set equal to $K = 2$. The locations of the initial centroids c_1 and c_2 are given below. List C_1 or C_2 under the numbers 1, 2, 3, 5 and 7 to indicate which cluster each number belongs to.



(a) Complete the two tables below.

centroid	c_1	c_2
iteration 0	1	4.25
1	1.5	5
2	2	6
3	2	6

point	1	2	3	5	7
iteration 0	c1	c1	c2	c2	c2
1	c1	c1	c1	c2	c2
2	c1	c1	c1	c2	c2
3	c1	c1	c1	c2	c2

(b) Compute the total SSE for the two initial clusters given above and the total SSE for the final two clusters the K-Means algorithm converges to. Which is smaller? Note that SSE (Sum of Square Error) is defined to be

$$SSE = \sum_{i=1}^K \sum_{x \in C_i} \text{dist}(x, c_i)^2$$

where K equals the number of clusters, C_i is the i th cluster, and $\text{dist}(x, c_i)^2$ is the squared distance between the data point x and the centroid of the i th cluster, c_i .

$$C1 = (1-1)^2 + (2-1)^2$$

$$C2 = (3-4.25)^2 + (5-4.25)^2 + (7-4.25)^2$$

$$\text{Initial SSE} = C1 + C2 = 10.6875$$

$$\text{Final SSE} =$$

$$(1-2)^2 + (2-2)^2 + (3-2)^2 + (5-6)^2 + (7-6)^2 = 4$$