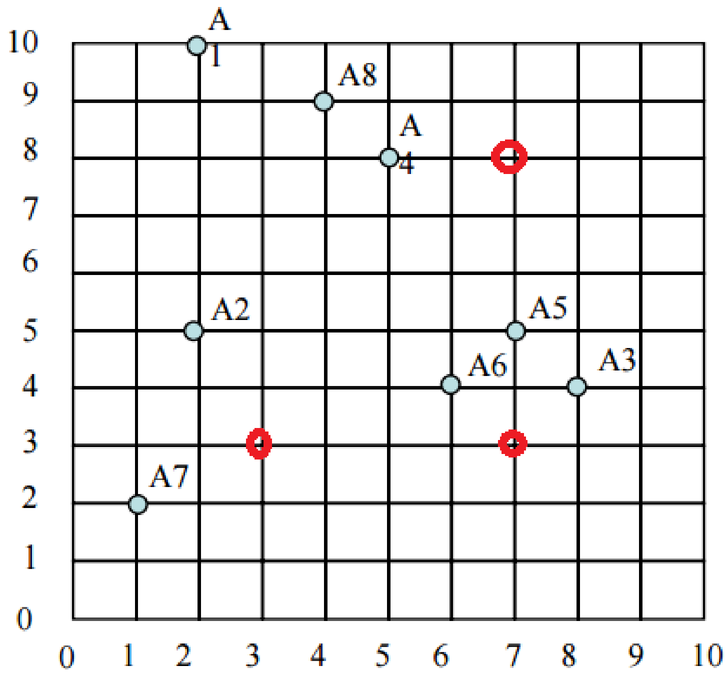
Exercise 1

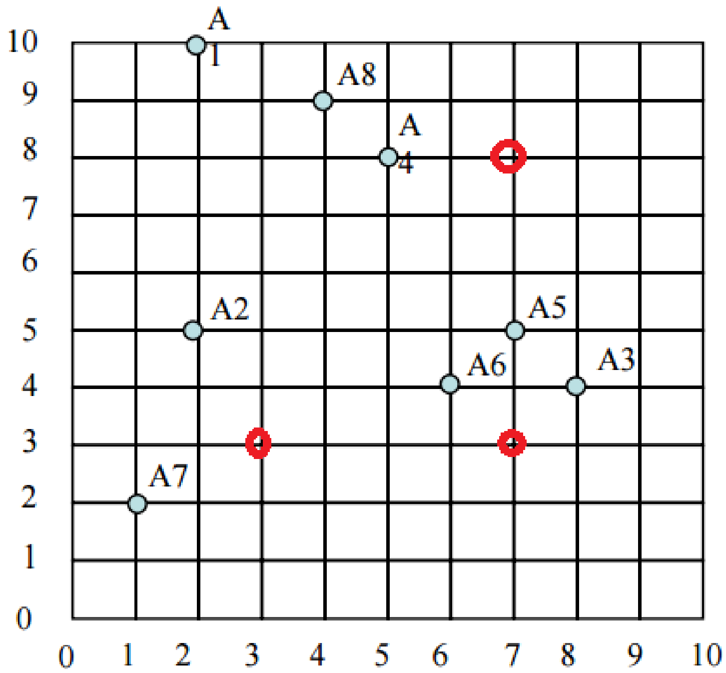
x	y
2	10
2	5
8	4
5	8
7	5
6	4
1	2
4	9



Compute complete partitionings of the data set into k = 3 clusters. As distance function use the Manhattan distance (L1 norm), which for two objects *x*, *y* is defined as:

$$L_1(x,y) = \sum_{i=1}^{d} |x_i - y_i|$$

Exercise 1



Compute complete partitionings of the data set into k = 3 clusters. As distance function use the Manhattan distance (L1 norm), which for two objects x, y is defined as:

$$L_1(x, y) = \sum_{i=1}^{d} |x_i - y_i|$$

a) Compute a partitioning into k = 3 clusters using the k-Means algorithm as introduced in the lecture (which is the Lloyd variant). As initial centroids use the red points. Draw the cluster assignments after this step, compute the new centroids. Run the procedure once again.

Exercise 1

- Initial centroids:

$$C_1 = (3,3) \qquad C_2 = (7,3) \qquad C_3 = (7,8)$$

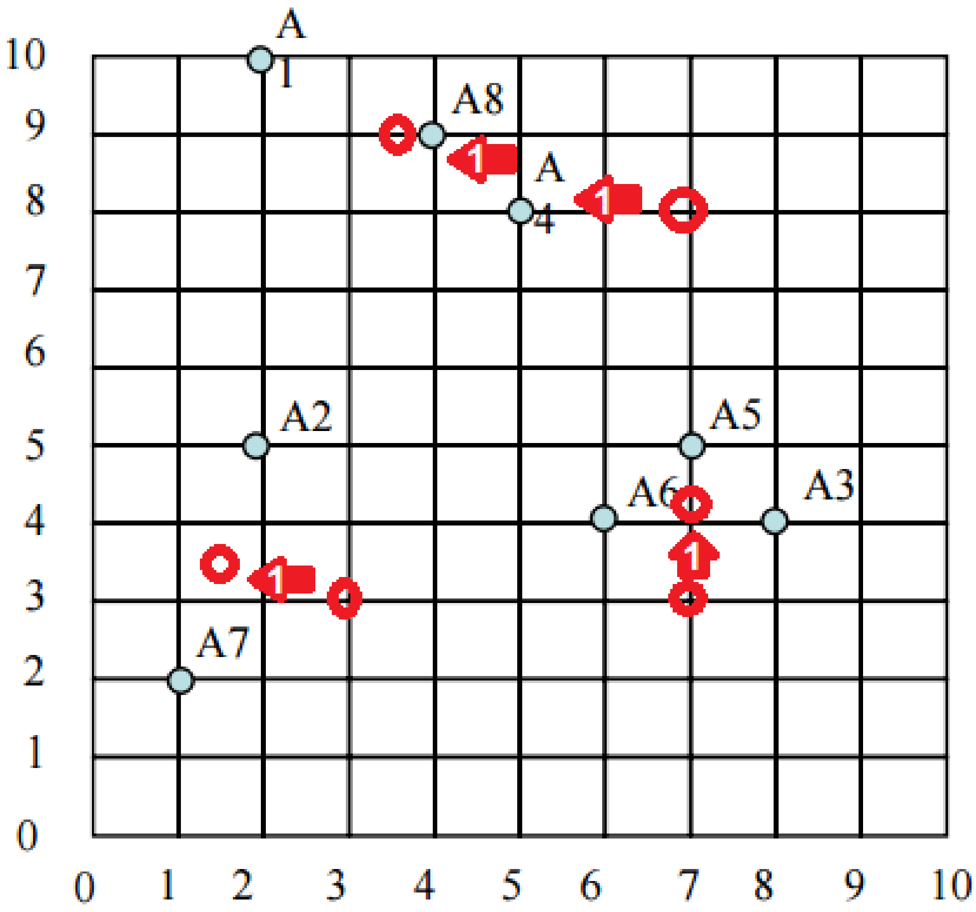
- 1. Iteration:
  - Assign datapoints to centroids:

$$C_1 \leftarrow A_2, A_7$$
$$C_2 \leftarrow A_3, A_5, A_6$$
$$C_3 \leftarrow A_1, A_4, A_8$$

  - Recalculate centroids:

$$C_1 = \left(\frac{1+2}{2}, \frac{2+5}{2}\right) = (1.5,3.5)$$
$$C_2 = \left(\frac{6+7+8}{3}, \frac{4+5+4}{3}\right) = (7,4.34)$$
$$C_3 = \left(\frac{2+4+5}{3}, \frac{10+9+8}{3}\right) = (3.67,9)$$

Exercise 1



Exercise 1

- Centroids:

$$C_1 = (1.5,3.5) \quad C_2 = (7,4.34) \quad C_3 = (3.67,9)$$

- 2. Iteration:
  - Assign datapoints to centroids:

$$C_1 \leftarrow A_2, A_7$$
$$C_2 \leftarrow A_3, A_5, A_6$$
$$C_3 \leftarrow A_1, A_4, A_8$$

➢ No change in the assignment

➢ Termination of k-Means