

Seminar 9

Algorithm 1 K-means($\{\vec{x}_1, \dots, \vec{x}_N\}, K, \text{stopping criterion}$)

```

1:  $(\vec{s}_1, \dots, \vec{s}_K) \leftarrow \text{SelectRandomSeeds}(\{\vec{x}_1, \dots, \vec{x}_N\}, K)$ 
2: for  $k \leftarrow 1$  to  $K$  do
3:    $\vec{\mu}_k \leftarrow \vec{s}_k$ 
4: end for
5: while stopping criterion has not been met do
6:   for  $k \leftarrow 1$  to  $K$  do
7:      $\omega_k \leftarrow \{\}$ 
8:   end for
9:   for  $n \leftarrow 1$  to  $N$  do
10:     $j \leftarrow \text{argmin}_{j'} |\vec{\mu}_{j'} - \vec{x}_n|$ 
11:     $\omega_j \leftarrow \omega_j \cup \{\vec{x}_n\}$  ▷ reassigning vectors
12:   end for
13:   for  $k \leftarrow 1$  to  $K$  do
14:     $\vec{\mu}_k \leftarrow \frac{1}{|\omega_k|} \sum_{\vec{x} \in \omega_k} \vec{x}$  ▷ recomputing centroids
15:   end for
16: end while
17: return  $\{\vec{\mu}_1, \dots, \vec{\mu}_K\}$ 

```

Exercise 1

Use the K -means algorithm with Euclidean distance to cluster the following $N = 8$ examples into $K = 3$ clusters: $A_1 = (2, 10)$, $A_2 = (2, 5)$, $A_3 = (8, 4)$, $A_4 = (5, 8)$, $A_5 = (7, 5)$, $A_6 = (6, 4)$, $A_7 = (1, 2)$, $A_8 = (4, 9)$. Suppose that the initial seeds (centers of each cluster) are A_1 , A_4 and A_7 . Run the K -means algorithm for 3 epochs. After each epoch, draw a 10×10 space with all the 8 points and show the clusters with the new centroids.

$d(A, B)$ denotes the Euclidean distance between $A = (a_1, a_2)$ and $B = (b_1, b_2)$. It is calculated as $d(A, B) = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2}$.

Take seeds $\vec{s}_1 = A_1 = (2, 10)$, $\vec{s}_2 = A_4 = (5, 8)$, $\vec{s}_3 = A_7 = (1, 2)$.

By 1 we count the alignment for epoch 1: $A_1 \in \omega_1$, $A_2 \in \omega_3$, $A_3 \in \omega_2$, $A_4 \in \omega_2$, $A_5 \in \omega_2$, $A_6 \in \omega_2$, $A_7 \in \omega_3$, $A_8 \in \omega_2$; and we get the clusters: $\omega_1 = \{A_1\}$, $\omega_2 = \{A_3, A_4, A_5, A_6, A_8\}$, $\omega_3 = \{A_2, A_7\}$.

Centroids of the clusters: $\vec{\mu}_1 = (2, 10)$, $\vec{\mu}_2 = ((8+5+7+6+4)/5, (4+8+5+4+9)/5) = (6, 6)$, $\vec{\mu}_3 = ((2+1)/2, (5+2)/2) = (1.5, 3.5)$.

After epoch 2 the clusters are $\omega_1 = \{A_1, A_8\}$, $\omega_2 = \{A_3, A_4, A_5, A_6\}$, $\omega_3 = \{A_2, A_7\}$ with centroids $\vec{\mu}_1 = (3, 9.5)$, $\vec{\mu}_2 = (6.5, 5.25)$ and $\vec{\mu}_3 = (1.5, 3.5)$. And finally after epoch 3, the clusters are $\omega_1 = \{A_1, A_4, A_8\}$, $\omega_2 = \{A_3, A_5, A_6\}$, $\omega_3 = \{A_2, A_7\}$ with centroids $\vec{\mu}_1 = (3.66, 9)$, $\vec{\mu}_2 = (7, 4.33)$ and $\vec{\mu}_3 = (1.5, 3.5)$.

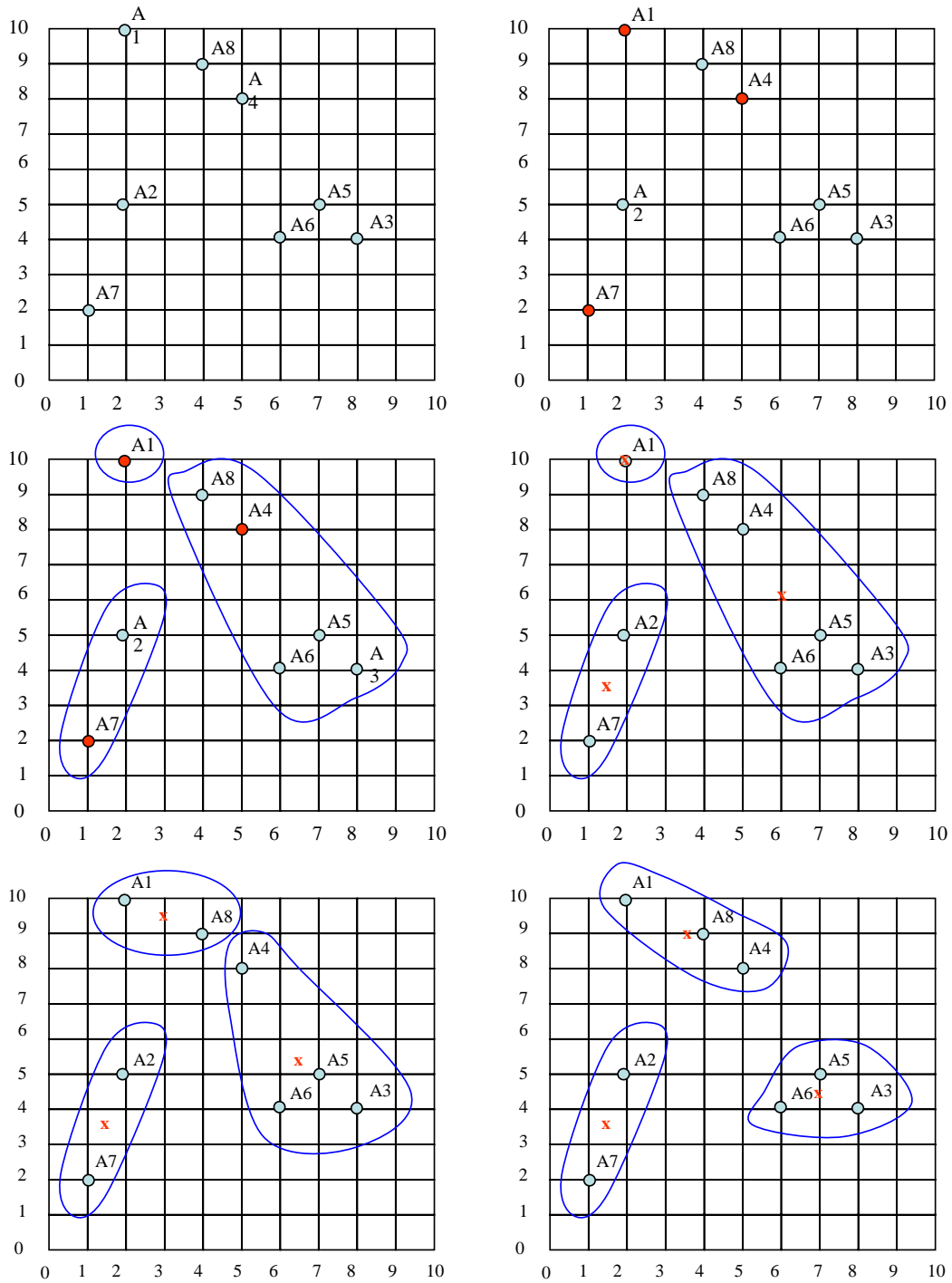


Figure 1: Visualization of K -means clustering algorithm.