Flat clustering (Chapter 16)

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

1: $(\vec{s}_1, \ldots, \vec{s}_K) \leftarrow$ SelectRandomSeeds($\{\vec{x}_1, \ldots, \vec{x}_N\}, K$)
2: **for** $k \leftarrow 1$ to K **do**
3: $\quad \vec{\mu}_k \leftarrow \vec{s}_k$
4: **end for**
5: **repeat**
6: $\quad$ **for** $k \leftarrow 1$ to K **do**
7: $\quad\quad \omega_k \leftarrow \{\}$
8: $\quad$ **end for**
9: $\quad$ **for** $n \leftarrow 1$ to N **do**
10: $\quad\quad j \leftarrow \text{argmin}_{j'} |\vec{\mu}_{j'} - \vec{x}_n|$
11: $\quad\quad \omega_j \leftarrow \omega_j \cup \{\vec{x}_n\}$ ▷ reassigning vectors
12: $\quad$ **end for**
13: $\quad$ **for** $k \leftarrow 1$ to K **do**
14: $\quad\quad \vec{\mu}_k \leftarrow \frac{1}{|\omega_k|} \sum_{\vec{x} \in \omega_k} \vec{x}$ ▷ recomputing centroids
15: $\quad$ **end for**
16: **until** a stopping criterion has been met
17: **return** $\{\vec{\mu}_1, \ldots, \vec{\mu}_K\}$

Exercise 16/1

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.

Exercise 16/2

What makes a good clustering? Give some clustering evaluation metrics.