Seminar 9

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

$(\vec{s}_1, \ldots, \vec{s}_K) \leftarrow \text{SelectRandomSeeds}(\{\vec{x}_1, \ldots, \vec{x}_N\}, K)$
for $k \leftarrow 1$ to K **do**
    $\vec{\mu}_k \leftarrow \vec{s}_k$
end for
repeat
    for $k \leftarrow 1$ to K **do**
        $\omega_k \leftarrow \{\}$
    end for
    for $n \leftarrow 1$ to N **do**
        $j \leftarrow \text{argmin}_{j'} |\vec{\mu}_{j'} - \vec{x}_n|$
        $\omega_j \leftarrow \omega_j \cup \{\vec{x}_n\}$ ▷ reassigning vectors
    end for
    for $k \leftarrow 1$ to K **do**
        $\vec{\mu}_k \leftarrow \frac{1}{|\omega_k|} \sum_{\vec{x} \in \omega_k} \vec{x}$ ▷ recomputing centroids
    end for
until a stopping criterion has been met
return $\{\vec{\mu}_1, \ldots, \vec{\mu}_K\}$

Exercise 9/1

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 9/2

Consider three points: $A_1 = [1, 1]$, $A_2 = [3, 1]$, $A_3 = [6, 1]$. Give an example of a point A_4 such that the K-means clustering algorithm with seeds $\{A_2, A_4\}$ and the agglomerative hierarchical clustering algorithm result in different clusterings of $\{A_1, A_2, A_3, A_4\}$ into 2 classes.

Exercise 9/3

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