
Algorithm 1 Unsupervised naive k-means clustering (Lloyd's algorithm)

Input: $\mathcal{D} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n \in \mathbb{R}^m\}$, number of clusters K and initial centroids $\{\mu_1, \mu_2, \dots, \mu_K \in \mathbb{R}^m\}$

Output: Cluster assignments $\mathcal{C} = \{c_1, c_2, \dots, c_K\}$ and updated cluster centroids $\{\mu_1, \mu_2, \dots, \mu_K\}$

flag \leftarrow false ▷ flag to indicate convergence: true for convergence and false otherwise

while flag is false **do**

for $\mathbf{x} \in \mathcal{D}$ **do** ▷ Iterate over all data points in \mathcal{D}

$c_i \leftarrow \arg \min_j \|\mathbf{x} - \mu_j\|^2$ ▷ Assign data point \mathbf{x} to the closest cluster centroid (Euclidean distance)

end for

$\hat{\mu} \leftarrow \mu$ ▷ Store the current best cluster centroids

for $j = 1$ to K **do** ▷ Iterate over all clusters

$\mu_j \leftarrow \frac{1}{|c_j|} \cdot \sum_{\mathbf{x} \in c_j} \mathbf{x}$ ▷ Update cluster centroid μ_j where $|c_j|$ is the number of data points in cluster c_j

end for

if $\|\mu - \hat{\mu}\| < \epsilon$ **then** ▷ Check for convergence: based on the change in cluster centroids

 flag \leftarrow true ▷ Set flag to true to terminate the algorithm

end if

end while
