# Goal

Cluster a dataset into *k* different clusters. The dataset is unlabeled (unsupervised learning).

Each sample is assigned to the cluster with the closest mean.

Optimization objective: Minimize the Euclidean distances between every instance and their assigned clusters.

# Iterative optimization

1. Initialize cluster centers randomly: *k* distinct instances are chosen randomly from the dataset and the centroids are placed at their locations, which will work better than completely random initialization.
2. Repeat until convergence (no more changes to the centroid positions):
   * Update cluster labels: Assign every instance to the nearest cluster center, the distance is computed using **Euclidean distance**.
   * Update cluster centers: Compute the mean of the instances for each cluster. Use the mean as the new cluster center for each cluster.