Algorithm 1: K-Means

Input: Set of observations X; Number of clusters K

Output: Final cluster assignments for the observations

Assign randomly a number, from 1 to K, to each observation

while *cluster assignments change* **do**

  for *each cluster* k **do**

    Calculate the centroid of cluster k:

$$c_k = \frac{1}{|C_k|} \sum_{x_i \in C_k} x_i$$

  for *each observation* x_i **do**

    Calculate the Euclidean distance to each centroid and assign x_i to the cluster whose centroid is closest:

$$k_i = \arg\min_k d_E(x_i - c_k)$$