

k-means Clustering Algorithm

For $i = 1$ to k

$c_i = \text{random } x \in D$

Repeat

For $i = 1$ to k

$S_i = \{ x \in D \mid \underset{i}{\operatorname{argmin}} \text{ distance}(c_i, x) \}$ /* subset of D closest to c_i */

For $i = 1$ to k

$c_i = \frac{\sum_{x \in S_i} x}{|S_i|}$ /* compute new “centroid” */

Until <stopping condition> /* example: until c_i 's don't change */