Given a training set = $\{x^{(1)}, \cdots x^{(n)}\}$ $x^{(i)} \in \mathbb{R}^d$

Algorithm:

1° initialize cluster centroids $\mu_1, \mu_2, \cdots \mu_k \in \mathbb{R}^d$ randomly

2° Repeat until converge {

For every i, set: $c^{(i)} = \arg\min_j \|x^{(i)} - \mu_j\|^2$ → 对各个样本到聚类中心的距离

For each j, set: $\mu_j = \frac{\sum_{i=1}^{n} I\{c^{(i)} = j\} x^{(i)}}{\sum_{i=1}^{n} I\{c^{(i)} = j\}}$ ← 重新计算聚类中心点

}