Step 1 给定参数 K 和 ε, 并生成 K 个初始中心点 $\{\mathbf{y}^{(i)}\}_{i=1}^{K}$, 这里 $\mathbf{y}^{(i)} \in R^n$.

Step 2 执行一次迭代. 具体步骤如下

2.1 按照以下公式确定每个样本所属的类别

中心

离哪个中心最近

$$c_i = \arg\min_{1\le j\le K} d(\mathbf{x}^{(i)}, \mathbf{y}^{(j)}), \quad i = 1, 2, \cdots, N, \tag{4.1.1}$$

其中 $c_i \in \{1, 2, \cdots, K\}$ 表示第 i 样本所属的类别. *(4.1.1)* 表明第 i 号样本归属于哪个类别, 取决于它与哪个中心点距离最小.

2.2 利用 *(4.1.1)* 确定 K 个样本子集 $\{X_i\}_{i=1}^{K}$, 计算公式为

同一个类

$$X_i = \{\mathbf{x}^{(j)} \in X : c_j = i\}, \quad i = 1, 2, \cdots, K. \tag{4.1.2}$$

即具有相同类别号的样本构成一个子集合.

Step 3 按以下公式重新确定中心点

更新中心

$$\overline{\mathbf{y}}^{(i)} := \mathbf{y}^{(i)}, \quad \mathbf{y}^{(i)} = c(X_i), \quad i = 1, 2, \cdots, K, \tag{4.1.3}$$

其中 $c(X_i)$ 的定义见 *(2.1.6)*.

Step 4 若条件

更新前后变化不大

$$\max_{1\le i\le K} d(\overline{\mathbf{y}}^{(i)}, \mathbf{y}^{(i)}) \le \varepsilon \tag{4.1.4}$$

成立, 则算法结束, 由 *(4.1.2)* 确定的 K 个集合即为聚类结果; 否则, 转至 *Step 2*.

1、初始K个中心.

2、计算每个点到每个中心距离，类别标记为最近的类

3、更新每个类的中心.

4、判断更新前后中心的变化.