**模糊均值（Fuzzy c-means,FCM）算法**

模糊均值聚类（FCM），即众所周知的模糊ISODATA，是用隶属度确定每个数据点属于某个聚类的程度的一种聚类算法。1973年，Bezdek提出了该算法，作为早期硬均值聚类(HCM）方法的一种改进。

类似于硬分类的情况，我们定义目标函数



（4.18）

其中r≥1是一个加权指数．

模糊均值算法的目标在于找到和使得最小。

下面，首先给出这个最小化问题的必要条件，然后根据此条件提出模糊均值算法。

**定理4.2** 若设， 为一个给定的数据集。给定和，假设对所有和有



（4.19）



（4.20）

则当满足（4.19）和（4.20）时，和才是的局部最小值。

**FCM算法描述如下**：

**步骤一：** 初始化。设总样本集，现在将样本集划分为类，记为的即，并初始化，设置迭代计数器和，并给定停止阈值。

**步骤二：**当迭代次数为时，计算聚类中心向量



（4.21）

其中。

**步骤三：** 根据下式将更新为，



（4.22）

**步骤四：** 比较和。为了确定是否已经达到收敛的聚类分区矩阵，我们选取一种矩阵范数，即行和范数，表示所有矩阵行向量绝对值之和的最大值（具体内容请参考方宝鎔，周继东，李医民的《矩阵论》，清华大学出版社，2004年）。若，则停止算法；否则令更新为，返回步骤二继续计算。

此算法也称为模糊ISODATA方法．应当注意的是，本方法要求，因此取初始分类时，遇到只有一个样本的类，要在聚类前先排除，待聚类后再加上该类，而参数一般常取。

**例4.5 设论域** 表示农业区域，已知每个区域的气候取决于4个指标：热量、水分、霜冻、霜雹。表4.4是一组气候的统计数据，要求把各区域气候数据划分为三类。

表4.4　气候分区统计表

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 指标 | 农业区域 | | | | | | | | | |
|  |  |  |  |  |  |  |  |  |  |
| 水分 | 1 | 2 | 2 | 3 | 3 | 5 | 6 | 6 | 5 | 4 |
| 水分 | 3.5 | 2.5 | 3.5 | 3 | 3 | 0.5 | 1.5 | 1.5 | 3 | 3 |
| 霜冻 | 1 | 2 | 1 | 3 | 1 | 5 | 4 | 4 | 2 | 1 |
| 霜雹 | 0 | 2 | 1 | 1 | 1 | 2 | 0 | 1 | 2 | 2 |

**解：**

**初始化**

设均值算法中加权数r=2，；令b=1，设定停止阈值；将统计数据分为、和三类,设初始隶属矩阵为：

；

得分类，，。

**计算聚类中心向量**

初始聚类中心的向量为

类： 

类： 

类： 

计算每个数据点到，，中心的距离

类似地，即可得

，，，，，，，，，；

，，，，，，，，，；

，，，，，，，，，；

**将更新为**

利用公式4.22计算数据点在类的中隶属度：

**．．．．．．**

（注：分母为0时，规定分式值等于无穷大）

类似地，即可得**:**

数据样本为类的隶属度为：

,,,,，，，，，；

数据样本为类的隶属度为：

,,,,，，，，，；

数据样本为类的隶属度为：

,,,,，，，，，；

得隶属矩阵为：

；

此时得分类结果为，，，可知分类结果相对于初始分类发生了改变。

**比较和**

，继续计算。

从步骤到步骤进行循环迭代：

第1次迭代得到；

第2次迭代得到；

**．．．．．．**

第11次迭代得到；

第12次迭代得到，至此停止迭代，得隶属矩阵为



得最终分类结果为，，。