FCM聚类算法开发文档

# 1 算法原理

## 1.1 算法概述

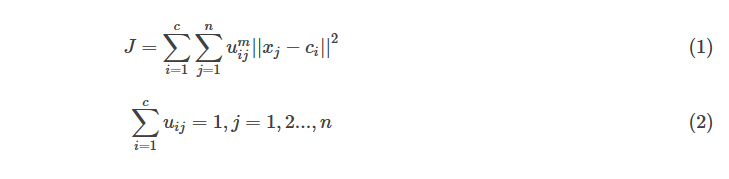
模糊C均值（Fuzzy C-means）算法简称FCM算法，是一种基于目标函数的模糊聚类算法，主要用于数据的聚类分析。理论成熟，应用广泛，是一种优秀的聚类算法。FCM算法是一种基于划分的聚类算法，它的思想就是使得被划分到同一簇的对象之间相似度最大，而不同簇之间的相似度最小。模糊C均值算法是普通C均值算法的改进，普通C均值算法对于数据的划分是硬性的，而FCM则是一种柔性的模糊划分。

## 1.2 算法组成与步骤

算法步骤：

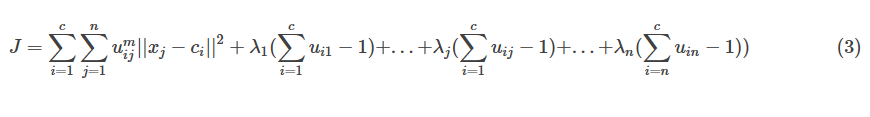
所谓模糊就是不确定，确定性的东西是什么那就是什么，而不确定性的东西就说很像什么。比如说把20岁作为年轻不年轻的标准，那么一个人21岁按照确定性的划分就属于不年轻，而我们印象中的观念是21岁也很年轻，这个时候可以模糊一下，认为21岁有0.9分像年轻，有0.1分像不年轻，这里0.9与0.1不是概率，而是一种相似的程度，把这种一个样本属于结果的这种相似的程度称为样本的隶属度，一般用u表示，表示一个样本相似于不同结果的一个程度指标。

基于此，假定数据集为X，如果把这些数据划分成c类的话，那么对应的就有c个类中心为C，每个样本j属于某一类i的隶属度为uij，那么定义一个FCM目标函数（1）及其约束条件（2）如下所示：

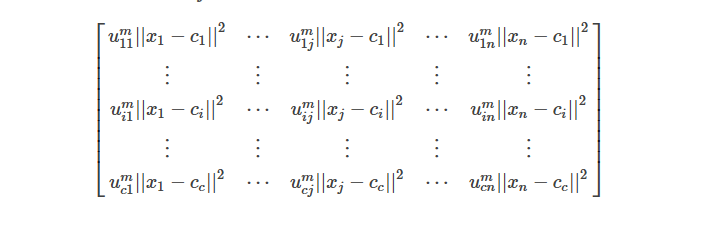


看一下目标函数（式1）而知，由相应样本的隶属度与该样本到各个类中心的距离相乘组成的，m是一个隶属度的因子，可理解为属于样本的轻缓程度，式（2）为约束条件，也就是一个样本属于所有类的隶属度之和要为1。观察式（1）可以发现，其中的变量有uij、ci，并且还有约束条件，那么求这个目标函数的极值：

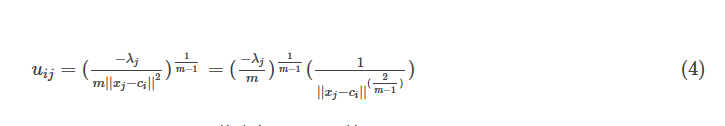
首先采用拉格朗日乘数法将约束条件拿到目标函数中去，前面加上系数，并把式（2）的所有j展开，那么式（1）变成下列所示：



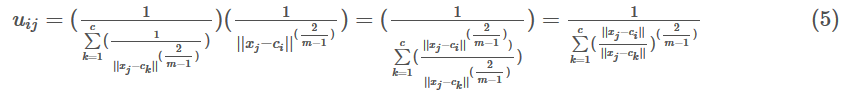
现在要求该式的目标函数极值，那么分别对其中的变量*uij*、*ci*求导数，首先对*uij*求导。分析式(3)，先对第一部分的两级求和的*uij*求导，对求和形式下如果直接求导不熟悉，可以把求和展开如下：



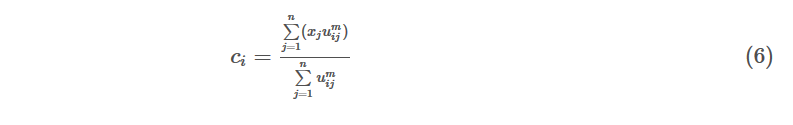
那么最终J对*uij*的求导结果并让其等于0就是：



最终的*uij*迭代公式：



下面在来求J对*ci*的导数。由公式（2）可以看到只有∑*i*=1*c*∑*j*=1*numij*||*xj*−*ci*||2这一部分里面含有*ci*，对其二级求和展开如前面所示的，那么它对*ci*的导数就是类中心的迭代公式：



程序开始的时候我们随便赋值给*uij*或者*ci*其中的一个，只要数值满足条件即可。然后就开始迭代，比如**一般的都赋值给*uij***，那么有了*uij*就可以计算*ci*，然后有了*ci*又可以计算*uij*，反反复复，在这个过程中还有一个目标函数J一直在变化，逐渐趋向稳定值。那么当J不在变化的时候就认为算法收敛到一个比较好的结了。可以看到*uij*和*ci*在目标函数J下似乎构成了一个正反馈一样，这一点很像EM算法，先E在M，有了M在E，在M直至达到最优。

**当目标函数J的大小和上一次迭代J的大小比较结果小于要求精度就退出迭代**。

上述公式中各个符合的意义   
- J     代表了目标函数   
- c     最后聚合的分类数目   
- n     数据集中的数据个数   
- uij   样本j属于类i的隶属度   
- xj     数据集j的位置   
- Ci     类i的中心位置   
- m     样本的轻缓程度

**总结FCM计算步骤**：

1.确定数据集、分类数、样本指数（轻缓程度）以及精度的值。

2.随机产生Uij

3.根据U计算聚类中心C的位置

4.再根据C去计算U

5. 计算目标函数J的大小和上一次迭代J的大小进行比较，小于要求精度就退出迭代。

6，计算每个点到所有类Uij的最大值来决定这个点属于哪个类

## 1.3 算法核心参考文献

<https://blog.csdn.net/on2way/article/details/47087201> 聚类之详解FCM算法原理及应用

https://blog.csdn.net/chenqianleo/article/details/78041476 聚类FCM算法

# 2 算法开发

## 2.1算法封装（输入与输出）

|  |  |  |
| --- | --- | --- |
|  | 名称 | 属性 |
| 输入：样本集 | x | double[][] |
| 输入：分类数 | knumOfClasses | int |
| 输入：精度 | E | double |
| 输入：FCM样本指数 | Q | double |
| 输出：确定性分类矩阵 | CertainClass | double[][] |

## 2.2算法核心函数及说明

组成关系图参考



## 2.3 算法组成与执行流程

首先输入数据集（X）、分类数、精度、FCM样本指数等参数，随机产生初始化初始隶属度矩阵R0，由特征指标矩阵X和初始分类矩阵R0和FCM公式，计算聚类中心的坐标矩阵VV，由特征指标矩阵X和聚类中心V修正隶属度矩阵RR，计算目标函数J的大小和上一次迭代J的大小进行比较，小于要求精度就退出迭代，最后计算确定性分类矩阵CertainClass

**注意计算距离有三种方式：**

1.Chebyshev切比雪夫距离

2. Euclid欧式距离

3. Hamming汉明距离

# 3 算法验证

## 3.1 验证算例说明

https://blog.csdn.net/keseliugeizuori/article/details/53162171测试数据

## 3.2 验证结果说明

输入的测试数据为：

1.0000 1.0000

2.0000 1.0000

1.0000 2.0000

2.0000 2.0000

3.0000 3.0000

8.0000 8.0000

8.0000 9.0000

9.0000 8.0000

9.0000 9.0000

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

---------------FCM聚类算法--------------

初始化初始隶属度矩阵R0

0.3804 0.4304 0.5465 0.2319 0.1444 0.2732 0.0579 0.9867 0.7591

0.6196 0.5696 0.4535 0.7681 0.8556 0.7268 0.9421 0.0133 0.2409

第1次迭代的聚类中心矩阵

6.6735 6.3060

4.4380 4.6451

第1次迭代的隶属矩阵

0.2938 0.2778 0.2706 0.2427 0.1635 0.8380 0.7783 0.7947 0.7584

0.7062 0.7222 0.7294 0.7573 0.8365 0.1620 0.2217 0.2053 0.2416

第2次迭代的聚类中心矩阵

7.6969 7.6863

2.2748 2.2807

第2次迭代的隶属矩阵

0.0352 0.0218 0.0216 0.0024 0.0232 0.9971 0.9772 0.9775 0.9635

0.9648 0.9782 0.9784 0.9976 0.9768 0.0029 0.0228 0.0225 0.0365

第3次迭代的聚类中心矩阵

8.4866 8.4864

1.8087 1.8087

第3次迭代的隶属矩阵

0.0115 0.0070 0.0070 0.0009 0.0450 0.9939 0.9945 0.9945 0.9949

0.9885 0.9930 0.9930 0.9991 0.9550 0.0061 0.0055 0.0055 0.0051

第4次迭代的聚类中心矩阵

8.4970 8.4970

1.7839 1.7839

第4次迭代的隶属矩阵

0.0108 0.0067 0.0067 0.0011 0.0467 0.9936 0.9945 0.9945 0.9952

0.9892 0.9933 0.9933 0.9989 0.9533 0.0064 0.0055 0.0055 0.0048

第5次迭代的聚类中心矩阵

8.4970 8.4970

1.7828 1.7828

第5次迭代的隶属矩阵

0.0108 0.0067 0.0067 0.0011 0.0467 0.9937 0.9945 0.9945 0.9952

0.9892 0.9933 0.9933 0.9989 0.9533 0.0063 0.0055 0.0055 0.0048

迭代了 5 次

-------------------聚类中心矩阵--------------------VV

8.4970 8.4970

1.7828 1.7828

---------------------隶属矩阵--------------------RR

0.0108 0.0067 0.0067 0.0011 0.0467 0.9937 0.9945 0.9945 0.9952

0.9892 0.9933 0.9933 0.9989 0.9533 0.0063 0.0055 0.0055 0.0048

------------------确定性分类矩阵--------------------CR

0.0000 0.0000 0.0000 0.0000 0.0000 1.0000 1.0000 1.0000 1.0000

1.0000 1.0000 1.0000 1.0000 1.0000 0.0000 0.0000 0.0000 0.0000

流程图参考

