# 模糊聚类分析在红酒数据集的运用

## 实验简述

本文章主要介绍基于模糊等价关系的聚类方法对经典红酒数据集进行模糊聚类分析。

实验的主要步骤为：

1. 导入数据
2. 确定评价标准
3. 数据规格化
4. 构造模糊相似矩阵
5. 构造模糊等价矩阵
6. 聚类

## 实验分析

### 导入数据

本实验随机选取红酒数据集中的10条红酒数据。



### 确定评价标准

因为红酒数据集有13个特征，每个特征代表红酒中的一种成分，因此使用13个特征来评价每一个红酒数据。

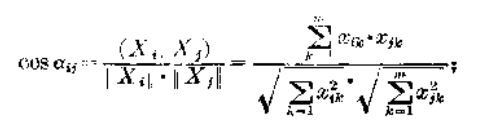
### 数据规格化

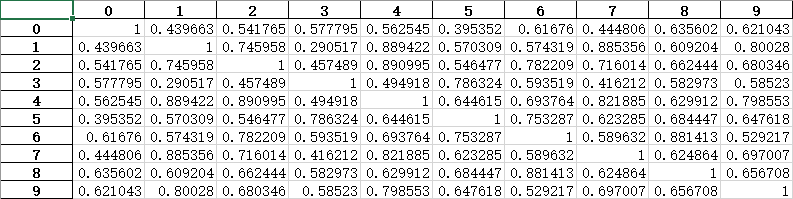
去量钢化，对数据进行归一化处理



### 构造模糊相似矩阵

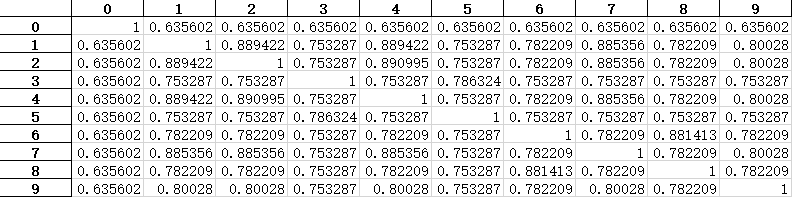
本实验使用余弦相似法构造相似矩阵（因为实验数据都是正数，所以相似度的取值范围在[0,1]，不会出现负数。





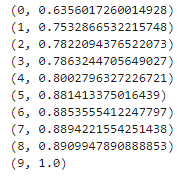
### 构造模糊等价矩阵

基于平方传递快速构建模糊等价矩阵，得到等价矩阵如下。



### 聚类

通过改变对模糊等价矩阵进行聚类。令大于等于的项设置为1，其他设置为0，对模糊等价矩阵进行统计。其中如下。



因为红酒数据集是有标注的，其分类数是三类，因此减少F检验部分，直接聚类成三类对比结果。根据对比分析，基于余弦相似度的模糊等价关系聚类准确率为60%



## 总结

本实验使用基于模糊等价关系的聚类方法对经典红酒数据集进行模糊聚类，其中构造模糊相似矩阵时使用的是余弦相似度方法。

实验结果并不太满意，其原因可能为余弦相似度并不太适合探索红酒数据集的相关性。可以改用相关系数法或者最大最小法类构建模糊相似矩阵对比结果。

因为本实验使用的是有标签的数据集，没有考虑统计量F检验去检测分类效果，直接使用分类数为3进行结果分析。