读书报告

09118119 黄一凡

# 自己提出的问题

#### 1.如何用二分查找快速找到最优的k？

当类别个数超过某个值后，平均直径就不再变化。所以我们只需求出最后那个直径不变的值所对应的k， 相当于求解一个函数的零点，这是由于函数单调递减，可以用二分查找方式求出该零点。

#### 2.层次聚类有什么具体的应用吗？（除了找初始中心）

可以对数据进行有层次，更为细致的分类。和K-means有着不同的着眼点。

# 别人提出的问题

#### 1.层次聚类与k均值聚类的差别是什么？有什么情况是只有层次聚类能进行处理的？

层次聚类会将类别进行一个分层，也就是可以得到一个阶梯式的分类结果。而k均值聚类是直接将样本分为k个类。 得到结果是平坦的，非层次化的。而且，k均值聚类k必须实现确定。同时k均值聚类必须用欧式距离来度量，而层次聚类可以选择多种度量方式。

#### 2.样本间距为什么选取欧式距离平方？如果选用欧式距离会有什么变化？

因为单单选择欧式距离还需要开根号，而k均值聚类时比较的是距离大小，同时平方并不会影响结果。

#### 3.聚合聚类将所有样本聚成一类不就没有意义了吗？

在过程中一层一层的聚类也是保留的，最后的结果不仅是最后的一大类。

#### 4.关于式14.21应如何理解？

这个式子描述了将n个样本分到m个不同的类的所有可能，这是一个第二斯特林数，可以用递归得到。

# 读书计划

#### 本周所读：

《统计学习方法》14.2-14.3

#### 下周计划：

《统计学习方法》15