1. （必填）自己提出的问题的理解（罗列全部）：
2. 提出的问题1：聚合聚类算法中，新类到原类的距离的定义是否欠妥？为何不考虑为类中各点中心到其他各类中的距离？

讨论后的理解： 在后续的k均值聚类算法中，衡量是否样本纳入某个类的标准就是类中心与样本的距离。

1. 提出的问题2：既然k均值聚类算法需要选择初值，初值的选择是否会对收敛性有影响？如果有，那么k均值聚类算法对于初值的选择很依赖吗？

讨论后的理解： 联想到某些数学知识，初值的选取很可能对最后的结果造成很大的影响。在实际应用中，此步可以使用人工选取的方式或者随机选择方式，尽量避免可能出现的不收敛的情况。

1. （必填）别人提出的问题的理解（选择几个问题罗列，并给出理解）：
2. 问题3：如何用二分查找快速找到最优的k？

自己的理解： 在实验中可以先将k赋为n/2看平均直径与k=1和k=n哪个更靠近，再继续二分。

1. 问题4：聚合聚类将所有样本聚成一类不就没有意义了吗？

自己的理解： 聚成一类时，对于分类而言没有实际意义，但是可以通过聚成一类计算平均直径等参数。也可以理解成这是一棵树自底向上求结点值的过程。所以每一层可以对应不同的层次。

5、问题5：为什么说k均值聚类中当k较大时平均直径将收敛？另外为什么拐点就是最优的k？

自己的理解：k取得越大，就可以认为是分类越精细。直观理解下，分类精细到一定程度时，若样本聚集程度较好，分类精细到一定程度时，每个类的直径就不再改变了。因为k越大可能使复杂度提升，而达到某种k时，平均直径不再明显改变，就取该k为最佳的k值选取。

1. （必填）读书计划

1、本周完成的内容章节：第十四章

2、下周计划：第十五章