说明：

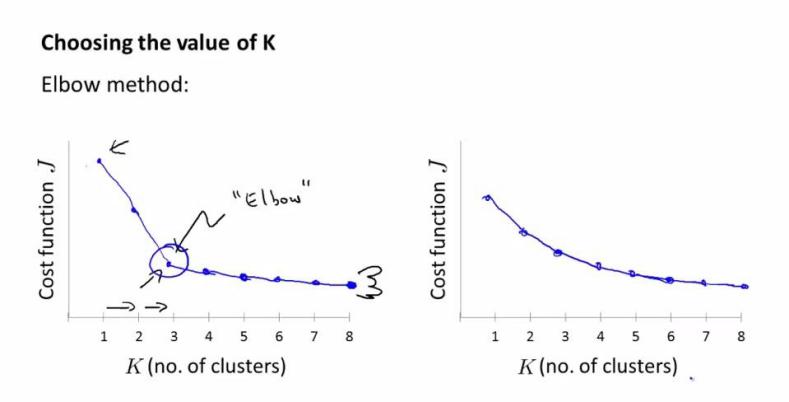
1. 第一、二、三部分必做，第四部分选做
2. 第一、二部分的问题数目不少于3个

读书报告内容：（复制下面部分到一个新文档，名称命名为自己的名字，填写完交给组长，没做第四部分可以不复制第四部分）

1. （必填）自己提出的问题的理解（罗列全部）：
2. 提出的问题1：为什么需要有链接方法？既然已经有了聚类，为什么又要把它们链接回去？

讨论后的理解：链接方法用于划分层次聚类中，在划分过程中，先从一个根节点就是一个聚类开始，运用链接方法，将聚类合并。初始的聚类并不是最终的聚类，不能说已经有了，所以不矛盾。

1. 提出的问题2：k均值算法中k究竟是怎么计算的？

讨论后的理解：Elbow method就是“肘”方法，对于n个点的数据集，迭代计算k from 1 to n，每次聚类完成后计算每个点到其所属的簇中心的距离的平方和，可以想象到这个平方和是会逐渐变小的，直到k==n时平方和为0，因为每个点都是它所在的簇中心本身。但是在这个平方和变化过程中，会出现一个拐点也即“肘”点，由图可以看到下降率突然变缓时即认为是最佳的k值。

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：为什么会出现空聚类的情况？

自己的理解：空聚类是一个比较特殊的情况。当当前有一个聚类点较少，而点又离其他聚类较劲、离所在聚类中心较远时，就可能导致在下一次或接下来几次的分配后，点都分配到其他聚类中，导致空聚类出现。

1. 问题4 分裂聚类时如何进行分裂的？与合并聚类有什么差异。

自己的理解：合并聚类是从一个个独立的点出发，一步步合并最相似的聚类来形成上一层聚类；分裂聚类是从一个包含所有点的根开始，把根节点一步步分成一个个子聚类。

如何将聚类分裂，我认为，可以先分出聚类每一个独立的点，然后运用k均值算法合并出聚类，这样就能达到分裂的目的

1. （必填）读书计划

1、本周完成的内容章节：第四章

2、下周计划：3-1~3-5

四、（选做）读书摘要及理解或伪代码的具体实现（读书摘要、伪代码的具体实现代码等可以写到这个部分）

1、读书摘要及理解（选做）

本章主要介绍了两种主要的聚类：划分聚类和层次聚类。划分聚类主要运用k均值算法进行划分。这个算法简便也好理解，不过也有很多缺点。因此我们在执行k均值算法时需要提前计算k值，通过计算确定种子点，编写排除异常点的算法。层次聚类是另外一种主要聚类方法，它是先将大聚类划分成微小聚类然后再一步步链接。这种也有广泛的运用，不过它的缺陷就是在计算大规模数据时较为低效。本章其余内容主要“数据处理”的理论学习，包括如何计算布尔属性和符号属性之间的距离，如何让数据标准化，处理混合属性。