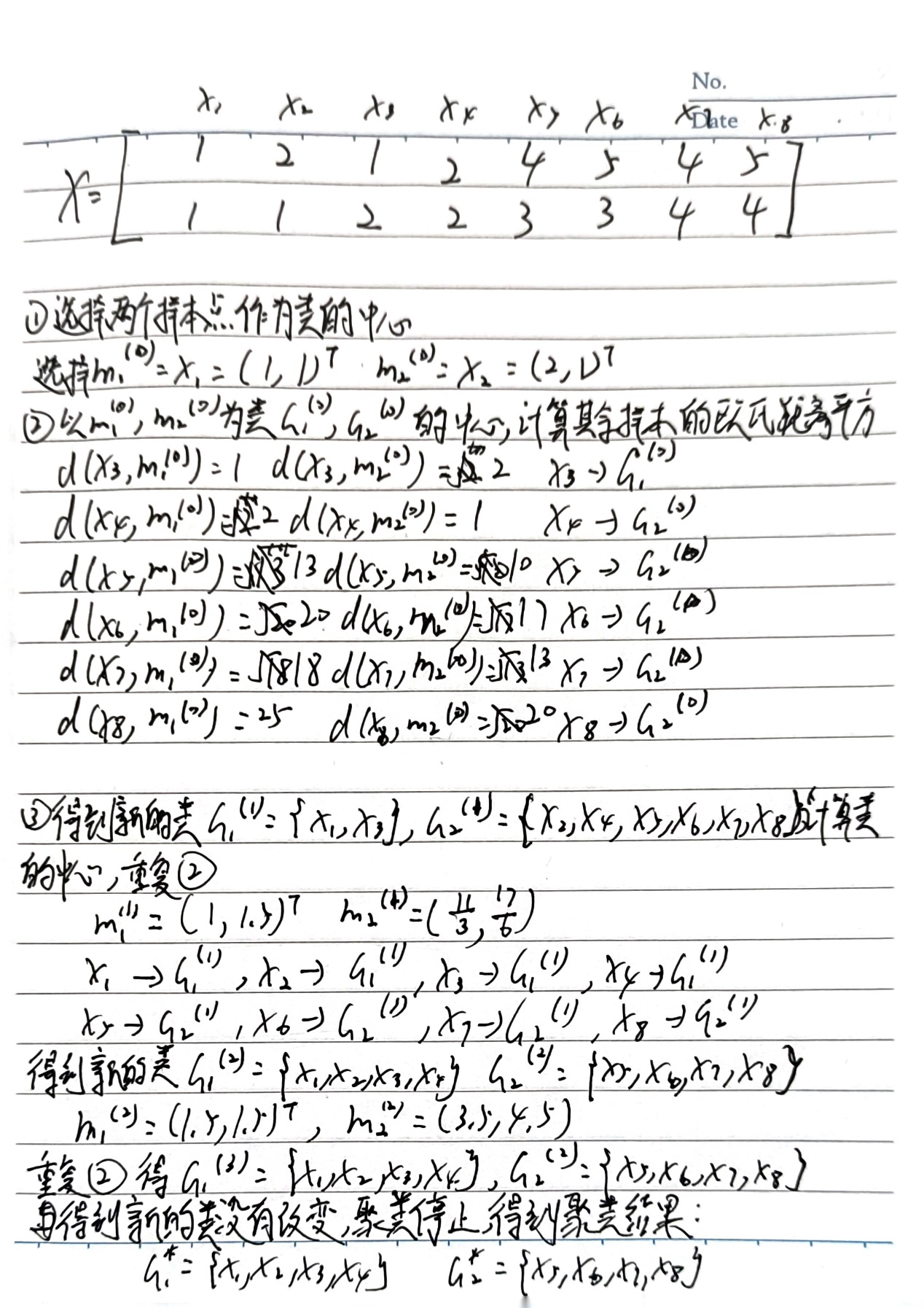
**聚类课后习题**

1. 试用k均值算法将下列样本聚到2个类中。

|  |  |  |
| --- | --- | --- |
| 样本 | 属性1 | 属性2 |
| 1 | 1 | 1 |
| 2 | 2 | 1 |
| 3 | 1 | 2 |
| 4 | 2 | 2 |
| 5 | 4 | 3 |
| 6 | 5 | 3 |
| 7 | 4 | 4 |
| 8 | 5 | 4 |





1. 简要概括密度聚类和层次聚类的算法特性，并指出这两种聚类算法分别适用于什么场景。

密度聚类的算法特性：1. 基于密度的层次聚类方法。2. 重点在于它可以将相关数据处理为几个高密度区域。3. 其中一项优势是无需定义聚类个数，而是根据样本分布自动划分。

层次聚类的算法特性：1. 对初始形态进行分割，逐步合并分割出来的子集。2. 把大量原始数据划分成少量的集合，并将这些集合形成一颗树形图。3. 适用于有先后依赖关系的数据，如文本文档中不同章节、故事或话题之间的关联性。

密度聚类的适应场景：当数据的空间距离具有显著的集群特性时，例如500个单一类别的样本可用密度聚类技术将它们划分成多个不同的子类。此外，在高斯混合模型中，也非常适用于对数据进行划分。

层次聚类的适应场景：当数据之间存在一定的嵌套特性时，可以使用层次聚类。层次聚类不需要事先定义子集个数，而是采用迭代方式来根据相似度来将样本逐步地分成更加详尽的子集。