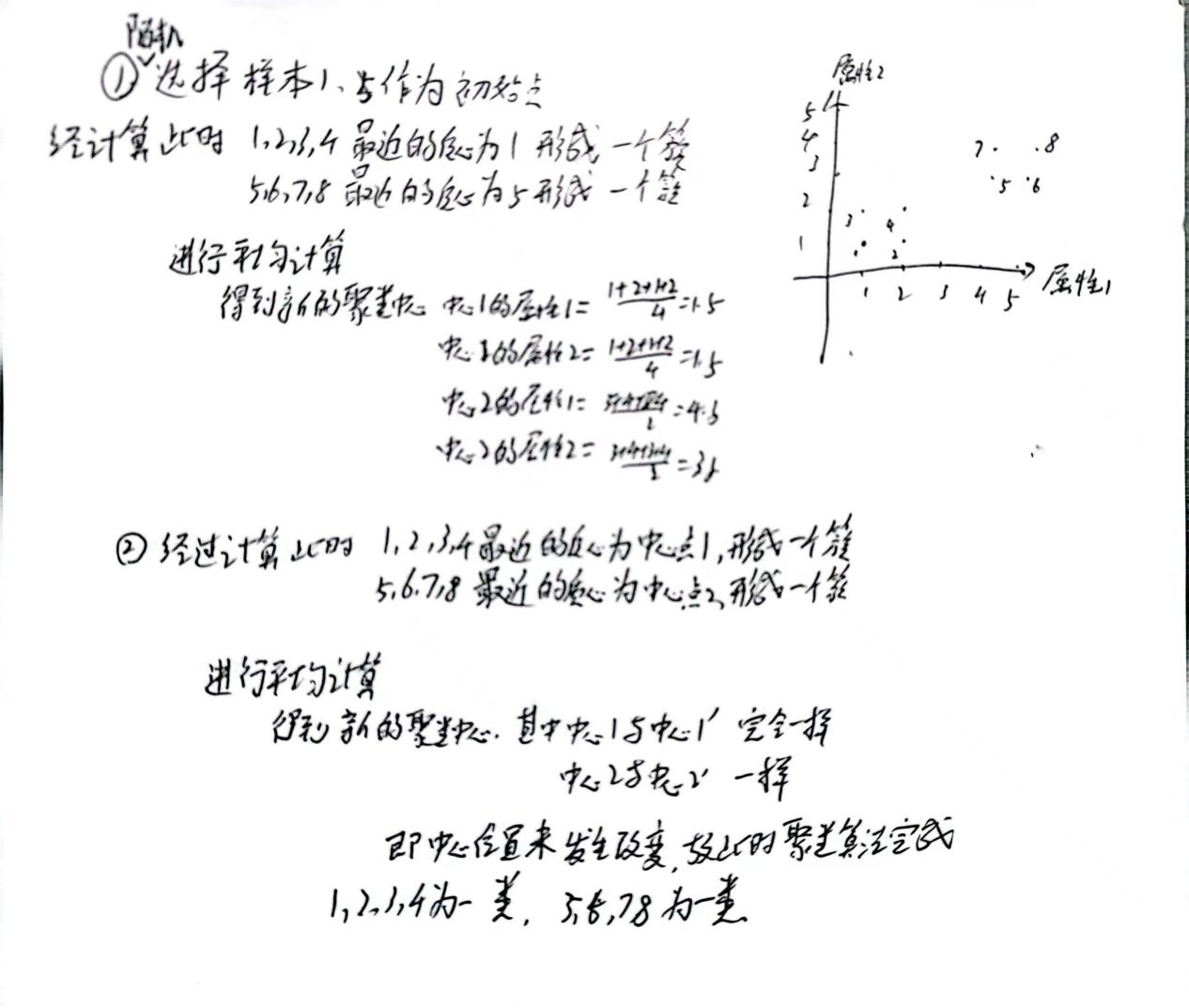
**1.**

****

**2.**

**密度聚类：**

从样本密度的角度考察样本的连接性，使密度相连的样本归结到一个簇，更符合直观认知，它将簇定义为密度相连的点的最大集合，能够把具有足够高密度的区域划分为簇，并在噪声的空间数据库中发现任意形状的聚类。

有两个超参数：扫描半径、最小包含点数来获得簇的数量，而不是猜测簇的数目。

将数据点分为三类：核心点、边界点、噪音点。

****优点：****  
1.对噪声不敏感；  
2.能发现任意形状的聚类。

****缺点：****  
1.但是聚类的结果与参数有很大的关系；  
2.用固定参数识别聚类，但当聚类的稀疏程度不同时，相同的判定标准可能会破坏聚类的自然结构， 即较稀的聚类会被划分为多个类或密度较大且离得较近的类会被合并成一个聚类。

适用场景：

包含大量聚类，可能是连接限制的，形状没有硬性要求的。

**层次聚类：**

假设簇之间存在层次结构，将样本聚到层次化的簇中。层次聚类有局和聚类、分裂聚类两种，因为每个样本只属于一个簇，所以层次聚类属于硬聚类。

开始将每个样本各自分到一个簇，之后将相距最近的两簇合并，建立一个新的簇，重复此操作直到满足停止条件。

优点：

1. 距离和规则的相似度容易定义，限制少。
2. 不需要预先制定聚类数；
3. 可以发现类的层次关系
4. 可以聚类成其它形状

缺点：

1. 计算复杂度太高；
2. 奇异值也能产生很大影响；
3. 算法很可能聚类成链状。

使用场景：

包含大量聚类，可能是连通性约束的，非欧几里得距离，少奇异值的，需要层次关系的，没有预先制定的聚类数的。