**聚类-习题**

**专业班级：**

**学号：**

**姓名：**

[计算题] **（40分）**1. 假设数据挖掘的任务是将如下的8个点（用（x，y）代表位置）聚类为3个簇。

A1(2,10)，A2(2,5)，A3(8,4)，B1(5,8)，B2(7,5)，B3(6,4)，C1(1,2)，C2(4,9)

距离函数是欧氏距离。假设初始我们选择A1、B1和C1分别为每个簇的中心，用k-means算法给出在第一轮执行后的3个簇中心。

[选择题] **（5分）**2. 对于单链层次聚类，两个簇的距离取决于不同簇中任意两点的（）。

1. 最小距离 B. 最大距离

[选择题] **（5分）**3. 对于单链层次聚类，两个簇的相似度取决于不同簇中任意两点的（）。

1. 最小相似度 B. 最大相似度

[选择题] **（5分）**4. 对于全链层次聚类，两个簇的距离取决于不同簇中任意两点的（）。

A.最小距离 B. 最大距离

[选择题] **（5分）**5. 对于全链层次聚类，两个簇的相似度取决于不同簇中任意两点的（）。

1. 最小相似度 B. 最大相似度

[选择题] （5分）6.聚类的目的是（）？

[A] 减少研究对象的数目。

[B] 方便对每个数据进行研究对比。

[C] 研究对象不足时，增加研究对象的数目。

[D] 方便储存。

[选择题] （5分） 7. K-means聚类的主要思想是（）

[A] 采用距离作为邻近度的评价指标，即认为两个对象的距离越近，其相似度就越大。

[B] 将多个变量综合成一个变量来分析。

[C] 通过迭代法找出使准则函数取极值的最好聚类结果。

[D] 从最小的类开始，逐步合并，最后成为一个最大的类。

[选择题] （5分） 8. K均值聚类中，初始聚类完成后，接下来的步骤是（）

[A] 计算新的簇中心，根据新的簇中心进行重新分配。

[B] 把每个簇中的样本综合成为一个样本。

[C] 在每个簇中再进行聚类。

[D] 算法结束。

[选择题] （5分） 9.下列描述符合层次聚类方法的是( )

[A] 采用迭代重定位技术，通过将对象从一个簇移到另一个簇来改善划分的质量。

[B] 由自顶向下或自底向上的方法创建给定数据对象集合的一个层次性分解。

[C] 根据邻域对象的密度或者根据某种密度函数来生成聚类。

[D] 将对象空间量化为有限数目的单元，形成网络结构，然后在网络结构上聚类。

[选择题] （5分）10.关于 K-means和 DBSCAN 的比较，以下说法不正确的是（ ）。

[A] K-means丢弃被它识别为噪声的对象，而 DBSCAN 一般聚类所有对象。

[B] K-means使用簇的基于原型的概念，而 DBSCAN 使用基于密度的概念。

[C] K-means很难处理非球形的簇和不同大小的簇，DBSCAN 可以处理不同大小和不 同形状的簇。

[D] K-means可以发现不是明显分离的簇，即便簇有重叠也可以发现，但是 DBSCAN 会合并有重叠的簇。

[选择题] （5分）11. 簇有效性的面向相似性的度量包括（ ）

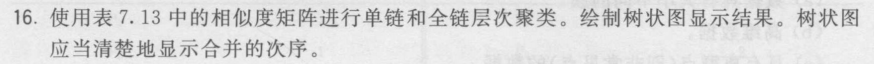
[A] 精度

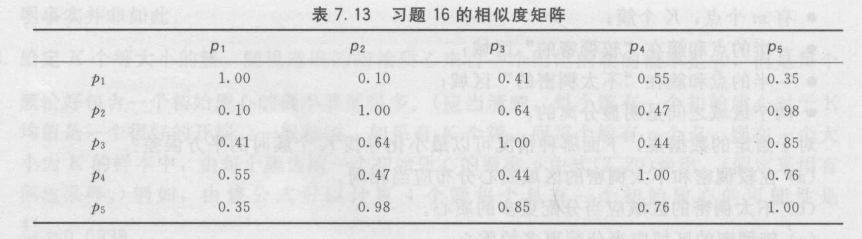
[B] Rand 统计量

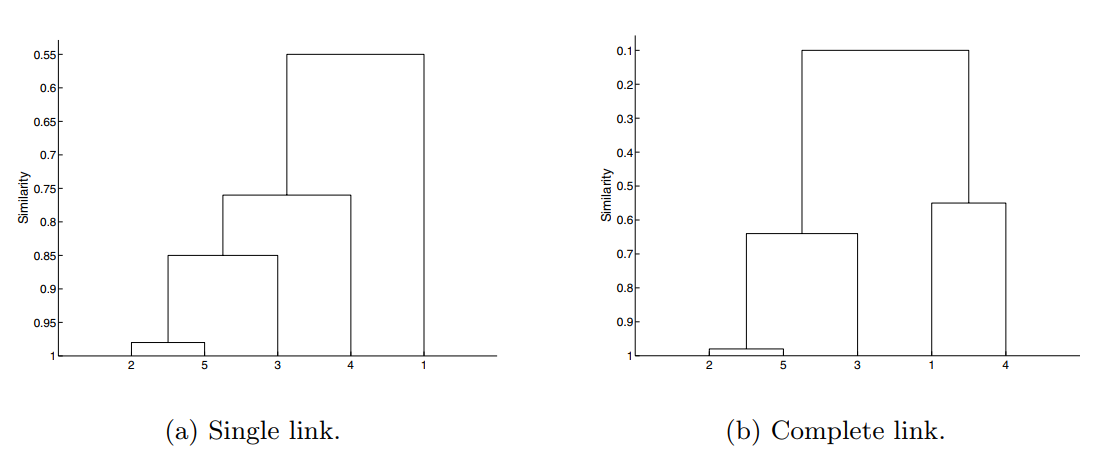
[C] Jaccard 系数

[D] 召回率

[计算题] **（40分）**12.



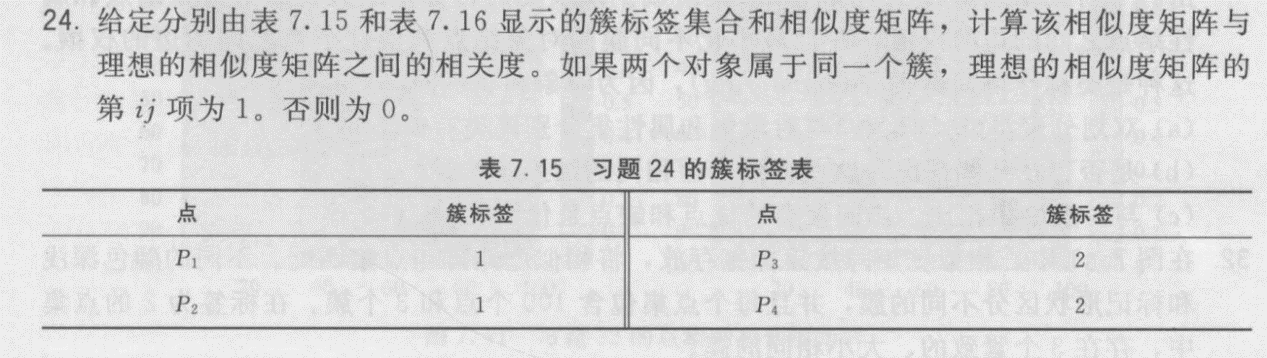


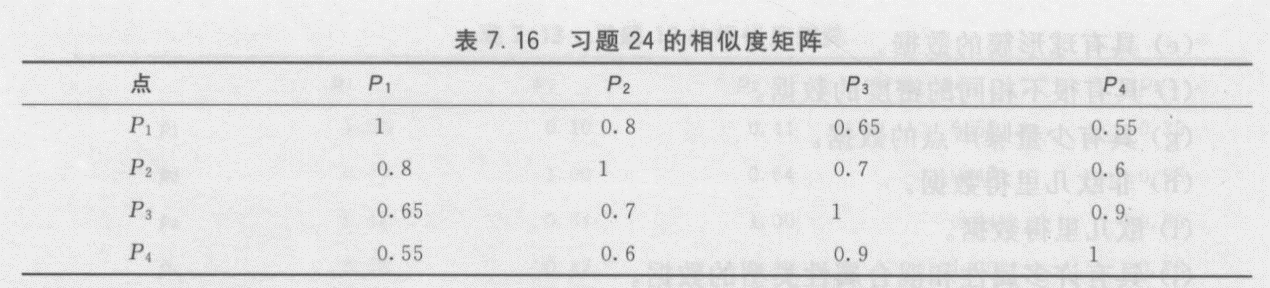


(1)根据画出的单链树状图（a），给出树中每两个簇合并处的高度的计算过程。

(2)根据画出的全链树状图（b），给出树中每两个簇合并处的高度的计算过程。

[计算题] **（30分）**13.





[计算题] **（20分）**14. 根据表7.14所示混淆矩阵，计算每个簇的纯度，和簇集的总纯度。

