# 《模式识别》课程

# 实 验 报 告



**姓 名： 金家耀**

**专 业：**  人工智能

**学 号： 1193210320**

**江南大学人工智能与计算机学院**

# 层次聚类

**1实验目的**

层次聚类法（Hierarchical Clustering Method）也称系统聚类法或分级聚类法，是实际应用中采用最多的方法之一。实验目的在于加深学生对层次聚类法原理的理解，掌握算法的实现过程，体会其在模式识别中的作用。

**2实验原理**

AGENS是一种自底向上聚合策略的层次聚类算法。它先将数据集中的每一个样本看作一个初始聚类，然后在算法运行的每一步找出距离最近的两个聚类簇进行合并，该过程不断重复，直至达到预设的聚类簇的个数。

这里的关键是如何计算聚类簇之间的距离。实际上，每个簇是一个样本集合，因此，只需要采用关于集合的某种距离即可。常用的簇间距离有：最小距离、最大距离、平均距离等。最小距离由两个簇的最近样本决定，最大距离由两个簇的最远样本决定，而平均距离则由两个簇的所有样本共同决定。

**3实验内容**

给出6个五维模式样本如下，按最短距离准则利用层次聚类方法进行聚类分析。



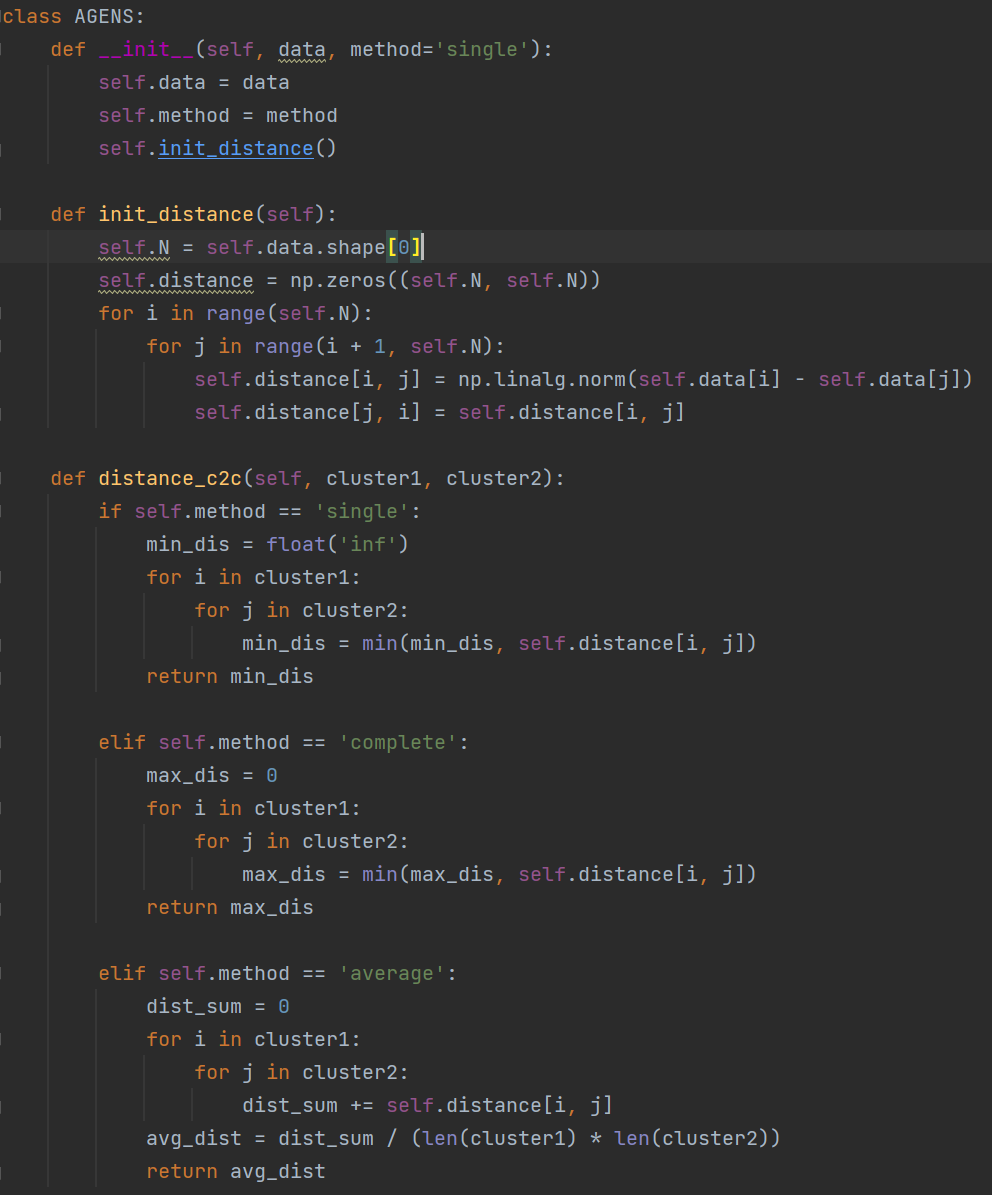
**4实验要求**

1. 若给定的阈值为，给出每次聚类时类别的合并以及最终的聚类结果；
2. 分别给定阈值，观察和分析聚类结果；
3. 绘制出层次聚类过程的树状图。

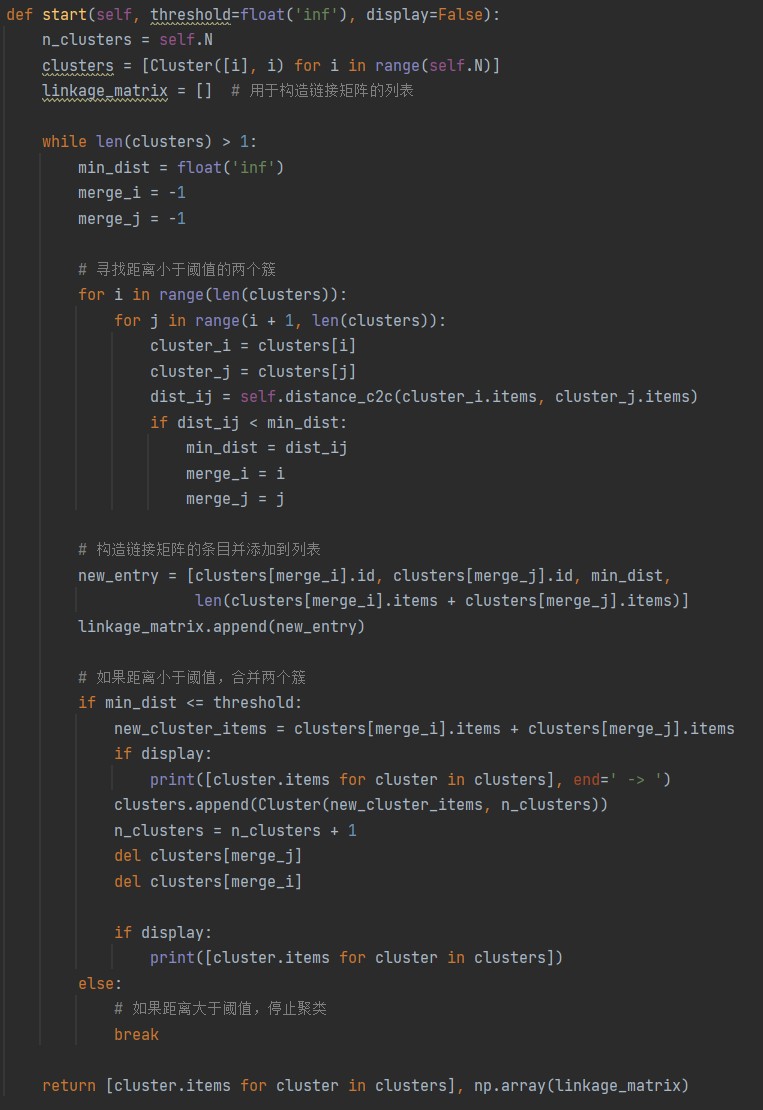
**5实验代码和结果**

0）代码已上传至

1）动手实现AGENS层次聚类



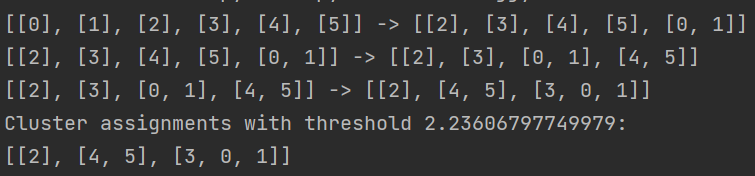
上图代码中自定义AGENS类，包含初始化类、初始化样本点之间的距离、计算两个簇之间的距离（虽然题目中只要求最短距离聚类，但为了锻炼个人能力，同时将三种不同的计算方式实现）。



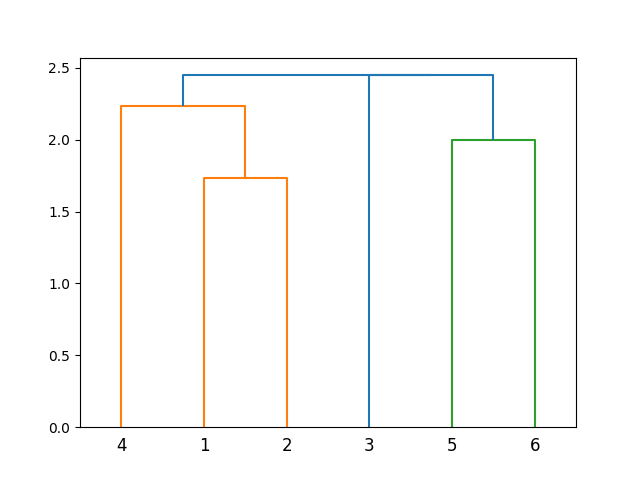
上图代码中实现的是层次聚类的过程，包括遍历所有的聚类簇对，计算需要合并的两个聚类簇，以及判断是否超过阈值。然后是合并两个簇，生成链接矩阵（记录本次两个簇的索引、两个簇之间的距离、合并后样本点的个数），同时该函数返回最终的聚类结果以及链接矩阵。

1. 阈值为时，聚类结果以及聚类过程



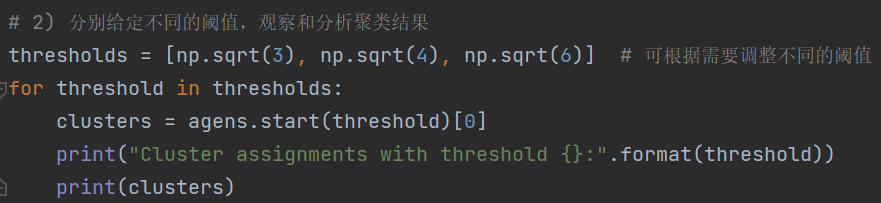


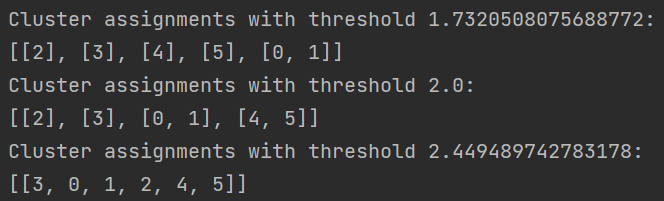
阈值为时，聚类结果和过程如上图所示。第一次0、1样本合并，第二次4、5样本合并，第三次3、[0, 1] 样本合并，最终得到 [2], [4, 5], [3, 0, 1]。



上图的橙色分支和绿色分支则代表聚类过程。

1. 分别给定阈值，观察和分析聚类结果；

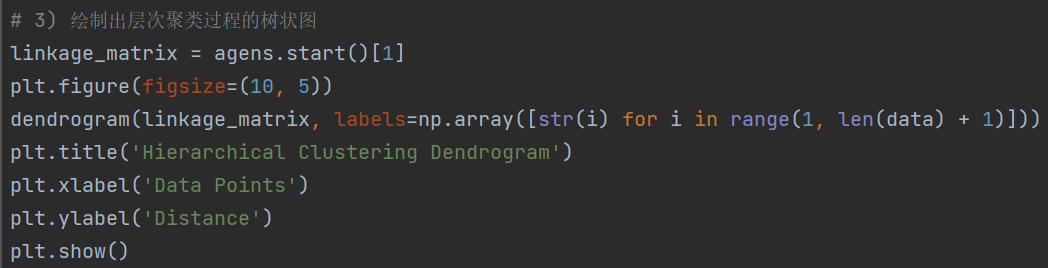


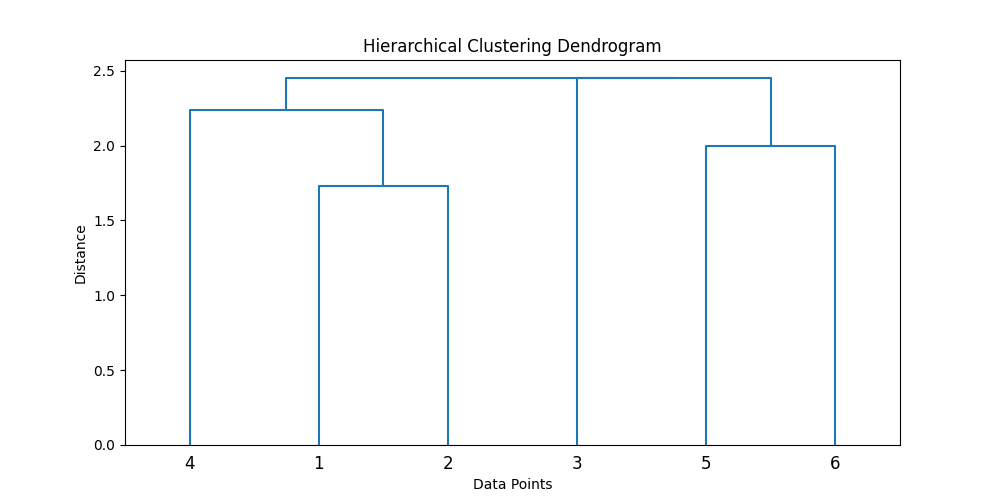


由上述代码可得上方图片结果，分别表示给定阈值。通过上述可视化结果，我们观察到随着给定的阈值逐渐增大，聚类的条件限制逐渐减弱，这导致越来越多的样本点被合并成更大的聚类簇，从而减少了最终的聚类数量。这种变化反映了层次聚类中的灵活性，使我们能够根据不同的阈值选择实现不同粒度的聚类结果。

这种参数化方法允许我们在数据分析中根据具体问题的需求，灵活地控制聚类结果的精细程度。较小的阈值可用于捕捉数据中的细微结构和相对较小的差异，而较大的阈值则可用于获得更抽象的聚类簇，强调更广泛的数据模式。

1. 绘制出层次聚类过程的树状图。





通过使用生成的链接矩阵和dendrogram()函数绘制的树状图，我们能够深入了解聚类过程的演化。从图中可以明显观察到，当阈值逐渐升高时，聚类的约束逐渐减小，允许更多的样本点被合并成更大的聚类簇。这种现象反映了在层次聚类中自由选择聚类粒度的灵活性。适用较小阈值可以捕捉数据中的细微结构和相对较小的差异，而较大的阈值则强调更宽泛的数据模式，从而导致聚类簇的数量逐渐减少。