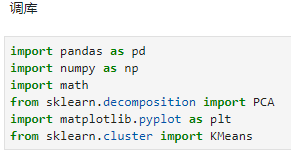
# 实验9\_任务3\_异常检测

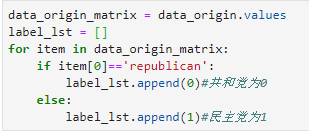
任务**3** 异常检测

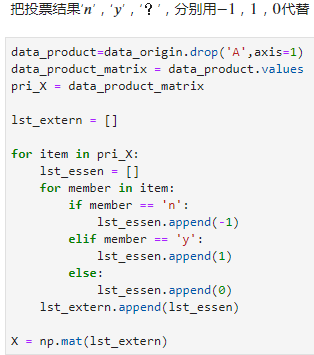
任务描述：针对上述两个任务中的某⼀个任务的数据，尝试使⽤某种异常检测⽅法分析数据的情况，判断数据集合中是否存在异常数据点。

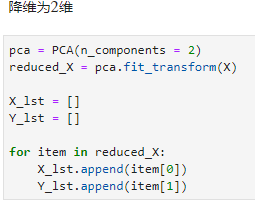
# 实验过程记录：



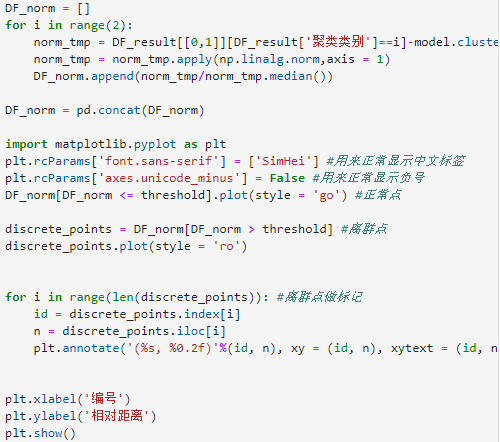


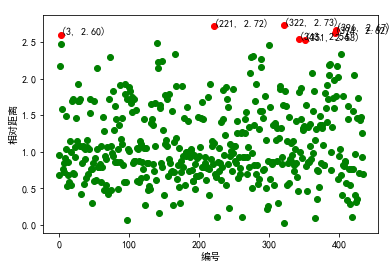




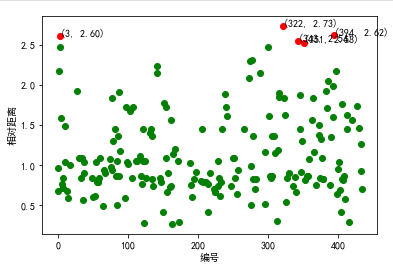




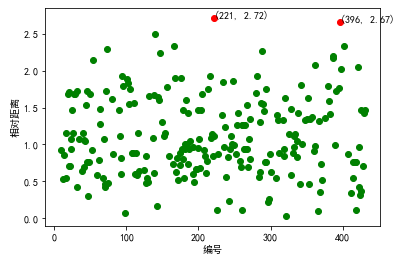












**问题1：**

此离散的检测的思路是：

1.进行聚类。选择聚类算法（如K-Means算法），将样本集聚K簇，并找到各簇的质心。

2.计算各对象到它的最近质心的距离。

3.计算各对象到它的最近质心的相对距离。（相对距离是点到质心的距离与簇中所有点到质心的距离的中位数之比）

4. 如果某对象距离大于阈值，就认为该对象是离散点。【1】

取定阈值为2.5，可以发现两个簇中均有离散点，但是，离散点数目较少，可以认为不是由于党派的差异，而是由于议员个人倾向的差异。

**参考资料：**

【1】：https://blog.csdn.net/sinat\_25873421/article/details/80727352

**提出问题：**

**1）在处理数据时，我采用了PCA降维，一定程度上，把离散的投票结果变成了连续的值，我认为这里缺乏科学性，但从聚类结果显示，似乎并没有产生很大影响。我对这里的原理不太理解，不知是不是因为对PCA的理解不到位产生的。**