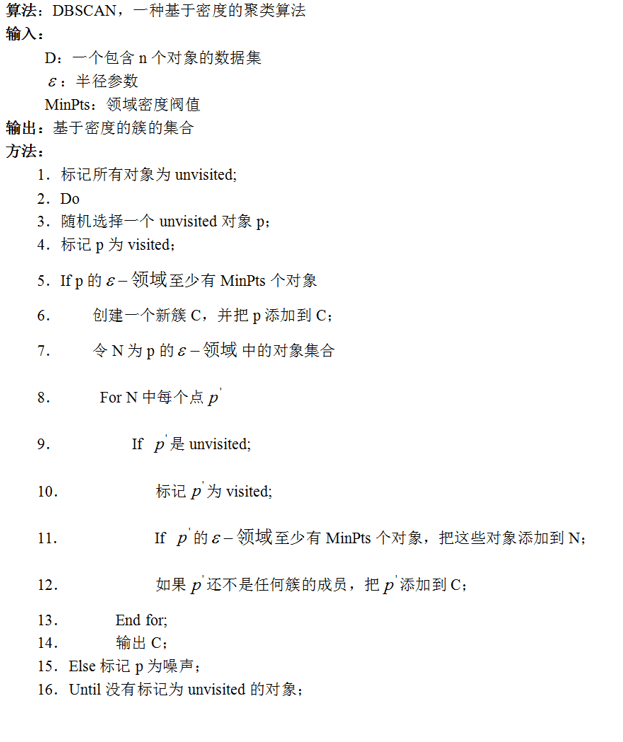
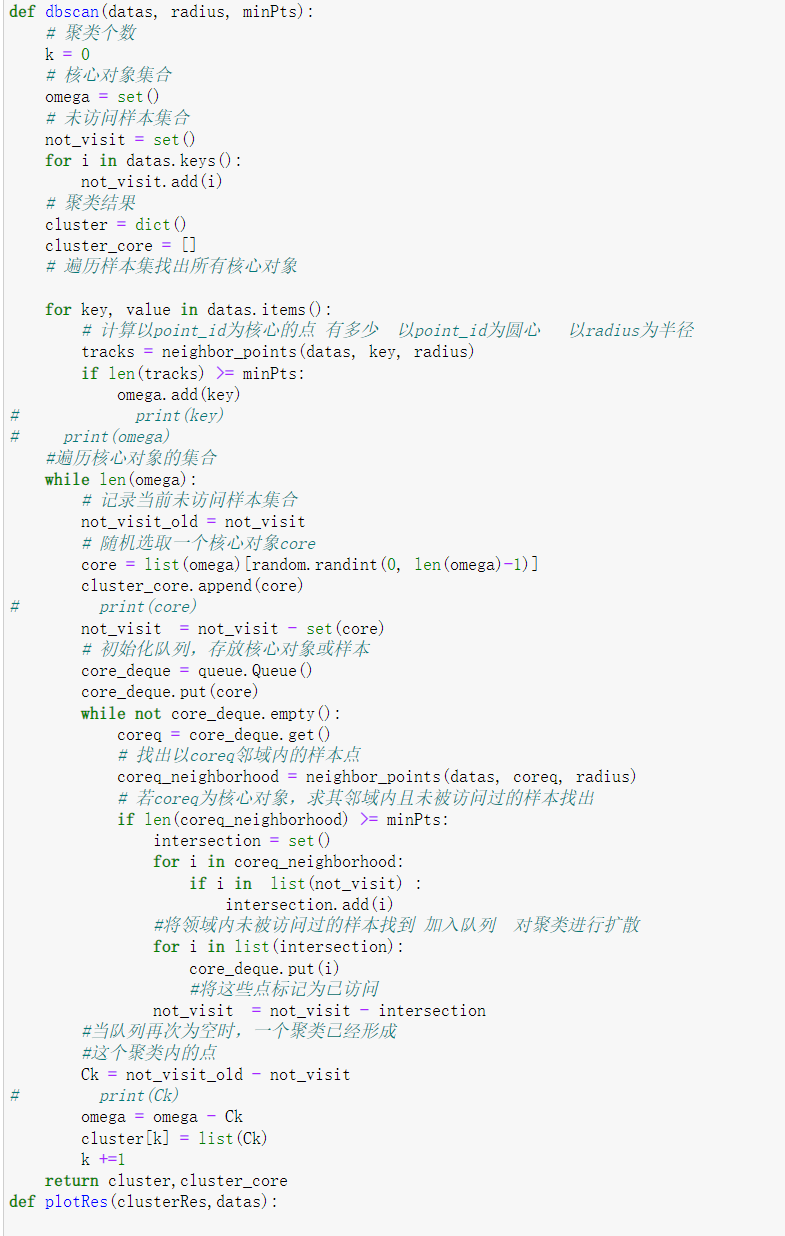
1. 写出伪代码和程序代码（标注主要过程及方法）

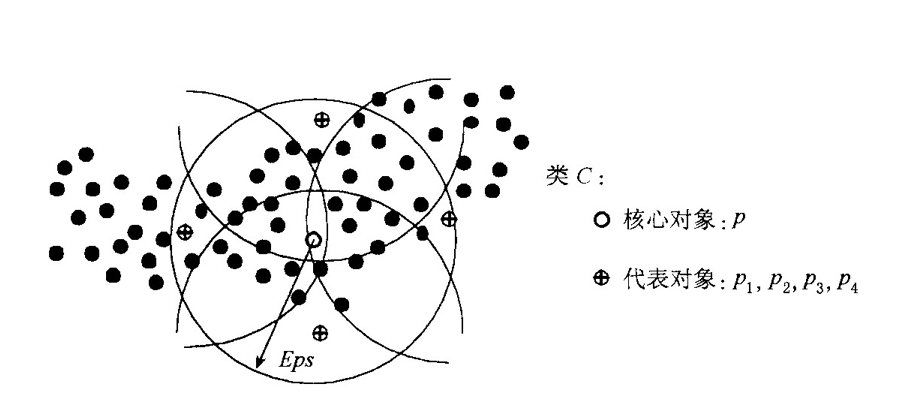
伪代码如下图：



算法源码如下图



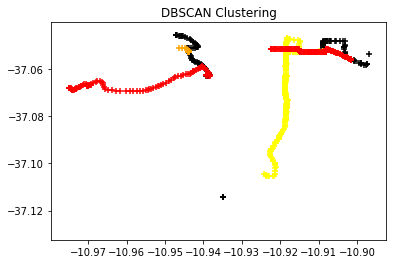
展示出20个数据对象的算法执行过程



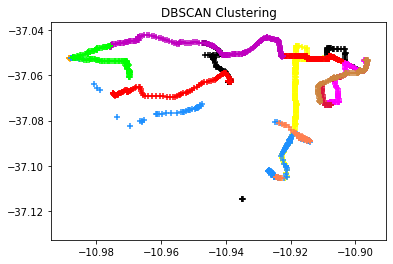
写出数据的分类结果和 数据划分的图形化展示结果

输入15条轨迹,半径0.5，minpot=2时，最终分为 4 个簇，簇 0 内的轨迹条数 6，簇 1 内的轨迹条数 3，簇 2 内的轨迹条数 3，簇 3 内的轨迹条数 1。

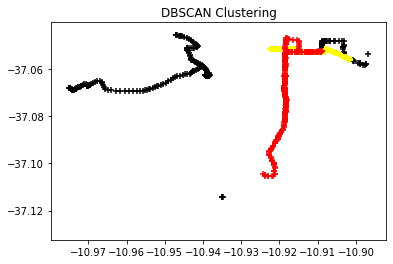
结果：{0: ['10', '18', '12', '1', '3', '14'],1: ['19', '13', '4'],2: ['20', '2', '21'],3: ['8']}图片如下：



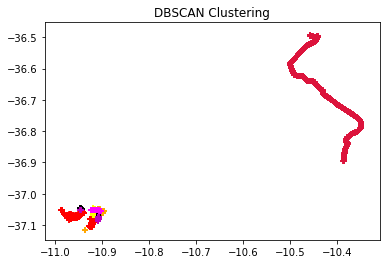
输入20条轨迹，半径0.5，minpot=1时，最终分为 11 个簇，簇 0 内的轨迹条数 7，簇 1 内的轨迹条数 3，簇 2 内的轨迹条数 3，簇 3 内的轨迹条数 1，簇 4 内的轨迹条数 1，簇 5 内的轨迹条数 1，簇 6 内的轨迹条数 1，簇 7 内的轨迹条数 1，簇 8 内的轨迹条数 1，簇 9 内的轨迹条数 1，簇 10 内的轨迹条数 1。结果：{0: ['10', '18', '12', '8', '1', '3', '14'],1: ['19', '13', '4'],2: ['20', '2', '21'], 3: ['25'], 4: ['22'],5: ['17'], 6: ['11'],7: ['26'],8: ['23'],9: ['24'],10: ['16']}图片如下



输入20条轨迹，半径0.4，minpot=2时，最终分为 3 个簇，簇 0 内的轨迹条数 7，簇 1 内的轨迹条数 2，簇 2 内的轨迹条数 3。结果：{0: ['10', '18', '2', '12', '1', '3', '14'],1: ['20', '21'],2: ['19', '13', '4']}，图片如下。



输入30条轨迹，半径0.4，minpot=2，最终分为 7 个簇，簇 0 内的轨迹条数 5，簇 1 内的轨迹条数 5，簇 2 内的轨迹条数 3，簇 3 内的轨迹条数 5，簇 4 内的轨迹条数 3，簇 5 内的轨迹条数 2，簇 6 内的轨迹条数 2。结果：{0: ['37', '28', '1', '11', '35'], 1: ['13', '4', '3', '19', '33'],2: ['2', '27', '26'],3: ['10', '18', '12', '14', '34'],4: ['36', '8', '38'],5: ['20', '21'],6: ['31', '30']}，图片如下。



运行效率的分析过程 算法优缺点的分析，及改进意见

优点：

（1）不需要提前设定K值大小，可以自适应的做聚类结果。

（2）对噪声不敏感。该算法能够较好地判断离群点，即使错判离群点，对最终的聚类结果也没什么影响

（3）能发现任意形状的簇。因为DBSCAN 是靠不断连接邻域的高密度点来发现簇的，只需要定义邻域大小和密度阈值，因此可以发现不同形状，不同大小的簇

（4）聚类结果没有偏倚，相对的，K-Means之类的聚类算法初始值对聚类结果有很大影响

缺点：

（1）对两个参数的设置敏感，即圈的半径 eps 、阈值 MinPts。

（2）DBSCAN 使用固定的参数识别聚类。当聚类的稀疏程度不同，聚类效果也有很大不同。即数据密度不均匀时，聚类质量较差。

（3）复杂度高，为降低复杂性若提前建立距离索引则又会占存储空间，如果数据样本集越大，收敛时间越长。

（4）调参相对于传统的K-Means之类的聚类算法稍复杂，主要需要对距离阈值ϵ，邻域样本数阈值MinPts联合调参，不同的参数组合对最后的聚类效果有较大影响。

改进

（1）针对数据样本集越大，收敛时间越长的问题。可以对搜索最近邻时建立的KD树或者球树进行规模限制来改进。

（2）针对DBSCAN算法在处理高维密度不均匀数据时会产生大量离群值的问题。可以引入密度检测的基本思想。提出的密度检测方法首先查询n个节点的Eps邻域，从而得到每个节点的cv值。利用cv值对数据分区进行分割。大致意思如下

①首先查询每个节点的Eps邻域，分别计算其密度；

②根据步骤①的结果计算每个节点的平均密度，进而得到每个节点的方差s2；

③将步骤②的结果代入式⑤中，计算各节点的cv值；

④基于等深度分块法的数据分割；

⑤自动设定Eps和MinPts值；

⑥在每个数据分区运行DBSCAN算法；

⑦ 得到聚类结果，结束整个算法。

7 程序源码，并对主要过程和方法进行注释

附件—聚类代码