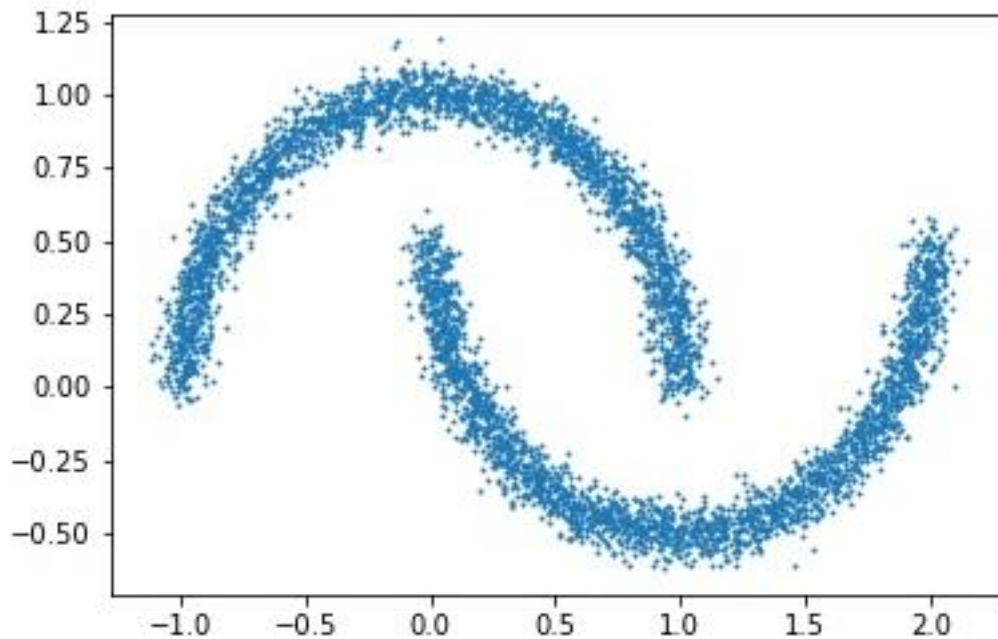


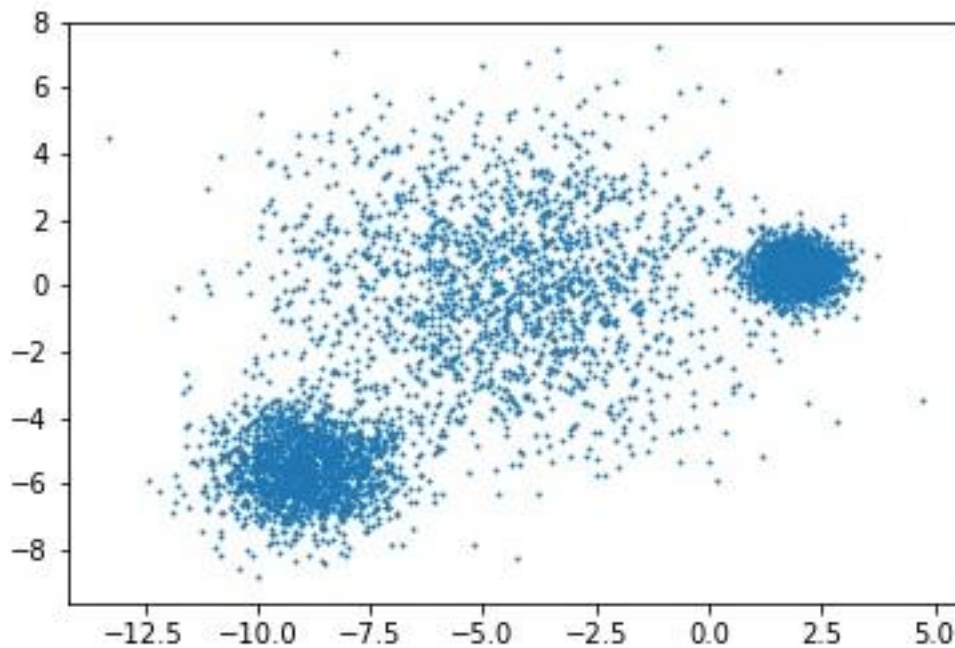
## 《系统工程导论》聚类分析作业

### 【题目】

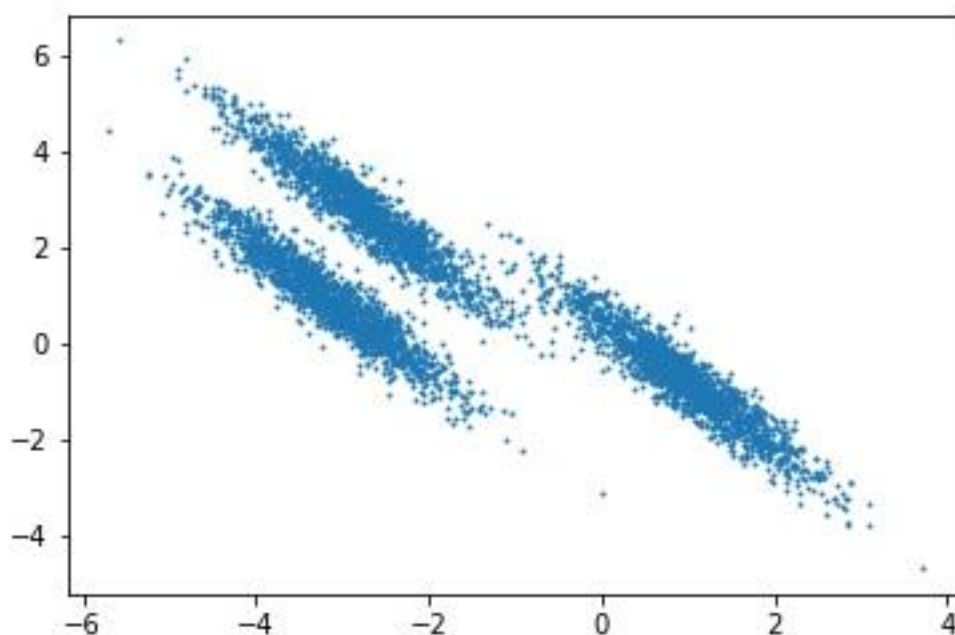
附件data.mat中包含三个数据集data1、data2、data3，每个数据集包含5000个二维平面上的点，维度为(5000, 2)。请根据课堂所学知识，编写相应的聚类算法对这些点进行聚类。这些点的分布情况如下：



data1数据分布



data2数据分布



data3数据分布

【具体要求】

1. 自行编写kmeans聚类算法，绘制以上3个数据集的聚类结果;(2分)
2. 利用数据集data2，对kmeans算法进行如下实验：
  - (1) 聚类数目从2类开始逐渐增加，分别进行计算并分析聚类结果，决定最合适的聚类数目并说明理由；(1分)
  - (2) 选择不同的初始点多次实验，观察初始点的选择对最终结果的影响，并分析原因；(1分)
3. 自行编写DBSCAN密度聚类算法，绘制以上3个数据集的聚类结果;(3分)
4. 利用数据集data3，对DBSCAN算法进行如下实验：
  - (1) 选择不同的 $\epsilon$ ，观察实验结果并分析原因；(1分)
  - (2) 选择不同的minPots，观察实验结果并分析原因；(1分)
5. 根据三个数据集的实验结果，对比分析kmeans和DBSCAN聚类算法。
6. (可选) 编写OPTICS密度聚类算法，绘制以上3个数据集的聚类结果。(2分)

【作业要求】

1. 本次作业需要编程完成，可选用python、matlab等语言，不可以使用现有的聚类函数。python可使用的函数库为matplotlib、numpy、scipy.io等；
2. 作业最后需要提交程序代码和作业包括，具体要求如下：
  - (1) 程序保证可运行，运行结果与报告中列举的结果一致，代码中需要必要的注释；
  - (2) 作业报告需要包含：kmeans聚类结果、kmeans探究结果和分析、DBSCAN聚类结果、DBSCAN探究结果和分析、两种聚类结果的对比分析等；
3. 程序和报告打包上交，命名格式为“姓名\_班级\_学号”；
4. 作业**独立完成**，按时提交。抄袭按照学校有关规定进行处理，迟交会进行一定的扣分。

### 【附录】

python可用scipy.io对mat文件进行数据读写，读写data.mat文件中三个数据集的代码如下：

```
import scipy.io as io
data=io.loadmat('data.mat')
data1,data2,data3=data['data1'],data['data2'],data['data3']
```