聚类算法：

一、算法步骤：

1．选取个中心点是在数据点值域内随机选取的。

2．计算n个数据点与K个中心点的距离，选取距离最小的中心点i，将该点附上属性i。

3．移动中心点：对于每个中心点i，计算所有属性为i的数据点的平均点。

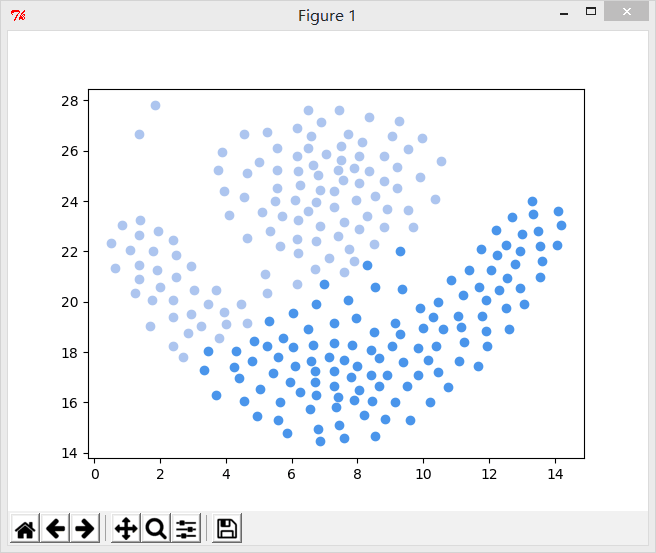
即,j是属性为i的数据点，numi为属性为i的数据点的个数。

4．重复2,3过程，直到中心点不再移动。

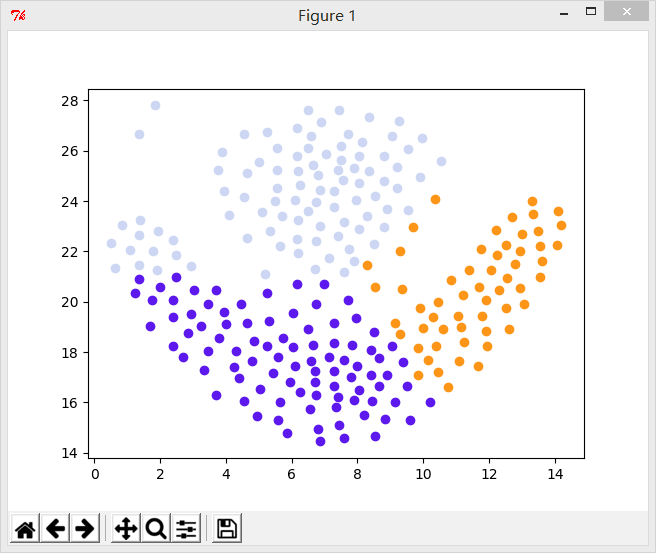
不足处：受到随机分布的中心点影响很大，每一次运行结果差很多。

运行结果：

当k=2时：

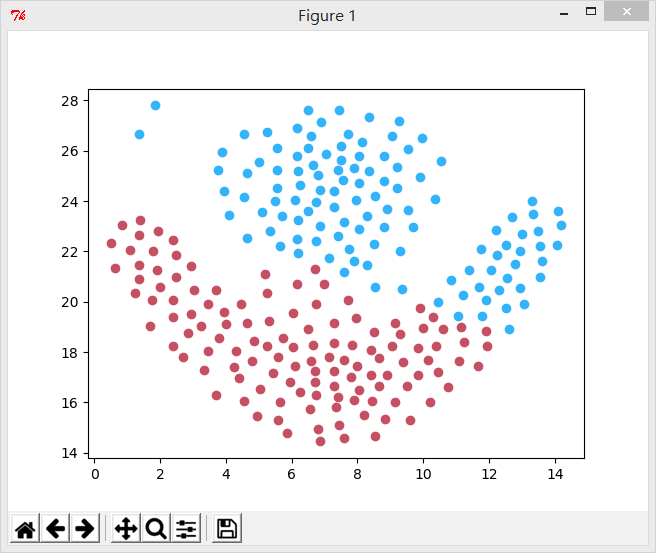


当k=3时：

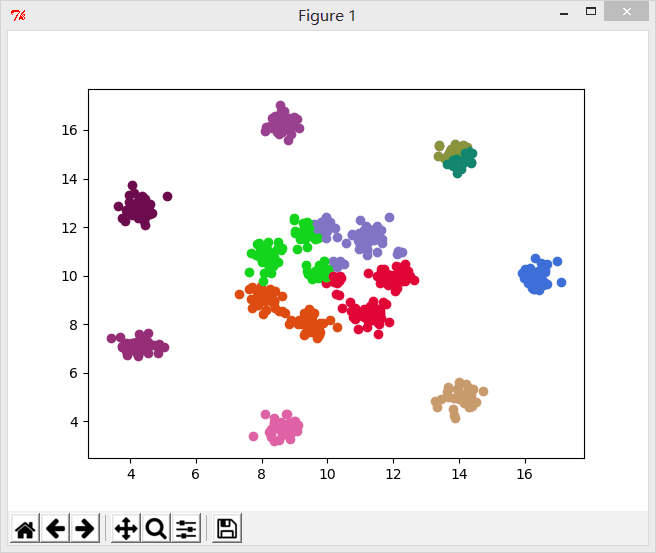
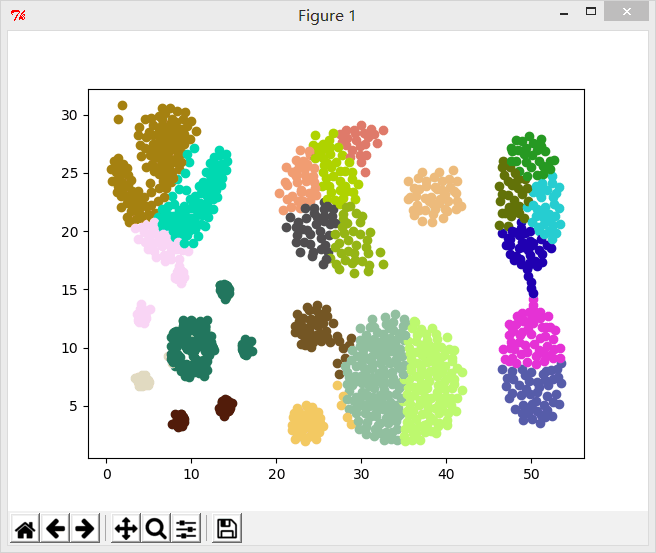
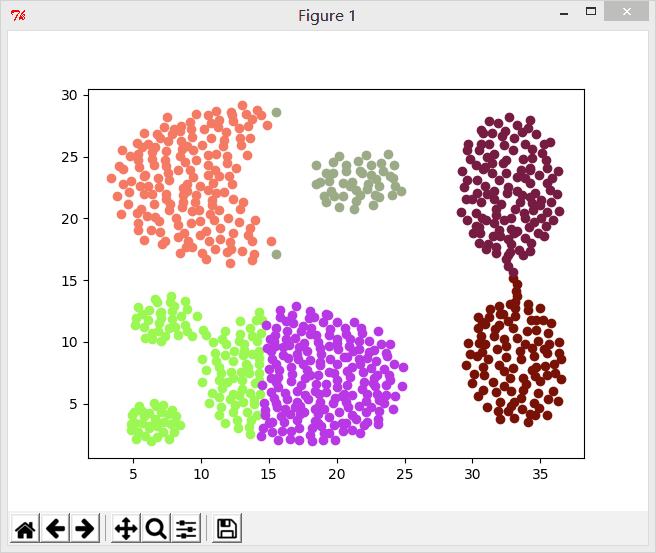


优化：本次实验尝试把随机取k个中心点的方式改成局部随机但是大致在图中均匀分布的方式取点。

效果如下：



其他点集的运行结果：



可以看出，Kmeans有3点不足：

1. 随机取的中心点对最后的聚类结果有很重要的影响，但是每张图的最佳选点方式是不同的，均匀取点并不能优化所有类型的图的聚类结果，甚至更糟。
2. 对于凹形图等有特别形状的点集，kmeans并不能很好地将这个形状的点集分出来，会把突出来的部分聚到它类。
3. 如果中心距离太近，会把一个类分为多类。
4. 不能有效分出噪音点。