Spectrum clustering——谱聚类算法：

一、算法步骤：

1. 得到数据的二维数组data[][] 行表示数据点，列表示维数。
2. 算出关系矩阵，此时关系矩阵算法是。

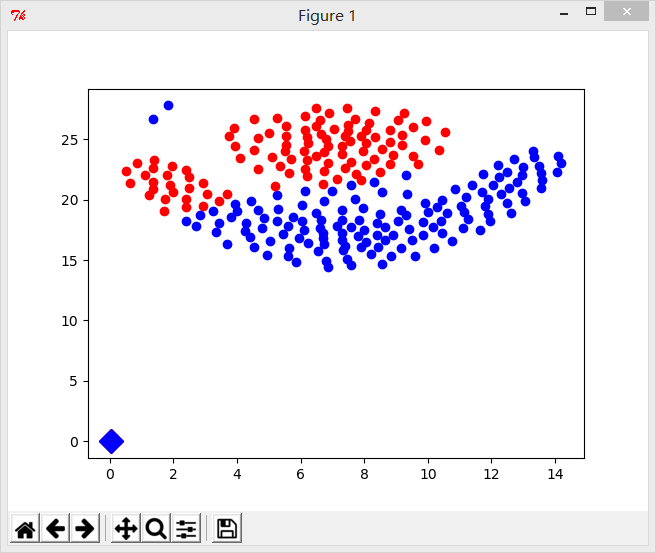
W：

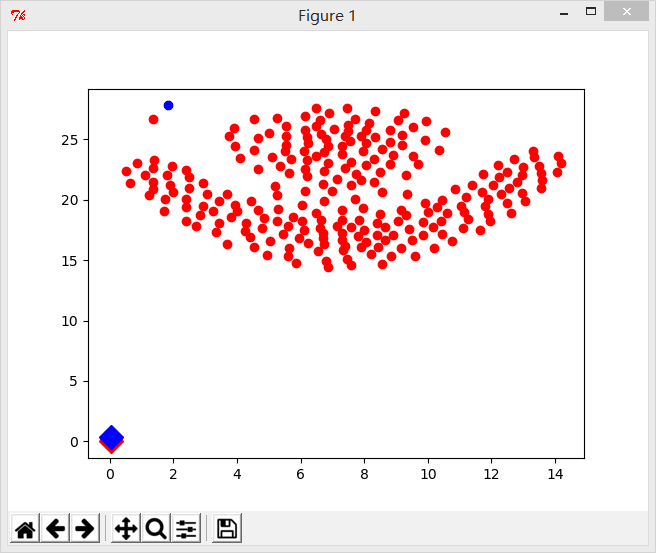
该关系矩阵对任意两点的距离进行了变换，使之聚类更明显。

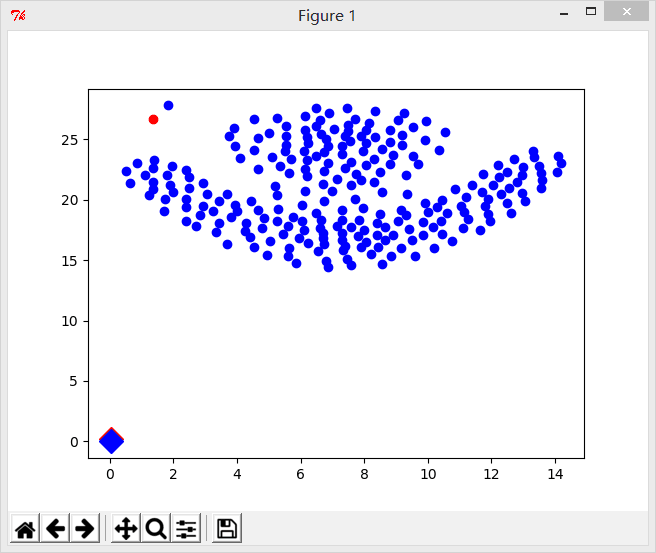
1. 算出度矩阵，度矩阵是一个对角函数：
2. 算出拉普拉斯矩阵：。
3. 算出的特征值和特征向量矩阵，将从大到小排列，取出前大特征值，然后将按此顺序取出最大特征值对应的个特征向量，得到矩阵，其中表示每个点的K1近邻点 其他点的距离视为无穷远。
4. 将正规化。
5. 输入到算法中，将其聚类。

算法选取个中心点是在数据点值域内随机选取的。

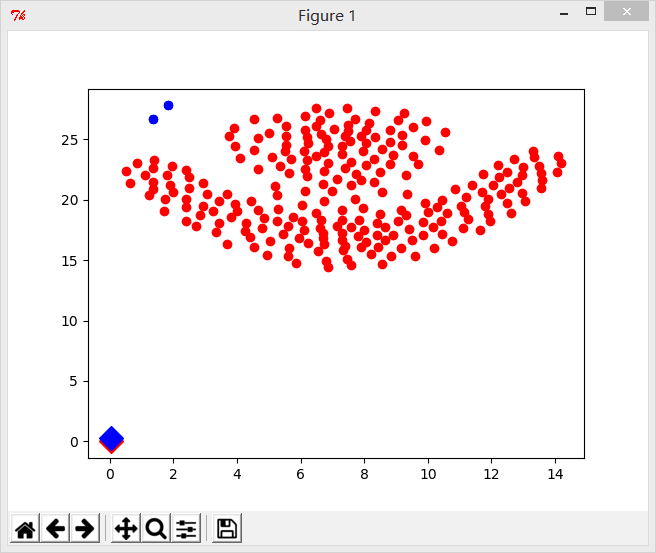
二、实验参数：（图片变形是因为选取的随机中心点是经过正规化的，位置都在0-1中）



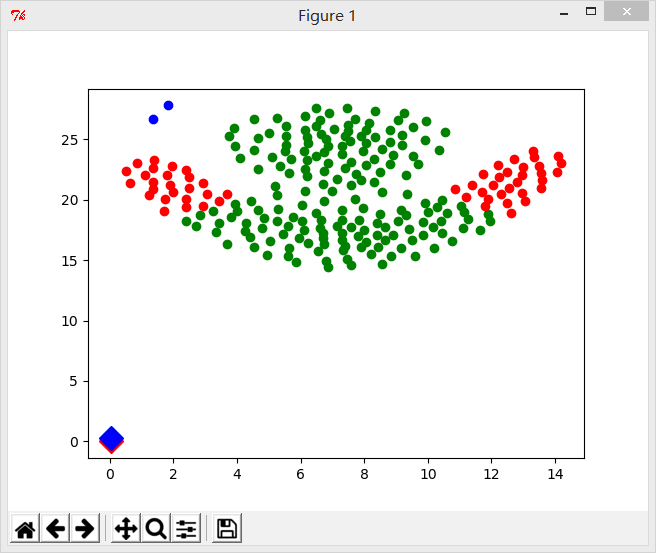


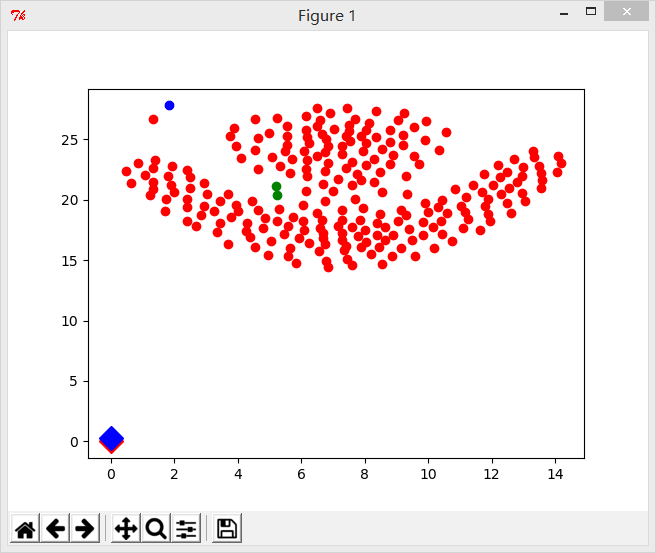


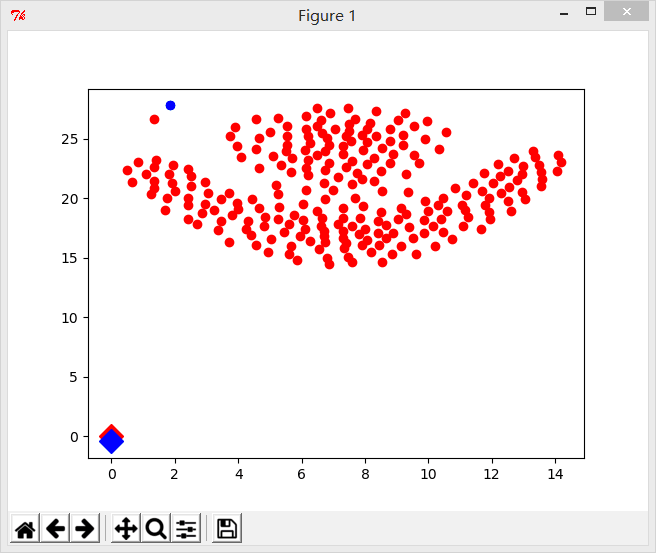
可以看出在这种参数下，聚类效果并不好，受到随机分布中心点的影响很大，但是比起K-means，在谱变形过后，可以更好地分出噪音点

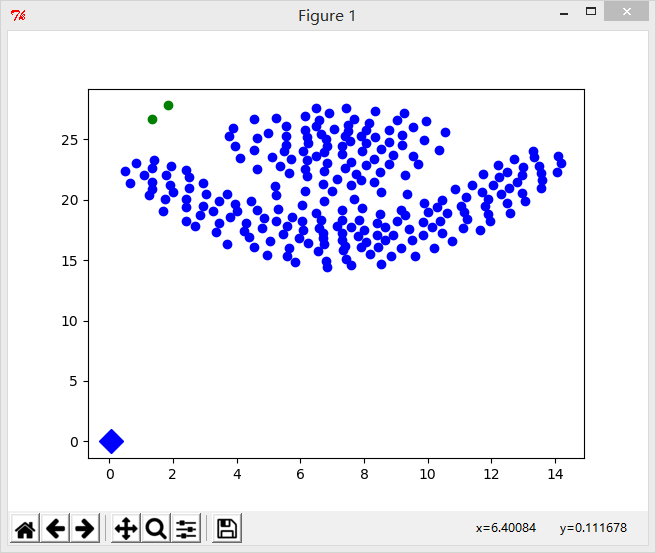


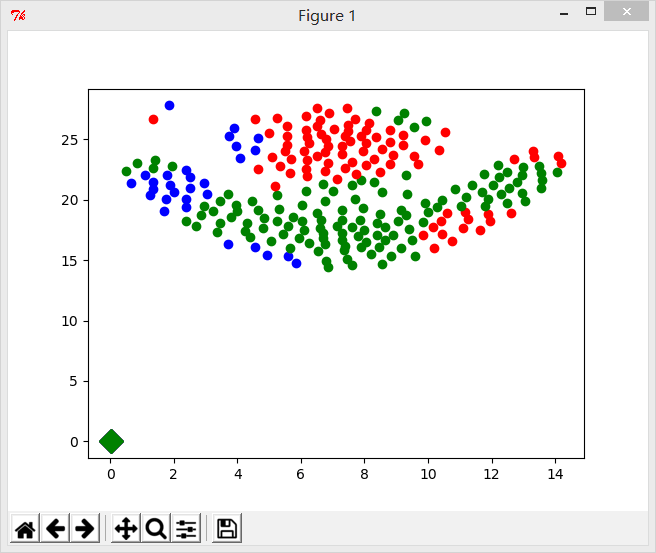
可以很好地分出噪音点

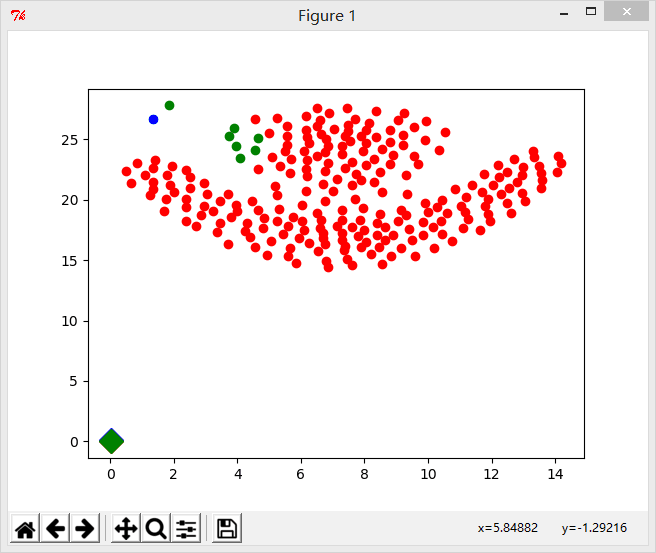


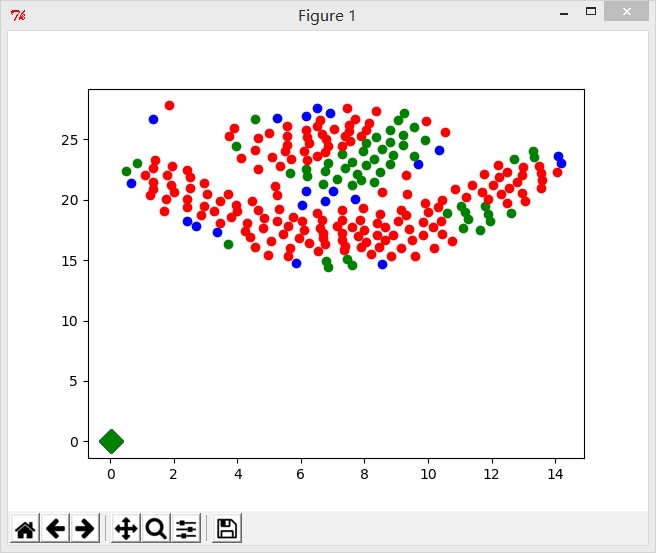






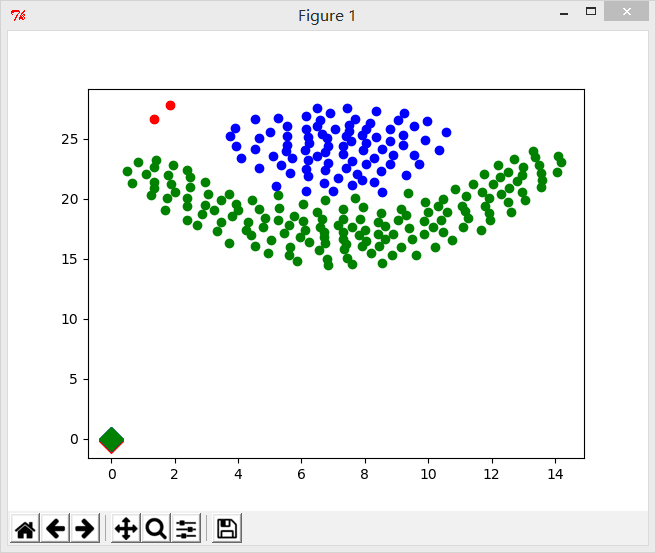




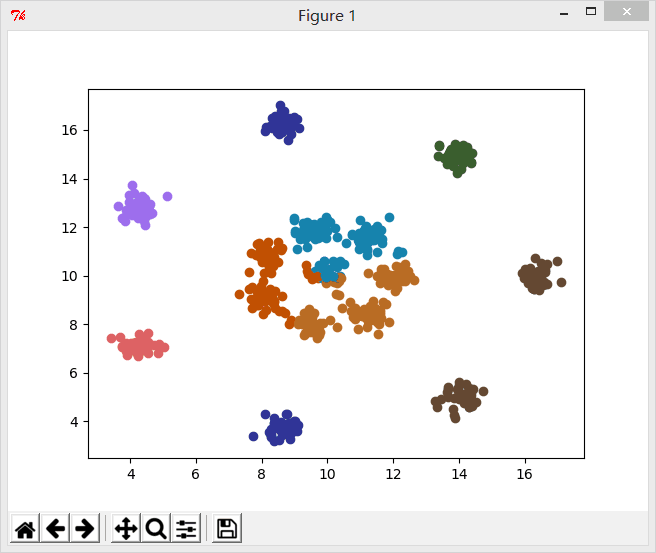
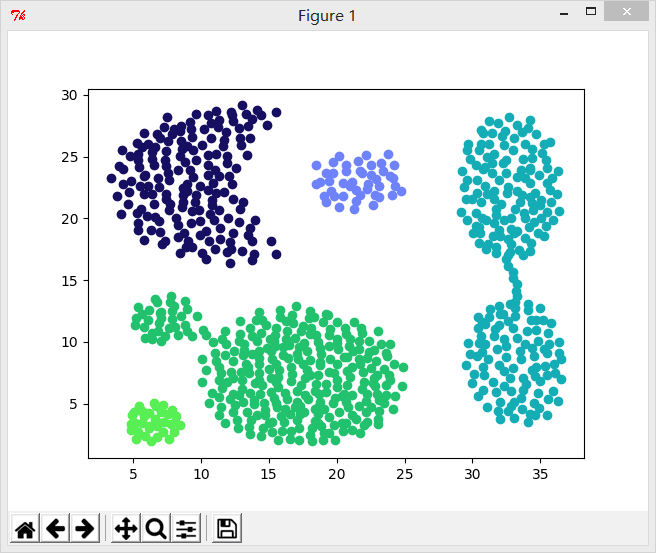
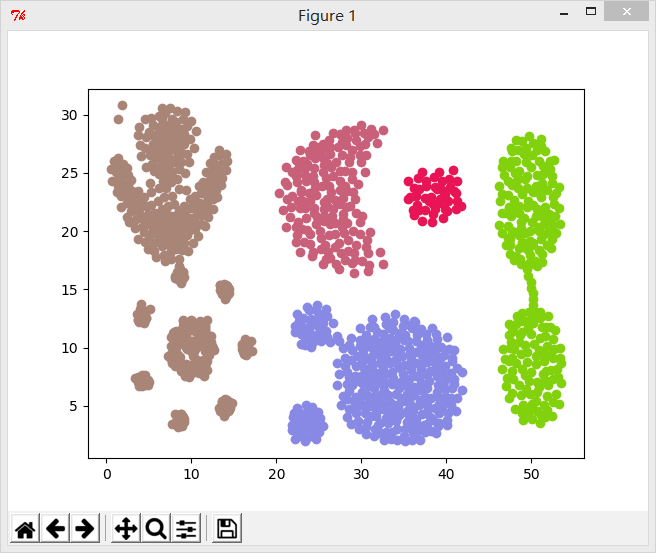


经过不懈的努力和尝试，在参数为

下 运行出结果：



在该参数下运行的其他三张图的效果：



谱聚类受影响很多，经试验sigma取1附近左右效果很好，k在5-10个左右，还有中心点个数，在不知道群数时目前采用随机算法。

在本次实验得出，特征向量矩阵是否正规化影响不大。

可以看出，谱聚类相比kmeans效果要好：

1. 它可以很好的把每一群分好类，不会把完整，分界清晰并且密集的一个类群分隔开，即使它存在凹形形状，但是两个群聚得太紧，如果随机中心点分布不佳，就很难分开。
2. 在某种程度上可以有效地分离噪音点。

待优化：

1. 每一张图的三个参数最优值是不同的，目前无法做到对于每张图给出不同的合适的参数。
2. 无法避免Kmeans算法随机中心点的选取影响大而每次运行结果不同，无法达到最佳效果的问题。
3. 运行时间太长，需要运算每个数据相互之间的关系矩阵，度矩阵从而求出拉布拉斯矩阵，还需要计算特征值，特征向量。数据量太大的数据需要运行很长时间。