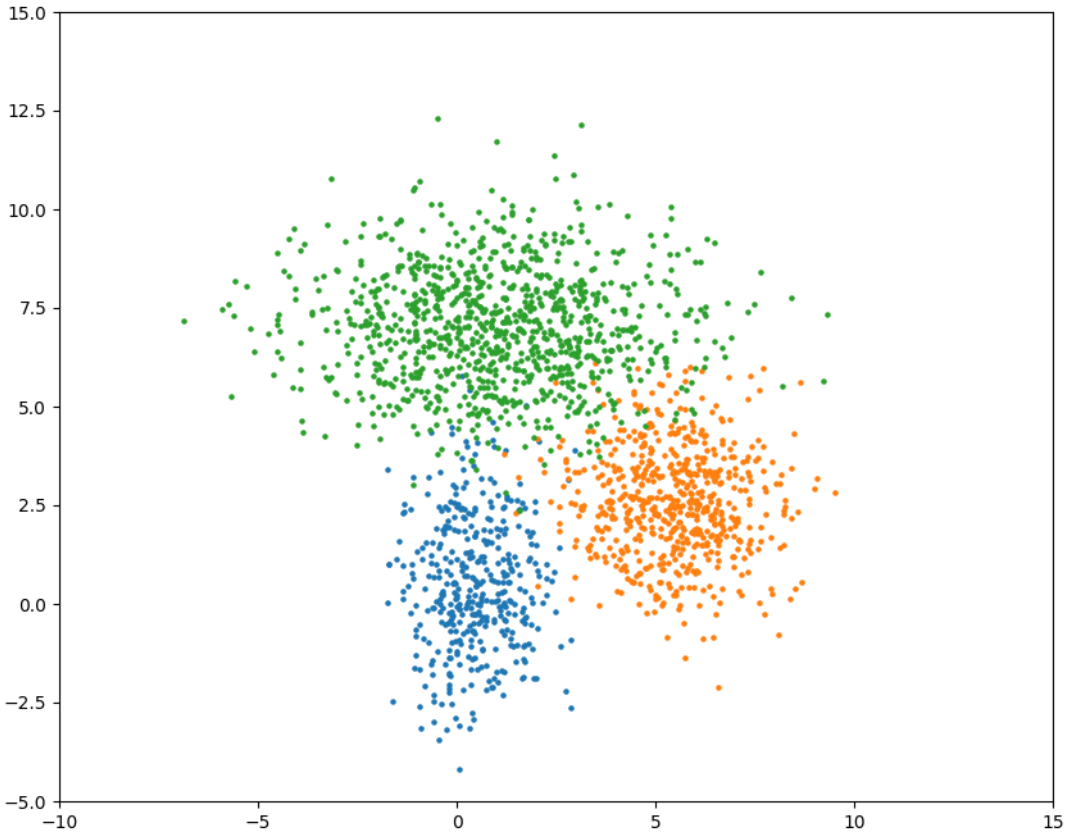
第三章作业

GT（适用于以下三种聚类方法的测试）：

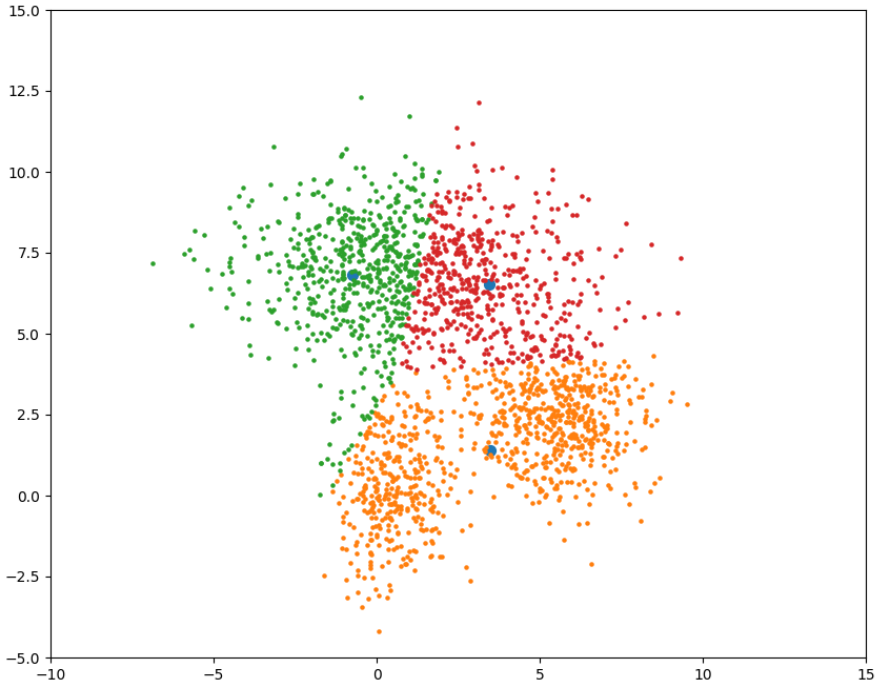


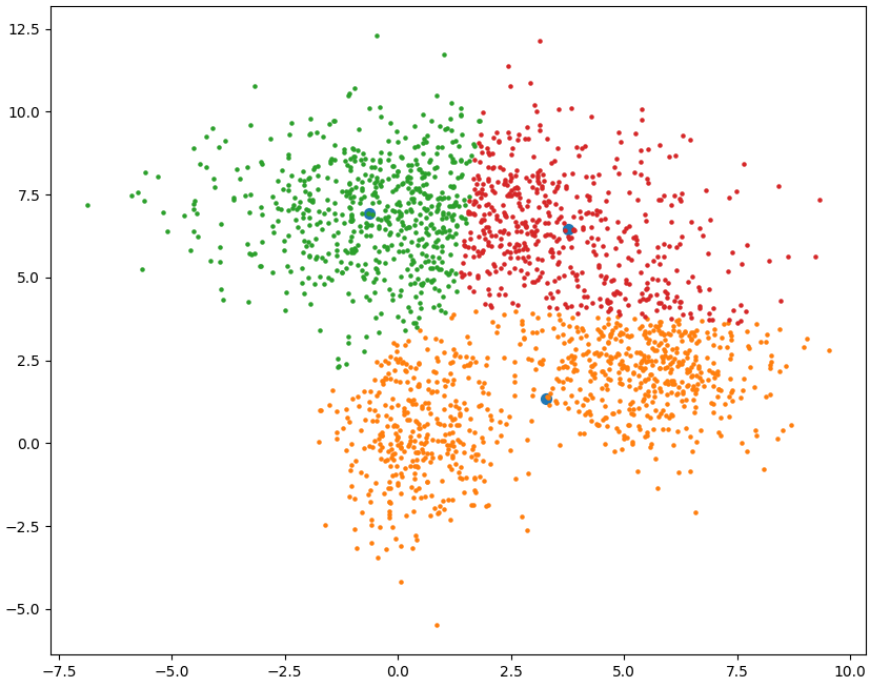
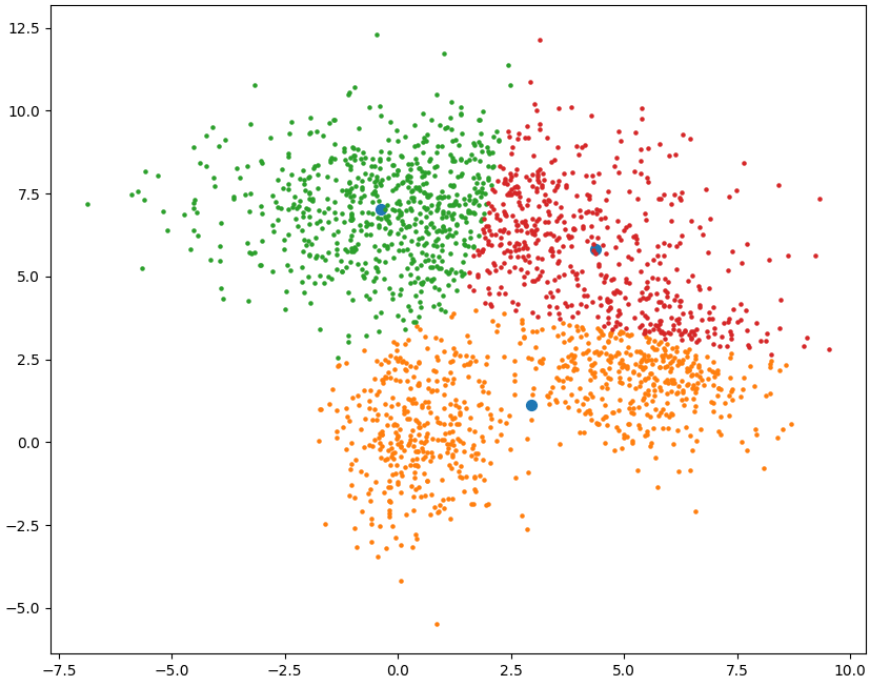
初始化方式：

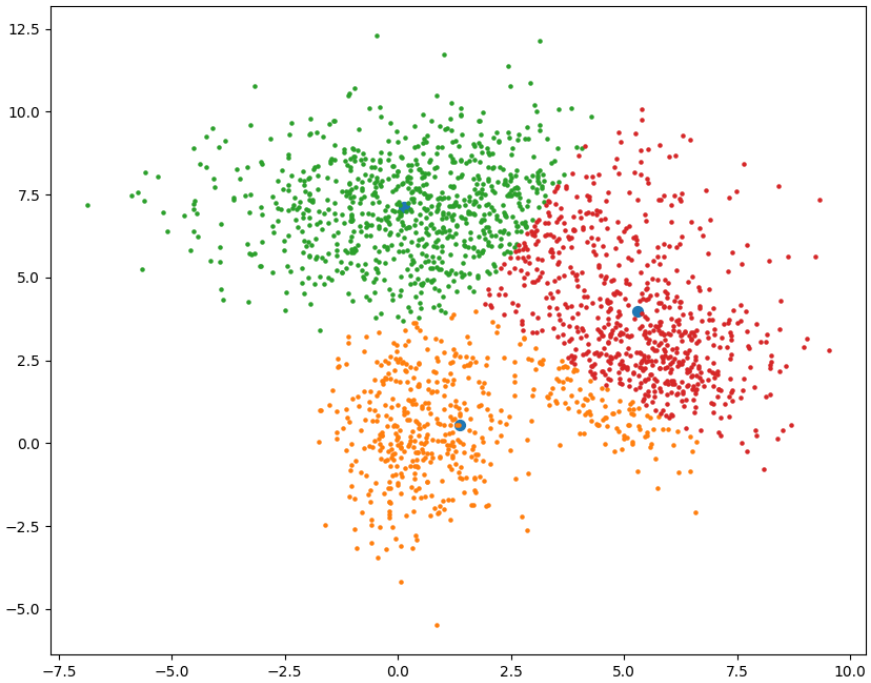
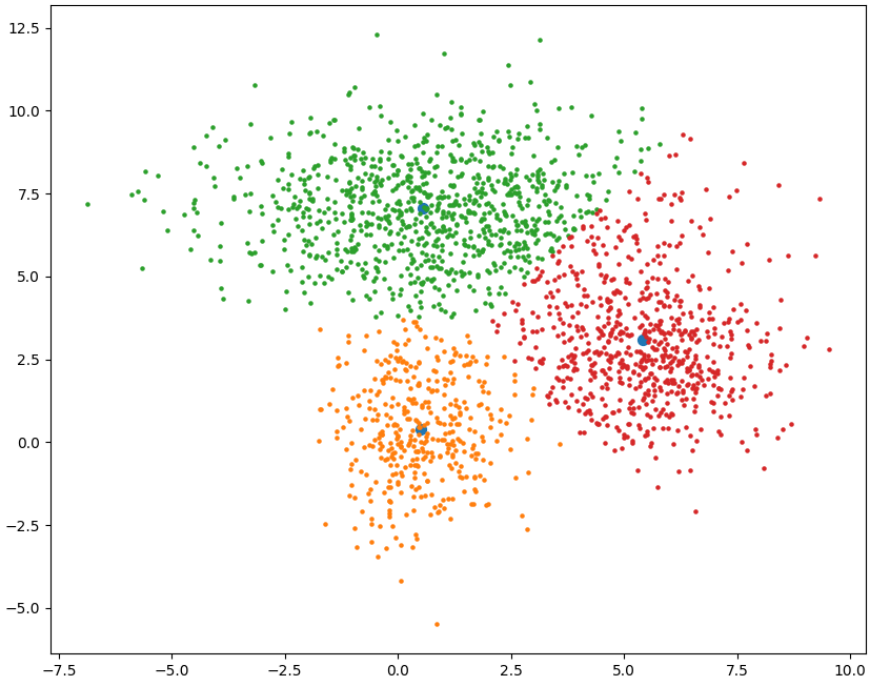
- Centers：从所有数据点中随机选择三个点
- 从初始化后开始的每个iteration中，保持每个点归属于离它距离最近的一个cluster

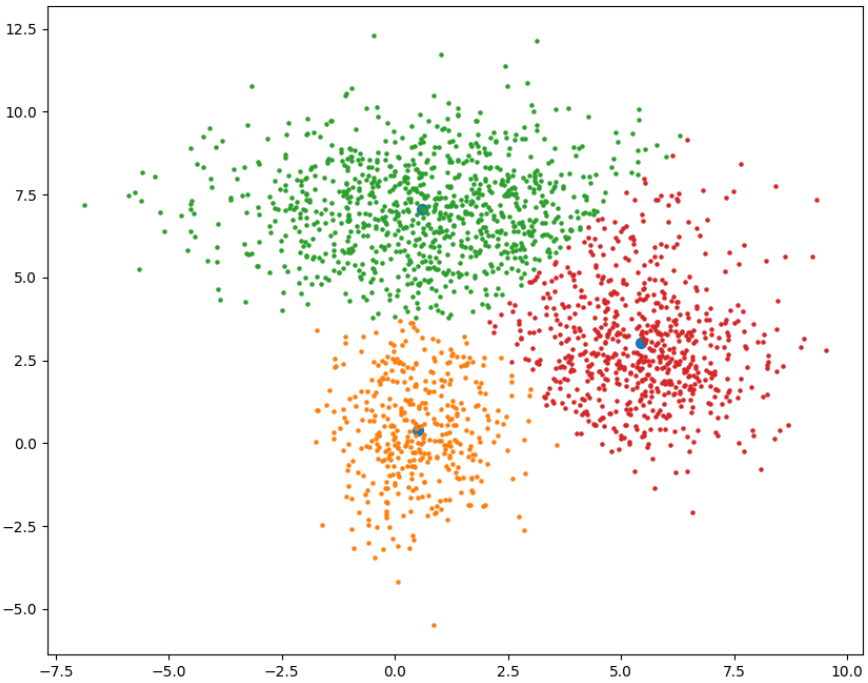
1. KMeans

（1）初始化较好的情况

NUMBER OF ITERATION	RESULT
Init(0)	

NUMBER OF ITERATION	RESULT
4	
8	

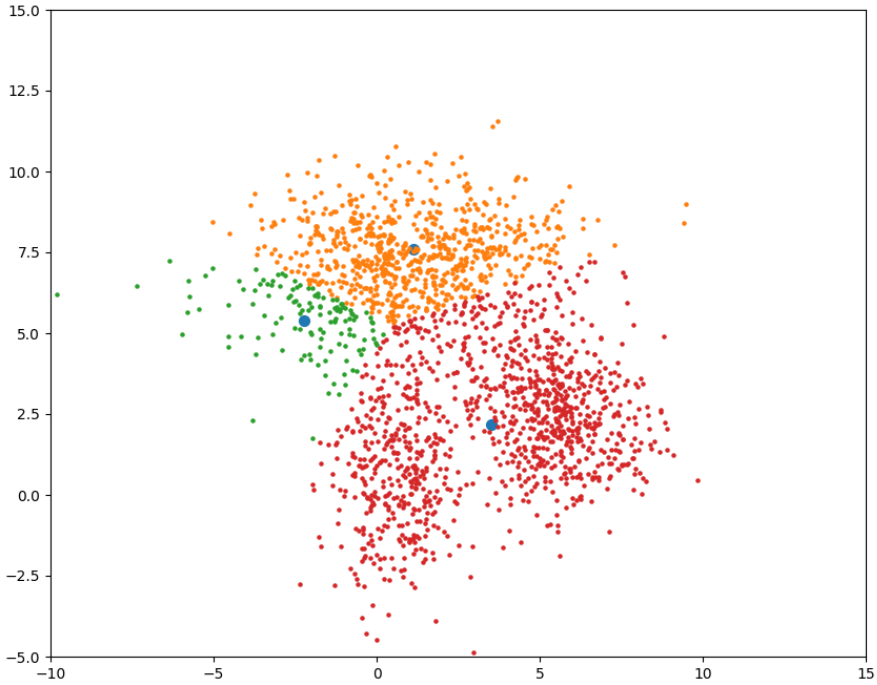
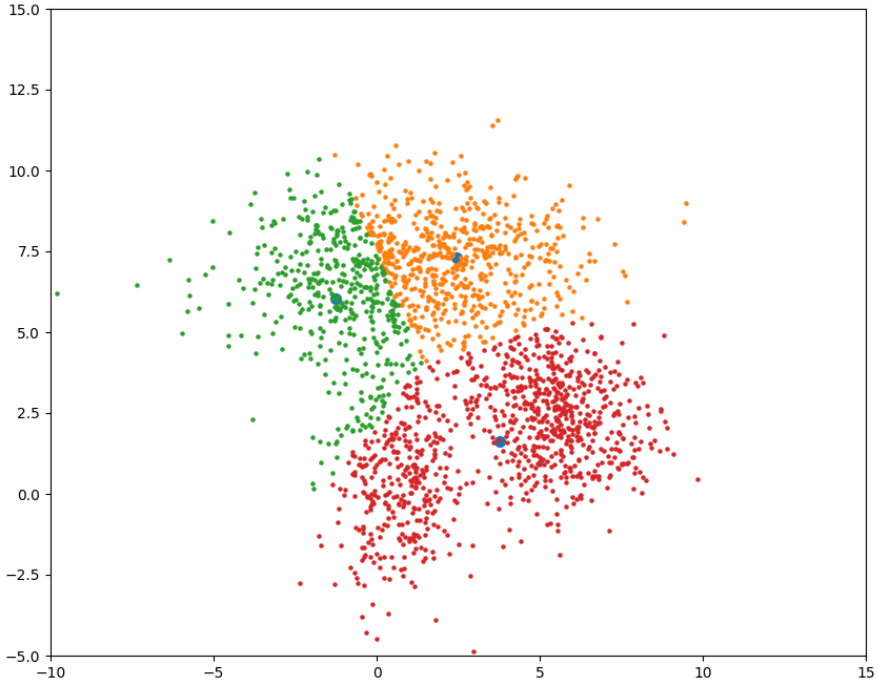
NUMBER OF ITERATION	RESULT
12	
16	

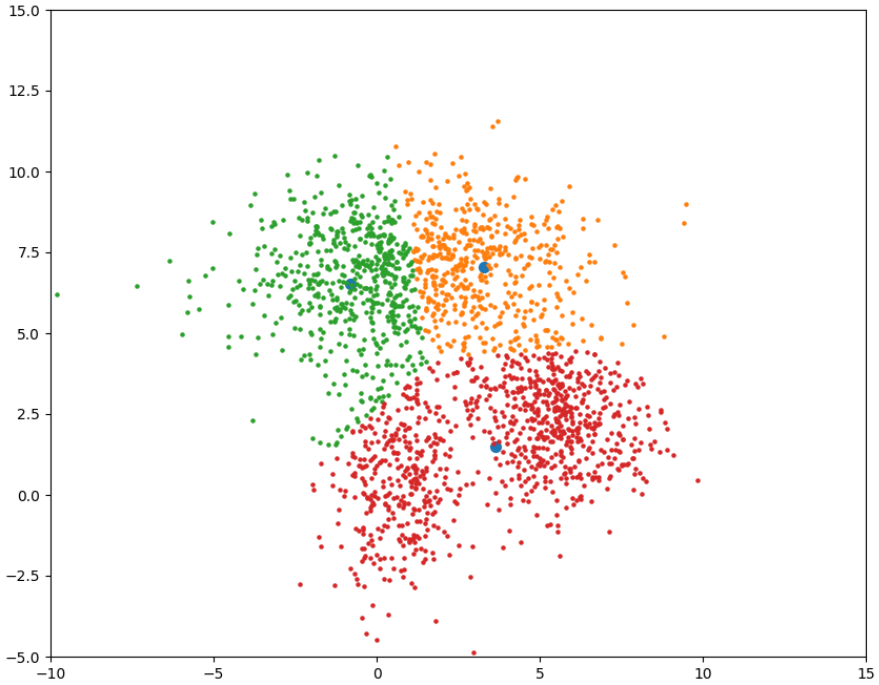
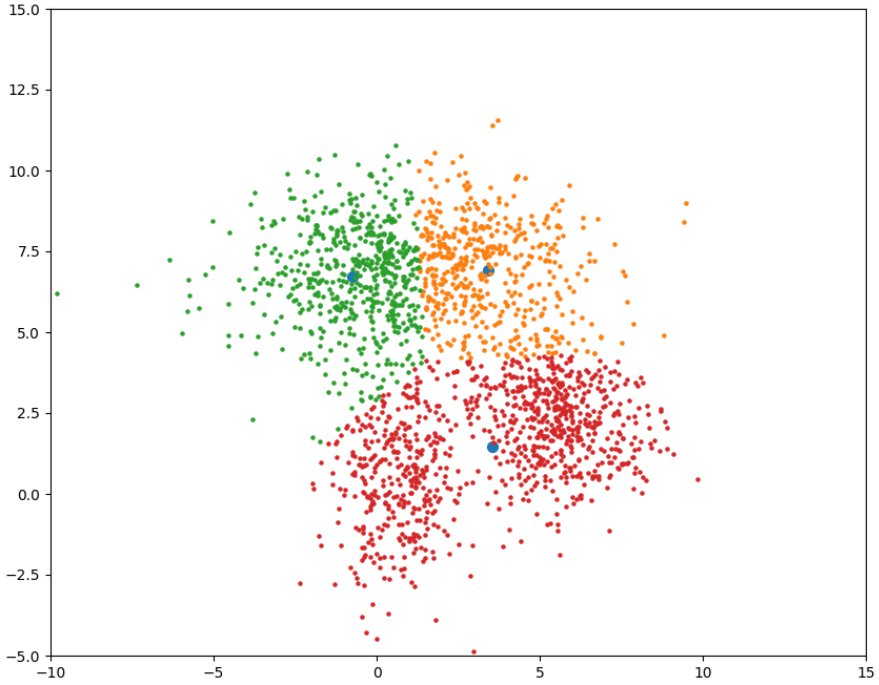
NUMBER OF ITERATION	RESULT
Last(18)	

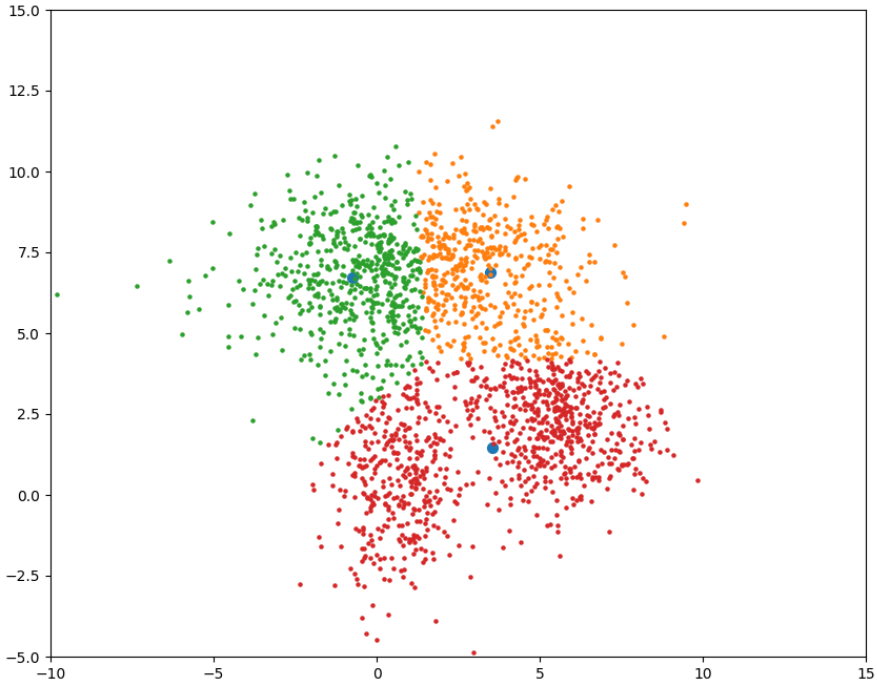
可以看到当初始化较好的时候，很快就可以收敛

（2）初始化不好的情况

NUMBER OF ITERATION	RESULT

NUMBER OF ITERATION	RESULT
Init(0)	
4	

NUMBER OF ITERATION	RESULT
8	
12	

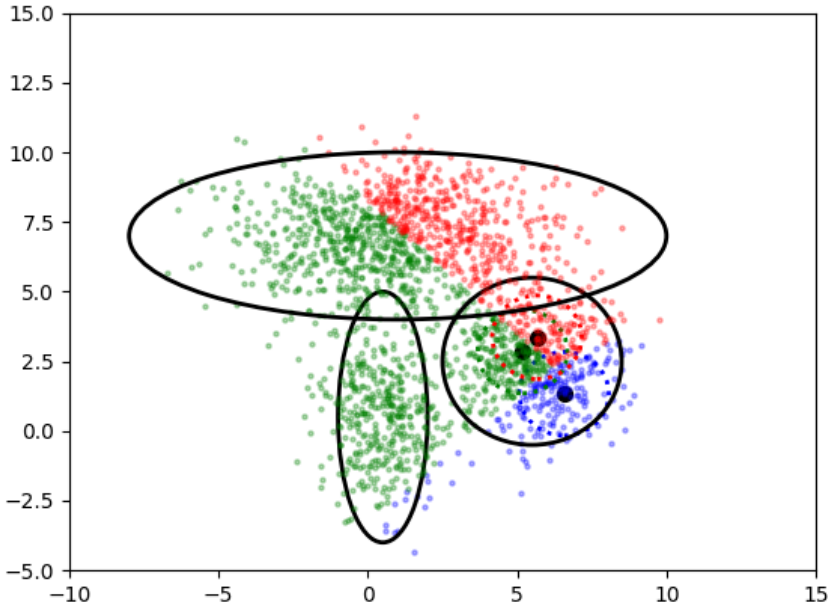
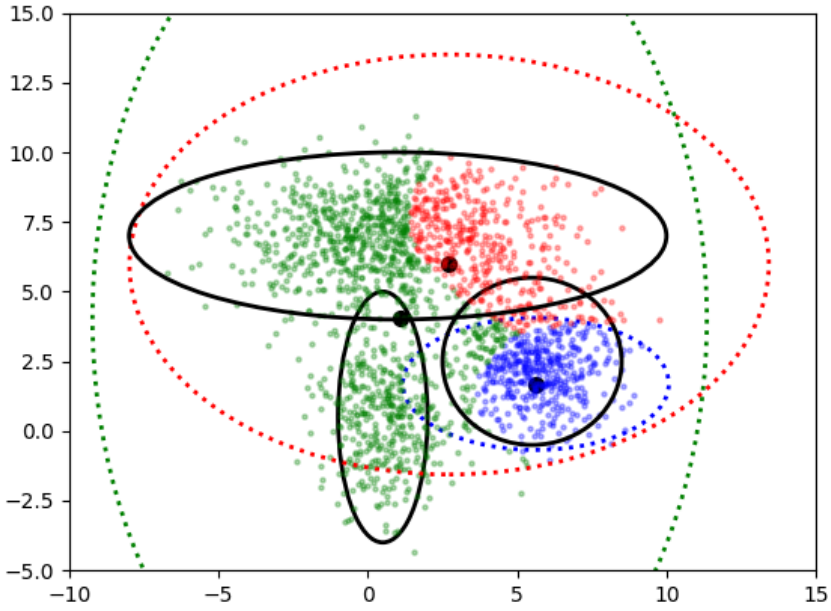
NUMBER OF ITERATION	RESULT
Last(15)	

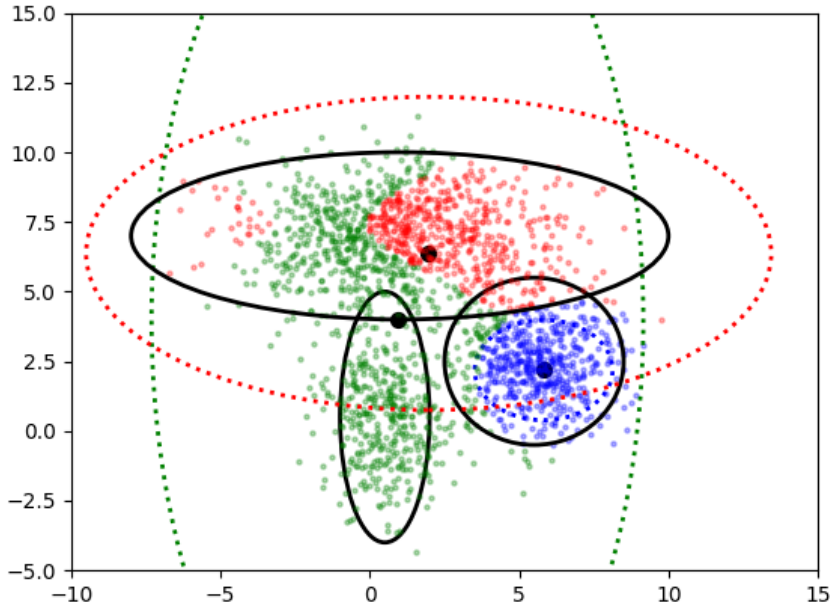
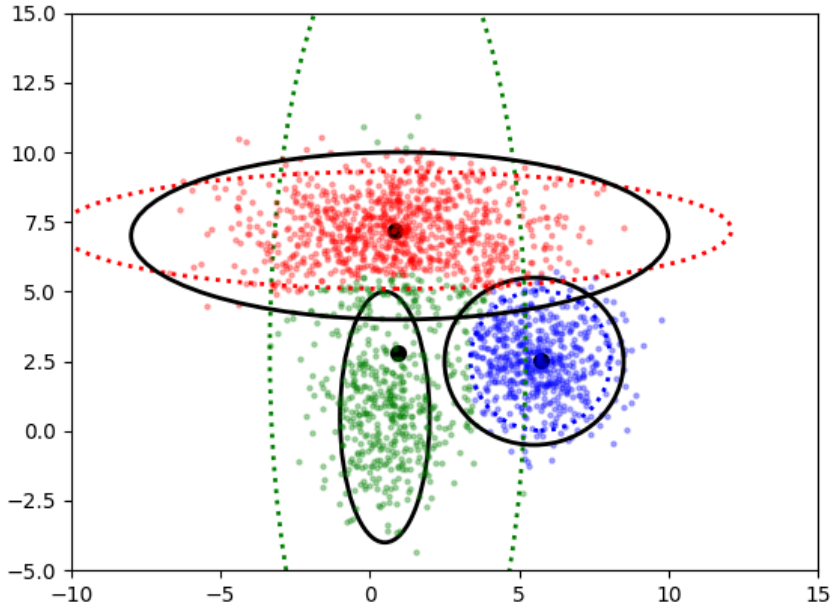
可以看到在初始化不良的情况下算法收敛在了一个局部最优解（为了方便测试tolerance设置的略大了，在tolerance足够小，iteration足够大的情况下有走出局部最优解的趋势）

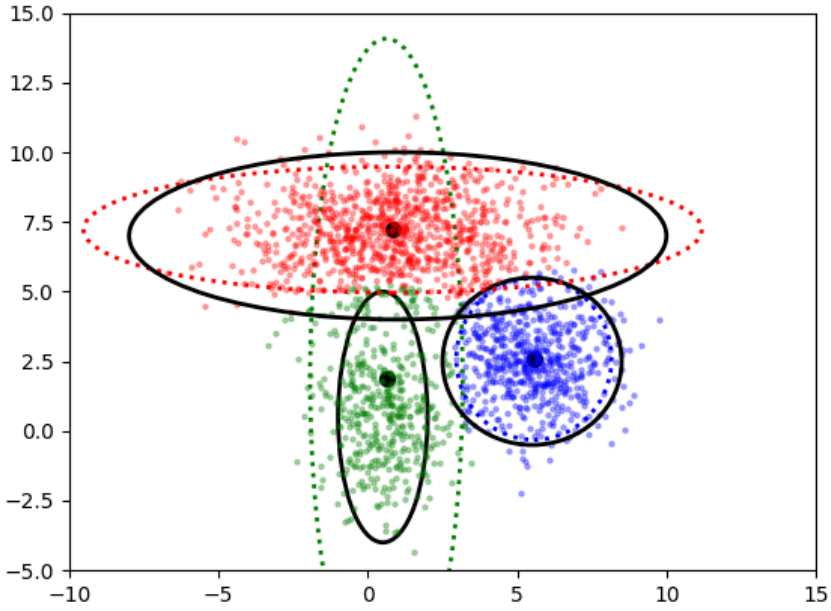
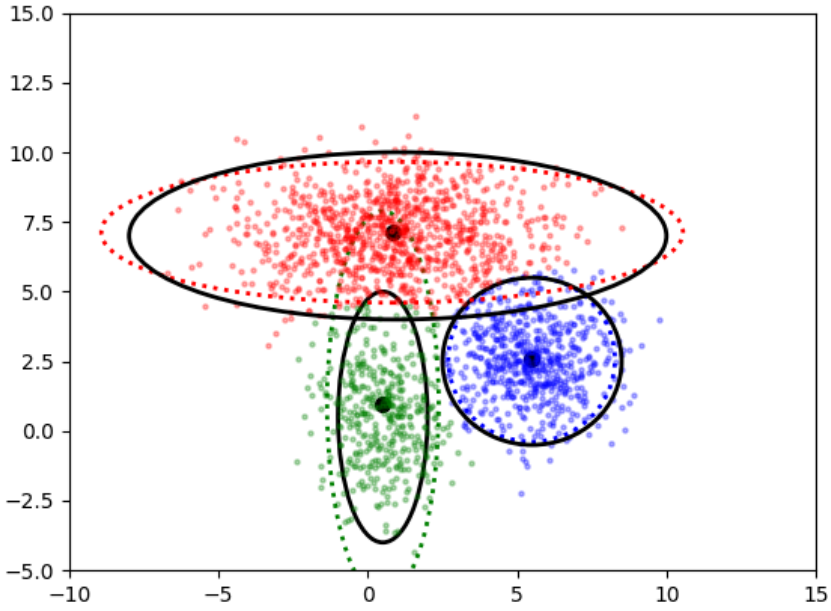
2. GMM

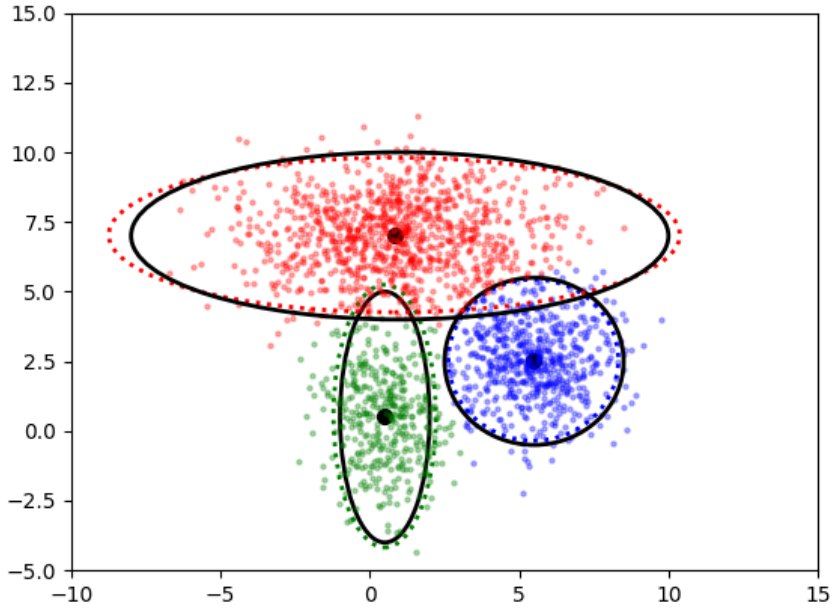
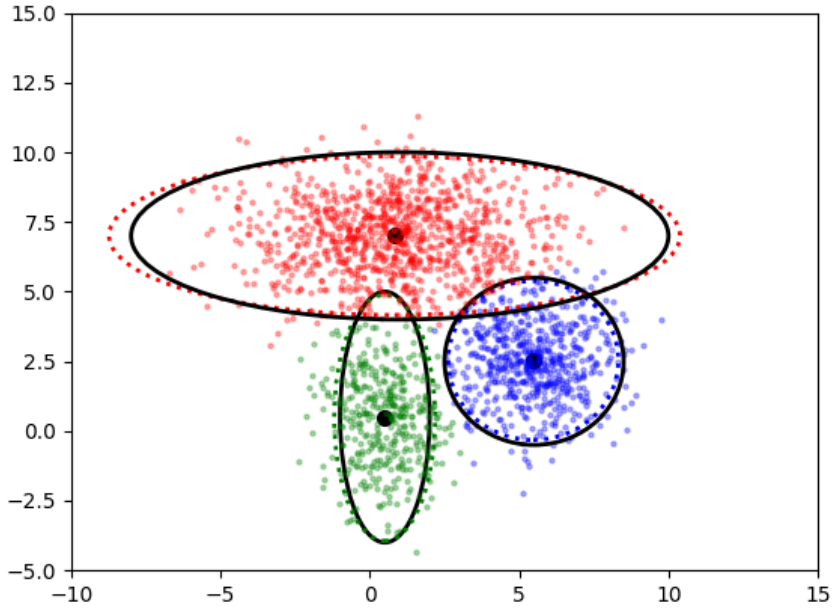
（1）初始化较好的情况

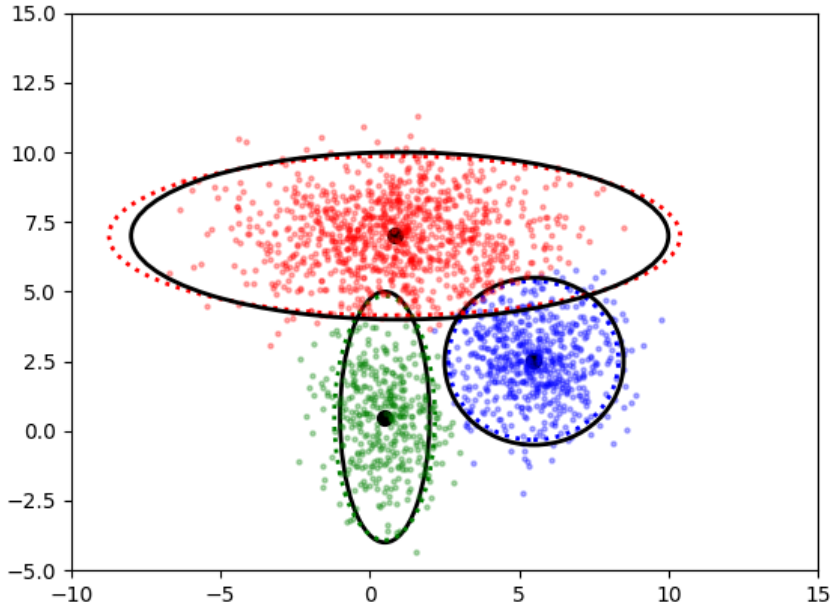
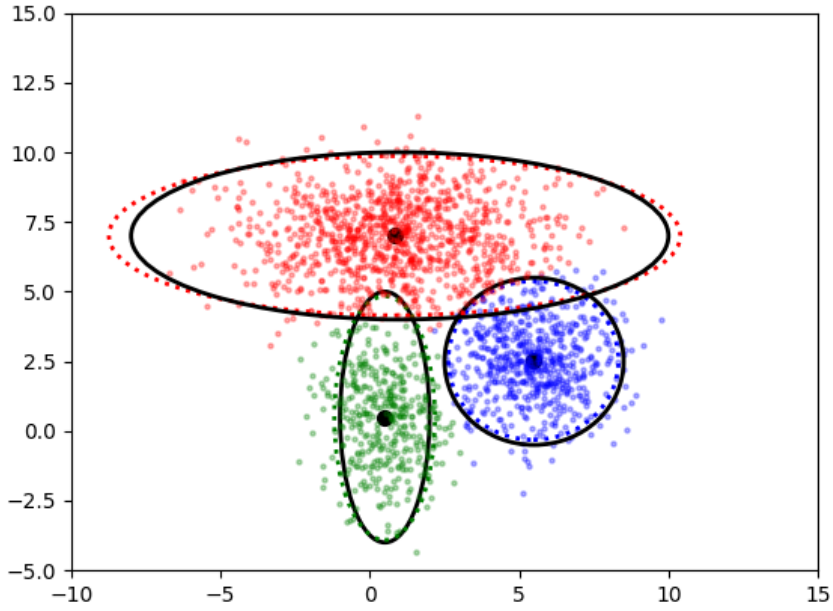
NUMBER OF ITERATION	RESULT

NUMBER OF ITERATION	RESULT
Init(0)	
3	

NUMBER OF ITERATION	RESULT
5	
10	

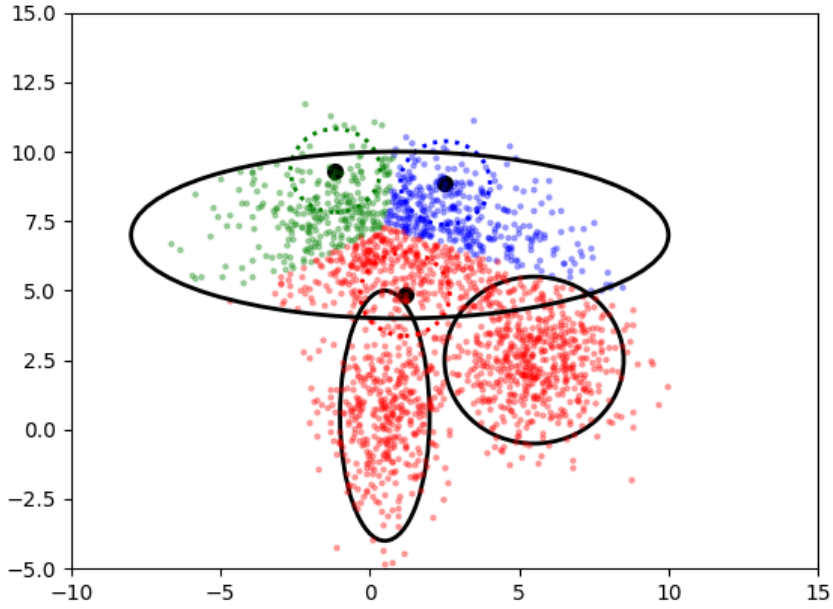
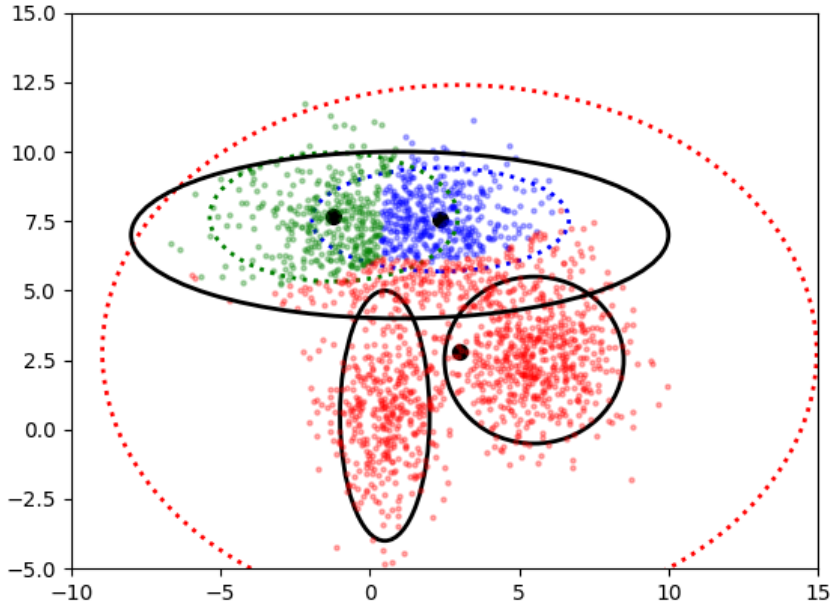
NUMBER OF ITERATION	RESULT
15	
20	

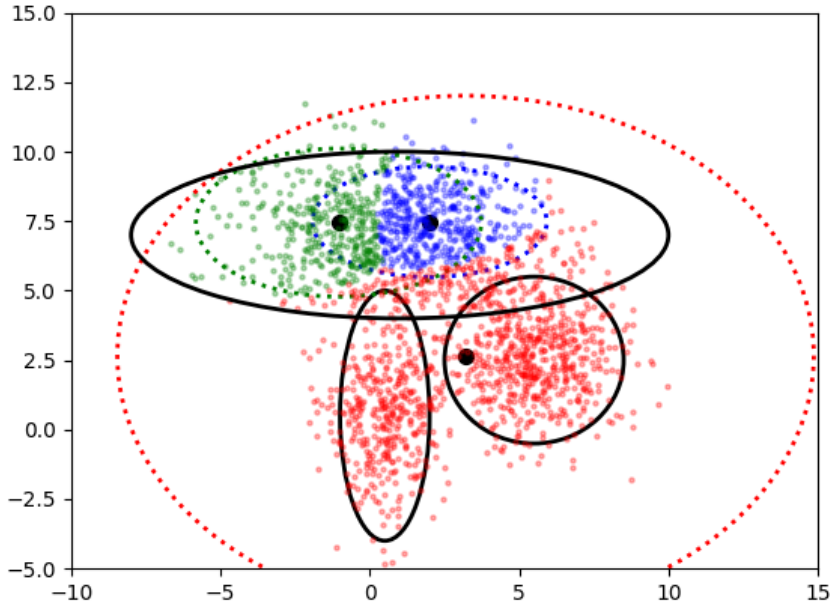
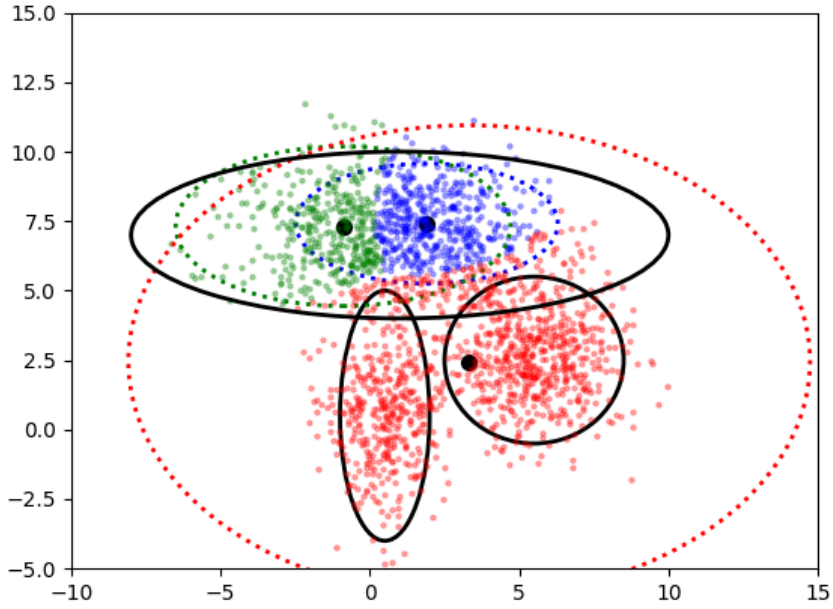
NUMBER OF ITERATION	RESULT
25	
30	

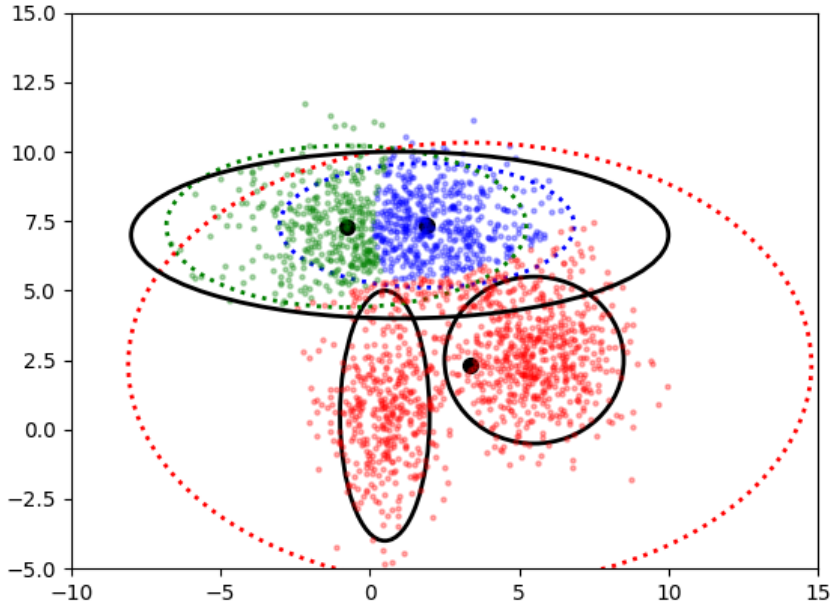
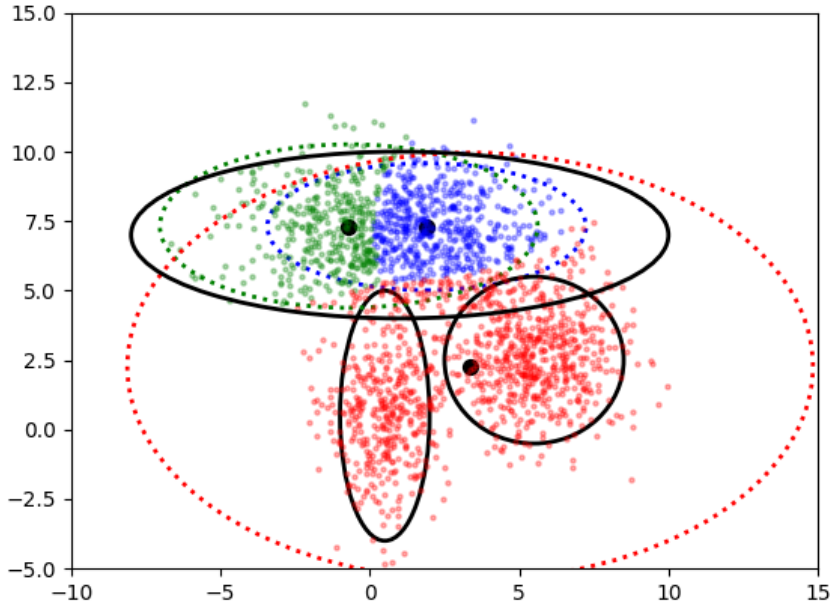
NUMBER OF ITERATION	RESULT
35	
Last(37)	

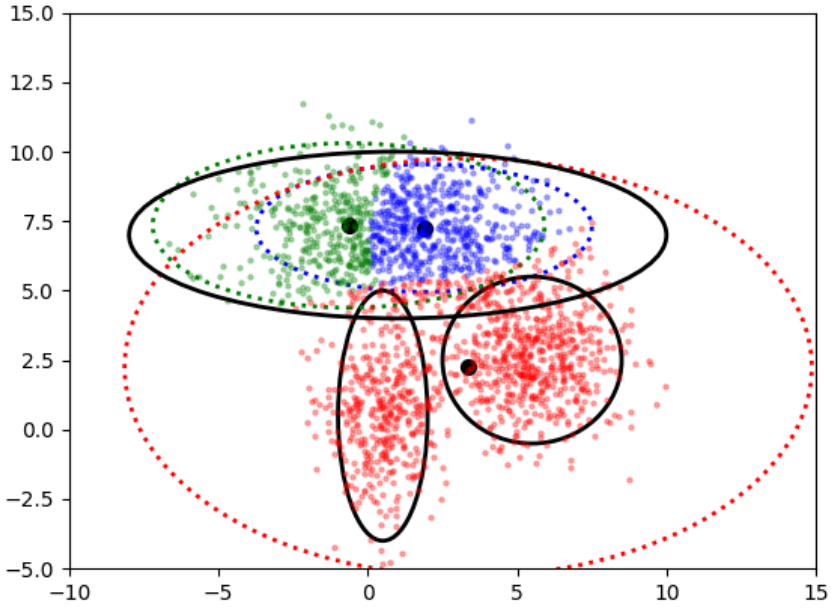
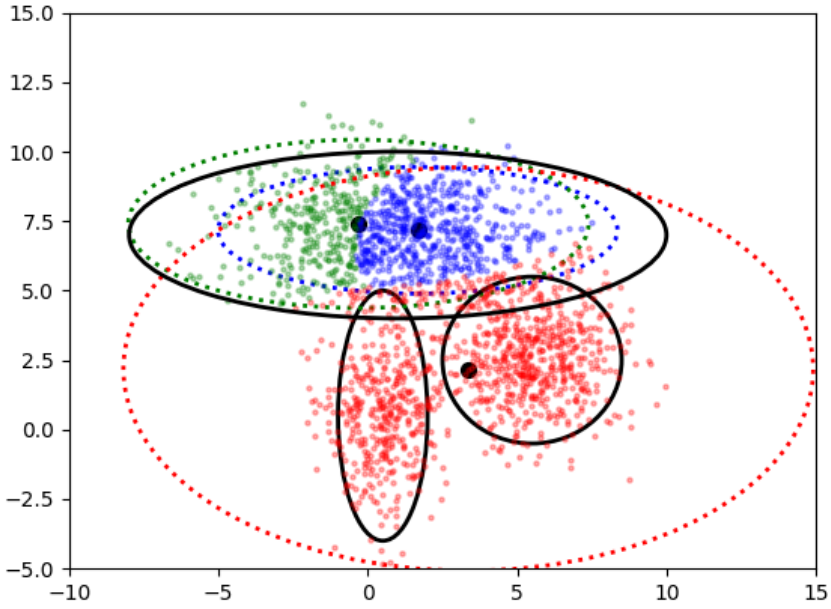
（2）初始化不良的情况

NUMBER OF ITERATION	RESULT

NUMBER OF ITERATION	RESULT
Init(0)	
3	

NUMBER OF ITERATION	RESULT
5	
10	

NUMBER OF ITERATION	RESULT
15	
20	

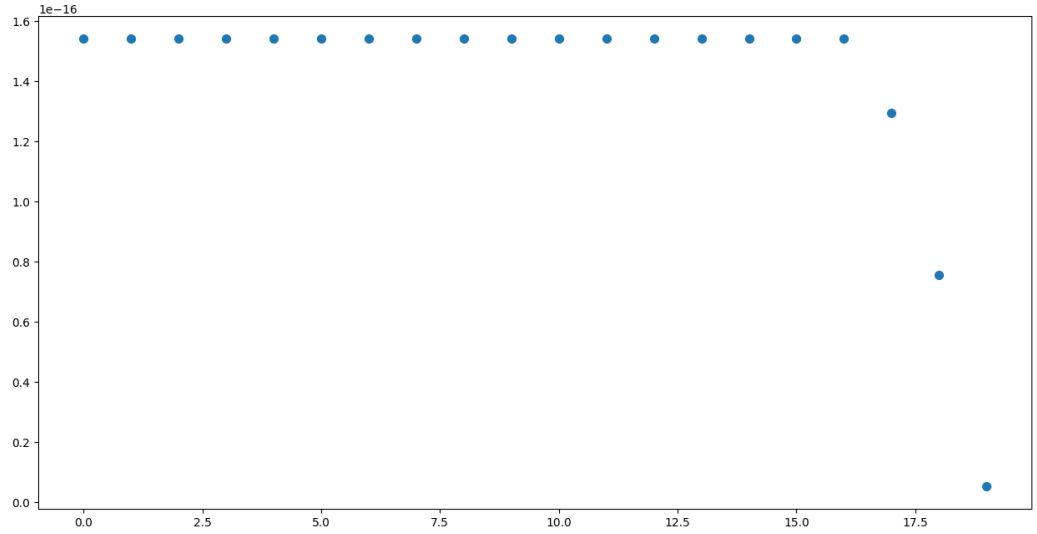
NUMBER OF ITERATION	RESULT
25	
Last(100)	

可以看出当初始化有两个中心非常靠近的时候导致算法陷入了局部最优解

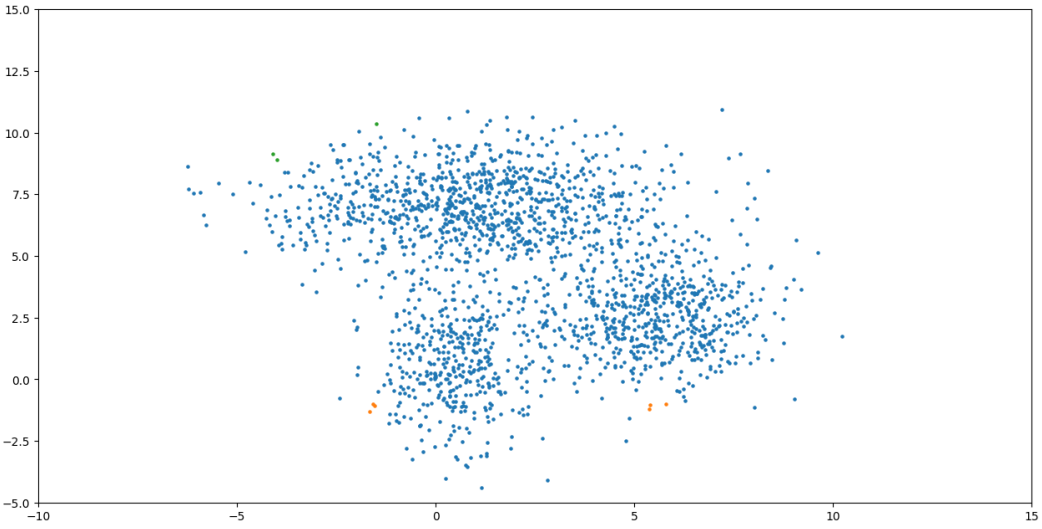
3. Spectral Clustering (unnormalized)

（1）Adjacency Matrix建立方式

- 对于w_{ij}的具体表示，采用高斯核函数RBF：$exp(-\frac{\|x_i-x_j\|_2^2}{2\sigma^2})$，测试发现这里的参数variance对算法表现影响很大，测试了从0.001~100的多个数值后发现var=0.05左右时算法表现较优
- Fully Connected

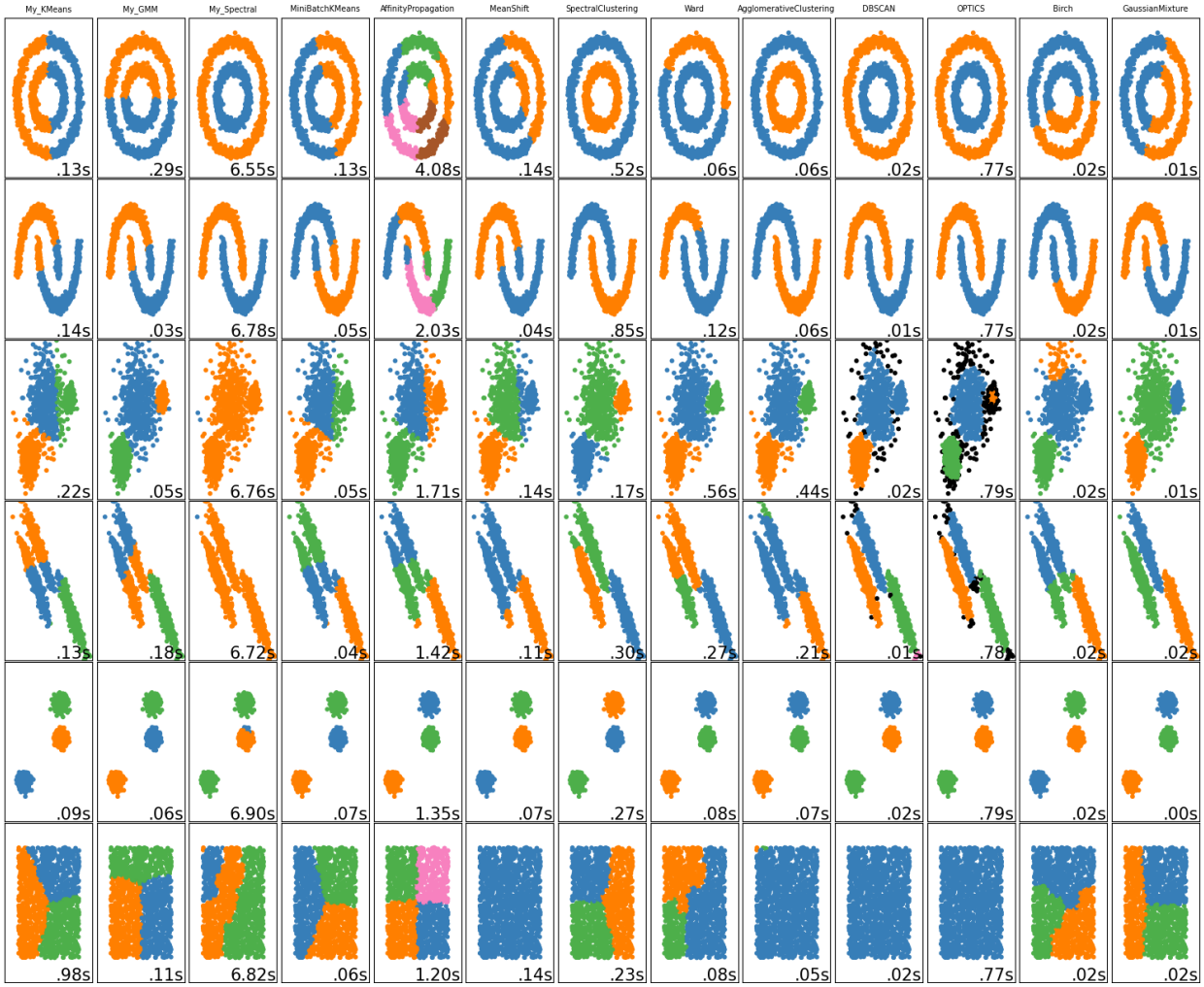


（2）Result



这个任务上的分类并不佳（同Clustering Comparison中的Task 3），考虑可能是variance的选择不好以及全连接的方式导致（拜托老师帮忙检查一下代码有没有问题？）

4. Clustering Comparison



可以看到My_Spectral（第三列）在前两个task上的表现达到了预期（variance=0.007），但在task3,4上表现不良，请问老师这是什么原因？