1、成员：

郑家民、华祖海

2、题目：

用C++实现k-means聚类算法，

1. 对实验二中的z-score归一化的成绩数据进行测试，观察聚类为2类，3类，4类，5类的结果，观察得出什么结论？

2. 由老师给出测试数据，进行测试，并画出可视化出散点图，类中心，类半径，并分析聚为几类合适。

样例数据(x,y)数据对：

3.45 7.08

1.76 7.24

4.29 9.55

3.35 6.65

3.17 6.41

3.68 5.99

2.11 4.08

2.58 7.10

3.45 7.88

6.17 5.40

4.20 6.46

5.87 3.87

5.47 2.21

5.97 3.62

6.24 3.06

6.89 2.41

5.38 2.32

5.13 2.73

7.26 4.19

6.32 3.62

找到聚类中心后，判断(2,6)是属于哪一类？

1. 环境

tset1.txt为样本数据

#include<iostream>

#include<cmath>

#include<vector>

#include<ctime>

#include<fstream>

4、总结

通过本次实验，收获了许多，K-Means是聚类算法中的最常用的一种，算法最大的特点是简单，好理解，运算速度快。一开始有点对分类和聚类的概念混淆。在经过两个实验后，终于理清了分类与聚类，它们之间最大的区别就是是否存在预定义；对于分类来说，在对数据集分类时，是知道这个数据集是有多少种类的，而对于聚类来说，在对数据集操作时，我们要做的，是将数据集中相似的数据归纳在一起。