**武汉大学遥感信息工程学院**

**2015年研究生复试计算机能力测试**

1. **编程实现**

选用K均值算法对一串整形数据（100行，100列）进行聚类。

输出两个结果文件：

1）第一个输出结果文件为cluster\_centers.txt，其中输出聚类得到的各区域（聚类）的中心，以及每个聚类区域像素个数。格式如下：

聚类序号 聚类平均横坐标x 聚类平均纵坐标y 聚类中心值 像素个数

1 X Y V N

......

2）第二个输出结果文件为cluster\_labels.txt，其中输出整串数据的聚类序号，即100X100的聚类序号图，每个元素的聚类序号用空格分开，每一行聚类序号对应每一行输入数据。格式如下：

1 1 1 1 1 1 ...... 1 1 1

2 2 1 1 2 2 ...... 2 2 3

.......

3 3 1 2 2 2 ...... 3 1 1

**二、数据说明**

整串数据存储在data.txt文件中，包括100行，100列，一共10000个整数，一行在txt文件中存储为一行，相邻元素之间用空格隔开。

1. **算法说明：**

**K均值算法实现步骤如下（本例中K=3）：**

设整串为一个模式样本集合{*x*}

1. 任选K个初始聚类中心；**(10,10), (70,30), (80,90)**
2. 逐个将影像中的每个像素按照最小距离原则分配给K个聚类中心，形成K个类群；
3. 由步骤2）聚类后计算新的聚类中心，即。式中，为第i个聚类域中的样本个数；
4. 若前后两次聚类中心不变，即，算法收敛，计算完毕，否则转步骤2）进行下一次迭代。
5. **编程要求**
6. 程序结构清晰
7. 模块划分合理
8. 注释清楚
9. **考试要求**
10. 考试时间2小时。
11. 在计算机E盘建立考试目录，目录名构成规则：考号，考生所有考试资料请放在该目录下，包括程序源代码，程序运行结果。考生本人承担一切由于错误存放考试资料而导致批阅老师不能正确评阅的责任；
12. 不得使用任何编程参考资料，包括MSDN，一经发现，以作弊论处。
13. 遵守考场纪律，不得相互讨论。