实验1 K均值聚类算法实现

* 1. 实验目的

通过编程和上机实验加深对无监督学习的理解和认识，掌握基于模型的聚类算法K均值算法的设计与实现；了解软件绘图的基本方法，掌握如何调用相关函数绘制二维平面图。

* 1. 实验环境

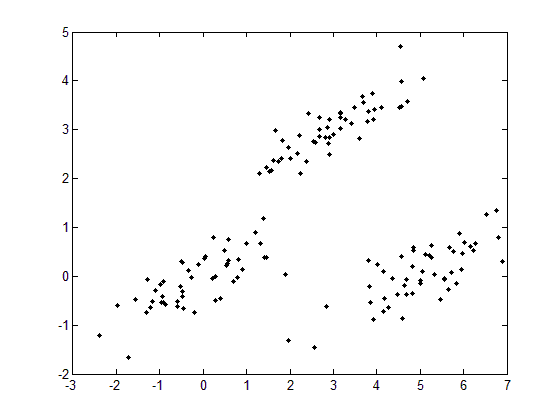
PC机，window操作系统，matlab2012编程环境（可选）。

* 1. 实验要求

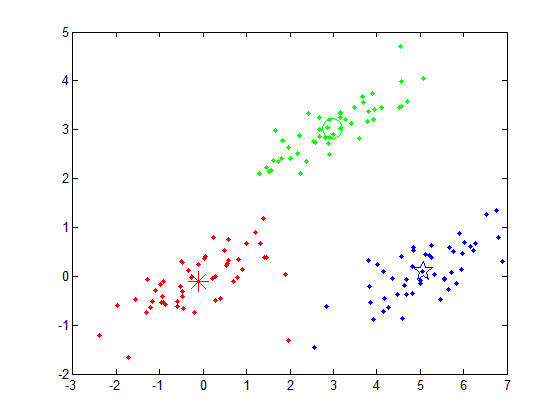
1. 掌握K均值聚类算法的原理；
2. 编程实现K均值聚类算法，并将实验结果显示出来。
   1. 实验步骤

说明：请同学按顺序依次完成（1）—（4）的编写工作，并向实验老师展示运行结果。这里以matlab中的函数为例，同学们可以选择其它的编程环境。

1. 调用mvnrnd函数生成三组满足高斯正态分布的二维数据，每组50个数据。第一个正态分布的参数为{均值=(0,0), 协方差矩阵=[0.9,0 .4; 0.4,0.3]}，第二个正态分布的参数为{均值= (5,0), 协方差矩阵= [0.9,0 .4; 0.4,0.3]}, 第三个正态分布的参数为{均值= (3,3), 协方差矩阵= [0.9,0 .4; 0.4,0.3]}。将三组数据合在一起，共150个数据，构成数据集合。将数据集合保存为data.mat文件，后续实验都在该数据集上完成。调用plot函数将150个数据显示在平面图中。



1. **编写程序实现K均值聚类算法**，设置聚类个数K=3，随机初始化聚类中心，调用plot函数显示聚类结果。不同的颜色代表不同的类别，聚类中心用标记符号（如五角星、十字、圆圈等）表示。



1. 在K=3时，再次运行程序，显示聚类结果。
2. 分别设置K=2、K=4，运行程序，显示聚类结果。
   1. 实验报告
3. 各个程序的源代码，要求对代码行进行**注释说明**；
4. 各个程序的运行结果；
5. 分析在聚类个数K=3时，多次运行程序的结果有何不同，并分析原因。
6. 分析当聚类个数K取不同值时，聚类结果有何不同，并分析原因。
   1. 注意事项及说明
7. 自己编写代码实现K均值函数，不允许直接调用K均值函数包；
8. 认真填写实验报告，不允许抄袭；
9. 遵守实验室各项制度，服从实验指导教师的安排，按规定的时间完成实验；
10. 建议学时数2学时。