

模式分类作业9_23

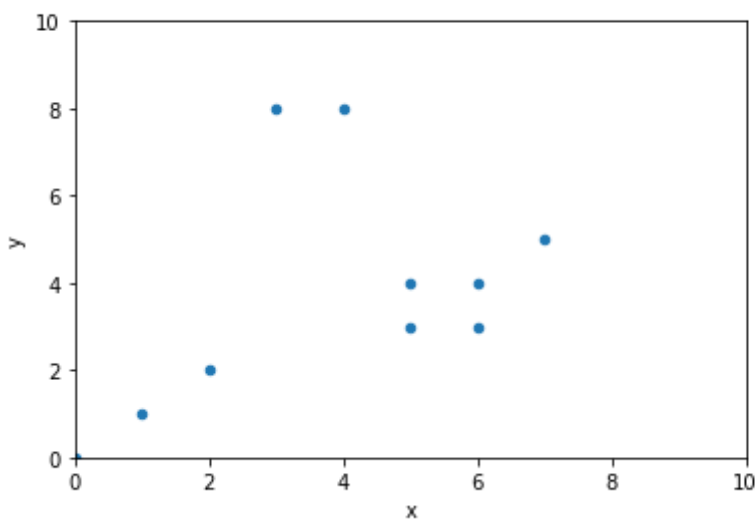
一. 题目描述

使用k-均值法对以下模式分布进行聚类分析，点集为：{X1[0, 0], X2[3, 8], X3[2, 2], X4[1, 1], X5[5, 3], X6[4, 8], X7[6,3], X8[5, 4], X9[6, 4], X10[7, 5]}。

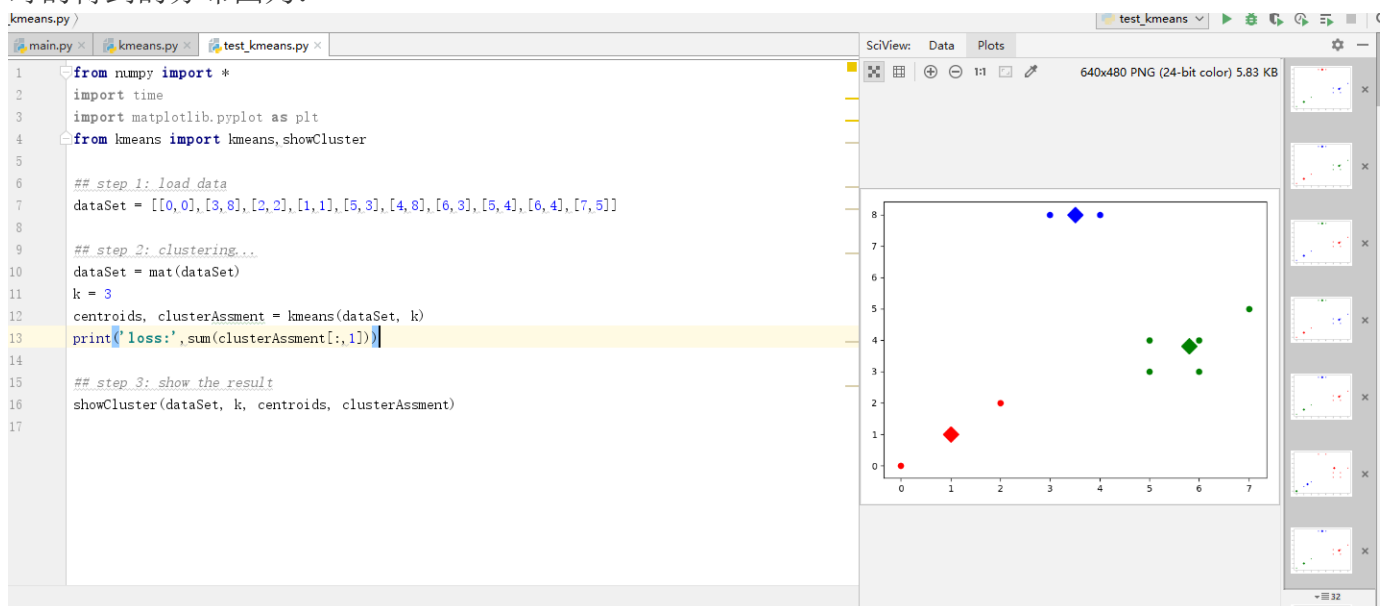
二. 数据分析

在坐标系上画出散点分布情况，如下图：

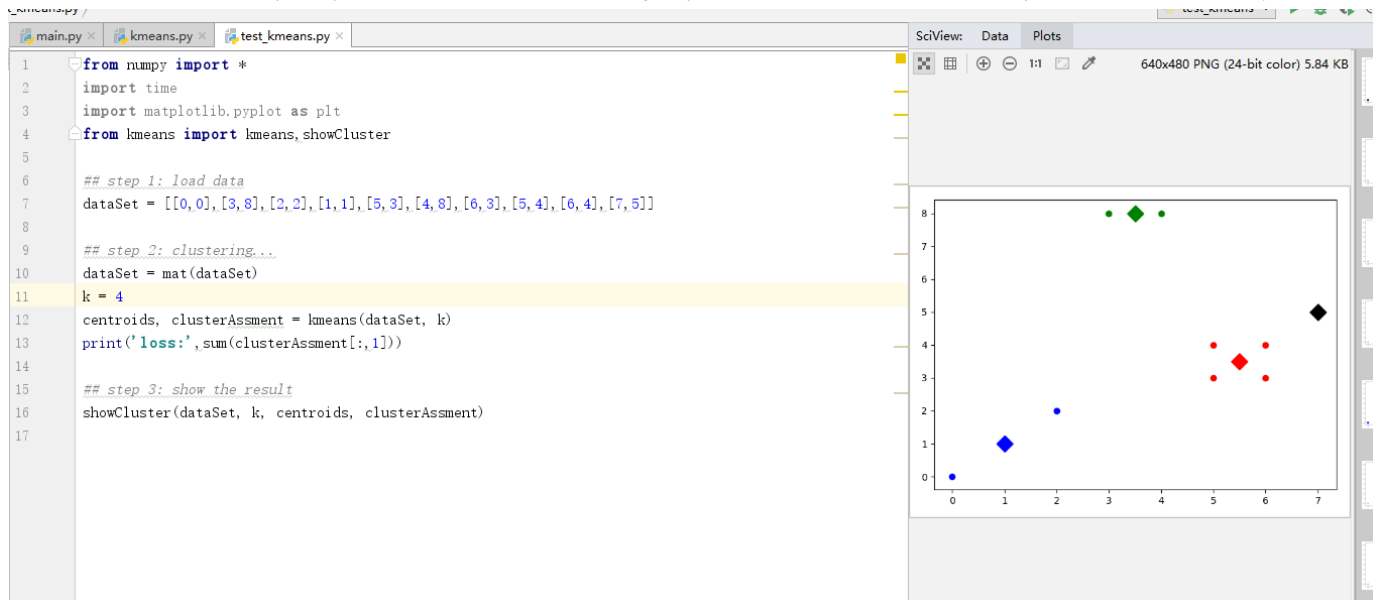
```
: data.plot.scatter(x='x', y='y', xlim=(0,10),ylim=(0,10)):
```



我们可以大致看出图上散点大致可分为3类到4类,通过附件中的代码运行分别计算当聚类数目为3时的得到的分布图为：



此时，他的准则函数的值为10.1。同理，当聚类数目为4时，他的准则函数为6.5，他的分布图为：



三. 手动计算过程及结果

1. $k=3, Z_1(1)=X_1=[0,0], Z_2(1)=X_6=[6,4], Z_3(1)=[3,8]$

2. 计算距离，聚类：

$$X_1 \begin{cases} D_1 = \|X_1 - Z_1(1)\| = 0 \\ D_2 = \|X_1 - Z_2(1)\| = \sqrt{52} \\ D_3 = \|X_1 - Z_3(1)\| = \sqrt{73} \end{cases} X_1 \in S_1(1)$$

同理可得到，

$$S_1(1) = \{X_1, X_3, X_4\}, S_2(1) = \{X_2, X_6\}, S_3(1) = \{X_4, X_7, X_8, X_9, X_{10}\}$$

3. 计算新的聚类中心：

$$Z_1(2) = \frac{1}{N_1} \sum_{X \in S_1(1)} X = \frac{1}{3} (X_1 + X_3 + X_4) = (1, 1)$$

$$Z_2(2) = \frac{1}{N_2} \sum_{X \in S_2(1)} X = \frac{1}{2} (X_2 + X_6) = (3.5, 8)$$

$$Z_3(2) = \frac{1}{N_3} \sum_{X \in S_3(1)} X = \frac{1}{5} (X_4 + X_7 + X_8 + X_9 + X_{10}) = (5.8, 3.8)$$

4. 判断，由于 $Z_j(2) \neq Z_j(1), j=1,2,3$ ，故返回第二步；

5. 重新计算距离，聚类：

$$X_1 \begin{cases} D_1 = \|X_1 - Z_1(2)\| = \sqrt{2} \\ D_2 = \|X_1 - Z_2(2)\| = \sqrt{55.25} \\ D_3 = \|X_1 - Z_3(2)\| = \sqrt{30.88} \end{cases} X_1 \in S_1(2)$$

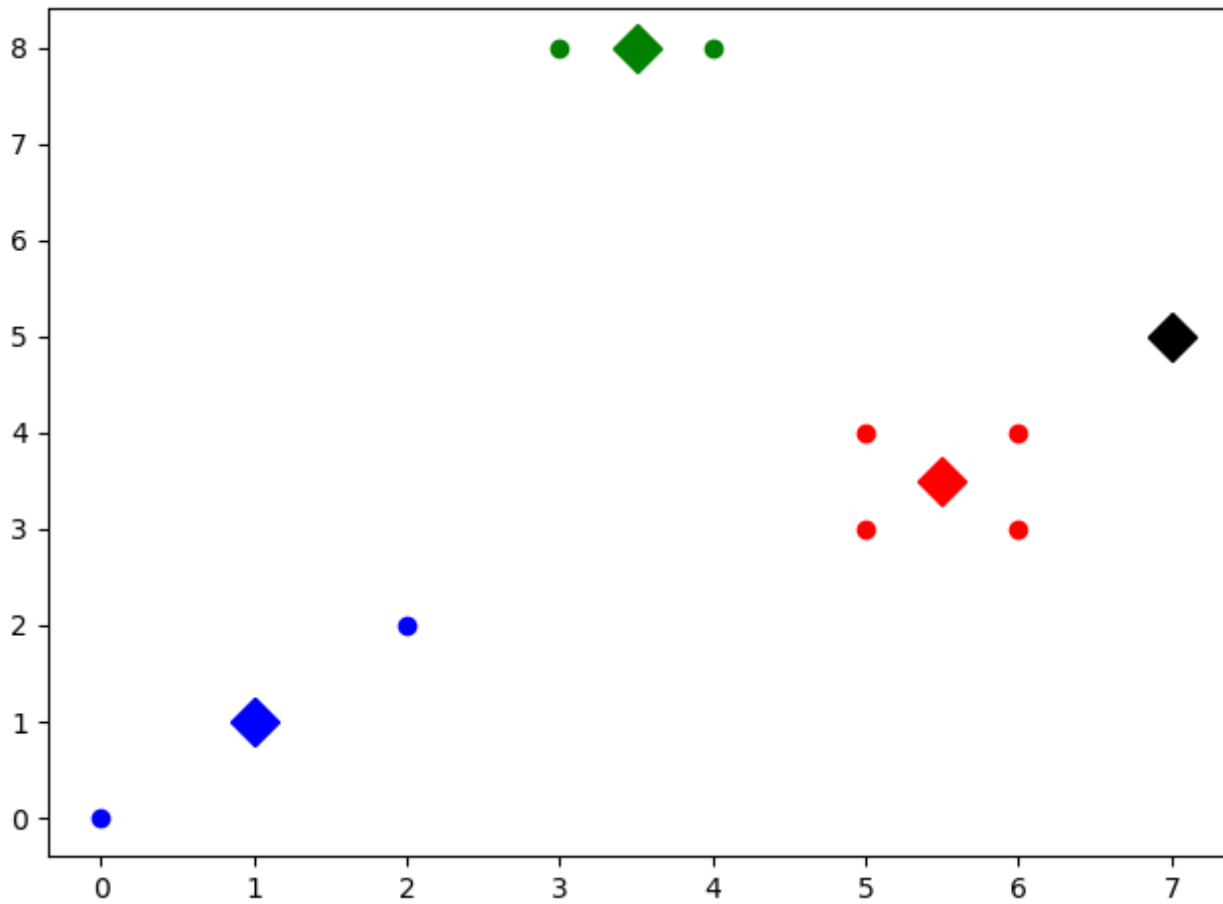
同理，可得到：

$$S_1(1) = \{X_1, X_3, X_4\}, S_2(1) = \{X_2, X_6\}, S_3(1) = \{X_4, X_7, X_8, X_9, X_{10}\}$$

6. 计算新的聚类中心，

$$Z_1(3) = (1, 1), Z_2(3) = (3.5, 8), Z_3(3) = (5.8, 3.8)$$

7. 判断，由于 $Z_j(2) = Z_j(1), j = 1, 2, 3$ ，故算法收敛，得到聚类中心 $Z_j(2) = Z_j(1), j = 1, 2, 3$ ，聚类结果如下图：



By:184395-韩璐璐

2018/9/23