**K-均值算法-实验报告**

1. **问题描述**

试用K均值算法对如下模式分布进行聚类分析。编程实现，编程语言不限：

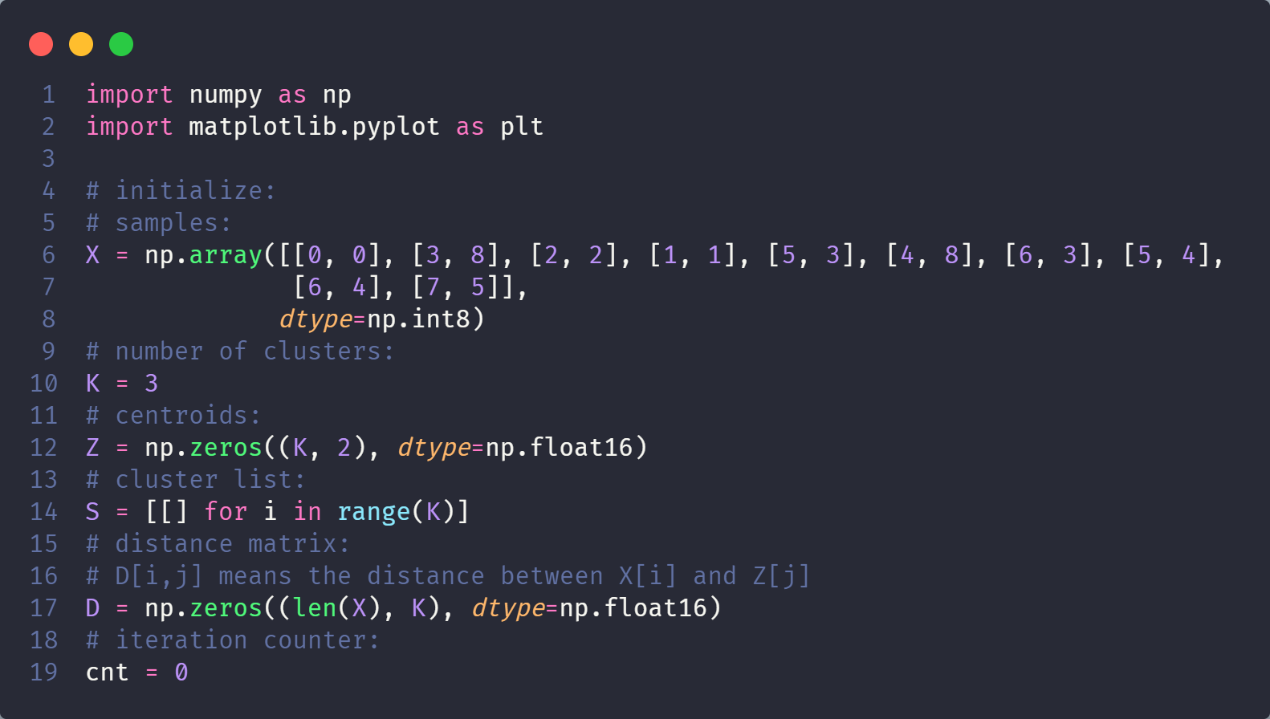
{x1(0, 0), x2(3,8), x3(2,2), x4(1,1), x5(5,3), x6(4,8), x7(6,3), x8(5,4), x9(6,4), x10(7,5)}

1. **算法描述 & 程序实现**

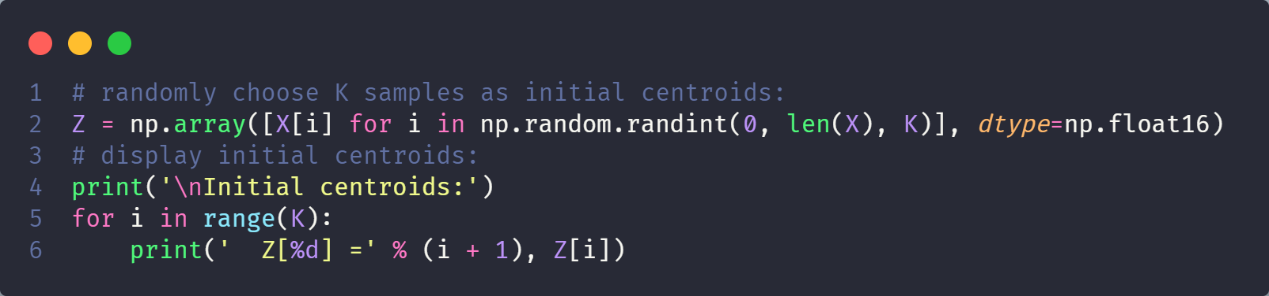
K-均值算法是根据函数准则进行分类的聚类算法，基于使聚类准则函数最小化。而使用的聚类准则函数是聚类集中每一个样本点到该聚类中心的距离平方和，即对第j个聚类集：

**算法的具体步骤如下：**

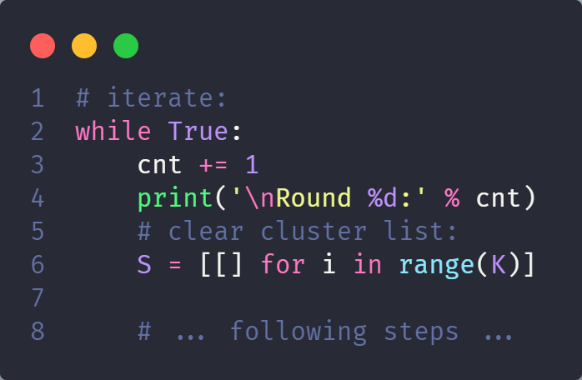
1. 库的导入、初始值的设定（包括样本、聚类数、聚类中心、聚类集、距离矩阵、迭代计数器）：



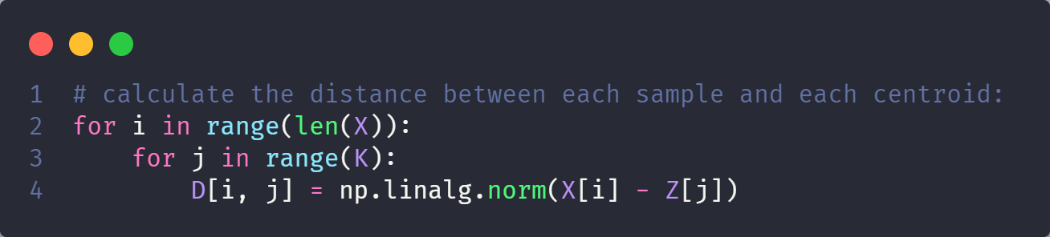
1. 从样本中任取K个，作为初始聚类中心（K<N），加入到Z中：



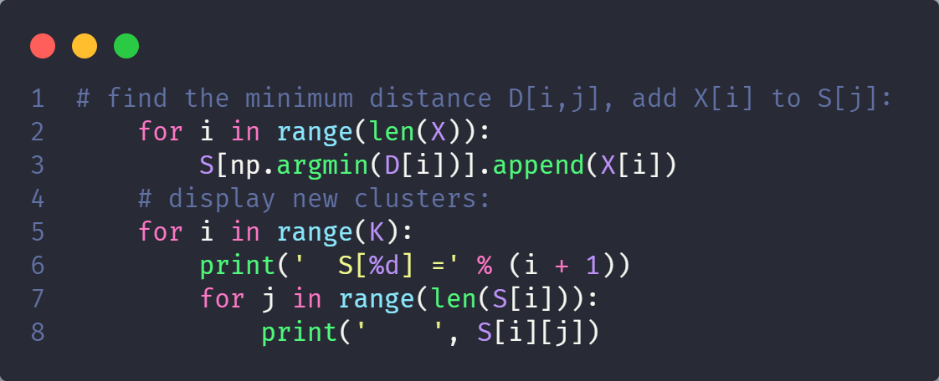
1. 开始迭代：
2. 初始化：迭代计数器自增1，清空聚类集（方便将样本分配在新的聚类中）：



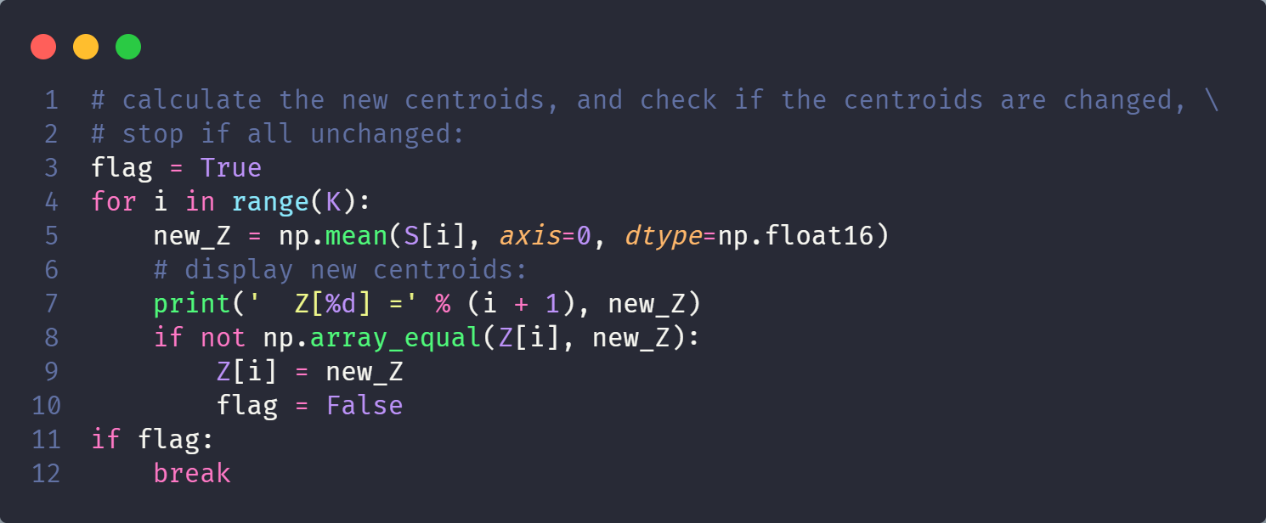
1. 计算距离矩阵D，D[i, j]表示样本Xi到聚类中心Zj的距离，所以要遍历样本和聚类中心，计算结果填入D中：



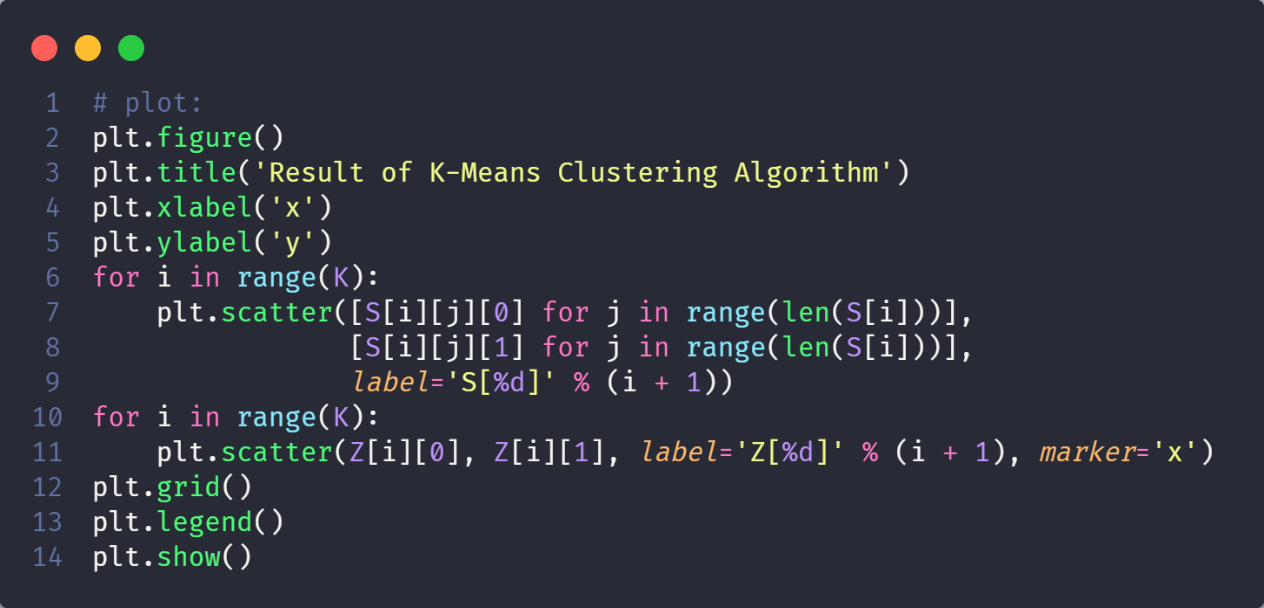
1. 对于样本Xi，找到有最小距离的聚类中心Zj，将样本分配到新的聚类集中（注意需要将样本从原本的聚类中移除，这里选择在每一轮迭代开始时中清空聚类集）：



1. 计算各聚类中样本均值作为新的聚类中心，并判断算法是否结束（当Z不再变化时结束）：



1. 算法结束后画图展示聚类结果：



1. **计算过程**

取K=3，运行，以下是程序在控制台的输出，展示了每一轮迭代中的聚类结果变化和聚类中心的变化：

Initial centroids:

Z[1] = [2. 2.]

Z[2] = [0. 0.]

Z[3] = [4. 8.]

Round 1:

S[1] =

[2 2]

[1 1]

[5 3]

[6 3]

[5 4]

[6 4]

S[2] =

[0 0]

S[3] =

[3 8]

[4 8]

[7 5]

Z[1] = [4.168 2.834]

Z[2] = [0. 0.]

Z[3] = [4.668 7. ]

Round 2:

S[1] =

[2 2]

[5 3]

[6 3]

[5 4]

[6 4]

S[2] =

[0 0]

[1 1]

S[3] =

[3 8]

[4 8]

[7 5]

Z[1] = [4.8 3.2]

Z[2] = [0.5 0.5]

Z[3] = [4.668 7. ]

Round 3:

S[1] =

[5 3]

[6 3]

[5 4]

[6 4]

[7 5]

S[2] =

[0 0]

[2 2]

[1 1]

S[3] =

[3 8]

[4 8]

Z[1] = [5.8 3.8]

Z[2] = [1. 1.]

Z[3] = [3.5 8. ]

Round 4:

S[1] =

[5 3]

[6 3]

[5 4]

[6 4]

[7 5]

S[2] =

[0 0]

[2 2]

[1 1]

S[3] =

[3 8]

[4 8]

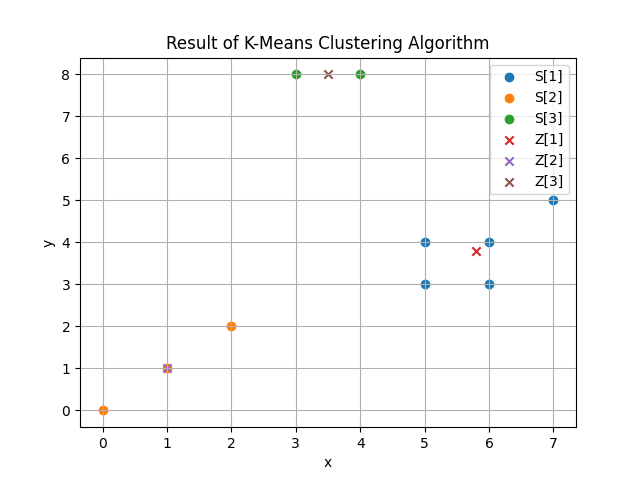
Z[1] = [5.8 3.8]

Z[2] = [1. 1.]

Z[3] = [3.5 8. ]

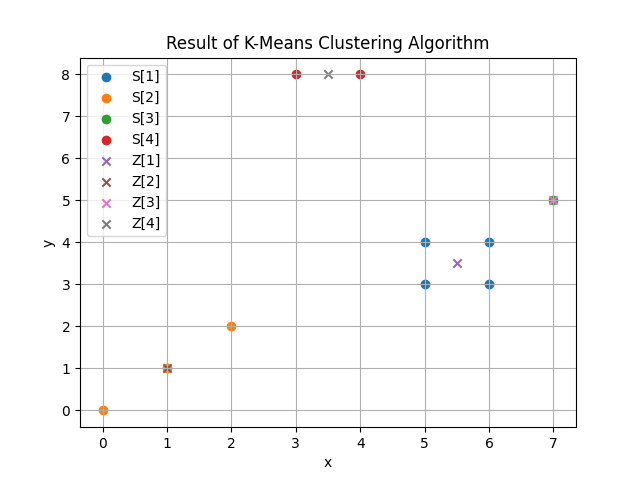
结果图的展示在下一部分“结果分析-1”。

1. **结果分析**
2. 取K=3时，聚类结果如下：



可以看到在图上看到聚类的结果是符合最小距离的聚类准则的。

1. 取K=4时，聚类结果如下：



同样符合判别准则。