# 机器学习实验报告

## 聚类分析

学院：航天学院

姓名： 石瑞河

学号：1180400510

**实验1：编程k-means聚类**

数据集：19个2维数据



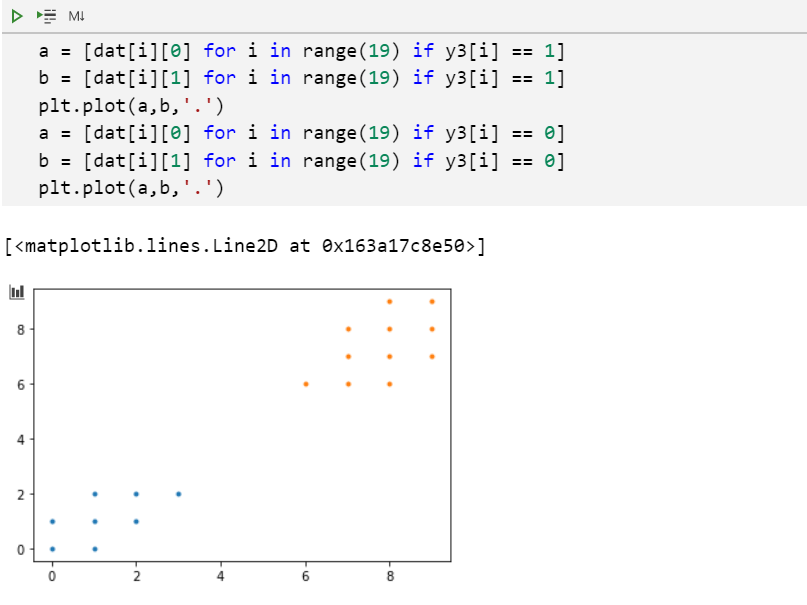
要求：

1. 编程实现k-means算法；
2. 将上列数据聚类为2类；



Kmeans求出两个簇的中心点为（7.81818182, 7.36363636）和（1.25 , 1.125 ）

聚类效果如图



**实验2：MNIST聚类**

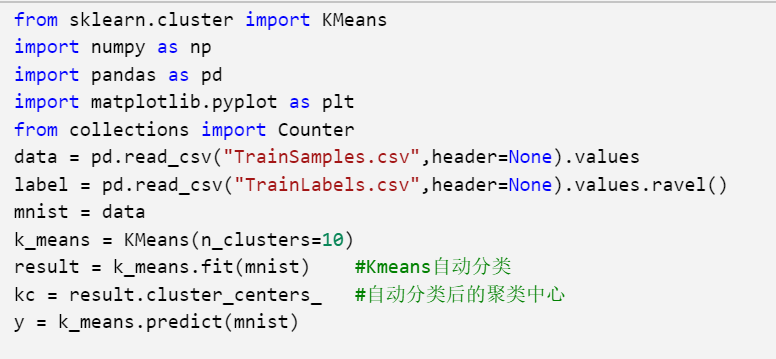
数据集：“Sample-0.csv”，“Sample-1.csv”，…，“Sample-9.csv”，10个数字的样本

要求：

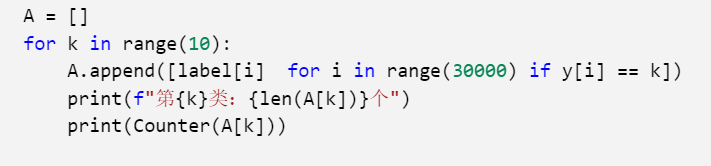
1. 编程实现k-means算法；
2. 将任意2个数字的数据聚类为2类；

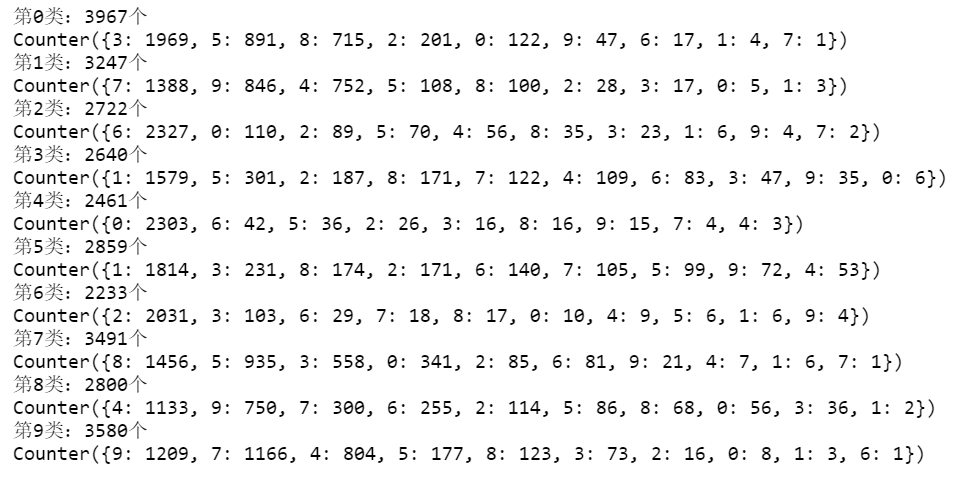
使用sklearn的kmeans算法

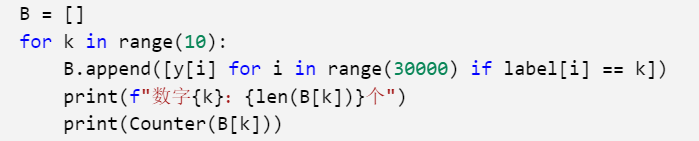
1. 直接使用kmeans划分为10个类

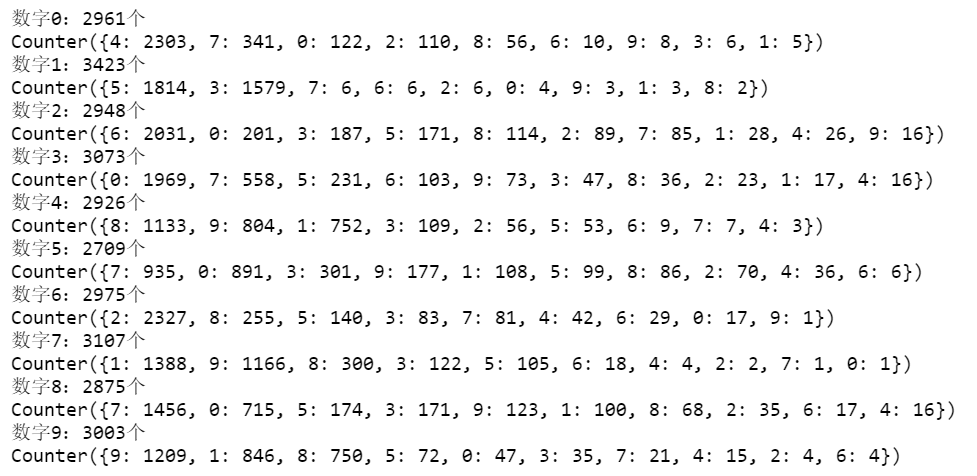


看第k类的成分









对比两者

数字0大多数被分为第4类，第4类绝大多数被识别为数字0

数字1，被分为第5类或第3类，第5类和第3类中绝大部分为数字1

数字2，基本上被分入第6类，第6类中基本上都是数字2

数字3，大部分被分入第0类，少部分为第7类或第5类，第0类中，约一半为3，还有相当数量的5、8

数字4，主要分在第8、9、1类中

数字5，主要分在第7、第0类

数字6，绝大多数分在第2类，第2类中绝大多数为6

数字7，主要分在1、9类中

数字8，大部分在第7类中，小部分在第0类中

数字9，主要分布在第9、1、8类中

比较确定的分类：

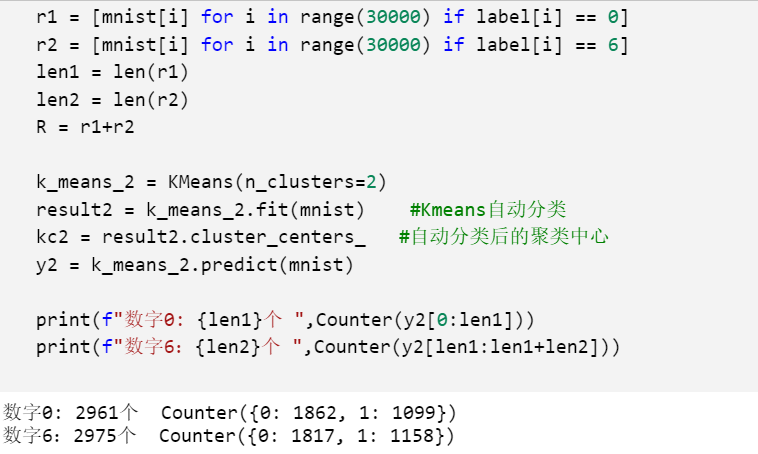
数字0 <==> 分类4

数字1 <==> 分类3、5

数字2 <==> 分类6

数字6 <==> 分类2

1. 对两个数字聚为两类



尝试过多种组合，效果都不理想，不能分成想要的两类。Kmeas以两者其他的特点完成了分类。