说明文档及相关截图

### 一、概述

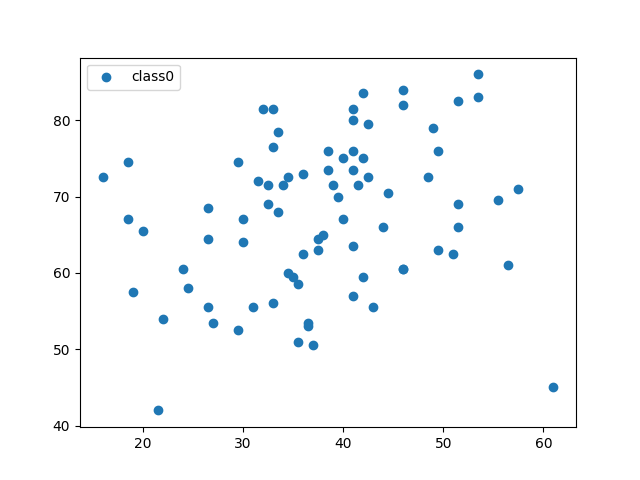
采用python进行代码编写，在代码中我通过课上所学的知识，手动实现了K-means聚类算法，距离计算使用了欧氏距离，经过尝试不同的k，最终发现了该算法的优越性。在代码中，我尝试了k=2,3,4,5的几种情况并做了可视化绘图用于观察。

另外，由于初始聚类中心的随机性，有较小的几率出现极端化的结果或数值报错，只需重新运行即可。

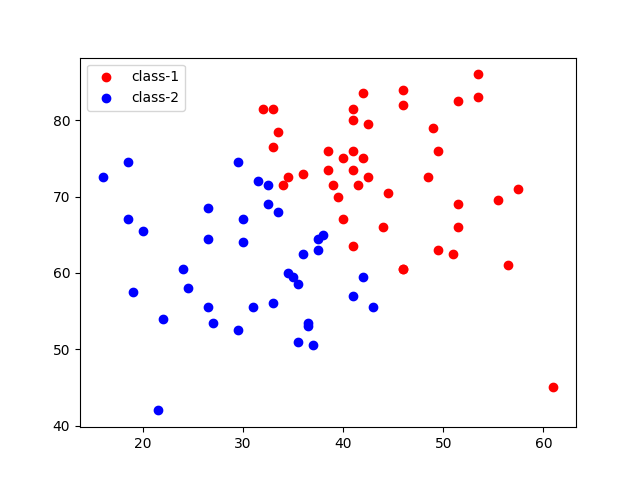
### 二、运行方法

代码用到的工具包有pandas（用于读取数据集）、numpy（用于数组运算、数组转换等）、matplotlib.pyplot（用于绘图实现可视化）。Python含有这三个包即可正常运行。

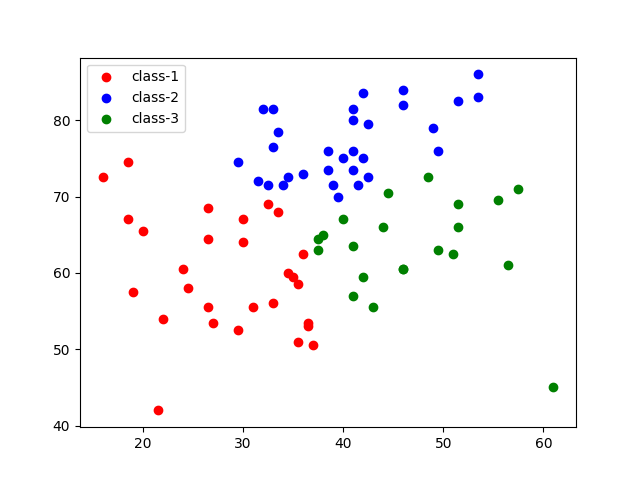
### 三、运行结果



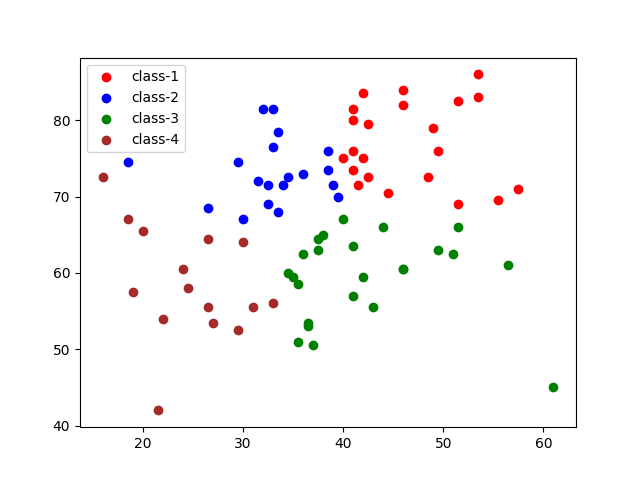
**图 1-未聚类时的散点图**



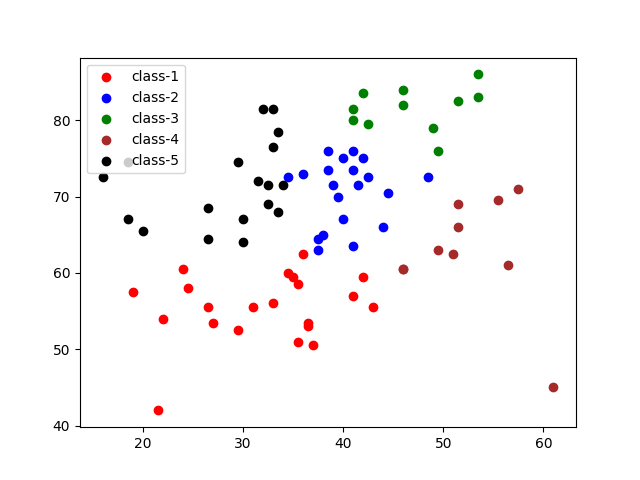
**图 2-k=2时的聚类结果**



**图 3-k=3时的聚类结果**



**图 4-k=4时的聚类结果**



**图 5-k=5时的聚类结果**

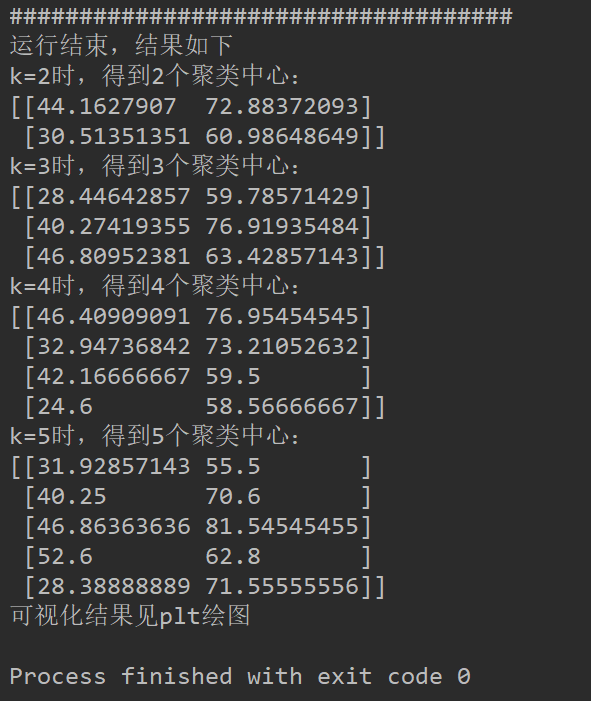


图 6-不同的k下得到的k个聚类中心坐标

### 四、分析

经过手动实现了K-means算法，我发现了该算法具有较强的鲁棒性，且运算速度高、效率高，对于未知标签的数据集效果尤其好，算法也很容易理解。但也有一定的局限性，所选取的k值直接影响了分类结果的好坏，且初始选择的随机的簇的中心不一定能很好的解决问题。对于噪声和非线性问题也有些吃力。

总而言之，该算法为无监督学习方法，有利有弊，其仍是一个值得学习的高效算法。