**聚类实验要求和内容**

一、 实验目的

1. 理解 k-means 算法的算法过程，并能够使用 Python 3 中 sklearn 包所提供的 KMeans 聚类包对数据集进行聚类
2. 理解层次聚类算法的算法过程，并能够使用 Python 3 中 scipy 包提供的

hierarchy 方法对 sklearn 中一个特定大的数据集 iris 进行聚类二、 算法过程

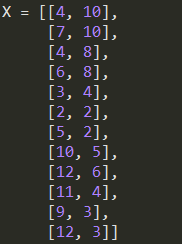
因为是直接调用 Python 的包，所以此部分略去，感兴趣同学可以仔细阅读其文档

三、 实验步骤 (一) K-means 聚类

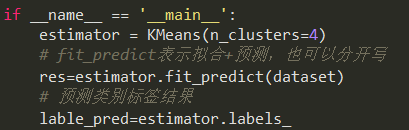
* 1. 由于需要使用 sklearn 包，所以需要先进行安装，可直接在 Python 3 运行终端输入语句：pip install sklearn
  2. 需要调用的包有两个，第一个是调用其聚类包 cluster，第二个是一个可视化工具，以完成如图 7-5 的工作

1530689662(1)

* 1. 该部分使用的数据集为书中例 7-2 所提供的数据集



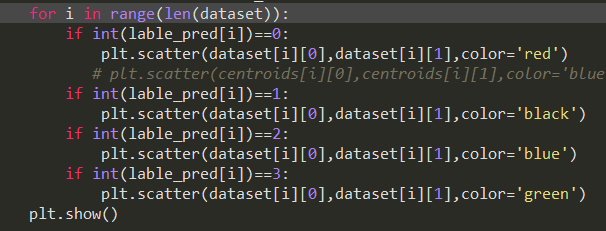
* 1. 调用 KMeans 包中的函数，并且可以设置类数以及对输入的数据集进行预测。这个调用方法可以自行上网搜索查询。



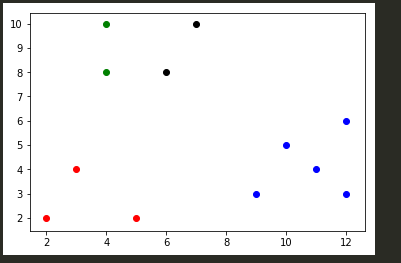
在这里可以知道，聚类的个数是4，且lable\_pred区分了不同聚类 （0,1,2,3），res存储了聚类结果。

* 1. 完成可视化工作，利用 matplotlib 中的 pyplot 包进行数据可视化

显示出结果。

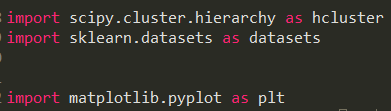


分别将不同聚类用不同颜色的点标记出来，可视化结果如下图：

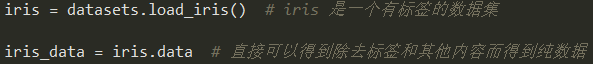


(二) 层次聚类

1. 需要调用两个包，第一个是层次聚类的包，第二个是数据集所在的包



1. 调用数据集的方法如下：

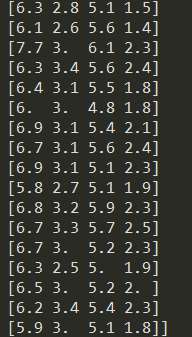


1. 层次聚类使用方法为：

1530690468(1)

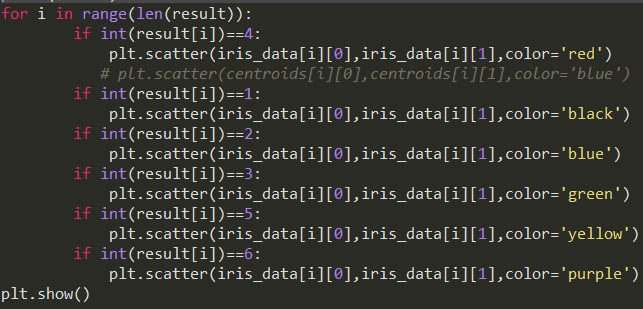
1. 完成可视化工作，利用 matplotlib 中的 pyplot 包进行数据可视化

显示出结果。（要求：将iris\_data前二维作为平面点的坐标在坐标系中表示出来，相同聚类节点颜色相同。）

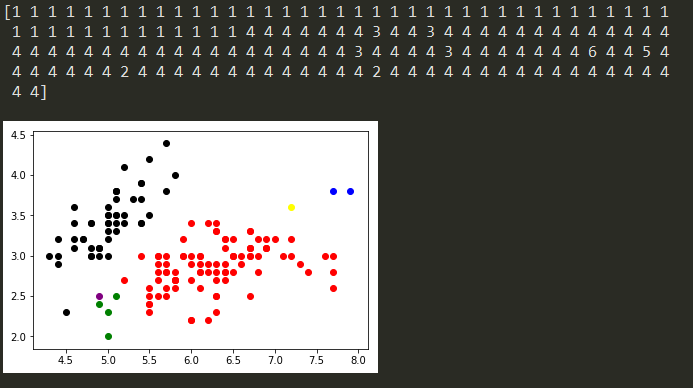


由上图可知，iris\_data 是一个四维的数据集，

因此为了进行可视化将前二维作为点的坐标显示出来，并用不同的颜色区分不同的聚类。代码如图：



最后的效果：



四、实验总结