# Lab\_1 基于PCA/LDA的数据降维和可视化

## 实习目的与要求

1、结合实际应用理解数据降维；

2、采用PCA或LDA两种方法对数据集的数据进行降维，并进行可视化；

3、通过进一步查阅文献，了解相关研究方向的最新研究进展。

## 二、实习题目

基于PCA/LDA的图像降维

#### 【实验数据】

feat\_before\_classifier\_set.npy ，形状为720\*1\*1024 ，720代表720个特征向量，每个特征向量为1\*1024大小

feat\_from\_resnet\_set.npy ，形状为720\*8\*1024 ，720代表720个特征向量，每个特征向量为8\*1024大小

label.npy ，形状为720，表示720个特征向量的类别。

#### 【过程描述】

##### 1、环境配置

安装sklearn库及其其他依赖库：

pip install numpy

pip install matplotlib

pip install scipy

pip install sklearn

注：直接利用anaconda软件，它是直接安装好了除sklearn库的其他库。

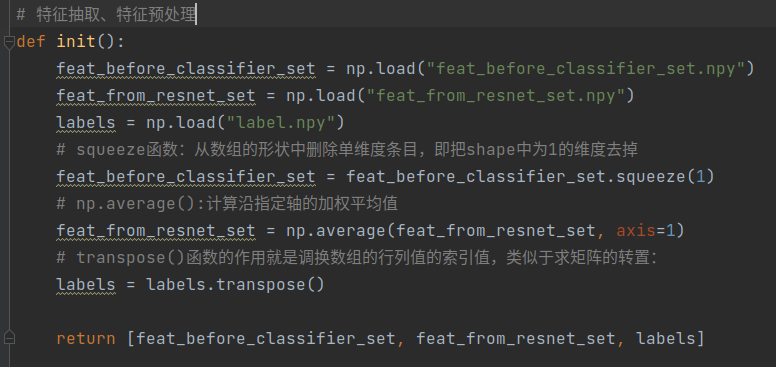
##### 2、数据特征提取和数据预处理

在进行降维之前，先对数据进行特征提取和数据预处理：

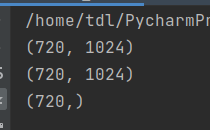
Squeeze函数，从数组的形状中删除单维度条目，即把shape中为1的维度去掉；

np.average()，计算沿指定轴的加权平均值；

transpose()函数的作用就是调换数组的行列值的索引值，类似于求矩阵的转置；



数据处理后，返回三个数据集的数组，并且shape分别为：



##### 3、数据特征工程：

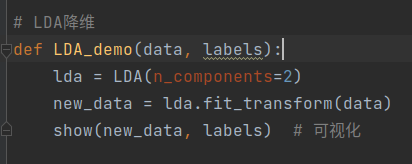
##### PCA降维

根据分析数据，将特征值进行降维，调用的是库函数，n-components参数的值，N为整数的时候，表示保留N个特征值，为小数的时候，表示特征值保留百分百比；



##### LDA降维

整体思路和PCA一样，不过使用的降维方法有所不同而已。



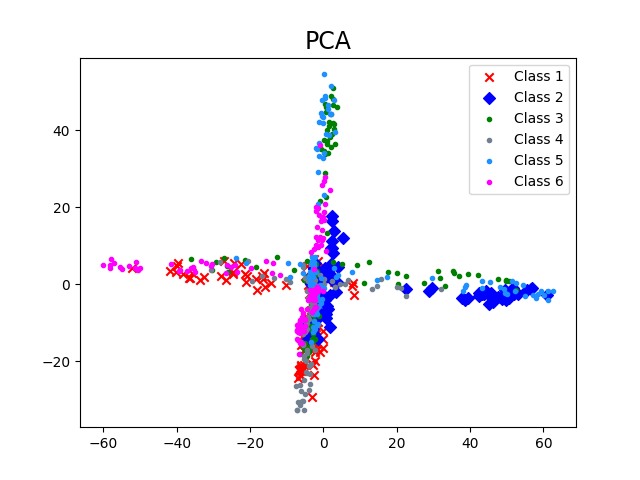
##### 4、数据可视化

使用 matplotlib进行数据可视化。

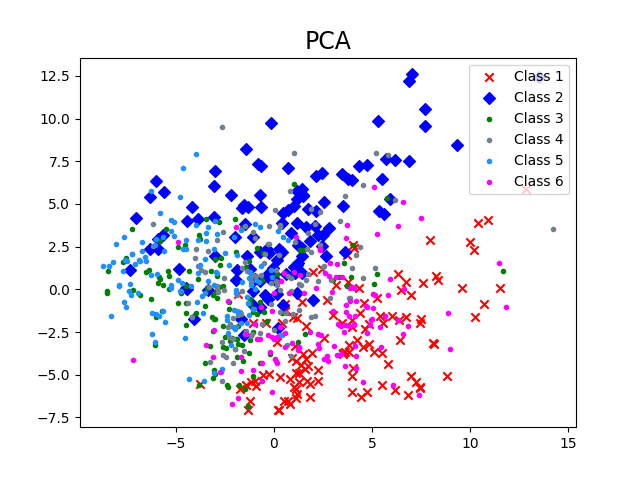


降维结果显示：

* feat\_before\_classifier\_set.npy数据集结果：

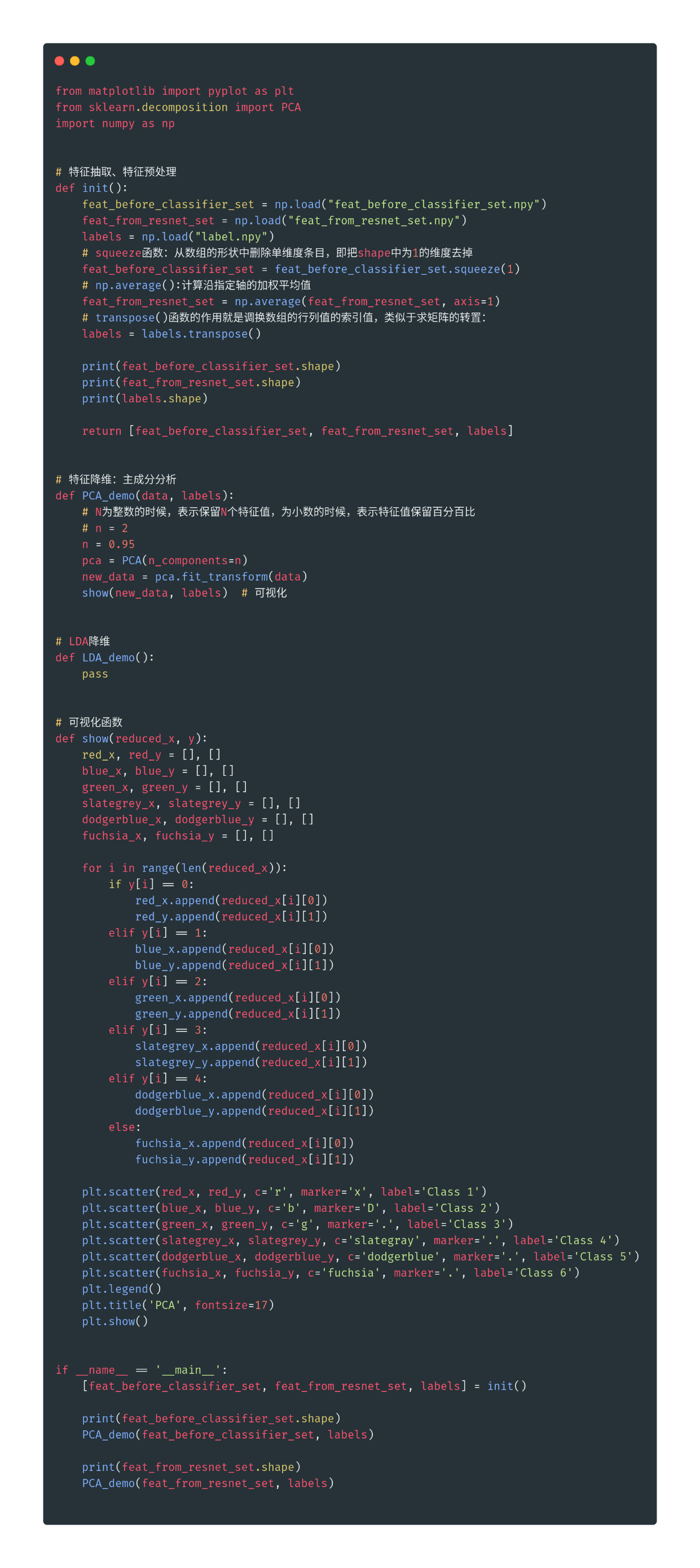


* feat\_from\_resnet\_set.npy数据集结果：



#### 【源代码】

##### 1.Home1.py



#### 【改进设想】

无

## 三、本次实习小结

通过这次的实习，亲自用代码实现了特征工程的各个流程，首先是数据抽取，然后是数据预处理，最后是特征降维。虽然实现的很简单，不过还是能够体会到机器学习的特征工程流程如何操作。在这个实习中，熟悉了numpy库、 matplotlib库、 sklearn库的使用。

=======================================================

我的联系方式：

电话：13349830890

邮箱：liuyy@cug.edu.cn