



《高级人工智能》

实 验 报 告

实验名称： PCA 人脸识别

学 号：

姓 名：

班 级：

**实验一 人脸识别——PCA算法**

**一 实验目的**

1)掌握主成分分析（PCA）算法的基本原理

2)利用PCA算法对数据进行降维

3)利用PCA处理后的数据进行人脸识别

**二 实验设备与环境**

已安装python3.6版本的计算机一台，主要使用的python库为OpenCV、NumPy 。

**三 预备知识**

**1 人脸识别的基本概念**

人脸识别是一项热门的计算机技术研究领域，它属于生物特征识别技术。广义的人脸识别实际包括构建人脸识别系统的一系列相关技术，包括人脸图像采集、人脸定位、人脸识别预处理、身份确认以及身份查找等；而狭义的人脸识别特指通过人脸进行身份确认或者身份查找的技术或系统。

**2 PCA算法的基本原理**

PCA由几个数学概念组成：标准差、方差、协方差。

2.1 Standard Deviation（标准差）

标准差就是用来描述一组数据与平均值得偏离程度，反映了一组数据的波动情况，平均值数学表达公式如下：

标准差的表达式：

2.2 Variance（方差）

方差是数据集中数据分布的另一种度量。实际上，它几乎与标准差相同。

方差的数学表达式如下

2.3 Covariance（协方差）

标准差与方差只针对一维数据进行衡量的指标，协方差是针对二维数据或者是更高维的数据进行的衡量指标，主要用来表示多维度与平均值的偏离程度。

协方差的数学公式

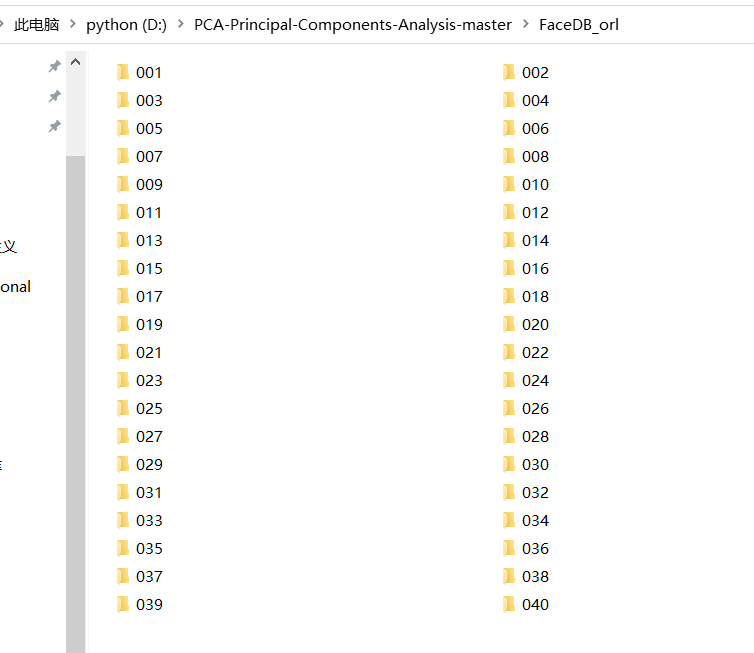
2.4 The covariance Matrix（协方差矩阵）

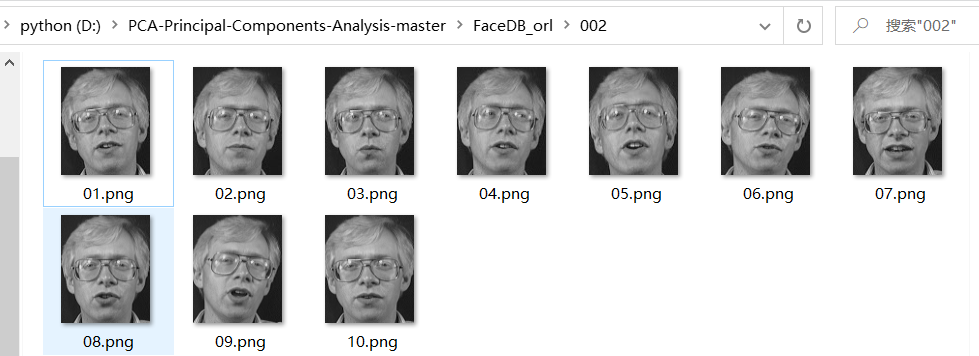
协方差矩阵主要是用于当数据的维度超过3或者更多的时候，我们可以通过一个矩阵来存储各个维度的协方差，这个矩阵就被称为“协方差矩阵”。

用数学方法来表示一个N为数据的协方差矩阵可以表示为：

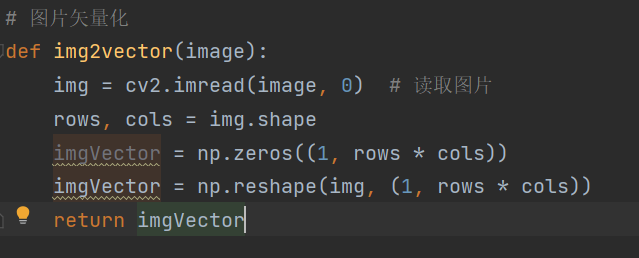
**四 实验内容**

根据所给图片集，图片集中包含四十个人，每个人有十张图片，像素为112\*92.在我的电脑中存放路径为D:/PCA-Principal-Components-Analysis-master/FaceDB\_orl，每个文件夹中存放着同一个人不同面部的10张图片

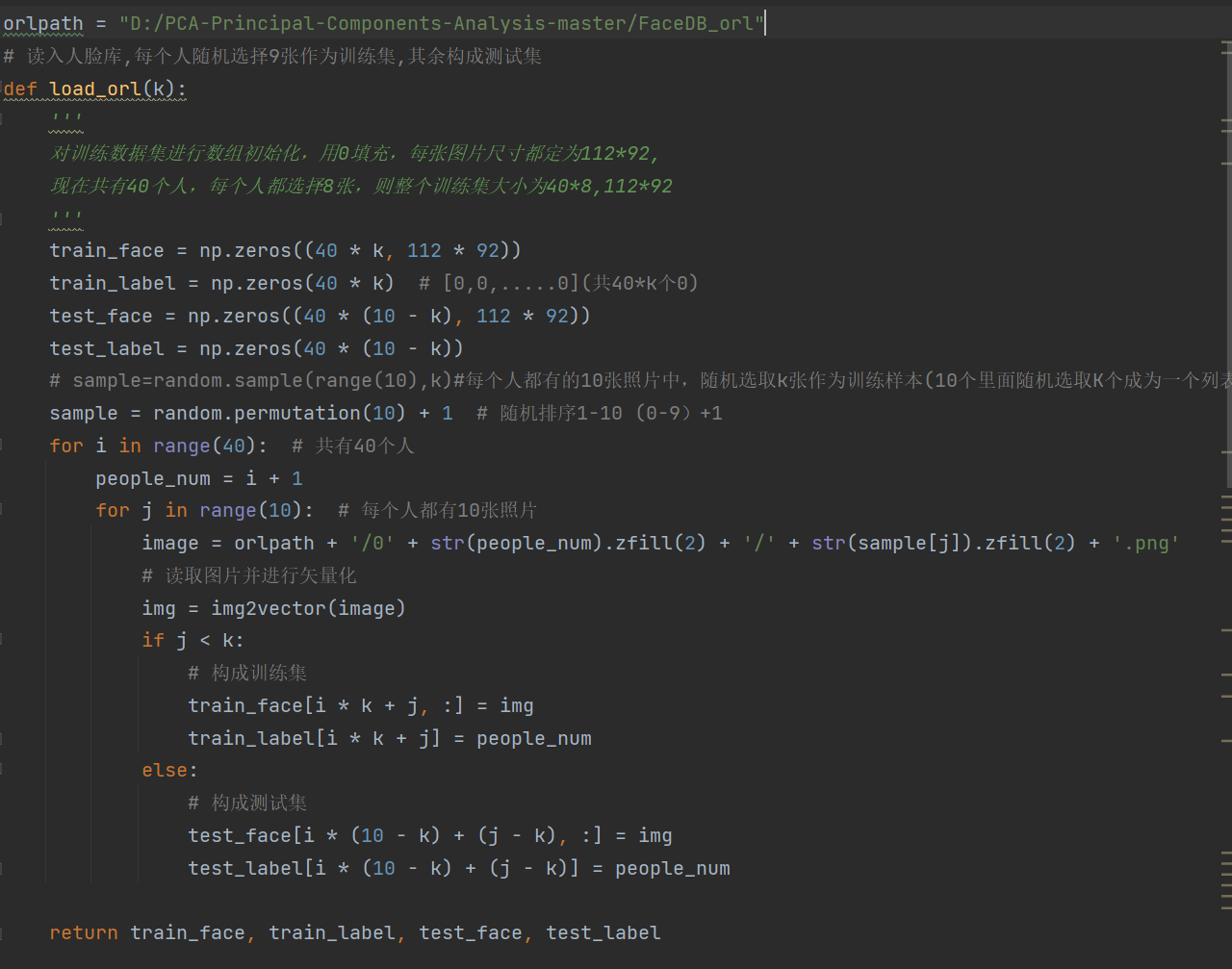




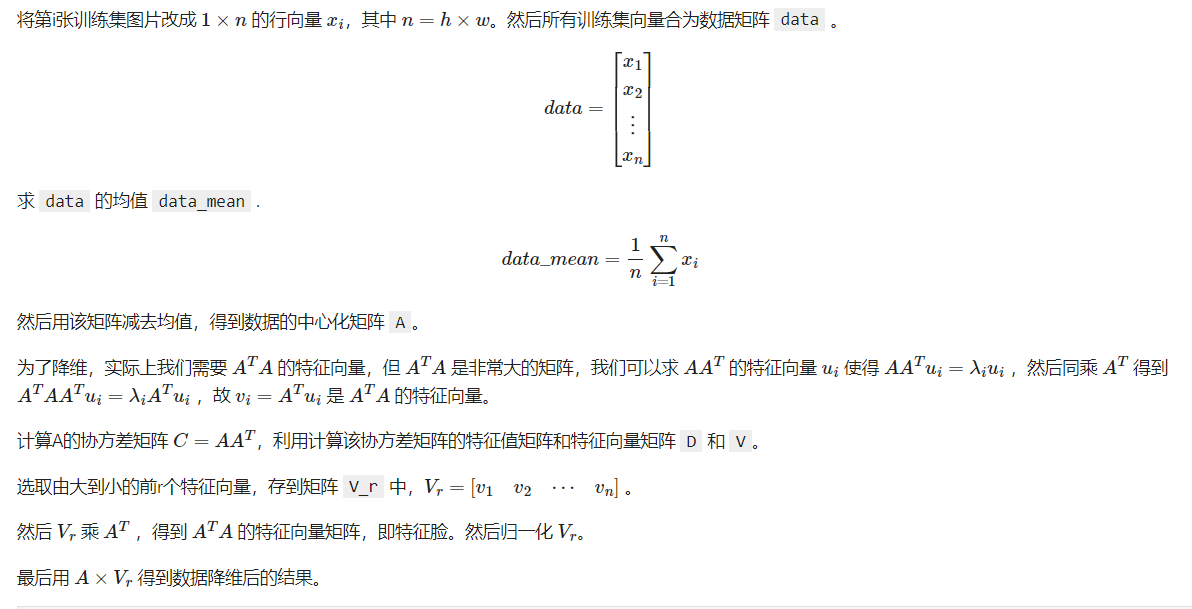
我们需要用到的第三方包有numpy主要用于科学计算，cv主要用于图像处理，训练结果展示。

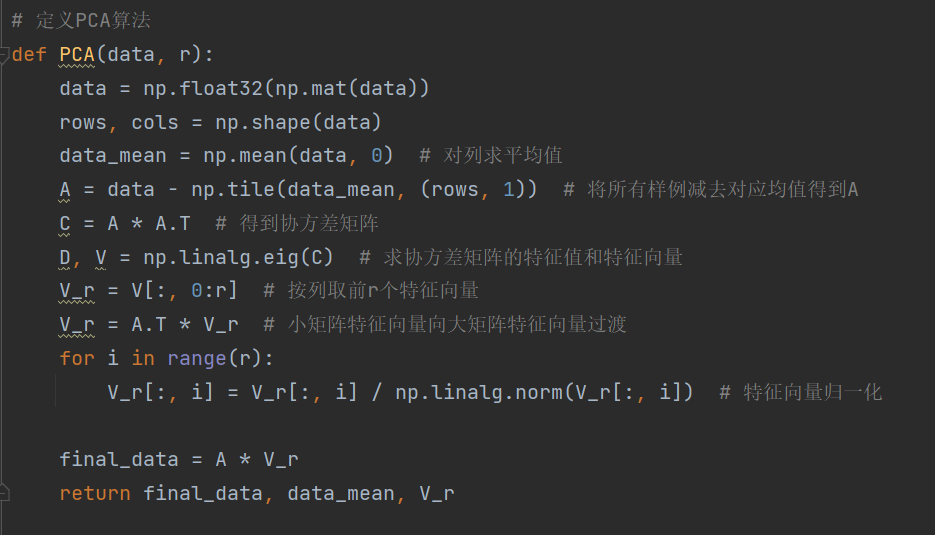


接下来定义一个函数用来选取训练图片，并对每张图片进行前面定义过的矢量化处理。

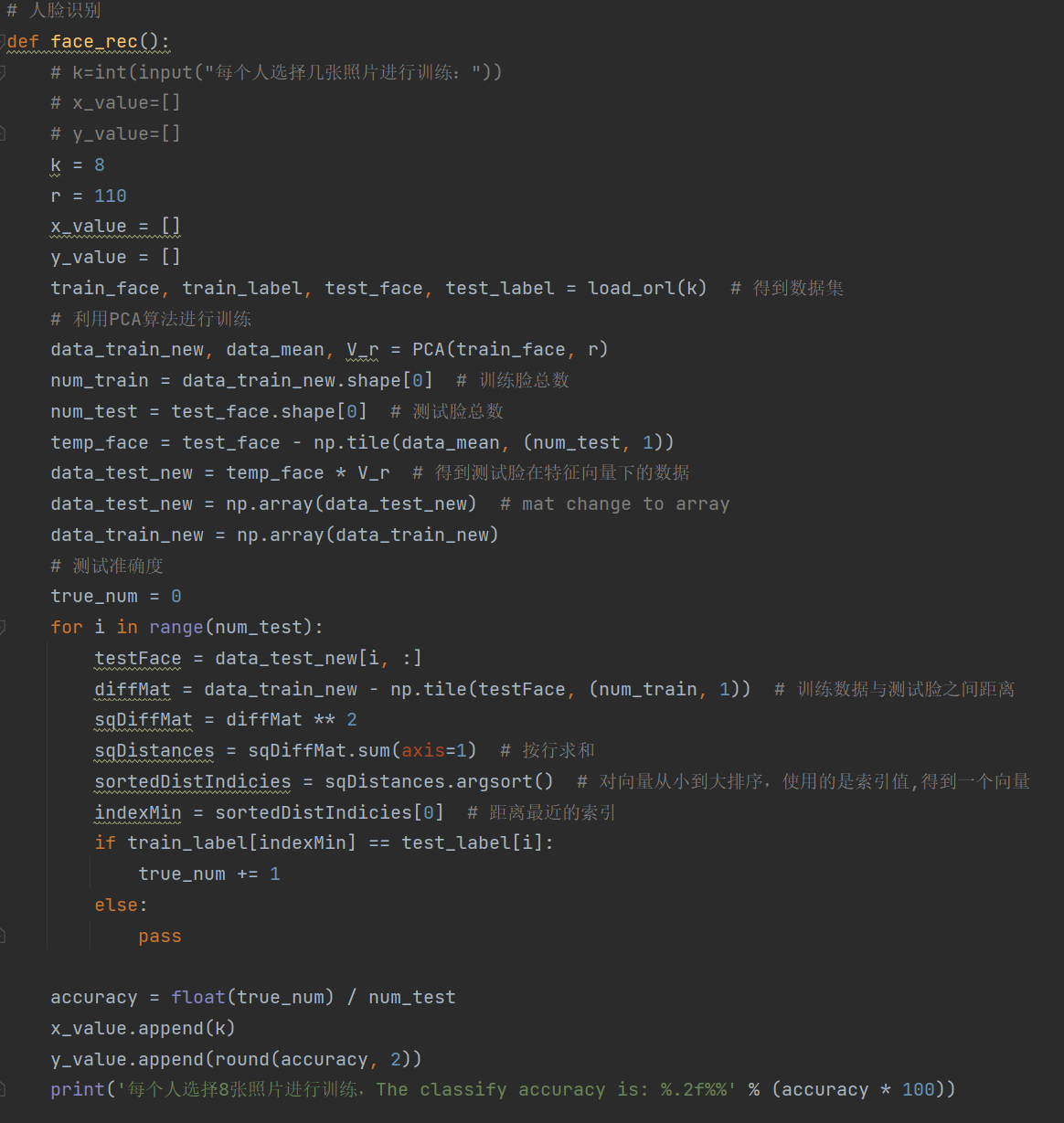


前期将所有训练图片矢量化之后，开始进行PCA算法的降维操作

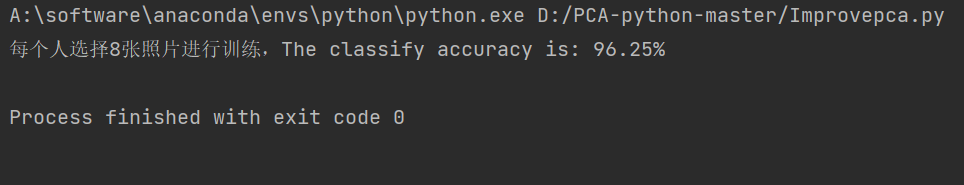




最后进行测试，为了准确方便，每组图片选取前8张做训练，2张用来做预测，并利用PCA降低维度为110维度，首先获取目标测试图片的向量，然后计算目标训练集与测试集的距离；将得到的距离进行平方处理，再对得到的矩阵按行求和，将得到的新一维矩阵进行从小到大排序，使用索引值，得到一个向量，最后选取距离最近的索引，观察该索引下的人物编号是否与测试人物编号相等



最后得到的训练结果为



**六 实验心得**

通过本次实验，我掌握了主成分分析（PCA）算法的基本原理，学会了利用PCA算法对数据进行降维，并利用处理后的数进行人脸识别。对我而言，本次实验较大的难点是对PCA原理的学习

经过多次试验总结发现，训练样本越多训练效果越好，训练维数越高效果越好，但并不是绝对的。110维是我认为比较合适的维度，所以在本实验中采取了该维度。