**

**HUNAN UNIVERSITY**

《DM&BID》

Report 13

|  |  |
| --- | --- |
| **报告名称：** | PCA主成分分析 |
| **学生姓名：** | 杨超然 |
| **学生学号：** | 202106060220 |
| **专业班级：** | 电商2102班 |
| **学 院：** | 工商管理学院 |
| **指导老师：** | 江资斌 |
| **日 期：** | 2023.4.24 |

目录

[一、PCA主成分分析代码实现 3](#_Toc2103145215)

[（一）二维空间降维 3](#_Toc2128282532)

[（三）三维空间降维 4](#_Toc1556890892)

[二、 案例：人脸识别模型 5](#_Toc1724466796)

[（一）读取人脸照片数据 5](#_Toc674140460)

[（二） 人脸数据处理：特征变量提取 5](#_Toc154989648)

[（三） 构造所有图片的特征变量 8](#_Toc13350125)

[（四） 提取目标变量 9](#_Toc1037251587)

[（五） 数据划分与降维 9](#_Toc1962660010)

[（六） 模型搭建与使用 10](#_Toc1077970150)

[三、 人脸识别外部接口调用 11](#_Toc1272264958)

[四、 知识小记 12](#_Toc462475927)

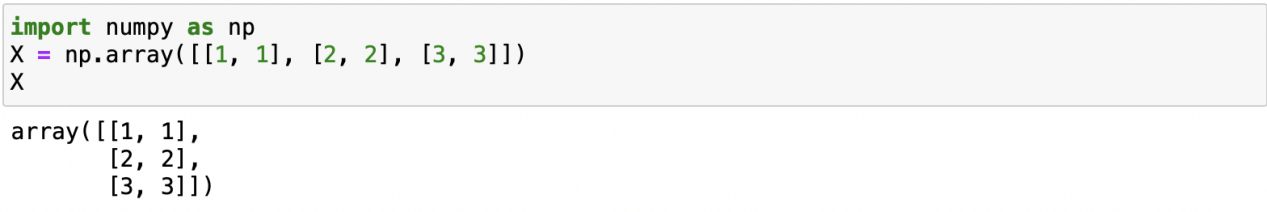
[（一） 数据降维 12](#_Toc1089586596)

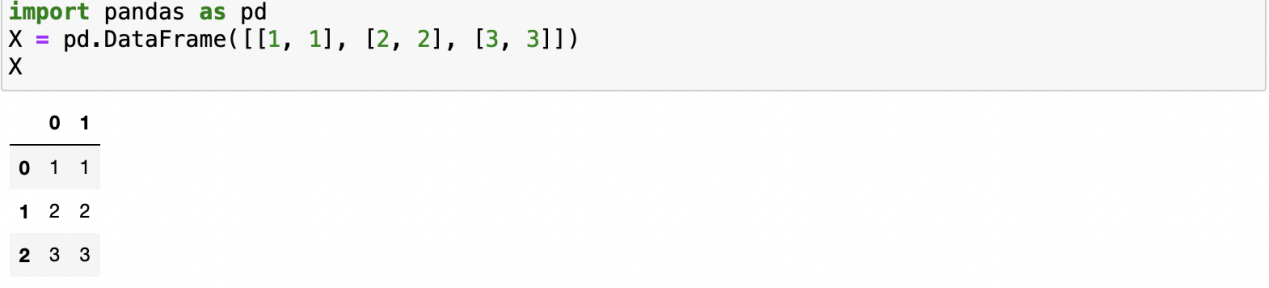
# 

# 一、PCA主成分分析代码实现

## （一）二维空间降维

首先通过numpy（或者pandas）构建数组：

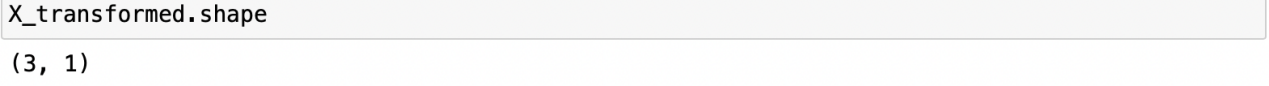




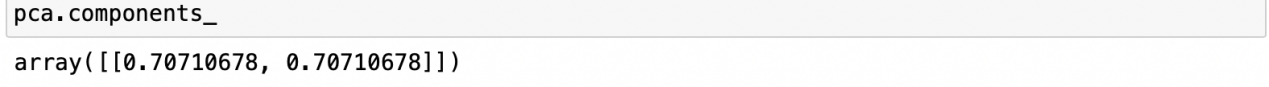
然后通过如下代码实行数据降维：



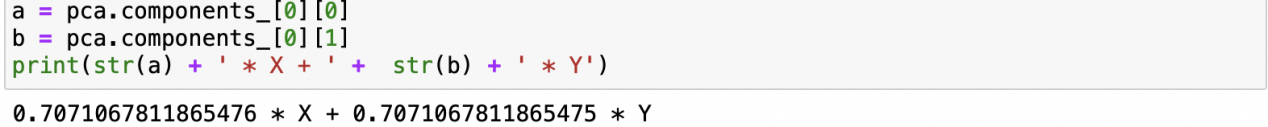
通过如下代码查看此时的维度：



通过如下代码查看降维系数：

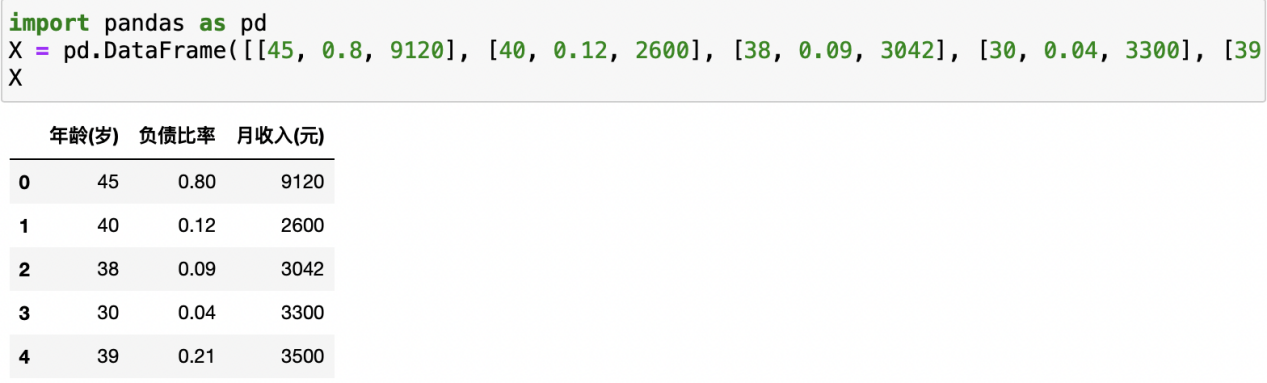


通过如下代码查看线性组合表达式：

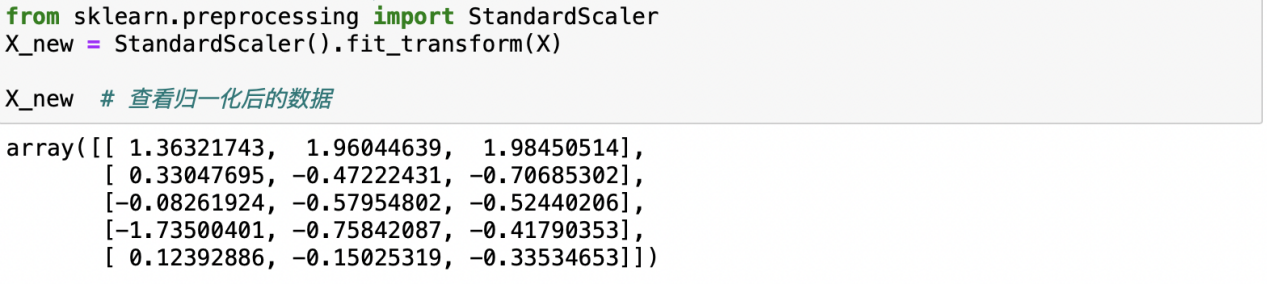


## （三）三维空间降维

首先，构造数组：

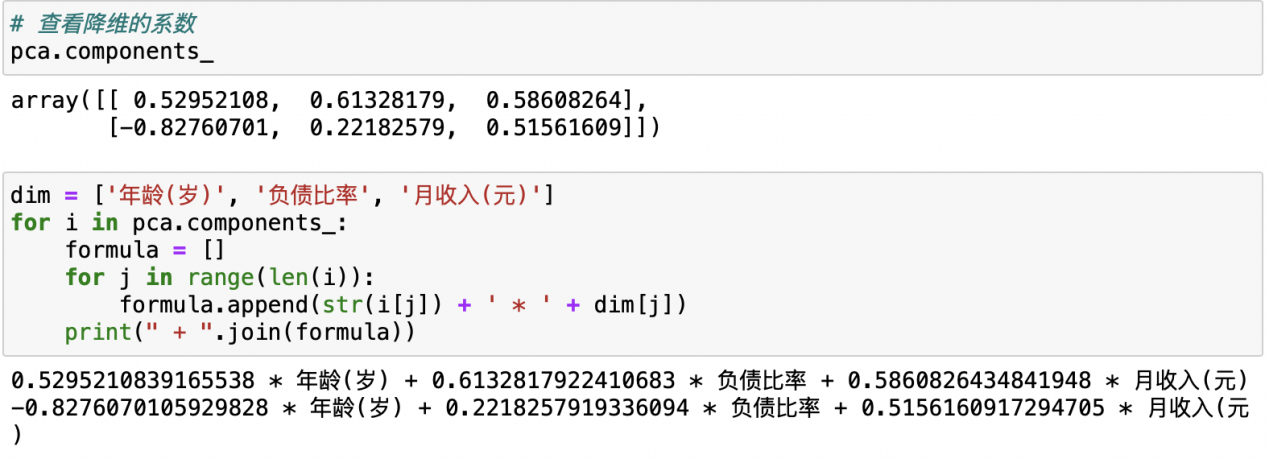


由于三个指标的数据量级差别较大，所以需要进行数据归一化处理：

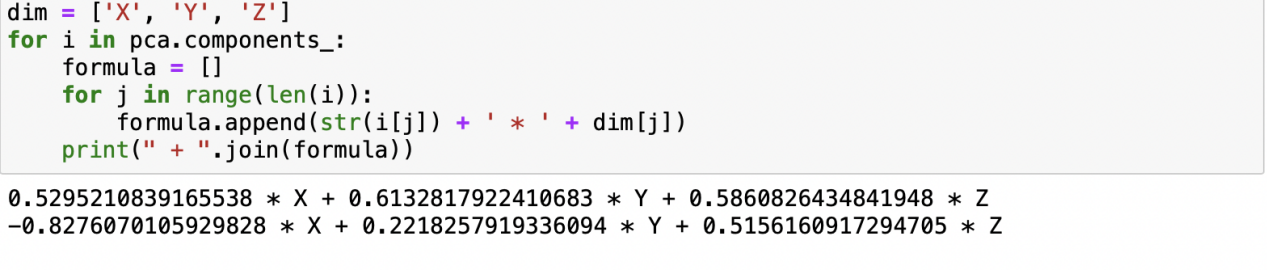


进行标准化处理后，再通过如下代码进行数据降维：





如果不想显示具体的具体的特征名称，可以采用如下的写法：



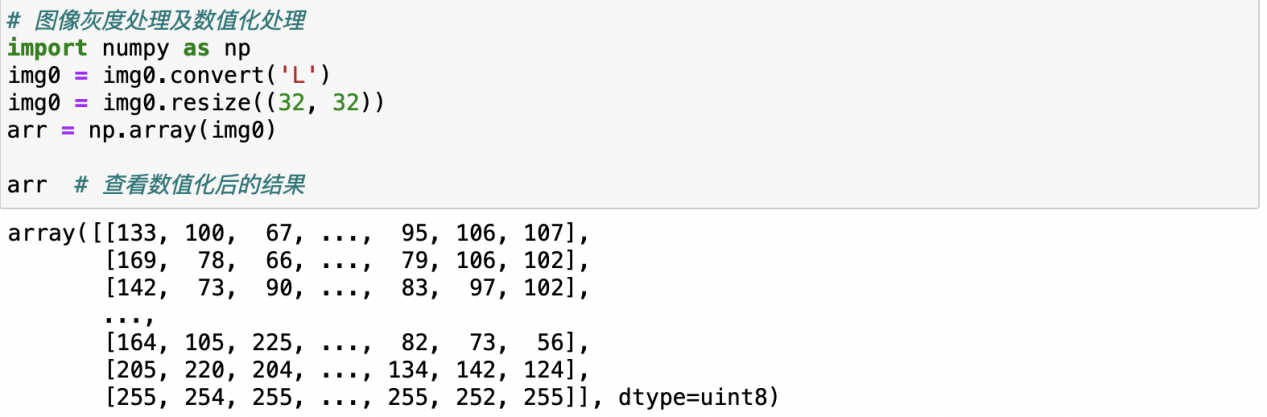
# 案例：人脸识别模型

## （一）读取人脸照片数据

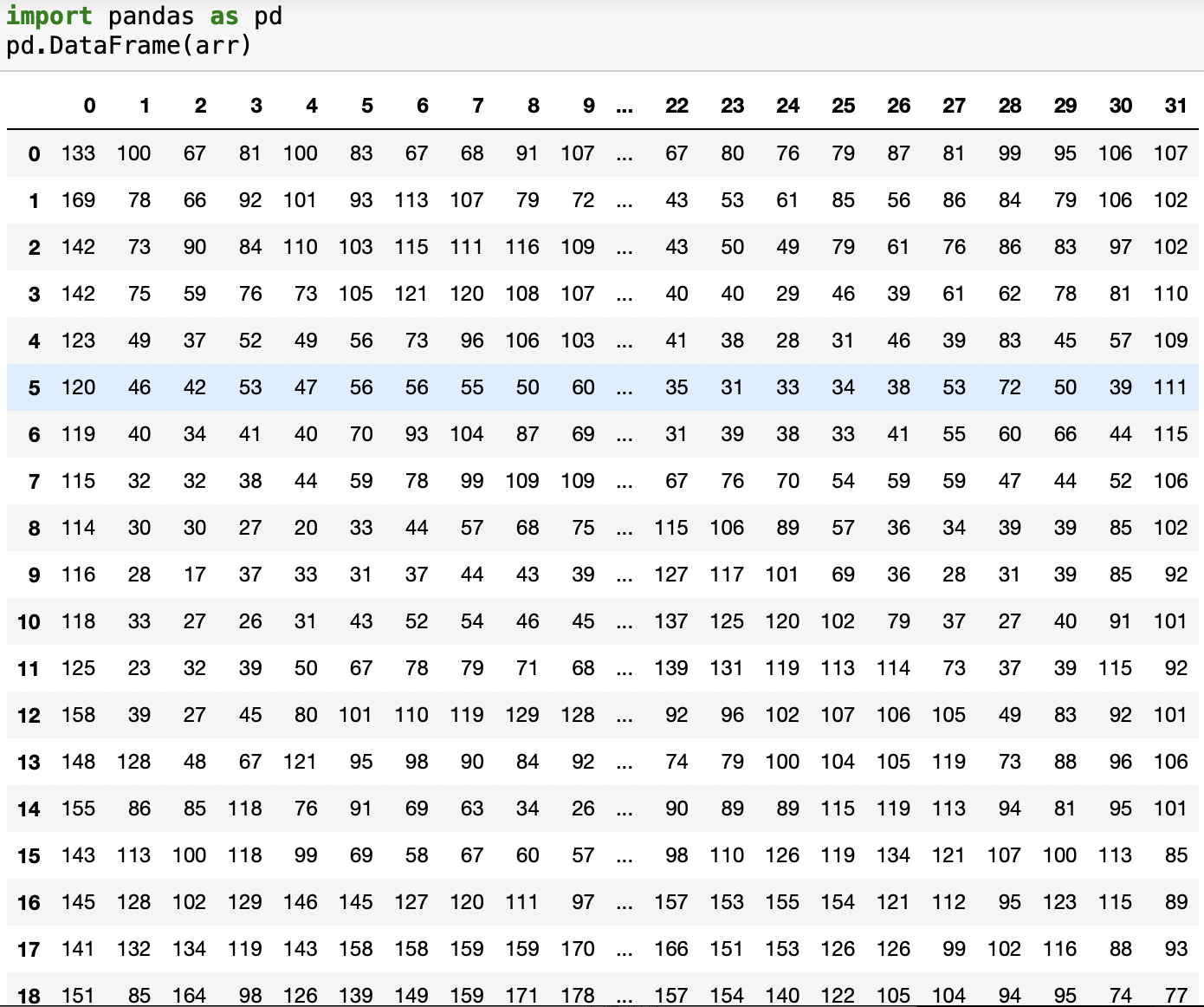


## 人脸数据处理：特征变量提取

通过如下代码进行图像灰度处理与数值化处理：



如果觉得numpy格式的arr不好观察，可以通过pandas将其转换为DataFrame格式，便于观察，部分输出结果如下：

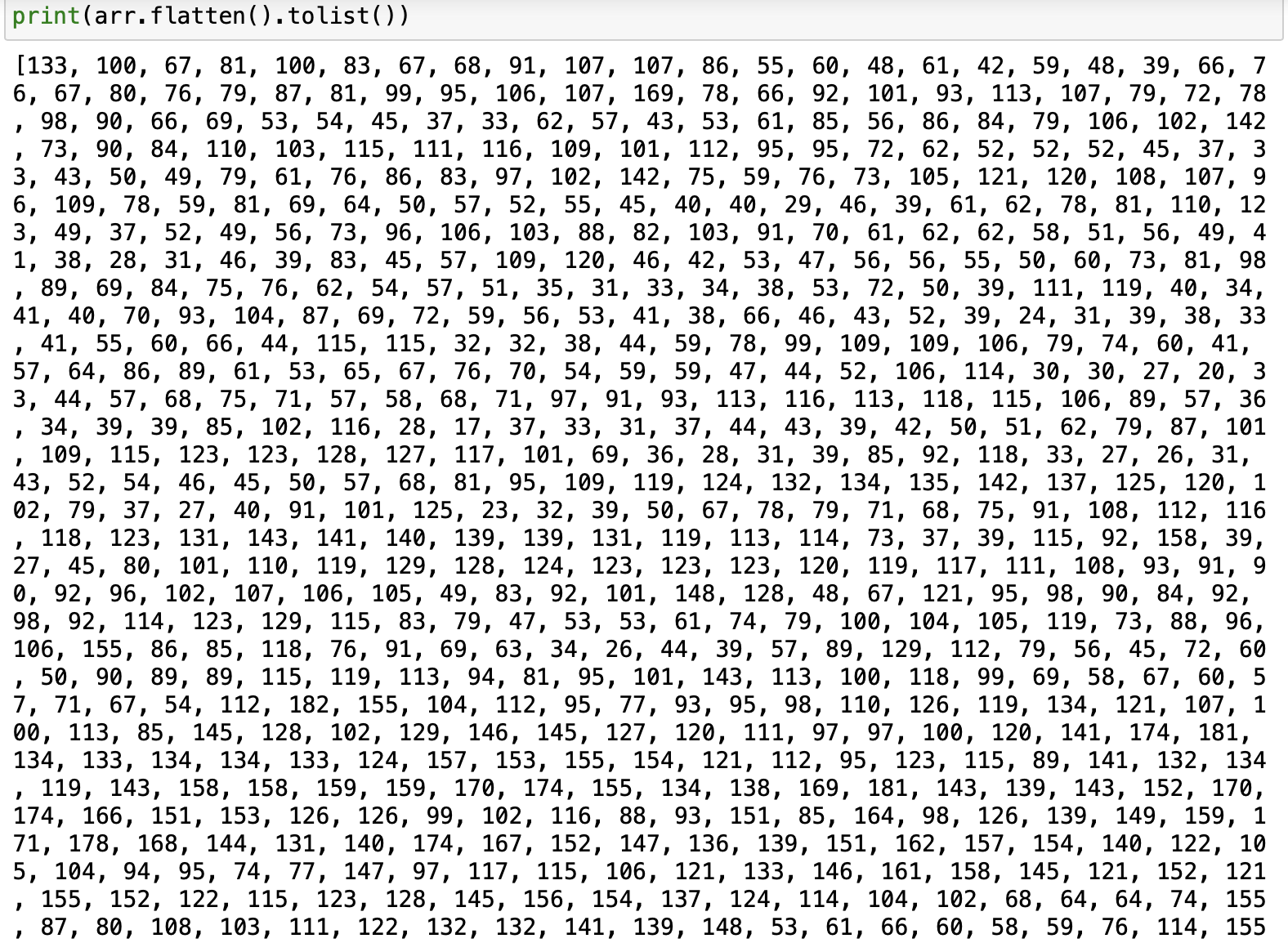


截屏2023-04-24 19.52.40

可见，输出获得的是一个32x32的二维数组，还不利于数据建模，所以我们还需要通过reshape(1, -1)方法将其转换成一行(若reshape(-1,1)则转为一列），也即1\*1024格式：

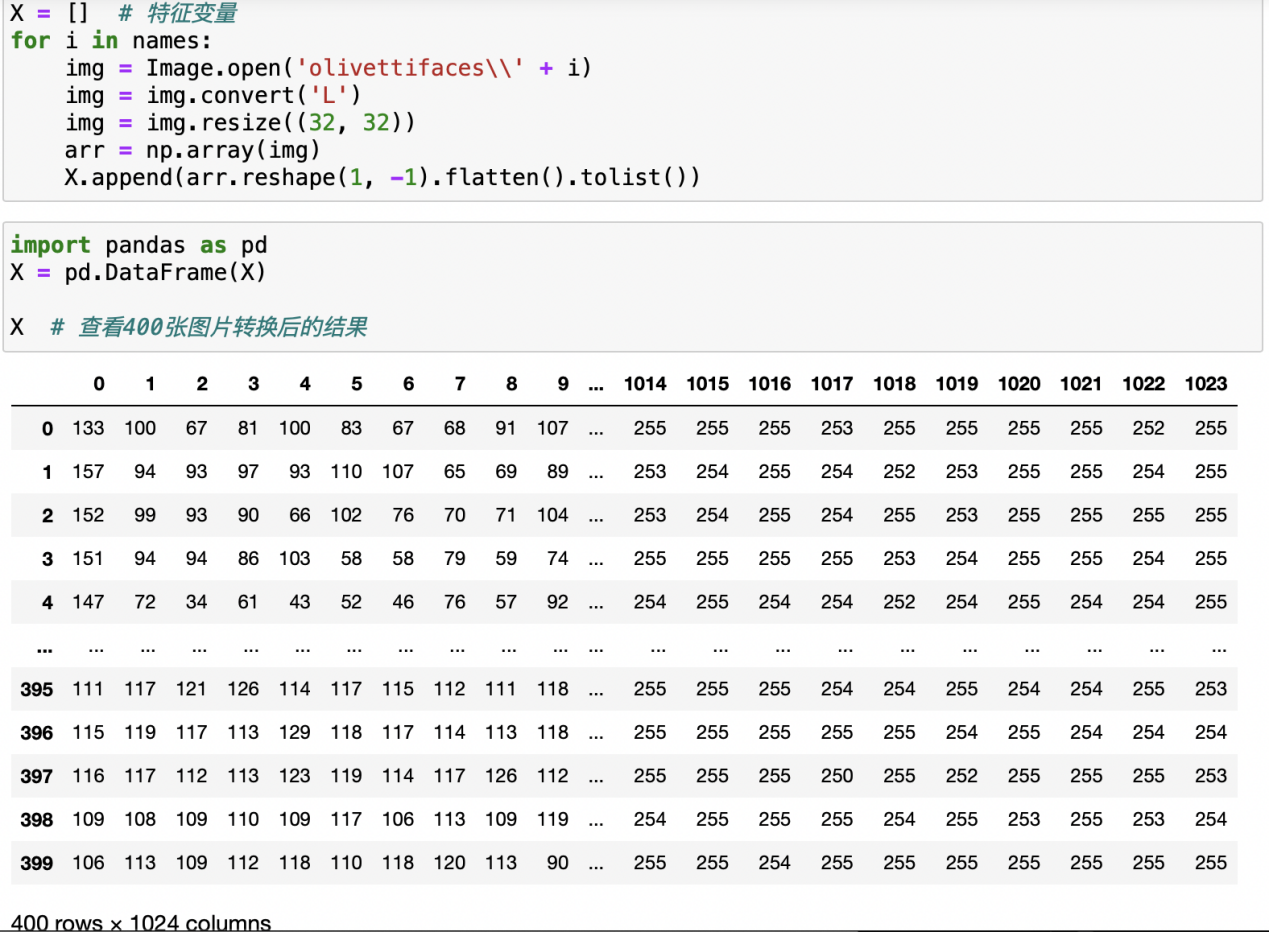


来通过print看一下人脸转换后的输出结果（对于看不懂的人来说，这也是一种加密方式）：



## 构造所有图片的特征变量

通过如下代码构造特征变量，查看转换结果：



## 提取目标变量

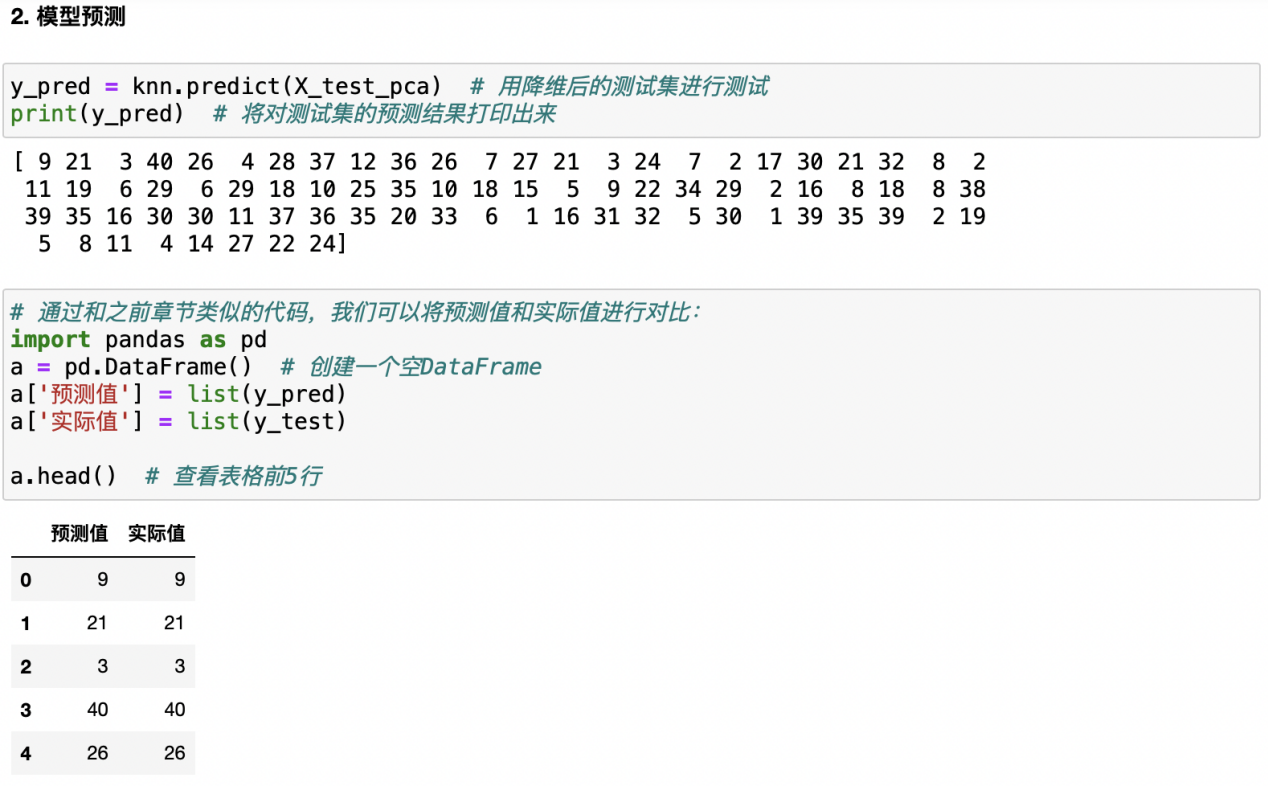


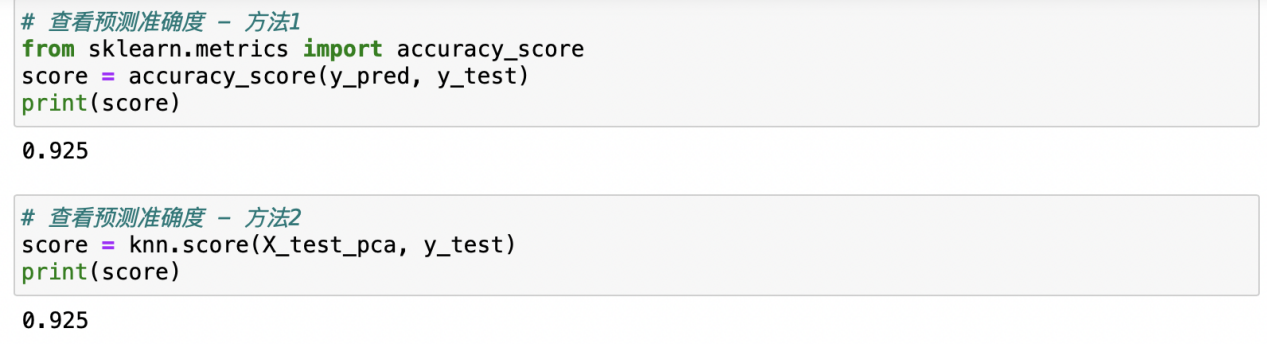
## 数据划分与降维

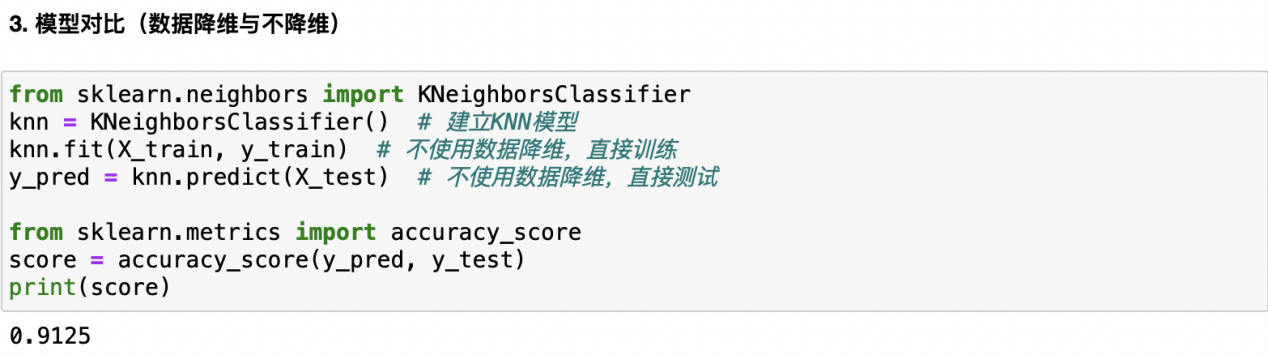


## 模型搭建与使用





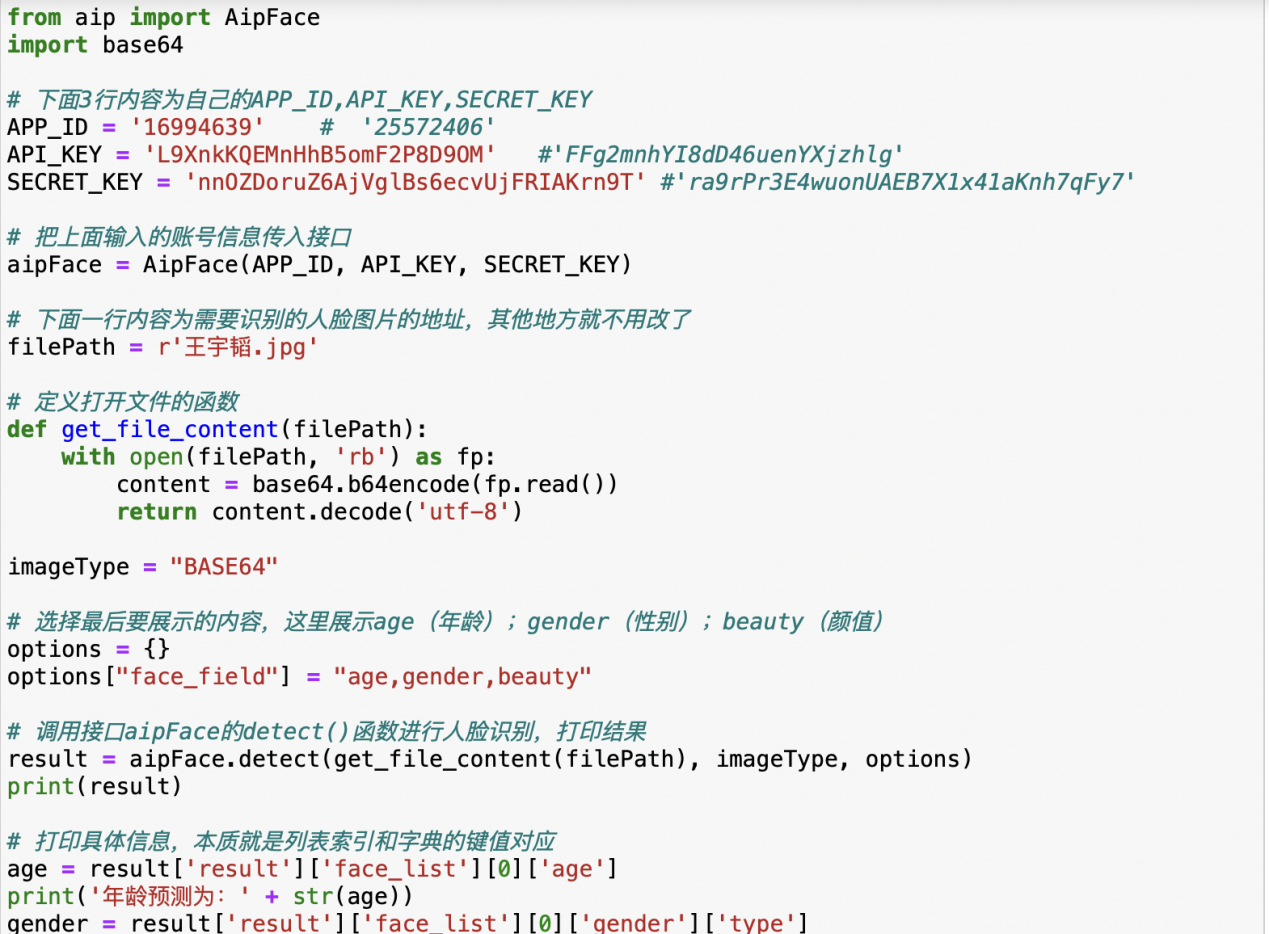


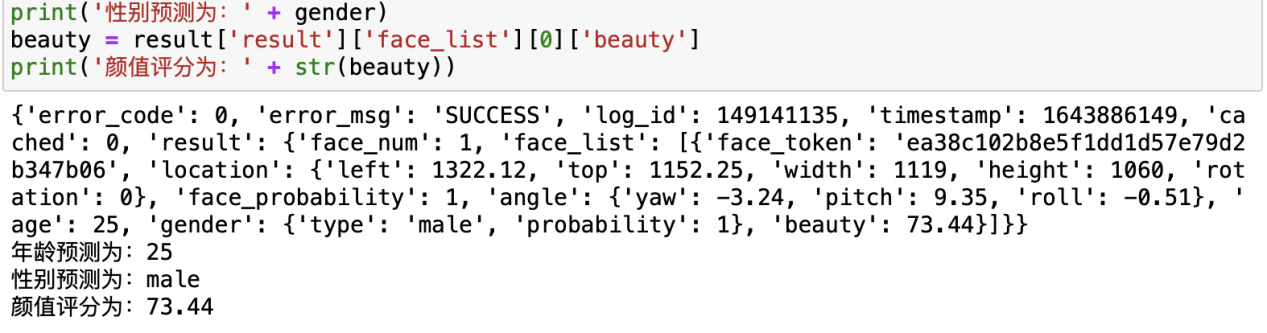


此时获得的准确度评分score为0.91，可以看到使用数据降维对提高模型预测效果还是有一些效果的，这里的数据量并不大，当数据量更大的时候，利用PCA主成分分析进行数据降维则会发挥更大的作用。

# 人脸识别外部接口调用

调用接口前，需使用pip命令安装baidu-aip库；然后调用接口进行人脸识别和打分：





# 知识小记

## 数据降维

降维的方法主要有两种：**选择特征**和**抽取特征**。选择特征即从原有的特征中挑选出最佳的特征，抽取特征即将数据由高维向低维投影，进行坐标的线性转换。PCA主成分分析即为典型的抽取特征的方法，它不仅是对高维数据降维，更重要的是经过降维去除噪声，发现数据中的模式。