**深 圳 大 学 实 验 报 告**

|  |
| --- |
| **课程名称 机器学习**  **项目名称 实验一：PCA算法**  **学 院 计算机与软件学院**  **专 业 软件工程（腾班）**  **指导教师 赖志辉**  **报 告 人 黄亮铭 学号 20221550258**  **实验时间 2024年3月8日至2024年3月17日**  **实验报告提交时间 2024年3月17日** |

**教务处制**

# 一、实验目的与要求

1. 熟悉机器学习的基础算法之一PCA；

2. 学习从零实现PCA算法；

3. 了解无监督学习算法；

# 二、实验内容与方法

在AR或COIL数据集上实现如下功能：

1. 简述PCA，并证明其投影矩阵等价于重构误差最小的模型的解。

2. 实现PCA算法的人脸重构，即用20,40,60,80,...,160个投影来重构图像的效果。

3. 实现PCA算法的人脸识别，给出 10,20,30,...,160维的人脸识别识别率。

4. 用PCA用来进行人脸（AR）与旋转目标（COIL）图像降维，对2个不同数据集1-2个类中图像实现二维和三维空间数据的可视化,并在图中对应的点显示对应的人脸/旋转目标图象。

5. 同时设计一个新的PCA算法，内容简要写在实验报告中，并与经典PCA比较。

# 三、实验步骤与过程

1. 简述PCA，并证明其投影矩阵等价于重构误差最小的模型的解。

① PCA简介：PCA（Principal Component Analysis）是一种常用的数据降维技术，它可以用来降低数据集的维度，并保留数据中最重要的信息。PCA的核心思想是找到数据中的主要变化方向，然后沿着这些方向进行投影，以实现降维。具体而言，PCA通过线性变换将原始数据投影到一个新的坐标系中，使得在新坐标系下数据的方差最大化。这些新的坐标轴称为主成分，它们是原始数据中方差最大的方向。通常情况下，只需保留数据中方差较大的主成分，即可实现有效的数据压缩和降维。

② PCA的工作流程如下：

1. 对原始数据进行标准化处理，使得每个特征的均值为0，方差为1。

2. 计算数据的协方差矩阵。

3. 对协方差矩阵进行特征值分解，得到特征值和对应的特征向量。

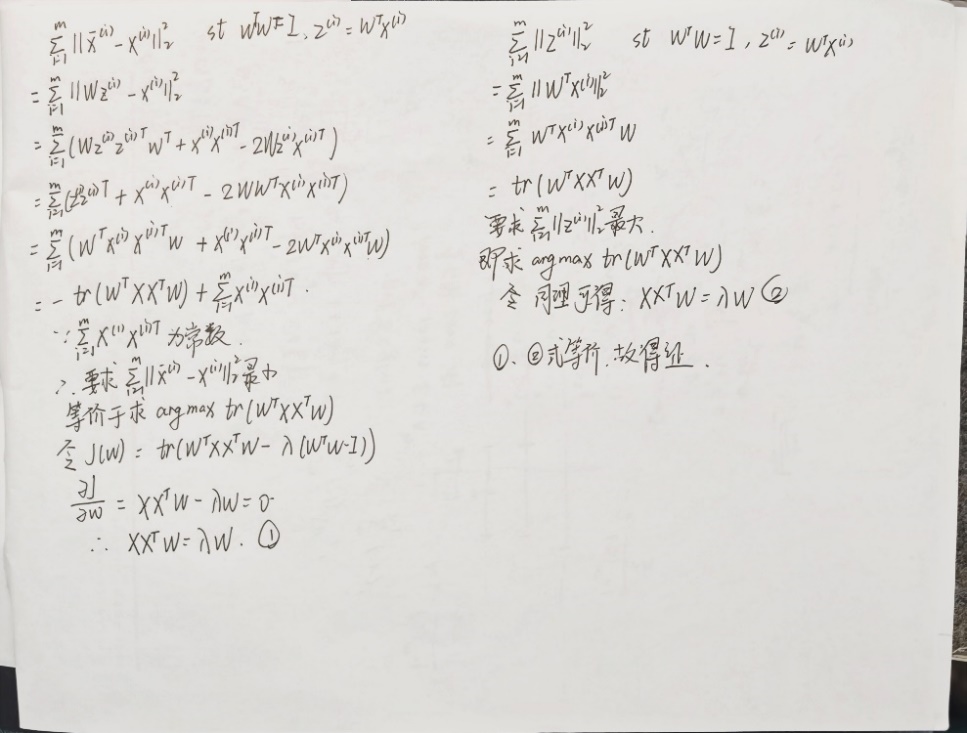
4. 选择最大的k个特征值对应的特征向量，这些特征向量构成了新的坐标系，也就是主成分。

5. 将原始数据投影到选定的主成分上，得到降维后的数据。

③ PCA的应用

PCA通常用于数据预处理、特征提取、数据可视化和去除数据中的噪声等任务。它被广泛应用于机器学习、模式识别、信号处理等领域。

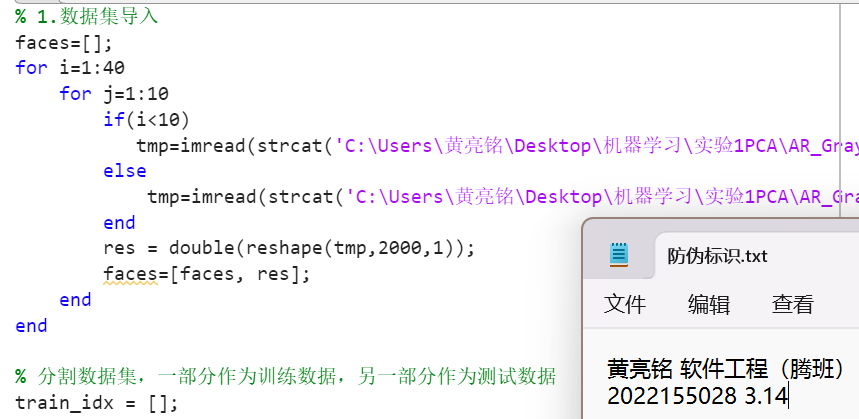
④ 证明其投影矩阵等价于重构误差最小的模型的解



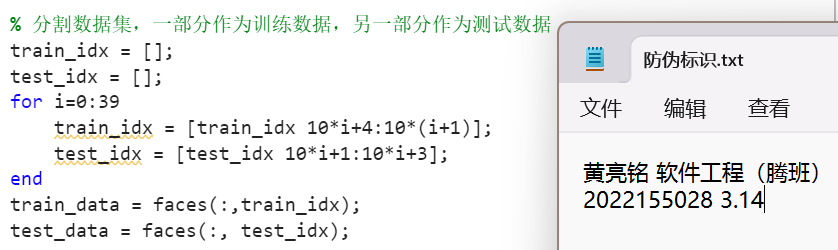
图：投影矩阵等价于重构误差最小

2. 实现PCA算法的人脸重构

1. 预处理工作
   1. 导入AR数据集



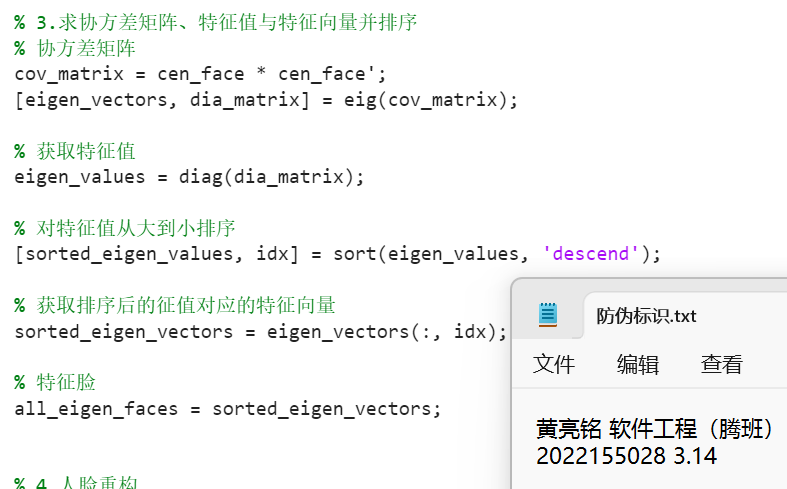
* 1. 分割数据集，一部分作为训练数据，另一部分作为测试数据



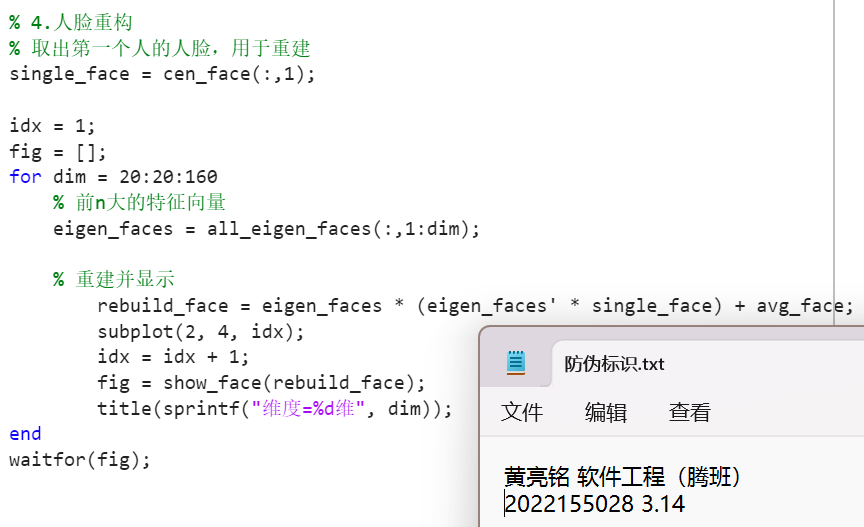
* 1. 对原始数据标准化处理



* 1. 对协方差矩阵进行特征值分解，得到特征值和对应的特征向量，并从中选取特征值最大的n个特征向量



② 人脸重构：选定n的取值，然后重构人脸。上述n的取值为[20:160:20]，即取值范围为[20, 160]，步长为20。

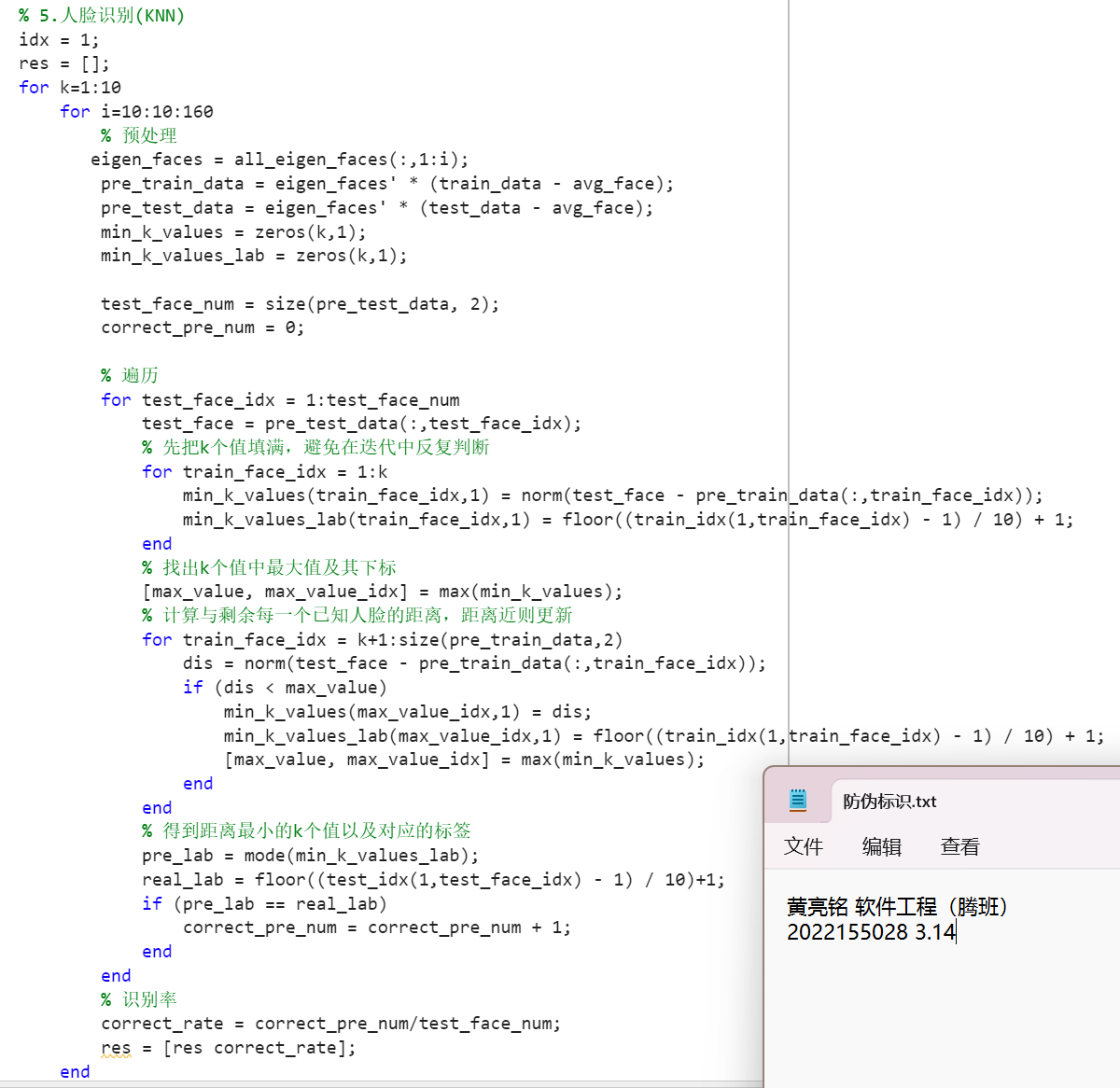


人脸重构代码



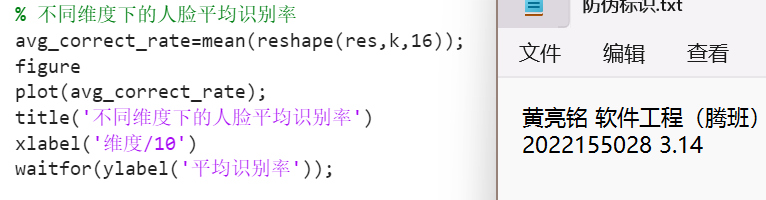
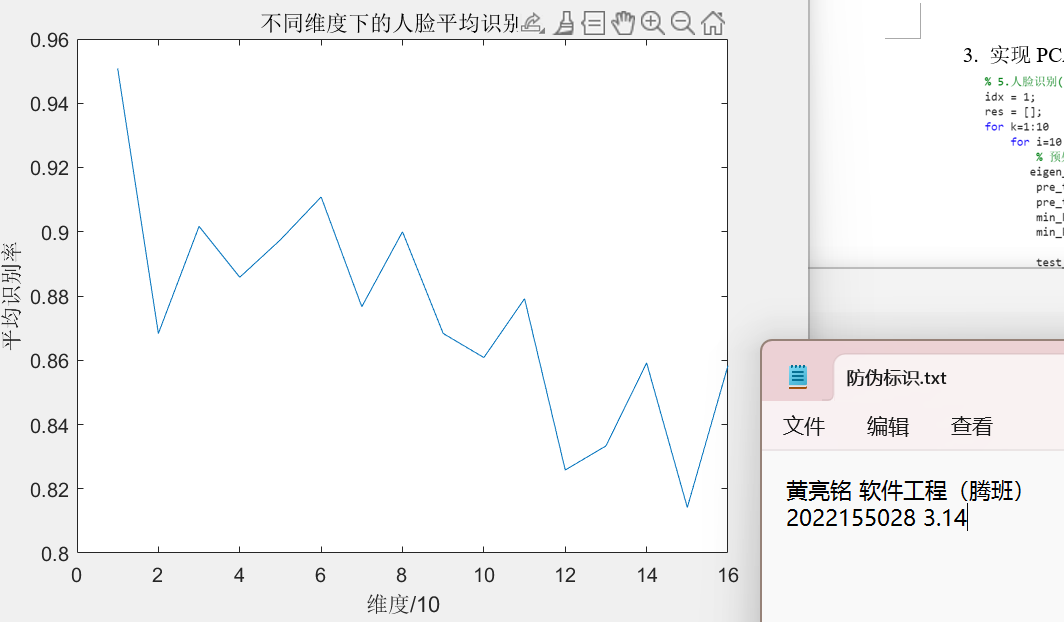
人脸重构结果显示

3. 实现PCA算法的人脸识别（KNN）



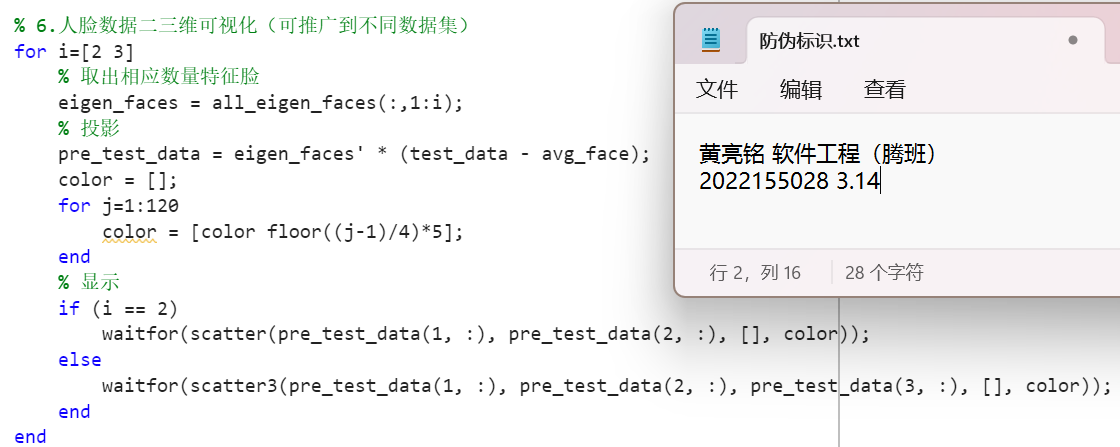
K邻近算法实现

人脸识别率可视化截图：垂直坐标轴为识别率，水平坐标为维度

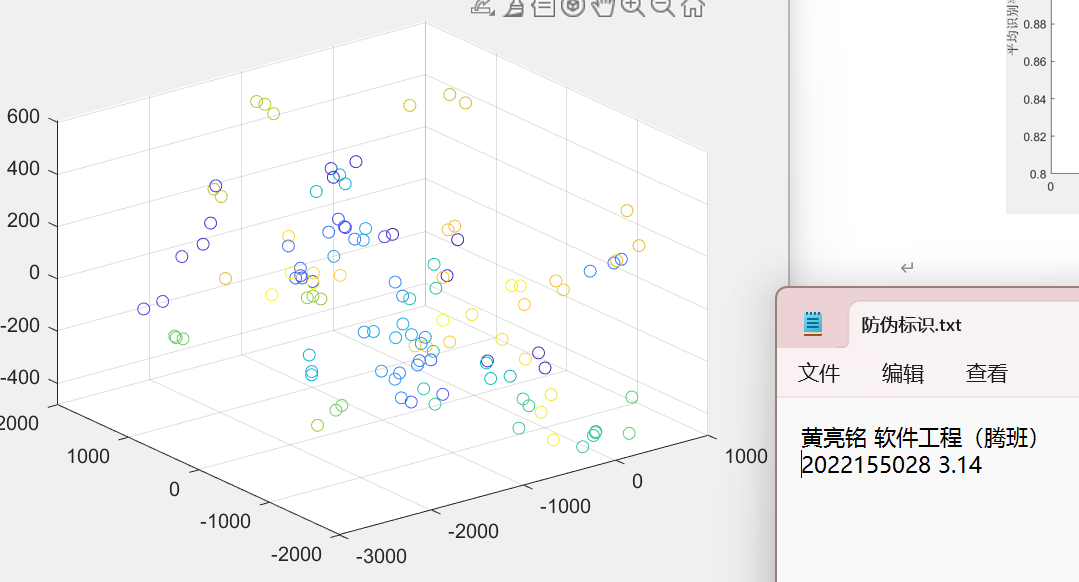
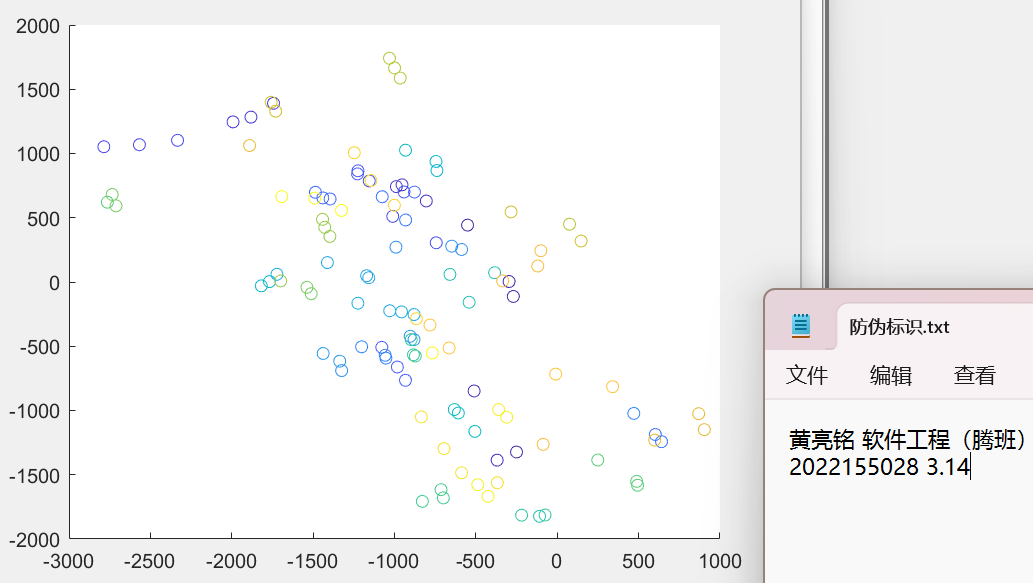
 

可视化结果

4. 实现二维和三维空间数据的可视化



实现代码



二维可视化 三维可视化

5. 设计一个新的PCA算法，内容简要写在实验报告中，并与经典PCA比较。

我设计了一个名为RPCA的算法，即：随机PCA算法。具体做法如下：在经典PCA算法开始之前，对原始数据进行随机降维，即随机生成一个低维投影矩阵，随机数满足高斯分布，然后将原始数据投影到这个低维空间中，作为经典PCA的原始数据。

与经典PCA相比，随机PCA在维数较高的时候复杂度相较低于经典PCA，但是随机PCA的精度低于经典PCA。

随机化的简单代码如下：

假设X是原始数据矩阵，n是样本数，m是特征维度，降维后的维度为k

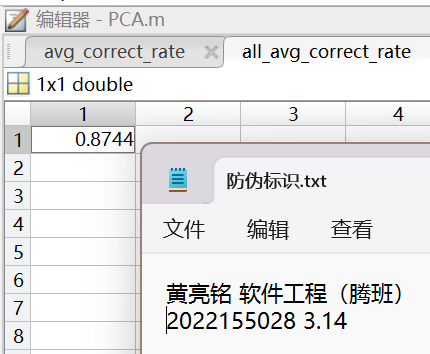
Matrix = randn(m, k);

Faces = X \* matrix;

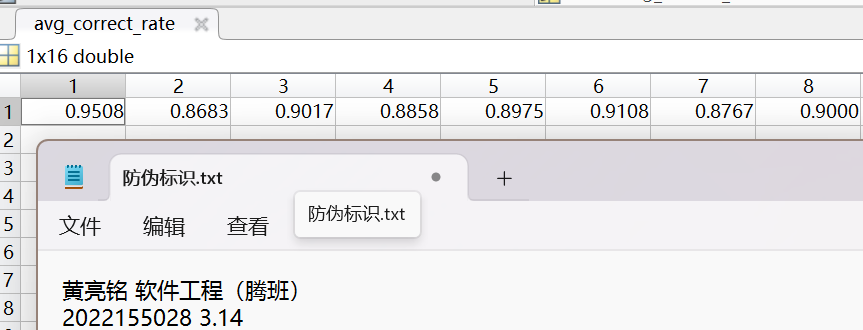
# 四、实验结论或体会

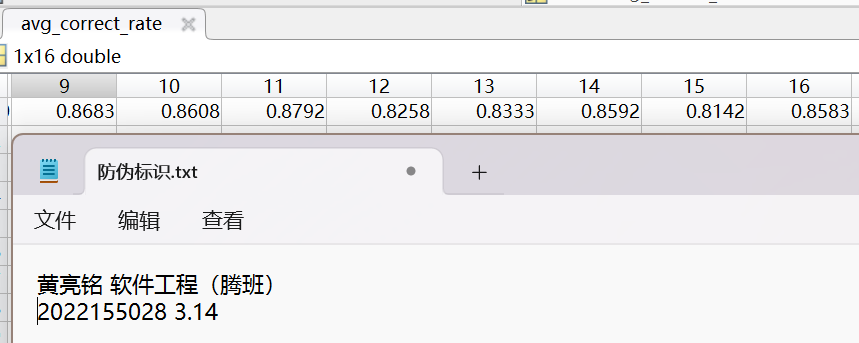
**1. 实验结论**

I．所有维数的平均识别率如下图：



II．不同维数的平均识别率如图所示：（维数=横坐标\*10）





不同维数平均识别率截图

**2.实验体会**

① 通过本次实验，我了解了PCA的原理以及应用；

1. 通过本次实验，我了解了KNN是如何应用于人脸识别技术上；
2. 我学会了发散自己的思维，简单地创新算法；
3. 通过本次实验，我意识自己代码能力的不足。虽然明白算法的原理，但是无法流畅编程，需要借助Matlab帮助、Google等工具编写代码。

|  |
| --- |
| 指导教师批阅意见：  成绩评定：  指导教师签字：  年 月 日 |
| 备注： |

注：1、报告内的项目或内容设置，可根据实际情况加以调整和补充。