《人工智能与模式识别》

实 验 报 告 书

班级： 计科 2002

学号： 202801108

姓名： 李蔚

指导教师： 翟婷婷

2022-2023 学年 第 二 学期

**实验名称： 特征降维算法应用**

**实验时间： 2023 年 4 月 21 日 第 10 周 星期 五**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 题目 | 二 | 三 | | | | | 总分 |
| 1 | 2(a) | 2(b) | 2(c) | 2(d) |
| 得分 |  |  |  |  |  |  |  |

一、实验目的

1. 掌握主成分分析算法PCA进行特征降维的步骤，并在Matlab中编程实现。

2. 通过对比样本集在使用PCA进行变换前和变换后的散点图，对PCA对数据的变换效果有最直观的了解。

3．熟悉PCA在人脸识别中的运用。

二、实验预习（预备知识的问题及回答）

1、 根据课件，详细写出PCA进行特征降维的步骤。(1分)

**解答：**

1.数据预处理：将原始数据进行进行零均值化，即减去这一行的均值。

2.计算协方差矩阵：将标准化后的数据集进行协方差矩阵的计算。

3.计算特征值和特征向量：通过对协方差矩阵进行特征值分解，得到其特征值和对应的特征向量。

4.选取主成分：按照特征值的大小，选取最大的k个特征值对应的特征向量，构成一个转换矩阵。

5.转换数据：将原始数据集乘以转换矩阵，得到降维后的数据集。

三、实验内容（包含实验所用命令或相关程序源代码）

1.使用matlab完成以下实验：

1. 使用n=2000; X = randn(n,2)\*[0.1  1; 1 0.7] 生成2000个2维的样本，使用scatter函数在2维的坐标系中画出这2000个样本点；用xlim([-5 5]) 设置图中x坐标轴的范围为[-5,5]，用ylim([-5 5])设置y坐标轴的范围也为[-5,5]。

**代码和运行结果截图：**(1分)

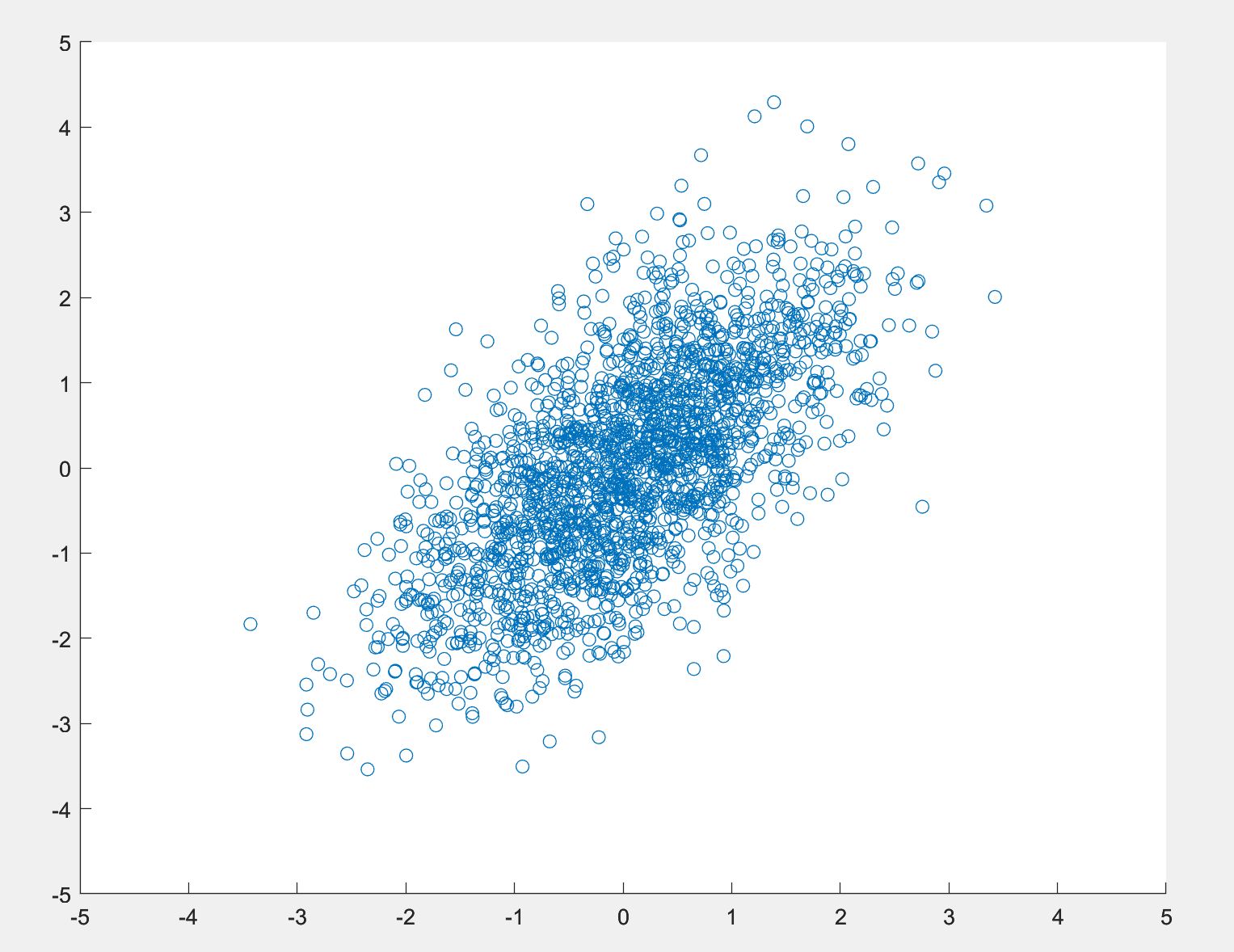
n = 2000;

X = randn(n,2)\*[0.1 1; 1 0.7];

scatter(X(:,1), X(:,2));

xlim([-5 5]);

ylim([-5 5]);



1. 编程实现PCA算法，注意不允许使用matlab自带的pca函数，只允许使用eig函数。利用你实现的PCA算法，对X中样本进行PCA变换，要求保留所有的2个维度，将对X进行PCA变换后的样本点用scatter函数画出，同样设置x和y轴的显示范围为[-5,5]。

**代码和运行结果截图**：(4分)

% 对样本数据进行PCA变换

[m, n] = size(X);

X\_std = (X - mean(X)) ./ std(X); % 对数据进行标准化处理

C = cov(X\_std); % 计算样本数据的协方差矩阵

[V, D] = eig(C); % 对协方差矩阵进行特征值分解，得到特征向量和特征值

[~, idx] = sort(diag(D), 'descend'); % 对特征值进行降序排列，得对应特征向量下标

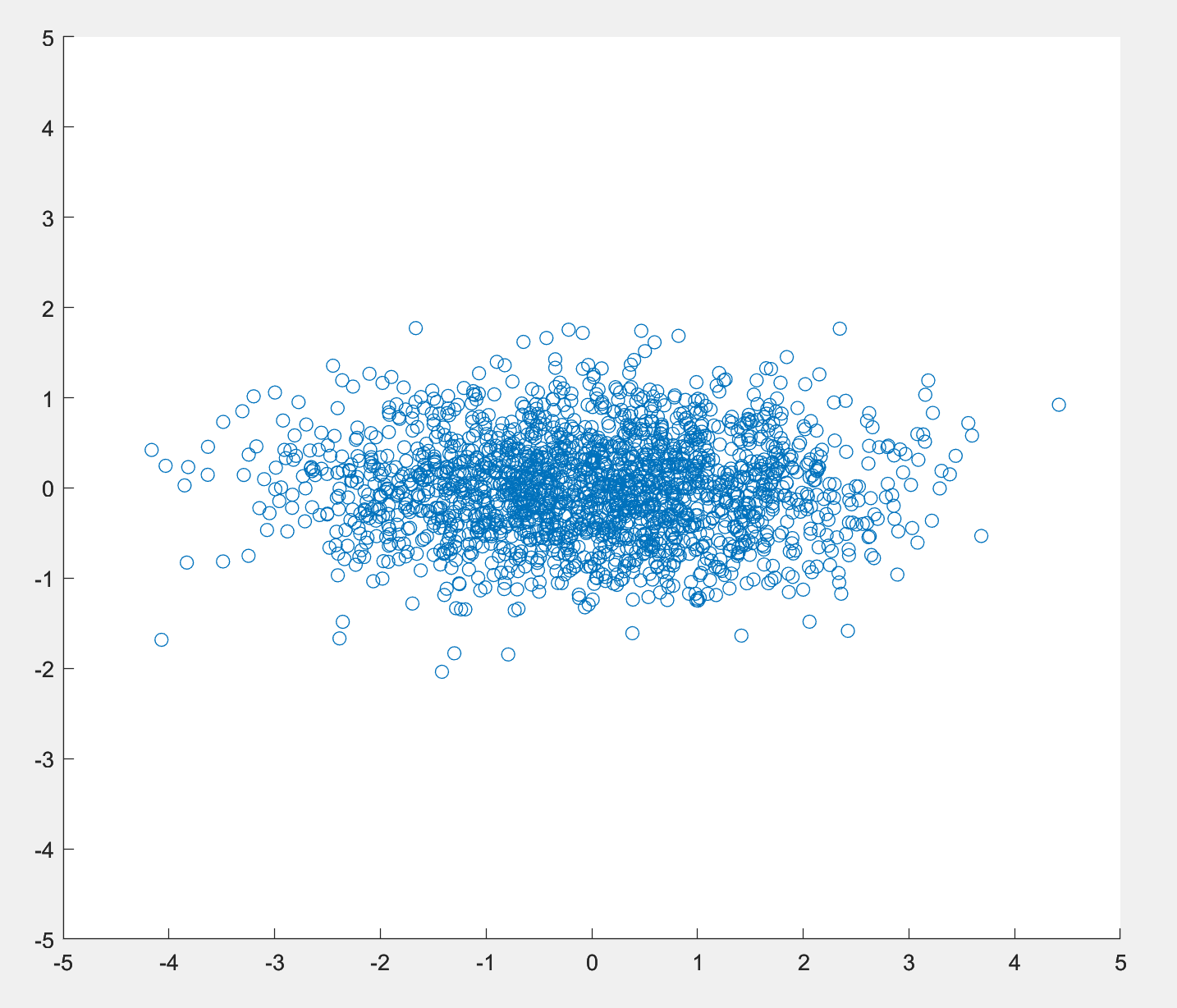
V = V(:, idx); % 根据下标获取对应的特征向量

X\_pca = X\_std \* V; % 将原始样本数据乘以特征向量得到新的PCA变换后的数据

scatter(X\_pca(:,1), X\_pca(:,2));

xlim([-5 5]);

ylim([-5 5]);



(c) 对比两幅scatter图，你将发现如果在PCA变换中保留数据的所有维度，则PCA变换等价于对原始的样本点做了一个旋转，那么旋转的作用是什么？（1分）

**解答：**

如果在PCA变换中保留数据的所有维度，PCA变换仅仅对原始的样本点做了一个旋转，保持了原始数据的所有维度不改变，没有减少任何信息。这种旋转的作用是调整数据坐标系的方向，让原始数据在新的坐标系下更容易被分析、解释和可视化。

2. 对faces\_for\_pca数据集中的28张112\*92的人脸图片进行如下处理：

(a) 对28张人脸图片依次进行如下处理：读取每一张图片，将图片的像素矩阵转换为一个特征向量，存储到数据data的一行，从而得到28\*10304的矩阵，矩阵每一行表示一个样本。在读取过程，使用subplot函数和imshow函数将28张原始的人脸图片显示在一张图中，每行显示7张图，共4行。(2分)

**代码和运行结果截图：**

faceData = imageDatastore('D:\faces\_for\_pca\faces\_for\_pca', 'IncludeSubfolders', true, 'FileExtensions', '.pgm');

data = zeros(28, 112\*92);

for i = 1:28

img = readimage(faceData, i); % 读取图片

img = im2double(img); % 将图像像素矩阵转换为 double 类型

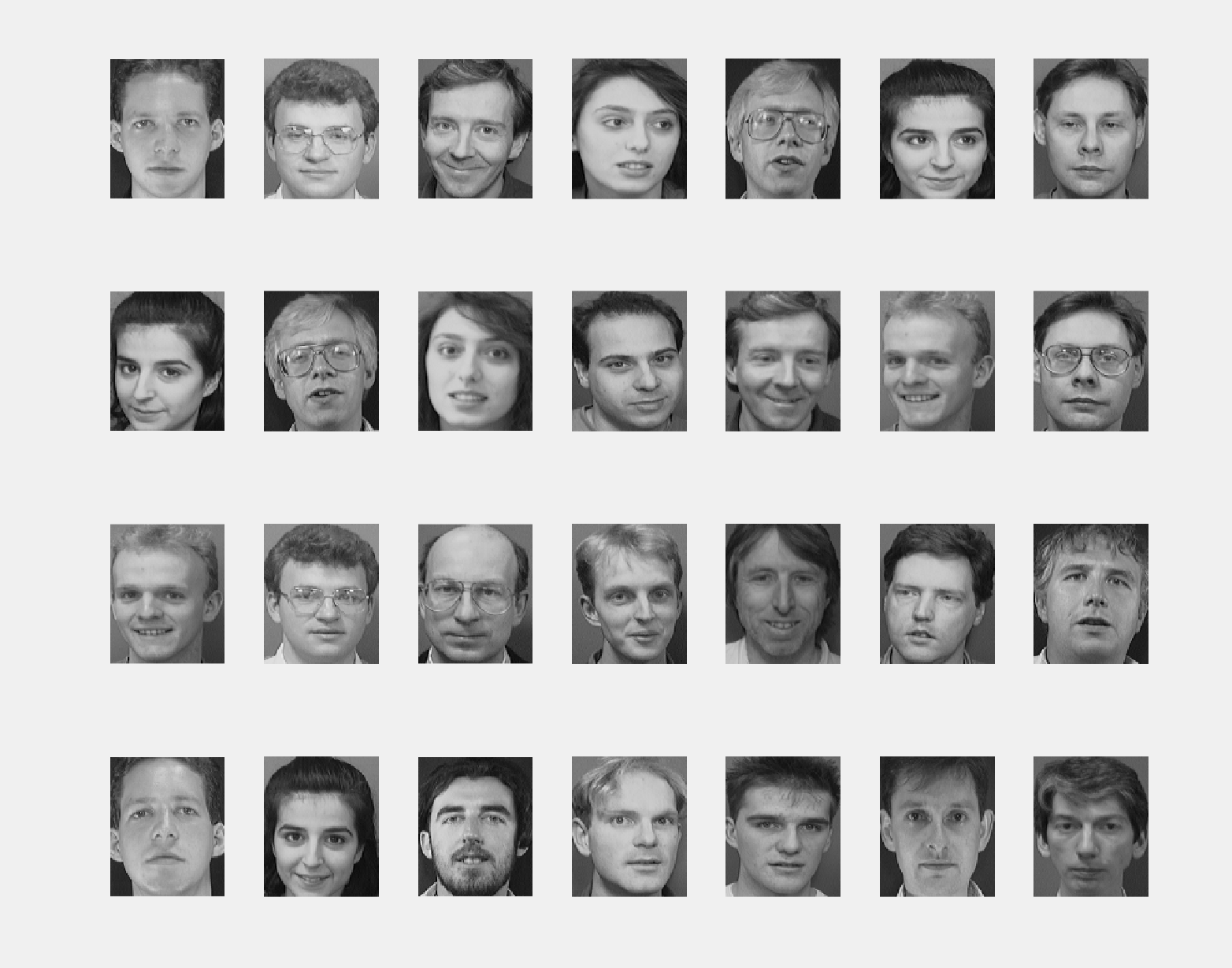
imgVec = reshape(img, 1, []);% 将图像像素矩阵拉伸为行向量

data(i,:) = imgVec;

subplot(4,7,i);

imshow(img);

end



(b) 计算这28个样本的均值向量，使用reshape函数将该均值向量重新调整为112\*92的矩阵，使用imshow函数显示该矩阵。在人脸识别中，均值向量称为平均脸，它描述了数据集中所有人脸平均长成什么样。(1分)

**代码和运行结果截图：**

% 计算均值向量并将其重新调整为112\*92的矩阵并显示

figure;

mean\_face = mean(data, 1);

mean\_face = reshape(mean\_face, [112, 92]);

imshow(mean\_face, []);



(c) 对data进行PCA变换，可以直接使用pca函数。在人脸识别中，每个主成分称为一个特征脸(eigenface)或主成分脸。使用subplot函数和imshow函数将所有的特征脸显示在一张图中，注意仍需要对每个主成分重新调整为112\*92的矩阵再显示。(4分)

**代码和运行结果截图：**

figure;

[COEFF,SCORE,latent,tsquare] = pca(data);

for i=1:27

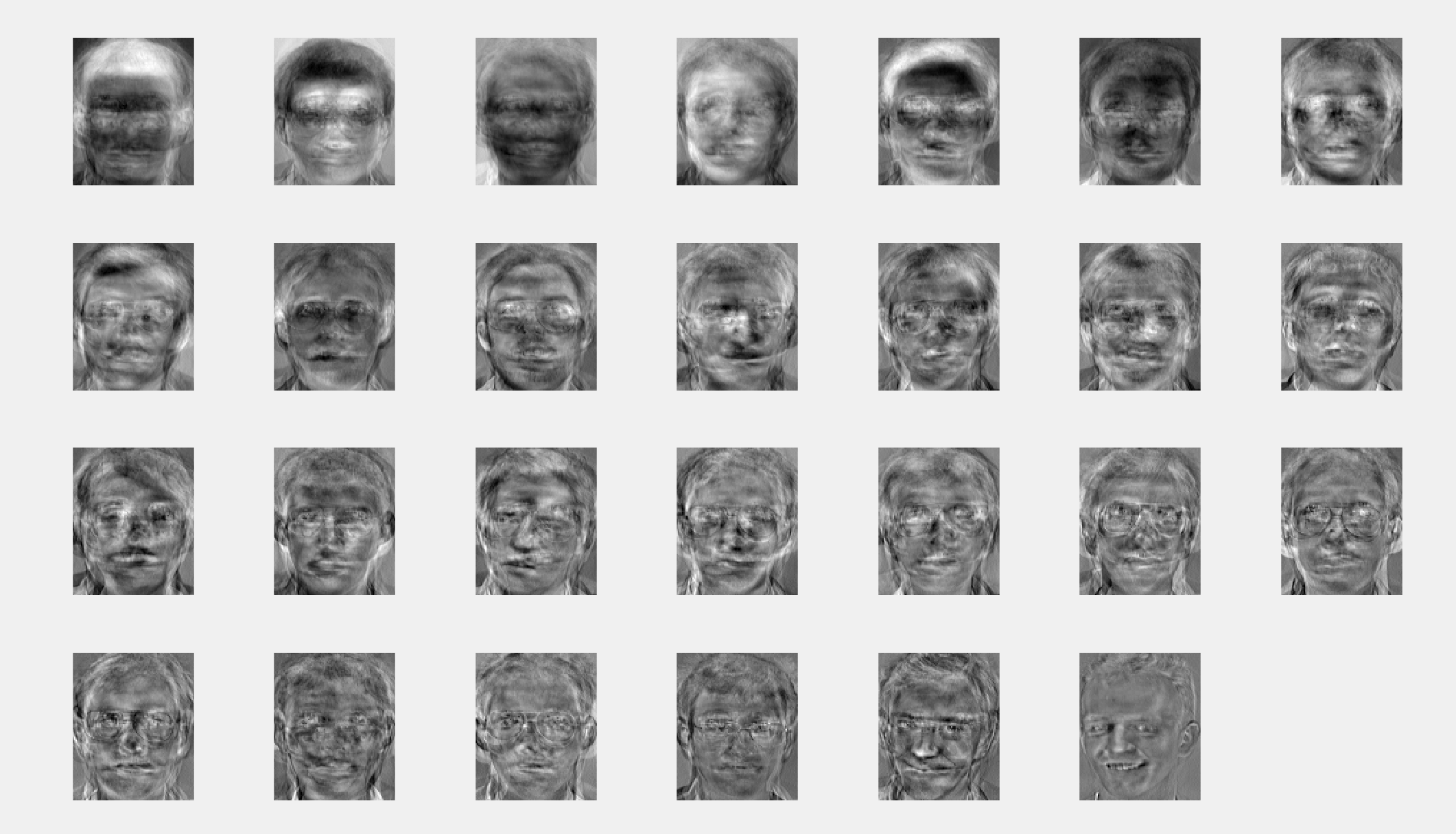
zimg=COEFF(:,i);

zimg=reshape(zimg,112,92);

subplot(4,7,i);

imshow(mat2gray(zimg));

end



(d) 观察原始数据进行PCA降维后的数据是多少维度的？ (1分)

**解答：**

原始数据是10304维，进行PCA降维后的数据是27维。