《人工智能与模式识别》

实 验 报 告 书

班级： 计科1801

学号： 181604124

姓名： 汤礽禾

指导教师： 翟婷婷

2020-2021 学年 第 二 学期

**实验名称： 特征降维算法编程实现**

**实验时间： 2021 年 4 月 21 日 第 8 周 星期 三**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 题目 | 二 | 三 | | | | | 总分 |
| 1 | 2 | 3 | 4 | 5 |
| 得分 |  |  |  |  |  |  |  |

一、实验目的

1．熟悉Matlab操作界面，掌握Matlab的help命令。

2. 掌握主成分分析PCA进行特征降维的步骤，并能在Matlab中编程实现，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、 请详细写出PCA进行特征降维的步骤。(0.5分)

**解答：**

1. **计算散布矩阵S**
2. **计算S的特征值和特征向量**
3. **将特征向量按相应的特征值从大到小排序**
4. **选择特征值最大的d’个特征向量作为投影向量，构成d\*d**’**维的投影矩阵Q**
5. **对于任意的d维样本x，用PAC降维向量为**

2、PCA算法的主成分e\_1, e\_2,..., e\_k之间有什么关系? 每个主成分的模是多少? (0.5分)

**解答：**

1. **他们是最优的投影矩阵，构成了一个新的K维坐标系，并组成了投影矩阵Q**
2. **每个主成分的模是1**

3、给定d维的样本集D={x\_1, x\_2,⋯, x\_n}，μ为样本集的均值向量，我们知道，样本的**协方差矩阵**定义为, 这里符号表示矩阵的转置运算。而在PCA中，**散布矩阵**定义为，可以看到散布矩阵是协方差矩阵的倍。试分析在PCA用协方差矩阵来代替散布矩阵，会有什么样的影响，会影响PCA降维的结果吗？(1分)

**解答：**

**协方差来代替散布矩阵得到的特征值可能偏小，因此难以找到较大的特征值，即难以辨别有用的较大的特征和数据噪音。但并不会影响到降维后的向量，结果并不会改变。**

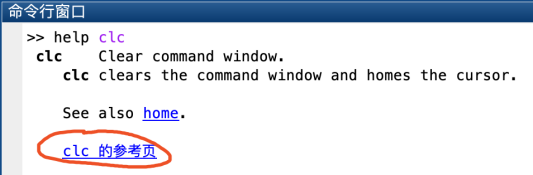
三、实验内容（包含实验所用命令或相关程序源代码）

1. 首先了解Matlab的运行环境，在Matlab的命令行窗口中输入clc，看看有什么效果。

**解答：**clc的作用是：(0.5分)

**清空命令行窗口，对工作环境中的全部变量无任何影响**

1. Matlab有非常棒的帮助文档，对于任何函数或命令，想要了解其功能是什么，其输入和输出是什么，只需要在命令行窗口中输入：help 命令或函数名。例如，想知道clc的作用，输入：help clc，就会在命令行窗口中显示clc的简介，如下图所示：



当点击最下方的**clc参考页，**还能打开Matlab详细的帮助文档，里面提供了大量的函数使用的例子，以供学习。现在要求用help 查询以下几个函数的功能和使用：scatter, randn, eig, mean, cov, repmat。(每个0.5分，共3分)

**解答：**

scatter的功能是：

**散点图**

1. **scatter([x](https://ww2.mathworks.cn/help/matlab/ref/scatter.html?searchHighlight=scatter&s_tid=srchtitle" \l "btrj9jn-1-x),[y](https://ww2.mathworks.cn/help/matlab/ref/scatter.html?searchHighlight=scatter&s_tid=srchtitle" \l "btrj9jn-1-y)) 在向量 x 和 y 指定的位置创建一个包含圆形的散点图。该类型的图形也称为气泡图。**
2. **scatter([x](https://ww2.mathworks.cn/help/matlab/ref/scatter.html?searchHighlight=scatter&s_tid=srchtitle" \l "btrj9jn-1-x),[y](https://ww2.mathworks.cn/help/matlab/ref/scatter.html?searchHighlight=scatter&s_tid=srchtitle" \l "btrj9jn-1-y),[sz](https://ww2.mathworks.cn/help/matlab/ref/scatter.html?searchHighlight=scatter&s_tid=srchtitle" \l "btrj9jn-1-sz)) 指定圆大小。要绘制大小相等的圆圈，请将 sz 指定为标量。要绘制大小不等的圆，请将 sz 指定为长度等于 x 和 y 的长度的向量。**
3. **scatter([x](https://ww2.mathworks.cn/help/matlab/ref/scatter.html?searchHighlight=scatter&s_tid=srchtitle" \l "btrj9jn-1-x),[y](https://ww2.mathworks.cn/help/matlab/ref/scatter.html?searchHighlight=scatter&s_tid=srchtitle" \l "btrj9jn-1-y),[sz](https://ww2.mathworks.cn/help/matlab/ref/scatter.html?searchHighlight=scatter&s_tid=srchtitle" \l "btrj9jn-1-sz),[c](https://ww2.mathworks.cn/help/matlab/ref/scatter.html?searchHighlight=scatter&s_tid=srchtitle" \l "btrj9jn-1-c)) 指定圆颜色。要以相同的颜色绘制所有圆圈，请将 c 指定为颜色名称或 RGB 三元组。要使用不同的颜色，请将 c 指定为向量或由 RGB 三元组组成的三列矩阵。**
4. **scatter(\_\_\_,['filled'](https://ww2.mathworks.cn/help/matlab/ref/scatter.html?searchHighlight=scatter&s_tid=srchtitle" \l "btrj9jn-1-filled)) 填充圆形。可以将 'filled' 选项与前面语法中的任何输入参数组合一起使用。**
5. **scatter(\_\_\_,[mkr](https://ww2.mathworks.cn/help/matlab/ref/scatter.html?searchHighlight=scatter&s_tid=srchtitle" \l "btrj9jn-1-mkr)) 指定标记类型。**
6. **scatter(\_\_\_,[Name,Value](https://ww2.mathworks.cn/help/matlab/ref/scatter.html?searchHighlight=scatter&s_tid=srchtitle" \l "namevaluepairarguments)) 使用一个或多个名称-值对组参数修改散点图。例如，'LineWidth',2 将标记轮廓宽度设置为 2 磅。**
7. **scatter([ax](https://ww2.mathworks.cn/help/matlab/ref/scatter.html?searchHighlight=scatter&s_tid=srchtitle" \l "btrj9jn-1-ax),\_\_\_) 将在 ax 指定的坐标区中，而不是在当前坐标区中绘制图形。选项 ax 可以位于前面的语法中的任何输入参数组合之前。**
8. **[s](https://ww2.mathworks.cn/help/matlab/ref/scatter.html?searchHighlight=scatter&s_tid=srchtitle" \l "btrj9jn-1-s) = scatter(\_\_\_) 返回 Scatter 对象。在创建散点图后，以后可使用 s 对其进行修改。**

randn的功能是：

**正态分布的随机数**

1. **X = randn 返回一个从标准正态分布中得到的随机标量。**
2. **X = randn([n](https://ww2.mathworks.cn/help/matlab/ref/randn.html?searchHighlight=randn&s_tid=srchtitle" \l "bufqhx9-n)) 返回由正态分布的随机数组成的 n×n 矩阵。**
3. **X = randn([sz1,...,szN](https://ww2.mathworks.cn/help/matlab/ref/randn.html?searchHighlight=randn&s_tid=srchtitle" \l "bufqhx9-sz1szN)) 返回由随机数组成的 sz1×...×szN 数组，其中 sz1,...,szN 指示每个维度的大小。例如：randn(3,4) 返回一个 3×4 的矩阵。**
4. **X = randn([sz](https://ww2.mathworks.cn/help/matlab/ref/randn.html?searchHighlight=randn&s_tid=srchtitle" \l "bufqhx9-sz)) 返回由随机数组成的数组，其中大小向量 sz 定义 size(X)。例如：randn([3 4]) 返回一个 3×4 的矩阵。**
5. **X = randn(\_\_\_,[typename](https://ww2.mathworks.cn/help/matlab/ref/randn.html?searchHighlight=randn&s_tid=srchtitle" \l "bufqhx9-typename)) 返回由 typename 数据类型的随机数组成的数组。typename 输入可以是 'single' 或 'double'。您可以使用上述语法中的任何输入参数。**
6. **X = randn(\_\_\_,'like',[p](https://ww2.mathworks.cn/help/matlab/ref/randn.html?searchHighlight=randn&s_tid=srchtitle" \l "bufqhx9-p)) 返回由 p 等随机数组成的数组；也就是与 p 同一对象类型。您可以指定 typename 或 'like'，但不能同时指定两者。**
7. **X = randn([s](https://ww2.mathworks.cn/help/matlab/ref/randn.html?searchHighlight=randn&s_tid=srchtitle" \l "mw_6c956948-f17a-472a-9564-996c7226a727),\_\_\_) 从随机数流 s 而不是默认全局流生成数字。要创建一个流，请使用 [RandStream](https://ww2.mathworks.cn/help/matlab/ref/randstream.html)。指定 s，后跟上述语法中的任意参数组合，但涉及 'like' 的组合除外。此语法不支持 'like' 输入。**

mean的功能是:

**数组的均值**

1. **M = mean([A](https://ww2.mathworks.cn/help/matlab/ref/mean.html?searchHighlight=mean&s_tid=srchtitle" \l "bt5b82t-1-A)) 返回 A 沿大小不等于 1 的第一个数组维度的元素的[均值](https://ww2.mathworks.cn/help/matlab/ref/mean.html" \l "bupom9u)。**

* **如果 A 是向量，则 mean(A) 返回元素均值。**
* **如果 A 为矩阵，那么 mean(A) 返回包含每列均值的行向量。**
* **如果 A 是多维数组，则 mean(A) 沿大小不等于 1 的第一个数组维度计算，并将这些元素 视为向量。此维度会变为 1，而所有其他维度的大小保持不变。**

1. **M = mean([A](https://ww2.mathworks.cn/help/matlab/ref/mean.html?searchHighlight=mean&s_tid=srchtitle" \l "bt5b82t-1-A),'all') 计算 A 的所有元素的均值。此语法适用于 MATLAB® R2018b 及更高版本。**
2. **M = mean([A](https://ww2.mathworks.cn/help/matlab/ref/mean.html?searchHighlight=mean&s_tid=srchtitle" \l "bt5b82t-1-A),[dim](https://ww2.mathworks.cn/help/matlab/ref/mean.html?searchHighlight=mean&s_tid=srchtitle" \l "bt5b82t-1-dim)) 返回维度 dim 上的均值。例如，如果 A 为矩阵，则 mean(A,2) 是包含每一行均值的列向量。**
3. **M = mean([A](https://ww2.mathworks.cn/help/matlab/ref/mean.html?searchHighlight=mean&s_tid=srchtitle" \l "bt5b82t-1-A),[vecdim](https://ww2.mathworks.cn/help/matlab/ref/mean.html?searchHighlight=mean&s_tid=srchtitle" \l "mw_7e557632-72ff-44be-b891-c8646646387d)) 计算向量 vecdim 所指定的维度上的均值。例如，如果 A 是矩阵，则 mean(A,[1 2]) 是 A 中所有元素的均值，因为矩阵的每个元素都包含在由维度 1 和 2 定义的数组切片中。**
4. **M = mean(\_\_\_,[outtype](https://ww2.mathworks.cn/help/matlab/ref/mean.html?searchHighlight=mean&s_tid=srchtitle" \l "bt5b82t-1-outtype)) 使用前面语法中的任何输入参数返回指定的数据类型的均值。outtype 可以是 'default'、'double' 或 'native'。**
5. **M = mean(\_\_\_,[nanflag](https://ww2.mathworks.cn/help/matlab/ref/mean.html?searchHighlight=mean&s_tid=srchtitle" \l "bt5b82t-1-nanflag)) 指定在上述任意语法的计算中包括还是忽略 NaN 值。mean(A,'includenan') 会在计算中包括所有 NaN 值，而 mean(A,'omitnan') 则忽略这些值。**

eig的功能是：

**特征值和特征向量**

1. **[e](https://ww2.mathworks.cn/help/matlab/ref/eig.html?searchHighlight=eig&s_tid=srchtitle" \l "btgapg5-1-e) = eig([A](https://ww2.mathworks.cn/help/matlab/ref/eig.html?searchHighlight=eig&s_tid=srchtitle" \l "btgapg5-1-A)) 返回一个列向量，其中包含方阵 A 的特征值。**
2. **[[V](https://ww2.mathworks.cn/help/matlab/ref/eig.html?searchHighlight=eig&s_tid=srchtitle" \l "btgapg5-1-V),[D](https://ww2.mathworks.cn/help/matlab/ref/eig.html?searchHighlight=eig&s_tid=srchtitle" \l "btgapg5-1-D)] = eig([A](https://ww2.mathworks.cn/help/matlab/ref/eig.html?searchHighlight=eig&s_tid=srchtitle" \l "btgapg5-1-A)) 返回特征值的对角矩阵 D 和矩阵 V，其列是对应的右特征向量，使得 A\*V = V\*D。**
3. **[[V](https://ww2.mathworks.cn/help/matlab/ref/eig.html?searchHighlight=eig&s_tid=srchtitle" \l "btgapg5-1-V),[D](https://ww2.mathworks.cn/help/matlab/ref/eig.html?searchHighlight=eig&s_tid=srchtitle" \l "btgapg5-1-D),[W](https://ww2.mathworks.cn/help/matlab/ref/eig.html?searchHighlight=eig&s_tid=srchtitle" \l "btgapg5-1-W)] = eig([A](https://ww2.mathworks.cn/help/matlab/ref/eig.html?searchHighlight=eig&s_tid=srchtitle" \l "btgapg5-1-A)) 还返回满矩阵 W，其列是对应的左特征向量，使得 W'\*A = D\*W'。**
4. **[e](https://ww2.mathworks.cn/help/matlab/ref/eig.html?searchHighlight=eig&s_tid=srchtitle" \l "btgapg5-1-e) = eig([A](https://ww2.mathworks.cn/help/matlab/ref/eig.html?searchHighlight=eig&s_tid=srchtitle" \l "btgapg5-1-A),[B](https://ww2.mathworks.cn/help/matlab/ref/eig.html?searchHighlight=eig&s_tid=srchtitle" \l "btgapg5-1-B)) 返回一个列向量，其中包含方阵 A 和 B 的广义特征值。**
5. **[[V](https://ww2.mathworks.cn/help/matlab/ref/eig.html?searchHighlight=eig&s_tid=srchtitle" \l "btgapg5-1-V),[D](https://ww2.mathworks.cn/help/matlab/ref/eig.html?searchHighlight=eig&s_tid=srchtitle" \l "btgapg5-1-D)] = eig([A](https://ww2.mathworks.cn/help/matlab/ref/eig.html?searchHighlight=eig&s_tid=srchtitle" \l "btgapg5-1-A),[B](https://ww2.mathworks.cn/help/matlab/ref/eig.html?searchHighlight=eig&s_tid=srchtitle" \l "btgapg5-1-B)) 返回广义特征值的对角矩阵 D 和满矩阵 V，其列是对应的右特征向量，使得 A\*V = B\*V\*D。**
6. **[[V](https://ww2.mathworks.cn/help/matlab/ref/eig.html?searchHighlight=eig&s_tid=srchtitle" \l "btgapg5-1-V),[D](https://ww2.mathworks.cn/help/matlab/ref/eig.html?searchHighlight=eig&s_tid=srchtitle" \l "btgapg5-1-D),[W](https://ww2.mathworks.cn/help/matlab/ref/eig.html?searchHighlight=eig&s_tid=srchtitle" \l "btgapg5-1-W)] = eig([A](https://ww2.mathworks.cn/help/matlab/ref/eig.html?searchHighlight=eig&s_tid=srchtitle" \l "btgapg5-1-A),[B](https://ww2.mathworks.cn/help/matlab/ref/eig.html?searchHighlight=eig&s_tid=srchtitle" \l "btgapg5-1-B)) 还返回满矩阵 W，其列是对应的左特征向量，使得 W'\*A = D\*W'\*B。**
7. **[\_\_\_] = eig([A](https://ww2.mathworks.cn/help/matlab/ref/eig.html?searchHighlight=eig&s_tid=srchtitle" \l "btgapg5-1-A),[balanceOption](https://ww2.mathworks.cn/help/matlab/ref/eig.html?searchHighlight=eig&s_tid=srchtitle" \l "btgapg5-1-balanceOption))（其中，balanceOption 为 'nobalance'）禁用该算法中的初始均衡步骤。balanceOption 的默认值是 'balance'，表示启用均衡步骤。eig 函数可以返回先前语法中的任何输出参数。**
8. **[\_\_\_] = eig([A](https://ww2.mathworks.cn/help/matlab/ref/eig.html?searchHighlight=eig&s_tid=srchtitle" \l "btgapg5-1-A),[B](https://ww2.mathworks.cn/help/matlab/ref/eig.html?searchHighlight=eig&s_tid=srchtitle" \l "btgapg5-1-B),[algorithm](https://ww2.mathworks.cn/help/matlab/ref/eig.html?searchHighlight=eig&s_tid=srchtitle" \l "btgapg5-1-algorithm))（其中，algorithm 为 'chol'）使用 B 的 Cholesky 分解计算广义特征值。algorithm 的默认值取决于 A 和 B 的属性，但通常是 'qz'，表示使用 QZ 算法。**
9. **[\_\_\_] = eig(\_\_\_,[eigvalOption](https://ww2.mathworks.cn/help/matlab/ref/eig.html?searchHighlight=eig&s_tid=srchtitle" \l "btgapg5-1-eigvalOption)) 使用先前语法中的任何输入或输出以 eigvalOption 指定的形式返回特征值。将 eigvalOption 指定为 'vector' 可返回列向量中的特征值，指定为 'matrix' 可返回对角矩阵中的特征值。**

cov的功能是：

**协方差**

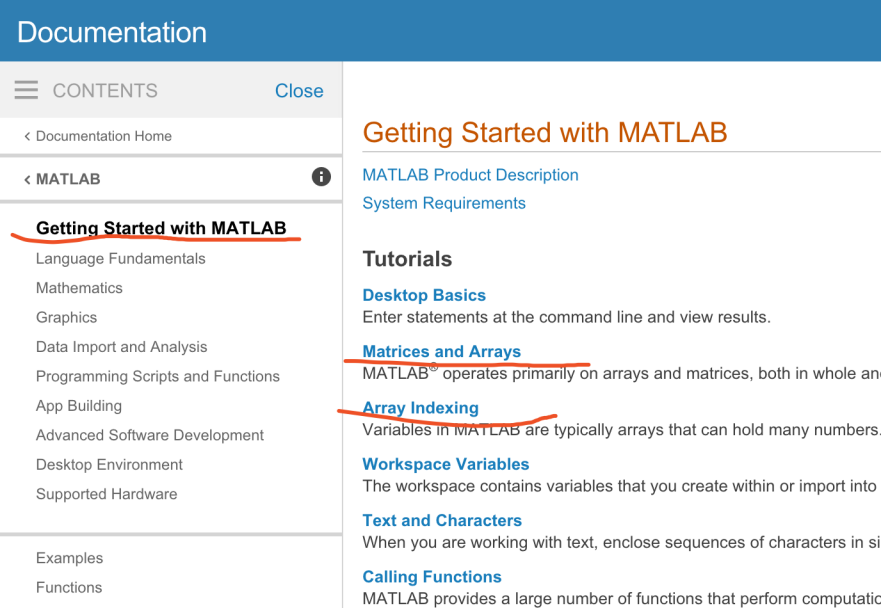
1. **[C](https://ww2.mathworks.cn/help/matlab/ref/cov.html?searchHighlight=cov&s_tid=srchtitle" \l "bumhxdc-1-C) = cov([A](https://ww2.mathworks.cn/help/matlab/ref/cov.html?searchHighlight=cov&s_tid=srchtitle" \l "bumhxdc-1-A)) 返回[协方差](https://ww2.mathworks.cn/help/matlab/ref/cov.html" \l "buo3tdl-1)。**
2. **C 按观测值数量 -1 实现归一化。如果仅有一个观测值，应按 1 进行归一化。**
3. **[C](https://ww2.mathworks.cn/help/matlab/ref/cov.html?searchHighlight=cov&s_tid=srchtitle" \l "bumhxdc-1-C) = cov([A](https://ww2.mathworks.cn/help/matlab/ref/cov.html?searchHighlight=cov&s_tid=srchtitle" \l "bumhxdc-1-A),[B](https://ww2.mathworks.cn/help/matlab/ref/cov.html?searchHighlight=cov&s_tid=srchtitle" \l "bumhxdc-1-B)) 返回两个随机变量 A 和 B 之间的协方差。**
4. **[C](https://ww2.mathworks.cn/help/matlab/ref/cov.html?searchHighlight=cov&s_tid=srchtitle" \l "bumhxdc-1-C) = cov(\_\_\_,[w](https://ww2.mathworks.cn/help/matlab/ref/cov.html?searchHighlight=cov&s_tid=srchtitle" \l "bumhxdc-1-w)) 为之前的任何语法指定归一化权重。如果 w = 0（默认值），则 C 按观测值数量 -1 实现归一化。w = 1 时，按观测值数量对它实现归一化。**
5. **[C](https://ww2.mathworks.cn/help/matlab/ref/cov.html?searchHighlight=cov&s_tid=srchtitle" \l "bumhxdc-1-C) = cov(\_\_\_,[nanflag](https://ww2.mathworks.cn/help/matlab/ref/cov.html?searchHighlight=cov&s_tid=srchtitle" \l "bumhxdc-1-nanflag)) 指定一个条件，用于在之前的任何语法的计算中忽略 NaN 值。例如，cov(A,'omitrows') 将忽略 A 的具有一个或多个 NaN 元素的所有行。**

repmat的功能是：

**重复数组副本**

1. **B = repmat([A](https://ww2.mathworks.cn/help/matlab/ref/repmat.html?searchHighlight=repmat&s_tid=srchtitle" \l "btzavfc-1-A),[n](https://ww2.mathworks.cn/help/matlab/ref/repmat.html?searchHighlight=repmat&s_tid=srchtitle" \l "btzavfc-1-n)) 返回一个数组，该数组在其行维度和列维度包含 A 的 n 个副本。A 为矩阵时，B 大小为 size(A)\*n。**
2. **B = repmat([A](https://ww2.mathworks.cn/help/matlab/ref/repmat.html?searchHighlight=repmat&s_tid=srchtitle" \l "btzavfc-1-A),[r1,...,rN](https://ww2.mathworks.cn/help/matlab/ref/repmat.html?searchHighlight=repmat&s_tid=srchtitle" \l "btzavfc-1-r1rN)) 指定一个标量列表 r1,..,rN，这些标量用于描述 A 的副本在每个维度中如何排列。当 A 具有 N 维时，B 的大小为 size(A).\*[r1...rN]。例如：repmat([1 2; 3 4],2,3) 返回一个 4×6 的矩阵。**
3. **B = repmat([A](https://ww2.mathworks.cn/help/matlab/ref/repmat.html?searchHighlight=repmat&s_tid=srchtitle" \l "btzavfc-1-A),[r](https://ww2.mathworks.cn/help/matlab/ref/repmat.html?searchHighlight=repmat&s_tid=srchtitle" \l "btzavfc-1-r)) 使用行向量 r 指定重复方案。例如，repmat(A,[2 3]) 与 repmat(A,2,3) 返回相同的结果。**

3. 打开Matlab详细的帮助文档，方法：Matlab界面中，菜单栏中帮助->文档，或者用help命令，在该文档中点击getting started with matlab，然后选择Matrices and Arrays，如下图所示：进入之后，通过Maltab提供的例子学习在Matlab中如何进行最基本的矩阵运算：矩阵/数组创建，矩阵/数组的加减法、乘法、连接操作，矩阵的转置运算，特别注意，乘(\*)和点乘(.\*)的区别。然后再进入Array Indexing，学习如何对矩阵/数组进行索引。



通过上述的学习，完成如下的练习：(每个0.5分，共2.5分)

(1)创建一个2\*3矩阵a，其第一列元素为1, 2，第二列为3, 4，第三列为5, 6；

(2) 将矩阵a与a的转置矩阵相乘，得到的结果赋值给b

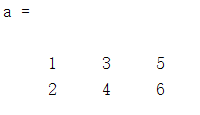
(3) 将矩阵a的第二列元素替换为矩阵b的第一列元素，得到的新矩阵赋值给c;

(4) 计算矩阵c每一行的元素的均值 (提示：用mean函数)

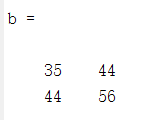
(5) 计算矩阵c每一列的元素的均值。

**在Matlab中运行结果的截图为：**

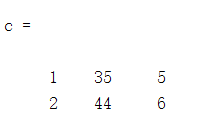
a = [1,2;3,4;5,6]' 或 a = [1 3 5; 2 4 6]



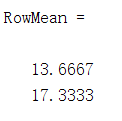
b = a \* a'



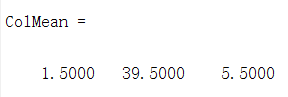
c = a;c(:,2) = b(:, 1)



RowMean = mean(c, 2)



ColMean = mean(c, 1)



1. 使用matlab完成以下实验：
2. 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函数画出这2000个样本；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]。

**scatter图截图：**(1分)

clear;clc;close all;

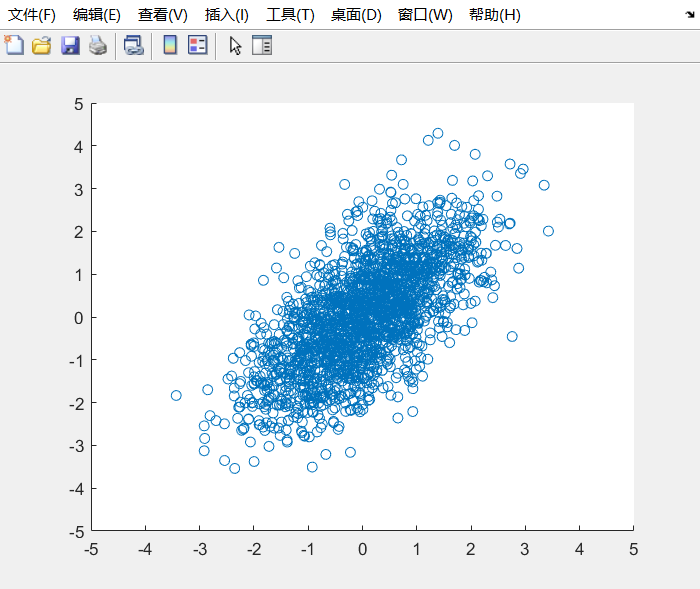
n=2000;

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

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

xlim([-5, 5]);

ylim([-5, 5]);



1. 编程实现对X中样本进行PCA变换，保留所有的2个维度，注意此处不允许使用matlab自带的pca函数直接得到结果，只允许使用eig函数。将对X进行PCA变换后的样本用scatter函数画出；同样设置x和y轴的显示范围为[-5,5]。

**PCA的代码**：(3分)

clear;clc;close all;

n=2000;

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

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

xlim([-5, 5]);

ylim([-5, 5]);

%%

mu = mean(X, 1);

% 1. 计算散布矩阵

S = (X - mu)' \* (X - mu);

% C = cov(X - mu);

% 2. 特征向量 V

[V, D] = eigs(S);

% 特征值 Lambda

Lambda = diag(D);

% 3. 将特征向量按相应的特征值从大到小排序

% B 排序后矩阵

% I 记录原位置

[B, I] = sort(Lambda, 'descend');

Q = V(:, I);

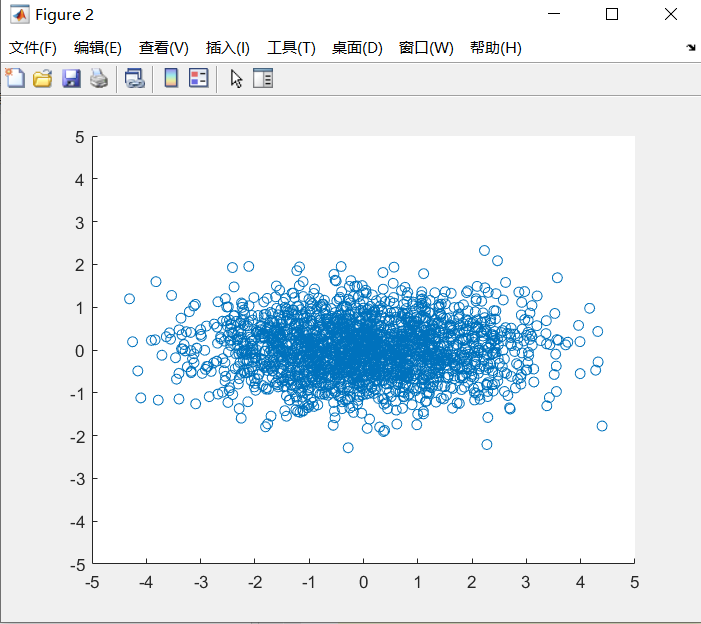
Y = (X - mu) \* Q;

figure,scatter(Y(:, 1), Y(:, 2));

xlim([-5, 5]);

ylim([-5, 5]);

**经过PCA变换后，样本集的scatter图截图：**（1分）



(c)对比两幅scatter图，你发现了什么？如果在PCA变换中保留所有的维度，则PCA变换是不是等价于对原始的样本集做了一个旋转，旋转的作用是什么？（1分）

**解答：**

1. **等价于作了一个旋转。**
2. **旋转的作用是：使得第一主成分方向变成ⅹ轴方向，将数据映射为一维数据，此时数据的方差最大。**

5. PCA在人脸识别中的应用。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。

(3分)

(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。(3分)

(c) 对data进行PCA变换，提示：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。观察特征脸和原始的人脸图像的区别。(3分)

(d) 观察原始数据进行PCA降维后的数据是多少维度的？ (2分)

**所有的Matlab代码贴上，并对运行结果进行截图：（总共3张图）**

clear;clc;close all;

%%

path = 'E:\扬大\模式识别\实验\测试\faces\_for\_pca';

img\_path = dir(strcat(path,'\\*.pgm'));

img\_num = length(img\_path);

data = [];

%%

%(a)

for i = 1:img\_num

% 查找图片

img\_name = img\_path(i).name;% 图片名

strcat(path,'\', img\_name)

picture1 = imread(strcat(path, '\', img\_name));%读进来是整数

picture = double(picture1)/255;%让其处于0-1之间

% 4行7列 显示图片

subplot(4,7,i);

imshow(picture);

figure(1);

hold on;

% 图片存储到data的一行

data = [data;picture(:)'];

end

%%

%(b)

data\_mean = mean(data, 1);%按行求mean

data\_mean = reshape(data\_mean, 112, 92);

figure;

imshow(data\_mean);%double 0-1 之间，0黑1白

%%

%(c)

pca\_picture = pca(data);

figure; %新建画布

for i = 1:img\_num-1

img = reshape(pca\_picture(:,i), 112, 92);

subplot(4, 7, i);

imshow(mat2gray(img));% 放大处理？？？？,不然太小，全黑

%imshow(img, []);

end

%%

%(d)

% 27维。查看pca\_picture里面维数是27列，因此27维。



图1：(a)题



图2：(b)题

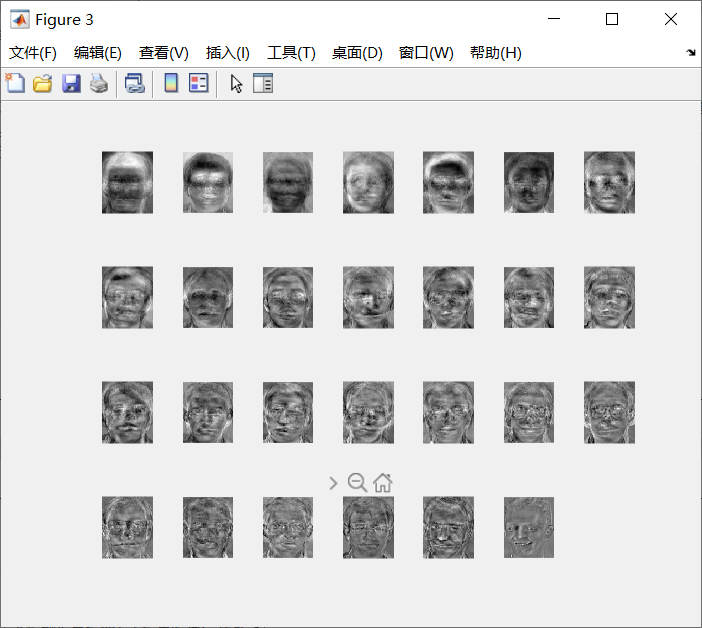


图3：(c)题