

# 学生实验报告

开课学院及实验室：机电学院 电子楼510

2018 年 5 月 24 日

学院	机电学院	年级、专业、班	电信151	姓名	苏伟强	学号	1507400051
实验课程名称	模式识别概论					成绩	
实验项目名称	实验三 主成分分析实验					指导老师	杨钊

## 1 实验目的

通过 matlab 简单的演示程序，理解 PCA 原理

## 2 实验原理

主成分分析 (principal component analysis PCA), 通过正交变换将一组可能存在相关性的变量转换成一组线性不相关的变量, 转换之后的这组变量叫做主成分。常用于提取数据主要特征变量, 常用于高维数据的降维。在用统计分析的方法研究多变量的时候, 变量的个数太多, 增加了复杂性, 由于变量之间是有一定的相关性的, 也就是说不同的变量所代表的全局信息是有重叠, 有冗余的。其算法流程如图 1所示:

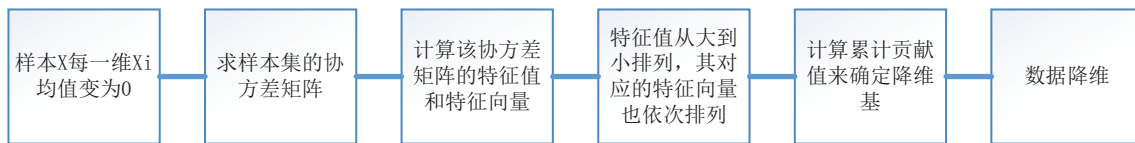


图 1: PCA 算法流程图

## 3 实验内容

如图 2所示绘制随机的 3 维点, 如图 3所示, 选取特征值最大的两个特征向量进行投影, 绘制其投影平面, 如图 4所示, 绘制出 3 维到 2 维的投影垂线以示意这个过程的原理。如图 5所示为最后投影得到的二维点在三维坐标中的位置, 可以看到, 这些点都位于一个平面上, 如图 6所示为二维点到一维点的投影线, 所选的投影基为特征值最大的特征向量, 如图 7所示, 绘制了二维到一维的投影垂线, 如图 8所示为二维降为一维的点。整个过程演示了一个三维特逐渐降为一维的过程, 特征有所损失, 但是仍能被区分。

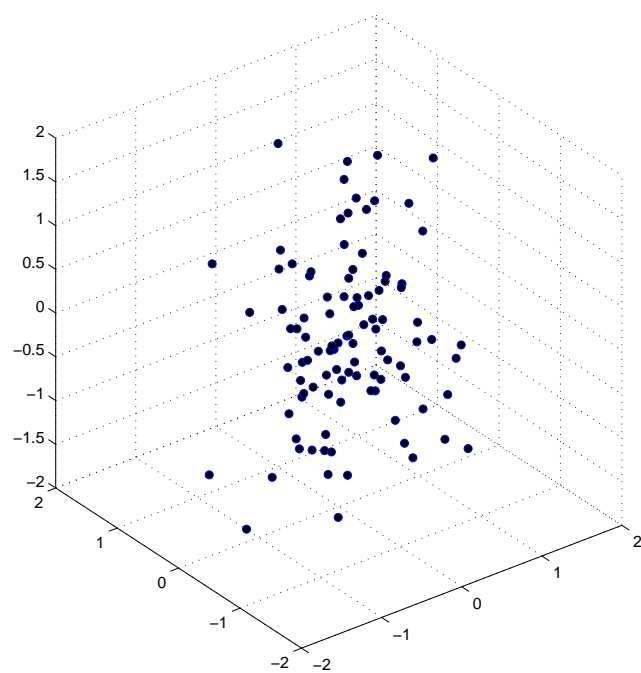


图 2: 随机三维点

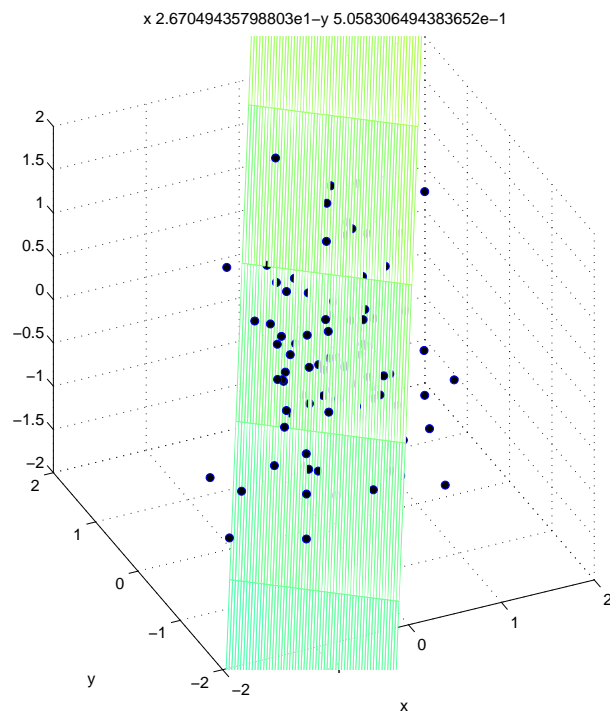


图 3: 三维到二维投影面

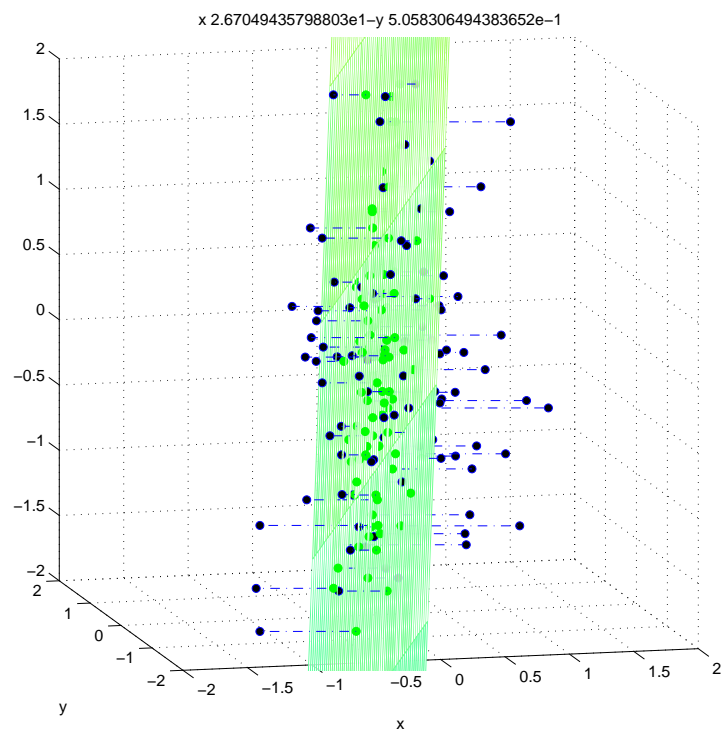


图 4: 三维到二维的投影过程

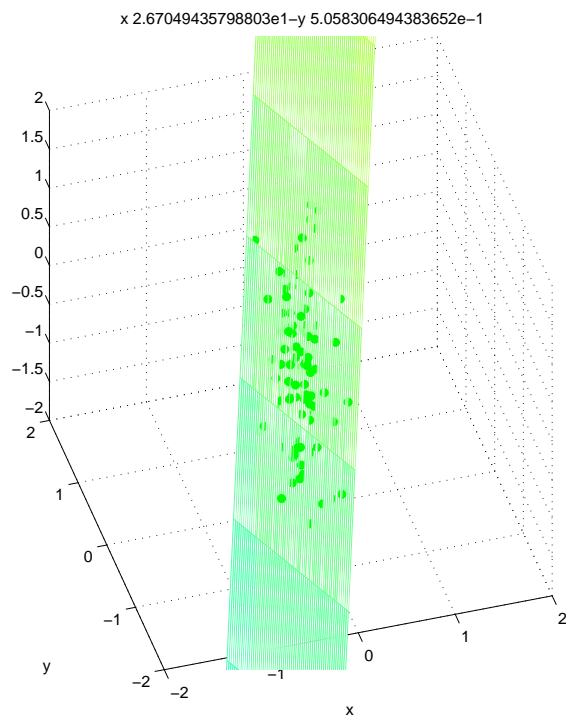


图 5: 三维到二维的投影点

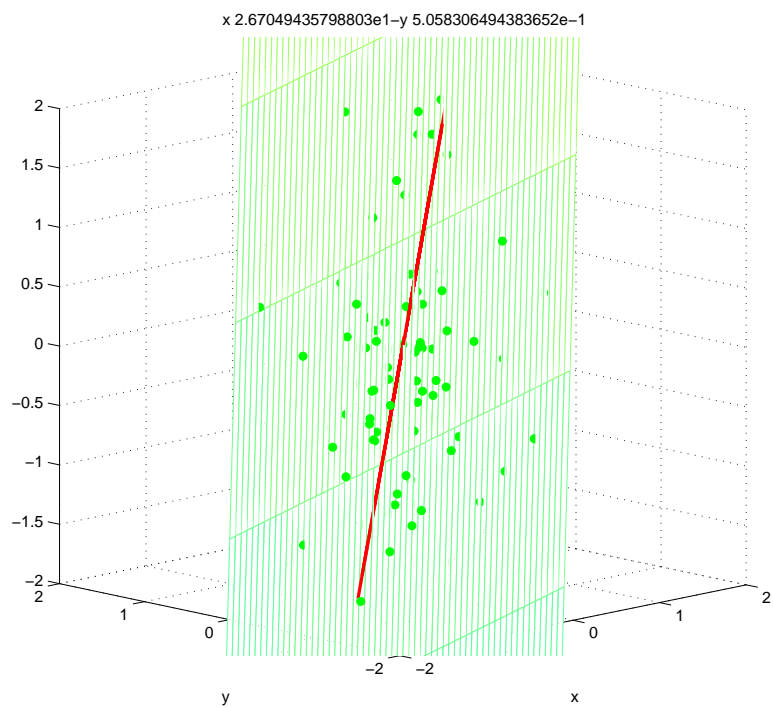


图 6: 二维到一维的投影线

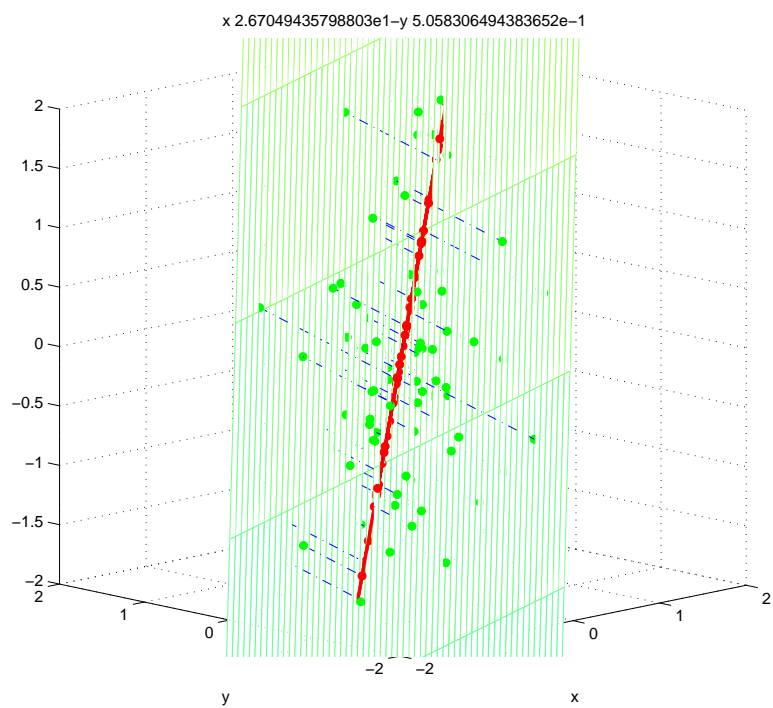


图 7: 二维到一维的投影过程

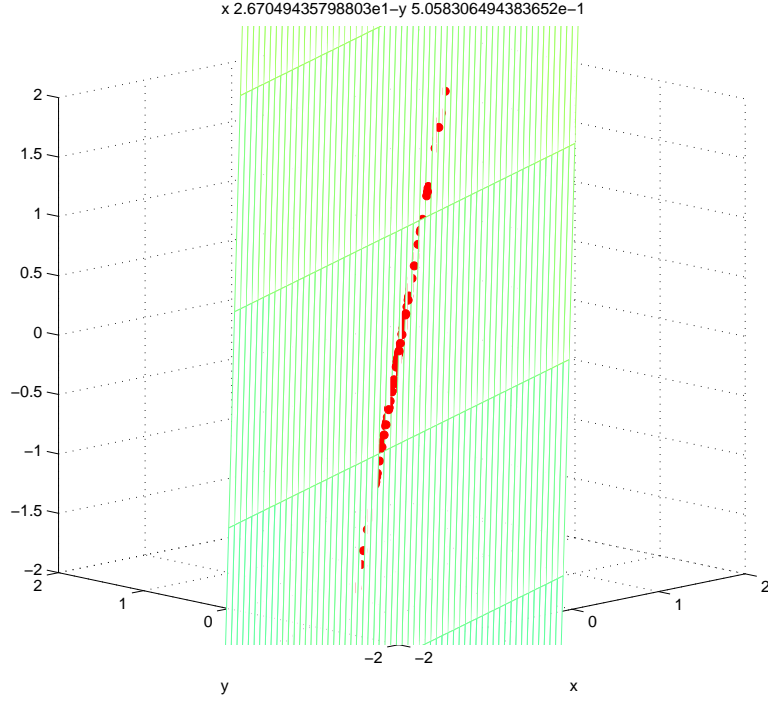


图 8: 二维到一维的投影点

## 4 心得体会

1. 使用 PCA 对数据进行降维的思路：将输入的样本看成一个随机向量，即

$$\vec{X} = \{X_1, X_2, X_3, \dots, X_N\} \quad N = 0, 1, 2, \dots \quad (1)$$

有  $M$  个样本，说明发生了  $M$  次  $\vec{X}$  事件， $\vec{X}$  有  $N$  维，每一维为单独的随机变量  $X_N$ ，若要将降维，降维后的需满足其内部各维之间的协方差为 0，已保证降维后的数据无冗余，每一位表示的信息不正交不重叠；同时各维的方差需要最大，因为方差大代表着数据集中性越差，说明越方便被辨别。随机向量协方差矩阵中，各维均值位于对角线，协方差位于其他非对角线位置，我们需要对角线值最大，其他位置值为 0，这不免让我们想到了矩阵的对角化。对角化完成的功能真是将矩阵除对角线为其他元素变成 0。矩阵的对角化表达式为

$$B = T^{-1}AT \quad (2)$$

其中  $A$  为待对角化矩阵， $T$  为  $A$  的特征向量组成的矩阵，这里我们将  $A$  看成是协方差矩阵，那么  $T$  就是这个协方差矩阵特征向量组成的矩阵。对角化的过程已经完成了协方差 0 的任务，接下来使各维的方差最大，即  $B$  的对角线元素最大。因为特征值大的特征向量是矩阵变换的主要方向，所以其占的变换成分应该比较大，对  $B$  的对角线元素影响也比较大，所以这里才有根据特征值来排序特征向量来组成矩阵  $T$  的方法。按照成分的贡献率，可以执行降维，维度贡献率不能少于 80%。

在使用式 3 计算协方差矩阵的时候， $A$  为样本数据集矩阵， $N$  为样本维数，需要注意的是样本数据集  $A$  必须去均值，即使样本各维均值为 0，这样使用该公式计算出来的才是协方差矩阵。

$$C = \frac{1}{N}AA^T \quad (3)$$

在执行数据降维的时候，从二维数据降至一维的变换为  $A$ ，得到的这个一维的数据是以  $A$  的基为基的坐标再乘以这个基的范数，若这个变换本身就同其基一致的话，得到的一维数据为其投影长度。*Matlab* 中计算特征向量的函数已经完成了特征向量单位化的操作，所以，将数据同其内积得到的便是投影长度。实验中涉及将这个投影长度在原来的维度中显示的问题，只要将这个投影长度乘以原来投影的基向量就行了。

*Matlab* 中如果要在上写文中索引当前 *axes*，进行绘图，只需要进行一次 *hold on*，就可以防止 *plot* 句柄的自动回收。

## 5 代码

PCAdemo.m

```
1 function PCAdemo
2 close all;clear;clc;
3 h_figure=figure;
4 mu=[0 0 0];
5 %协方差矩阵， 对角为方差值0.3， 0.35
6 dim=3;
7 var=[0.2 0 0;0 0.5 0;0 0 0.8];
8 samNum=100;
9 data=mvrnd(mu,var,samNum);
10 h_plotMain=plot3(gca,data(:,1),data(:,2),data(:,3),'o',...
11     'MarkerSize',5,'MarkerFaceColor','k');
12 xlim([-2 2]);ylim([-2 2]);zlim([-2 2]);
13 grid on;axis square;hold on;
14 pause();
15 covdata = cov(data);%求协方差矩阵
16 [eigVector, eigValue]=eig(covdata);%求协方差矩阵的特征值和特征向量
17 eigValue=diag(eigValue)';
18 [eigValue_sort,IX]=sort(eigValue,2,'descend');
19 eigVector_sort=[];
20 for i=1:dim
21 tempIX=IX(i);
22 eigVector_sort=[eigVector_sort eigVector(:,tempIX)];
23 end
24 base_3to2=eigVector_sort(:,1:2);%三维降2维
25 base_3to1=eigVector_sort(:,1);%三维降1维
26 base_3to2=base_3to2';
27 base_3to1=base_3to1';
28 %%%%%%绘制3维变成2维度的投影平面%%%%%%%%
29 A=base_3to2(1,:);%新基i帽(相对原基)
30 B=base_3to2(2,:);%新基j帽(相对原基)
31 C=[0 0 0];
32 syms x y z;
33 D=[ones(4,1),[[x,y,z];A;B;C]]; %由空间解析几何的内容知道D的行列式等于零就是平面方程。
34 detd=det(D);
35 z=solve(detd,z);
36 ezmesh(z,[-2, 2, -2, 2]);
37 pause();
38 %%%%%%%%%%
39 data_proj3_2=base_3to2*data';%投影到新基坐标(相对新基)
40 data_proj3_1=base_3to1*data';%投影到新基坐标(相对新基)
41
42 data_proj3_2=data_proj3_2';
```

---

```

43 data_proj3_1=data_proj3_1';
44
45 data_proj_respect_to_orienbasis3_2=[];%新基坐标用原基表示
46 data_proj_respect_to_orienbasis3_1=[];%新基坐标用原基表示
47 for i=1:samNum
48 data_proj_respect_to_orienbasis3_2=[data_proj_respect_to_orienbasis3_2;...
49 data_proj3_2(i,1)*base_3to2(1,:)+data_proj3_2(i,2)*base_3to2(2,:)];
50 data_proj_respect_to_orienbasis3_1=[data_proj_respect_to_orienbasis3_1;...
51 data_proj3_1(i)*base_3to1(1,:)];
52 end
53
54 %绘制3维降为2维后的数据
55 h_plot3_2=plot3(gca,data_proj_respect_to_orienbasis3_2(:,1),...
56 data_proj_respect_to_orienbasis3_2(:,2),data_proj_respect_to_orienbasis3_2(:,3),...
57 'o','MarkerSize',5,'MarkerEdgeColor','g','MarkerFaceColor','g');
58
59 %绘制3维降为2的投影虚线
60 hl3_2=[];
61 hl3_1=[];
62 for i=1:samNum
63 dd=[data_proj_respect_to_orienbasis3_2(i,:);data(i,:)];
64 hl3_2(i)=plot3(gca,dd(:,1),dd(:,2),dd(:,3),'-','markersize',10);
65 end
66 pause();
67 delete(h_plotMain); delete(hl3_2);
68 pause();
69 %绘制3维降为1的基准线
70 h_zhixian=plot3(gca,data_proj_respect_to_orienbasis3_1(:,1),...
71 data_proj_respect_to_orienbasis3_1(:,2),data_proj_respect_to_orienbasis3_1(:,3),...
72 'color','r','linewidth',2);
73 pause();
74 %绘制3维降为1的数据
75 h_plot3_1=plot3(gca,data_proj_respect_to_orienbasis3_1(:,1),...
76 data_proj_respect_to_orienbasis3_1(:,2),data_proj_respect_to_orienbasis3_1(:,3),...
77 'o','MarkerSize',5,'MarkerEdgeColor','r','MarkerFaceColor','r');
78 %绘制3维降为1的投影虚线
79 for i=1:samNum
80 dd=[data_proj_respect_to_orienbasis3_1(i,:);data_proj_respect_to_orienbasis3_2(i,:)];
81 hl3_1(i)=plot3(gca,dd(:,1),dd(:,2),dd(:,3),'-','markersize',10);
82 end
83 pause();
84 delete(h_plot3_2); delete(hl3_1); delete(h_zhixian);
85 end

```

---