

# PCA实现

——陈汝丹

# 1.数据来源

Andrew Ng在coursera上的机器学习公开课第八课两个数据集

a.Dataset1是二维数据集，图1表示从ex7data1.mat中读取的数据，图2是将这组数据画成点图

```
Terminal
Visualizing example dataset for PCA.
X =
  3.3816  3.3891
  4.5279  5.8542
  2.6557  4.4120
  2.7652  3.7154
  2.8466  4.1755
  3.8907  6.4884
  3.4758  3.6328
  5.9113  6.6808
  3.9289  5.0984
  4.5618  5.6233
  4.5741  5.3977
  4.3717  5.4612
  4.1917  4.9547
  5.2441  4.6615
  2.8358  3.7680
  5.6353  6.3121
  4.6863  5.6652
  2.8505  4.6265
  5.1102  7.3632
```

图1. 读取dataset1

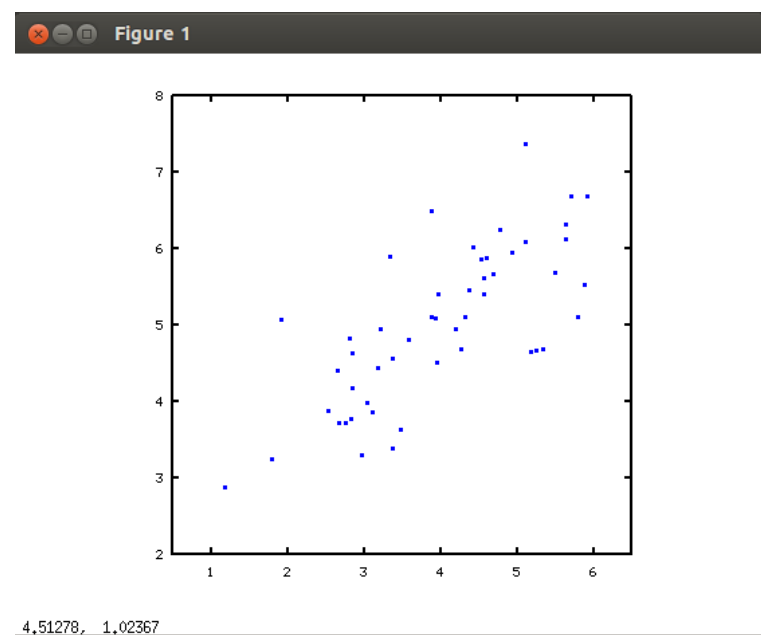


图2. dataset1的2D数据

# 1.数据来源

b. Dataset2人脸数据集ex7faces.mat, 每一张为32\*32的灰度图, 共5000张图, 每张图当成1024维的数据, 图3显示前100张图。输入X的矩阵大小为5000\*1024



图3. 前100张example

## 2.pca实现

### a.计算主成分

1.特征归一化，输入X先减均值再除以标准差

2.计算数据的协方差， $\Sigma = \frac{1}{m} X^T X$

3.然后计算协方差的特征向量，此处是由matlab提供的svd()函数进行奇异值分解，得到主成分U

## 2.pca实现

### b.降维与还原

- 1.用输入 $X$ 与主成分 $U$ 相乘，将输入映射到主成分的坐标系中为 $Z$
- 2.然后将 $Z$ 与主成分转置相乘，恢复到原始数据 $X$

### 3. 执行过程及实验结果

执行ex7\_pca.m

a. Dataset1

load载入数据，plot画出点图  
（图2），然后进行特征归一化，  
执行pca.m，计算出主成分，然  
后将主成分方向在图2上表示出  
来，如图4所示：

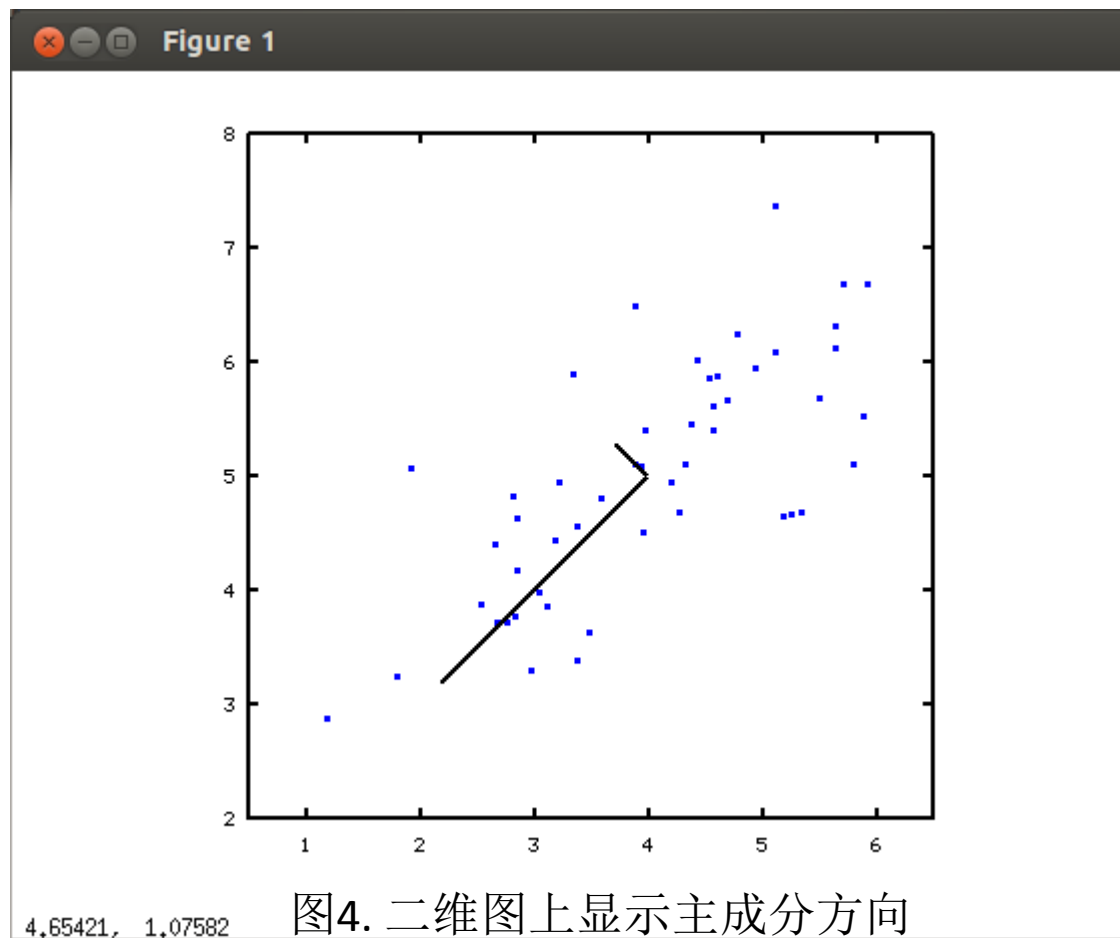
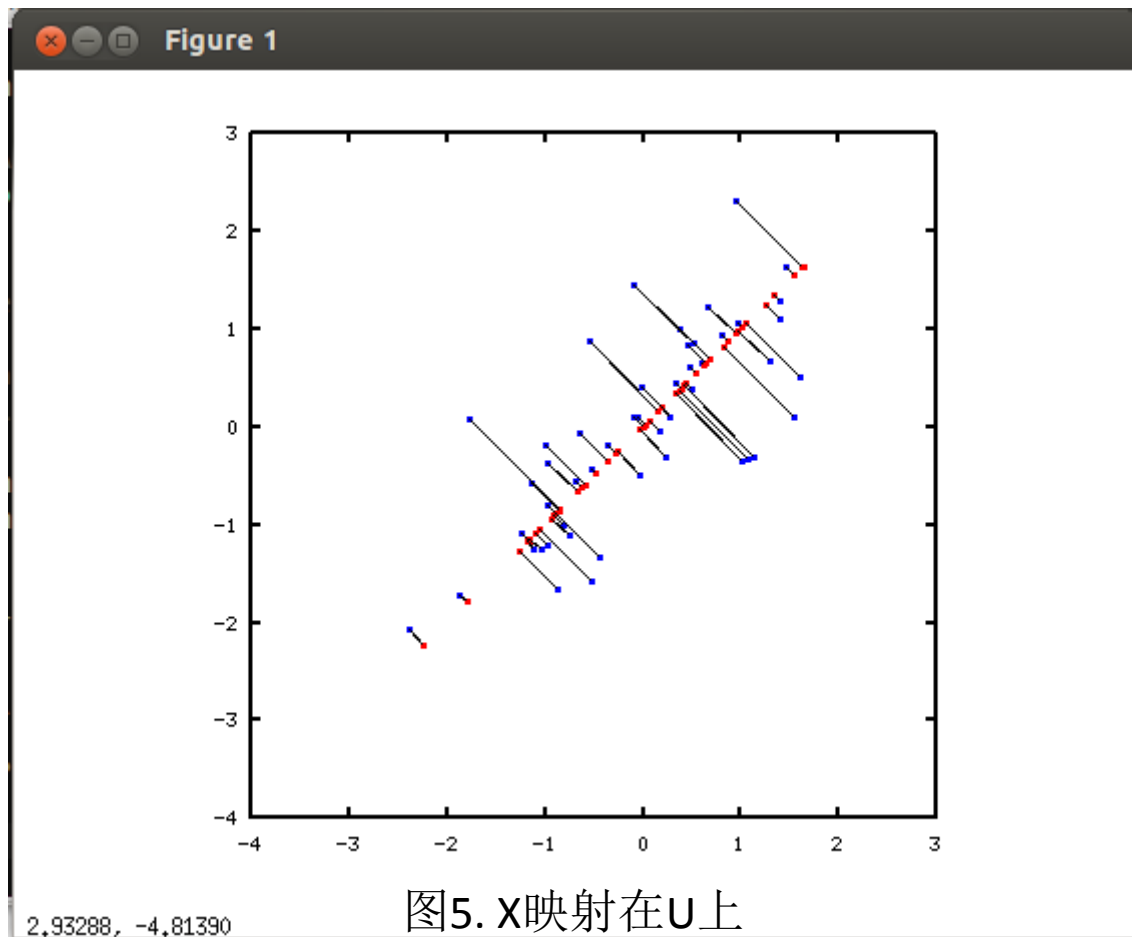


图4. 二维图上显示主成分方向

### 3. 执行过程及实验结果

#### a. Dataset1

图5显示的是将原始数据集 $X$ 映射到 $U$ 方向上的过程，其中蓝色点表示 $X$ ，红色点表示在第一主成分上的对应投影点



### 3. 执行过程及实验结果

#### b. Dataset2

load函数载入人脸数据 $X$ 后，同样经过特征归一化，执行pca.m，计算出主成分 $U$ ，图6显示了前面36个主成分，并选取了前100个主成分进行后续计算

Figure 1



图6. 人脸前36个主成分



### 3. 执行过程及实验结果

#### b. Dataset2

将 $X$ 与 $U$ 的前100项做矩阵乘法得到 $Z$ ，再计算 $Z * U^T$ 得到了重构的人脸，图7左是用于训练原始人脸，右是重构得到的人脸。

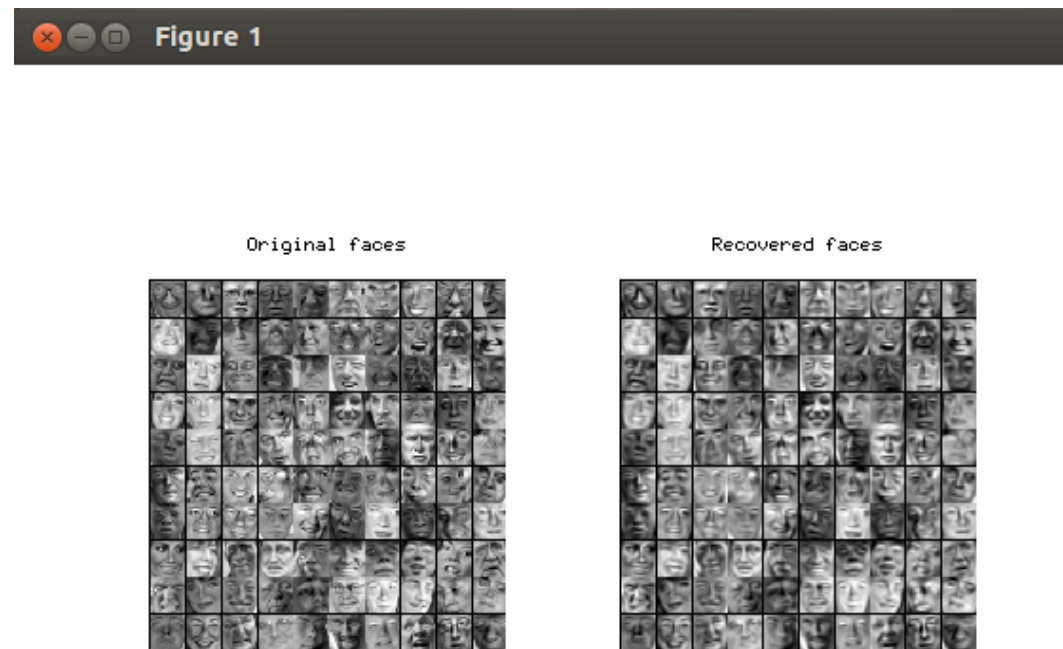


图7. 原始人脸与pca重构人脸

## 4.小结

PCA的目的就是降维，对数据进行压缩处理，通过正交变换映射到低维子空间，保留了原数据的信息。当保留了映射变换的主成分，可以重构出原始信息。