

PCA 学习

王泽民

School of Software
Sun-Yat Sen University

October 27, 2014



- 1 降维与 pca
- 2 PCA 的过程与原理
 - 一起来研究 PCA
 - 重新梳理
- 3 PCA 的应用与实现
 - princomp 介绍
 - 在图形学的其它一些应用
- 4 PCA 的进一步讨论

1 降维与 pca

2 PCA 的过程与原理

- 一起来研究 PCA
- 重新梳理

3 PCA 的应用与实现

- princomp 介绍
- 在图形学的其它一些应用

4 PCA 的进一步讨论



维度引发的问题

- 大多数机器学习和数据挖掘算法的效率和复杂度和维度相关
- 维度信息的冗余，即某些维度之间有明显的关联

Example

拿到一个汽车的样本，里面既有以“千米/每小时”度量的最大速度特征，也有“英里/小时”的最大速度特征

- 一个样本，特征非常多，而样例特别少，这样用回归去直接拟合非常困难，容易过度拟合
- 在信号传输过程中，由于信道不是理想的，信道另一端收到的信号会有噪音扰动，那么怎么滤去这些噪音呢？



维度引发的问题

- 大多数机器学习和数据挖掘算法的效率和复杂度和维度相关
- 维度信息的冗余，即某些维度之间有明显的关联

Example

拿到一个汽车的样本，里面既有以“千米/每小时”度量的最大速度特征，也有“英里/小时”的最大速度特征

- 一个样本，特征非常多，而样例特别少，这样用回归去直接拟合非常困难，容易过度拟合
- 在信号传输过程中，由于信道不是理想的，信道另一端收到的信号会有噪音扰动，那么怎么滤去这些噪音呢？



维度引发的问题

- 大多数机器学习和数据挖掘算法的效率和复杂度和维度相关
- 维度信息的冗余，即某些维度之间有明显的关联

Example

拿到一个汽车的样本，里面既有以“千米/每小时”度量的最大速度特征，也有“英里/小时”的最大速度特征

- 一个样本，特征非常多，而样例特别少，这样用回归去直接拟合非常困难，容易过度拟合
- 在信号传输过程中，由于信道不是理想的，信道另一端收到的信号会有噪音扰动，那么怎么滤去这些噪音呢？



维度引发的问题

- 大多数机器学习和数据挖掘算法的效率和复杂度和维度相关
- 维度信息的冗余，即某些维度之间有明显的关联

Example

拿到一个汽车的样本，里面既有以“千米/每小时”度量的最大速度特征，也有“英里/小时”的最大速度特征

- 一个样本，特征非常多，而样例特别少，这样用回归去直接拟合非常困难，容易过度拟合
- 在信号传输过程中，由于信道不是理想的，信道另一端收到的信号会有噪音扰动，那么怎么滤去这些噪音呢？



PCA (Principal Component Analysis):

通过线性变换将原始数据变换为一组各维度**线性无关**的表示，可用于提取数据的**主要特征分量**，常用于高维数据的**降维**



1 降维与 pca

2 PCA 的过程与原理

- 一起来研究 PCA
- 重新梳理

3 PCA 的应用与实现

- princomp 介绍
- 在图形学的其它一些应用

4 PCA 的进一步讨论



要搞清楚的问题

- 1 什么是主成分？
- 2 如何分析？
- 3 怎么降维？



1 降维与 pca

2 PCA 的过程与原理

- 一起来研究 PCA
- 重新梳理

3 PCA 的应用与实现

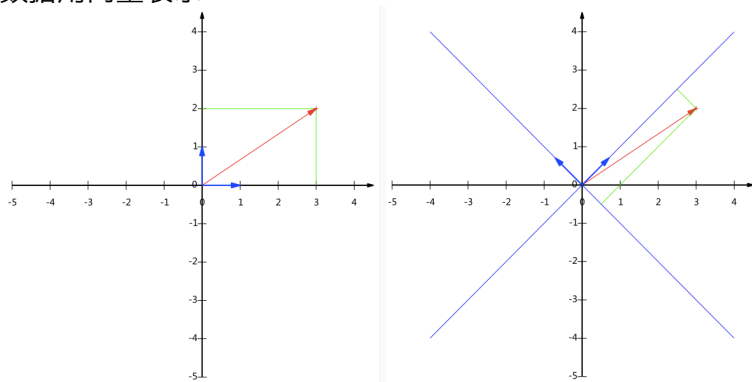
- princomp 介绍
- 在图形学的其它一些应用

4 PCA 的进一步讨论



维度与基变化

维度 = 基的数目
数据用向量表示



$$\begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 3 \\ 2 \end{pmatrix} = \begin{pmatrix} 5/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix}$$



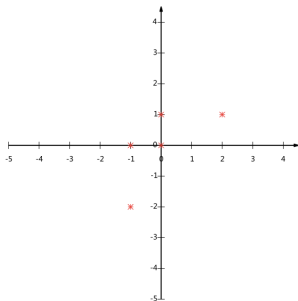
如果我们有 M 个 N 维向量, 将其变换到 R 个 N 维向量表示的新空间中
 p, a 都是 $n \times 1$

$$\begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_R \end{pmatrix} (a_1 \quad a_2 \quad \cdots \quad a_M) = \begin{pmatrix} p_1 a_1 & p_1 a_2 & \cdots & p_1 a_M \\ p_2 a_1 & p_2 a_2 & \cdots & p_2 a_M \\ \vdots & \vdots & \ddots & \vdots \\ p_R a_1 & p_R a_2 & \cdots & p_R a_M \end{pmatrix}$$



先假定数据只有二维，即只有两个变量，它们由横坐标和纵坐标所代表；因此每个观测值（一个观测值就是图中的一个点）都有相应于这两个坐标轴的两个坐标值，我们数据有 5 条记录

$$\begin{pmatrix} -1 & -1 & 0 & 2 & 0 \\ -2 & 0 & 0 & 1 & 1 \end{pmatrix}$$



现在问题来了：如果我们必须使用一维来表示这些数据，又希望尽量保留原始的信息，你要如何选择？
一种直观的看法是：希望投影后的投影值尽可能分散。



而这种分散程度，可以用数学上的方差来表述

$$\text{Var}(\mathbf{a}) = \frac{1}{m} \sum_{i=1}^m (\mathbf{a}_i - \mu)^2$$

在实际应用时，我们可以先将每个字段均值化为 0（将每个数据预先减去均值）

$$\text{Var}(\mathbf{a}) = \frac{1}{m} \sum_{i=1}^m \mathbf{a}_i^2$$



对于上面二维降成一维的问题来说，找到那个使得方差最大的方向就可以了。

不过对于更高维，如何选择多个方向？

还是单纯只选择方差最大的方向？

tip: 让两个字段尽可能表示更多的原始信息，我们是不希望它们之间存在（线性）相关性的，因为相关性意味存在重复表示的信息。



数学上可以用两个字段的协方差表示其相关性

$$\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

由于已经让每个字段均值为 0,

$$\text{Cov}(\mathbf{a}, \mathbf{b}) = \frac{1}{m} \sum_{i=1}^m a_i b_i$$

在字段均值为 0 的情况下，两个字段的协方差简洁的表示为其内积除以元素数 m 。

当协方差为 0 时，表示两个字段不相关



协方差矩阵

我们看到，最终要达到的目的与字段内方差及字段间协方差有密切关系。假设我们只有 **a** 和 **b** 两个字段，那么我们将它们按行组成矩阵 **X**

$$X = \begin{pmatrix} a_1 & a_2 & \cdots & a_m \\ b_1 & b_2 & \cdots & b_m \end{pmatrix}$$

然后我们用 **X** 乘以 **X** 的转置，并乘上系数 $1/m$

$$\frac{1}{m}XX^T = \begin{pmatrix} \frac{1}{m} \sum_{i=1}^m a_i^2 & \frac{1}{m} \sum_{i=1}^m a_i b_i \\ \frac{1}{m} \sum_{i=1}^m a_i b_i & \frac{1}{m} \sum_{i=1}^m b_i^2 \end{pmatrix}$$

这个矩阵对角线上的两个元素分别是两个字段的方差，而其它元素是 **a** 和 **b** 的协方差。两者被统一到了一个矩阵的。

这个结论很容易被推广到一般情况：设 $C = \frac{1}{m}XX^T$ ，则 **C** 是一个对称矩阵，其对角线分别个各个字段的方差，而第 *i* 行 *j* 列元素表示 *i* 和 *j* 两个字段的协方差



但我们希望的是，投影到该维度上方差最大，维度间无关，即协方差为 0，那么这样的矩阵就是对角矩阵即除对角线外的其它元素化为 0，并且在对角线上将元素按大小从上到下排列，这样我们就达到了优化目的



设原始数据矩阵 X 对应的协方差矩阵为 C

而 P 是一组基按行组成的矩阵

设 $Y=PX$ ，则 Y 为 X 对 P 做基变换后的数据。 Y 的协方差矩阵为 D 。

我们推导一下 D 与 C 的关系

$$\begin{aligned} D &= \frac{1}{m} YY^T \\ &= \frac{1}{m} (PX)(PX)^T \\ &= \frac{1}{m} PXX^T P^T \\ &= P\left(\frac{1}{m} XX^T\right)P^T \\ &= PCP^T \end{aligned}$$



协方差矩阵 C 是一个是对称矩阵，在线性代数上，实对称矩阵有一系列非常好的性质：

- 1 实对称矩阵不同特征值对应的特征向量必然正交。
- 2 设特征向量 λ 重数为 r ，则必然存在 r 个线性无关的特征向量对应于 λ ，因此可以将这 r 个特征向量单位正交化。



由上面两条可知，一个 n 行 n 列的实对称矩阵一定可以找到 n 个单位正交特征向量，设这 n 个特征向量为 $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$ ，我们将其按列组成矩阵：

$$E = (\mathbf{e}_1 \quad \mathbf{e}_2 \quad \cdots \quad \mathbf{e}_n)$$

则对协方差矩阵 C 有如下结论：

$$E^T C E = \Lambda = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}$$

其中 Λ 为对角矩阵，其对角元素为各特征向量对应的特征值（可能有重复）



到这里，我们发现我们已经找到了需要的矩阵 P

$$P = E^T$$

P 是协方差矩阵的特征向量单位化后按行排列出的矩阵，其中每一行都是 C 的一个特征向量。

如果 P 按照 Λ 中特征值的从大到小，将特征向量从上到下排列，则用 P 的前 K 行组成的矩阵乘以原始数据矩阵 X ，就得到了我们需要的降维后的数据矩阵 Y



总结一下 PCA 的算法步骤：设有 m 条 n 维数据。

步骤1

将原始数据按列组成 n 行 m 列矩阵 X

Example

这里以上文提到的数据

$$\begin{pmatrix} -1 & -1 & 0 & 2 & 0 \\ -2 & 0 & 0 & 1 & 1 \end{pmatrix}$$



步骤2

将 X 的每一行（代表一个属性字段）进行零均值化，即减去这一行的均值

Example

因为这个矩阵的每行已经是零均值，这里我们直接求协方差矩阵：



步骤3

求出协方差矩阵 $C = \frac{1}{m}XX^T$

Example

因为这个矩阵的每行已经是零均值，
这里我们直接求协方差矩阵：

$$C = \frac{1}{5} \begin{pmatrix} -1 & -1 & 0 & 2 & 0 \\ -2 & 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} -1 & -2 \\ -1 & 0 \\ 0 & 0 \\ 2 & 1 \\ 0 & 1 \end{pmatrix}$$
$$= \begin{pmatrix} \frac{6}{5} & \frac{4}{5} \\ \frac{4}{5} & \frac{6}{5} \end{pmatrix}$$



步骤4

求出协方差矩阵的特征值及对应的特征向量

Example

求解后特征值为：

$$\lambda_1 = 2, \lambda_2 = 2/5$$

其对应的特征向量分别是：

$$\mathbf{c}_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \mathbf{c}_2 \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$



步骤5

将特征向量按对应特征值大小从上到下按行排列成矩阵，取前 k 行组成矩阵 P

Example

降维，取一行

$$P = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$$



步骤6

$Y=PX$ 即为
降维到 k 维
后的数据

Example

$$Y = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} -1 & -1 & 0 & 2 \\ -2 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} =$$
$$\begin{pmatrix} -3/\sqrt{2} & -1/\sqrt{2} & 0 & 3/\sqrt{2} & -1/\sqrt{2} \end{pmatrix}$$



1 降维与 pca

2 PCA 的过程与原理

- 一起来研究 PCA
- 重新梳理

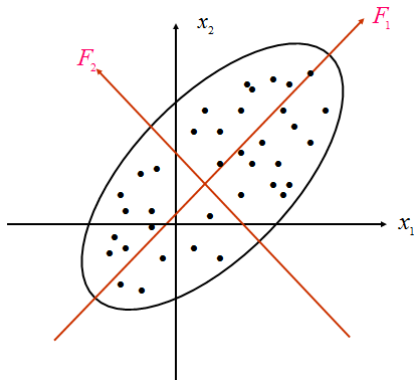
3 PCA 的应用与实现

- princomp 介绍
- 在图形学的其它一些应用

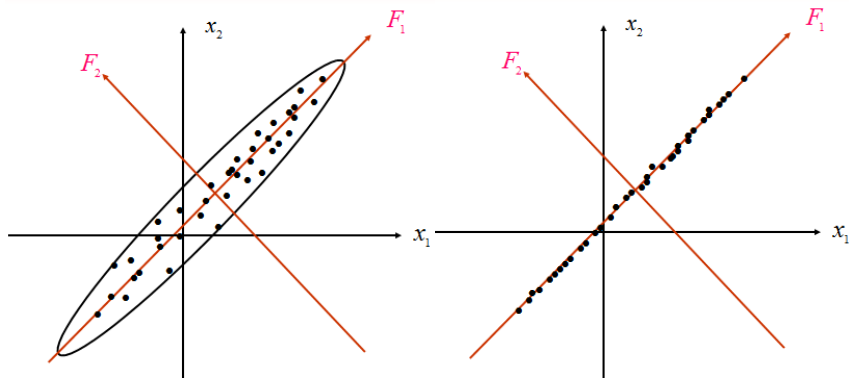
4 PCA 的进一步讨论



以二维为例：数据形成一个椭圆形状的点阵



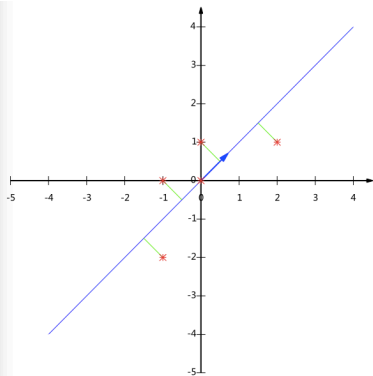
椭圆有一个长轴和一个短轴。在短轴方向上，数据变化很少；
在极端的情况，短轴如果退化成一点，那只有在长轴的方向才能够解释
这些点的变化了；
这样，由二维到一维的降维就自然完成了。



首先把高维椭圆的主轴找出来，再用代表大多数数据信息的最长的几个轴作为新变量；

和二维情况类似，高维椭圆的主轴也是互相垂直的。这些互相正交的新变量是原先变量的线性组合，叫做主成分

数据的主要成分 (即特征向量) 与它们的权值 (即特征值)



1 降维与 pca

2 PCA 的过程与原理

- 一起来研究 PCA
- 重新梳理

3 PCA 的应用与实现

- princomp 介绍
- 在图形学的其它一些应用

4 PCA 的进一步讨论



人脸数据集: [mit CBCL 人脸数据库](#)
人脸保存为灰度图像, 为 pgm 格式, 19*19
代码为 matlab



```
numOfPic = 10;  
trainSet = zeros( numOfPic, 19*19 );  
for i = 1 : numOfPic  
    str=sprintf( 'face/face%05d.pgm', i );  
    img = imread( str );  
    img = double(img);  
    trainSet(i,:) = img(:);  
end
```

图像分辨率: 19*19

文件名: faceXXXXX.pgm trainSet 维数: 10x361



```
meanValue = mean( trainSet );

trainSet_norm = bsxfun(@minus,trainSet , meanValue );
sigma = std(trainSet_norm); %每一列的标准差
trainSet_norm = bsxfun(@rdivide ,
                        trainSet_norm , sigma );
imwrite( uint8(
                        reshape(meanValue,19,19)), 'mean.bmp' );
```

数据标准化



标准化

对原始数据进行缩放处理，限制在一定的范围内。一般指正态化，即均值为 0，方差为 1。即使数据不符合正态分布，也可以采用这种方式方法，标准化后的数据有正有负

目的是为了进行数据无量纲化，解决数据的可比性
一般灰度图像无需此处理



协方差矩阵与特征值

```
X = trainSet_norm;  
[m, n] = size(X);  
Cov = 1/m*X*X';  
[U, S, V] =svd(Cov);
```

奇异值与特征值

对称矩阵特征值与奇异值差正负号，特征向量和奇异向量张成空间一样。从谱分解来看，对称矩阵可以酉对角化， $A = Q * B * Q^t$ ，B 对角线特征值，只要将负号加到 Q 的列上，使 B 对角线大于 0，得到的就是奇异值分解

svd 得到的奇异值是按降序排列的




```
a =
```

```
1    1    1
1    2    1
1    1    3
```

```
>> [ u, s, v] = svd(a)
```

```
u =
```

```
-0.3971    0.2332    0.8877
-0.5207    0.7392   -0.4271
-0.7558   -0.6318   -0.1721
```

```
s =
```

```
4.2143         0         0
         0    1.4608         0
         0         0    0.3249
```

```
>> [b , c ] =eig(a)
```

```
b =
```

```
0.8877    0.2332    0.3971
-0.4271    0.7392    0.5207
-0.1721   -0.6318    0.7558
```

```
c =
```

```
0.3249         0         0
         0    1.4608         0
         0         0    4.2143
```



降到多少维

```
eigValue = diag(S);  
AllSum = sum(eigValue);  
cur_Sum = 0;  
p = 0;  
while( cur_Sum/ AllSum < 0.9)  
    p = p + 1;  
    cur_Sum = sum(eigValue(1:p));  
end;  
P = U(:, 1:p);
```

选择主轴成分比重 0.9 以上，比重是可以自己定的，一般大于 0.85，丢失的信息才不多



- 1 降维与 pca
- 2 PCA 的过程与原理
 - 一起来研究 PCA
 - 重新梳理
- 3 PCA 的应用与实现
 - `princomp` 介绍
 - 在图形学的其它一些应用
- 4 PCA 的进一步讨论



`[COEFF,SCORE,latent,tsquare] = princomp(X)`

1. 输入参数 X 是一个 n 行 p 列的矩阵。每行代表一个样本观察数据，每列则代表一个属性，或特征

2. $COEFF$ 就是所需要的特征向量组成的矩阵，是一个 p 行 p 列的矩阵，每行经常也称为（协方差矩阵的）特征向量。并且是按照对应特征值降序排列的。所以，如果只需要前 k 个主成分向量，可通过：`COEFF(:, 1:k)` 来获得。

3. $SCORE$ 表示原数据在各主成分向量上的投影。但注意：是原数据经过中心化后在主成分向量上的投影 $SCORE = x_0 * COEFF$ 求得。其中 x_0 是中心平移（减去均值）后的 X

4. $latent$ 是一个列向量，表示特征值，并且按降序排列。



```
>> load hald  
>> ingredients
```

```
ingredients =
```

7	26	6	60
1	29	15	52
11	56	8	20
11	31	8	47
7	52	6	33
11	55	9	22
3	71	17	6
1	31	22	44
2	54	18	22
21	47	4	26
1	40	23	34
11	66	9	12
10	68	8	12



```

latent =

517.7969
67.4964
12.4054
0.2372

>> [pc, score, latent, tsquare] = princomp(ingredients)

pc =

-0.0678    -0.6460     0.5673     0.5062
-0.6785    -0.0200    -0.5440     0.4933
 0.0290     0.7553     0.4036     0.5156
 0.7309    -0.1085    -0.4684     0.4844

```



```
>> cov_in=cov(ingredients)
```

```
cov_in =
```

```
34.6026    20.9231   -31.0513   -24.1667
20.9231   242.1410   -13.8782  -253.4167
-31.0513  -13.8782   41.0256    3.1667
-24.1667 -253.4167    3.1667   280.1667
```

```
>> [V, D] =eig(cov_in)
```

```
V =
```

```
0.5062    0.5673    0.6460   -0.0678
0.4933   -0.5440    0.0200   -0.6785
0.5156    0.4036   -0.7553    0.0290
0.4844   -0.4684    0.1085    0.7309
```

```
D =
```

```
0.2372         0         0         0
         0  12.4054         0         0
         0         0  67.4964         0
         0         0         0  517.7969
```

```
>> [U, S, V] = svd(cov_in)
```



meanValue = mean(ingredients)

meanValue =

7.4615	48.1538	11.7692	30.0000
--------	---------	---------	---------

a = bsxfun(minus,ingredients,meanValue);

1/12*a'*a

ans =

34.6026	20.9231	-31.0513	-24.1667
20.9231	242.1410	-13.8782	-253.4167
-31.0513	-13.8782	41.0256	3.1667
-24.1667	-253.4167	3.1667	280.1667



- 1 降维与 pca
- 2 PCA 的过程与原理
 - 一起来研究 PCA
 - 重新梳理
- 3 PCA 的应用与实现
 - princomp 介绍
 - 在图形学的其它一些应用
- 4 PCA 的进一步讨论



降维找数据的分布特征

分析网格数据（如对人物的网格进行分析从而发现其骨架拓扑）时，有时会得到一堆散乱分布在关节附近的顶点数据，利用主元分析对这些散乱的顶点进行降维，如从 3 维降到 2 维，则可以发现这些顶点数据的分布特征。



三角形网格的有向包围盒 (OBB)

OBB 是一种对模型进行视锥裁剪与碰撞检测的有效方式，然而通常网格模型表示在一个与世界坐标系平行的框架中，可以利用主元分析求得模型所在的各个自然轴，如一个斜放着的圆柱其第一个自然轴是有圆柱底面中心指向圆柱顶面中心的方向。求得各个自然轴后，通过各顶点坐标与单位自然轴的点积，即可获得各顶点在自然轴上的分布范围，进而得到 OBB。



例如数据捕捉采集到的人体步行动画通常由不同时刻各个关节（约 40 60 个）的角度来表示，当数据量很大时，对数据的存储与处理都是一件耗费资源的事情。而普通的步行动作中，当一个人左臂向前摆动时右臂总向后摆动，这就是说这些数据有耦合，从而可以利用主元分析进行特征提取与压缩。由此也可以想象得到主元分析另一个强大的地方：对于一些高维度的数据，很难用三维图形的方式去展现，因此很难观察出其分布特征，而主元分析计算则可以发现这些特征。



1 降维与 pca

2 PCA 的过程与原理

- 一起来研究 PCA
- 重新梳理

3 PCA 的应用与实现

- princomp 介绍
- 在图形学的其它一些应用

4 PCA 的进一步讨论



根据上面对 PCA 的数学原理解释，我们可以了解到一些 PCA 的能力和限制。PCA 本质上是将方差最大的方向作为主要特征，并且在各个正交方向上将数据“离相关”，也就是让它们在不同正交方向上没有相关性。

因此，PCA 也存在一些限制，

- 它可以很好的解除线性相关，但是对于高阶相关性就没有办法了

解决办法

对于存在高阶相关性的数据，可以考虑 Kernel PCA，通过 Kernel 函数将非线性相关转为线性相关，关于这点就不展开讨论了

- PCA 假设数据各主特征是分布在正交方向上，如果在非正交方向上存在几个方差较大的方向，PCA 的效果就大打折扣了



根据上面对 PCA 的数学原理解释，我们可以了解到一些 PCA 的能力和限制。PCA 本质上是将方差最大的方向作为主要特征，并且在各个正交方向上将数据“离相关”，也就是让它们在不同正交方向上没有相关性。

因此，PCA 也存在一些限制，

- 它可以很好的解除线性相关，但是对于高阶相关性就没有办法了

解决办法

对于存在高阶相关性的数据，可以考虑 Kernel PCA，通过 Kernel 函数将非线性相关转为线性相关，关于这点就不展开讨论了

- PCA 假设数据各主特征是分布在正交方向上，如果在非正交方向上存在几个方差较大的方向，PCA 的效果就大打折扣了



本文档使用或参考了以下来源内容:

[melory 博客主成份分析](#)

[pca-tutorial](#)

[线代启示录 -主成分分析 \(网站需要翻墙\)](#)

