

New begin, new life

I don't know if i choose the right way, but i will go on. Hello Oracle.

[目录视图](#)[摘要视图](#)[RSS](#) [订阅](#)

个人资料



watkins

访问: 740903次

积分: 10948

等级: [BLOG > 7](#)

排名: 第683名

原创: 244篇 转载: 66篇
译文: 4篇 评论: 546条

文章搜索

文章分类

[Algorithms](#) (47)
[Design Pattern](#) (29)
[c++](#) (56)
[Linux](#) (45)
[Python](#) (8)
[opencv](#) (18)
[Android](#) (10)
[NLP](#) (15)
[ML](#) (7)
[Nodejs](#) (2)

文章存档

[2015年12月](#) (1)
[2015年11月](#) (2)
[2015年09月](#) (1)
[2015年08月](#) (1)
[2015年07月](#) (8)

[展开](#)

阅读排行

[PCA降维算法总结以及m](#)
(49076)
[AdaBoost中利用Haar特](#)
(27147)
[AdaBoost中利用Haar特](#)
(20188)

学院APP首次下载, 可得50C币! 有奖试读—增长黑客, 创业公司必知的“黑科技” 当讲师? 爱学习? 投票攒课吧 CSDN
2015博客之星评选 感恩分享

PCA降维算法总结以及matlab实现PCA(个人的一点理解)

标签: [PCA](#) [降维](#) [样本](#) [主成分分析](#)

2012-11-28 16:55

49077人阅读

[评论\(64\)](#)

[收藏](#)

[举报](#)



分类:

[Algorithms \(46\)](#)

版权声明: 本文为博主原创文章, 未经博主允许不得转载。

[目录\(?\)](#)

[\[+\]](#)

转载请声明出处。by watkins song

鉴于本文比较混乱, 所以写了一个新的PCA的详细介绍, 请参见: <http://blog.csdn.net/watkinsong/article/details/38536463>
两篇文章各有侧重, 对照看效果更加 o(∩∩)o..

PCA的一些基本资料

最近因为最人脸表情识别, 提取的gabor特征太多了, 所以需要用PCA进行对提取的特征进行降维。

本来最早的时候我没有打算对提取的gabor特征进行降维, 但是如果一个图像时64*64,那么使用五个尺度八个方向的gabor滤波器进行滤波, 这样提取的特征足足有64*64*5*8这么多, 如果图像稍微大一点, 比如128*128的图像, 那么直接提取的特征就会几十万, 所以不降维的话直接用SVM训练分类器是非常困难的。

所以在这段时间我就学习了一下PCA降维的基本原理和使用方法, 网上给出的资料都比较乱, 而且很不清楚, 经过这几天的学习和测试, 终于把调理弄清楚了, 给大家分享一下, 下面只是我对于PCA的个人理解, 肯定有不对的地方, 还请各位大牛多多指教。

下面先给出一下PCA的资料地址, 都是我收集的:

<http://hi.baidu.com/yicomrdztxbeiwid/item/913f28c05cf7ebc4994aa06f>

<http://blog.sciencenet.cn/blog-265205-544681.html>

<http://blog.csdn.net/mpbchina/article/details/7384425>

http://blog.sina.com.cn/s/blog_6833a4df0100pvk7.html

<http://stackoverflow.com/questions/4991343/matlab-principal-component-analysis-eigenvalues-order>

<http://stackoverflow.com/questions/10400230/what-is-score-in-princomp>

http://www.mathworks.com/matlabcentral/newsreader/view_thread/152608

<http://stats.stackexchange.com/questions/27572/matlab-princomp-latent>

<http://www.nlpc.org/pca-principal-component-analysis-matlab.html>

<http://www.matlabsky.com/thread-11751-1-1.html>

<http://stackoverflow.com/questions/10818718/principal-component-analysis>

<http://www.mathworks.cn/cn/help/stats/princomp.html>

<http://www.mathworks.cn/cn/help/stats/pca.html#bti6n7k-2>

<http://lovelittlebean.blog.163.com/blog/static/116582186201181213911729/>

在MFC中通过opencv显示	(19486)
rvm的安装, 使用rvm, 安装	(13968)
Android(安卓)开发通过N	(9585)
gabor滤波器的几种实现:	(9585)
在Android中使用JNI调用	(8340)
在MFC控件中显示openc	(8232)
C#设计模式之1——简单	(8078)

评论排行	
在MFC中通过opencv显示	(84)
Android(安卓)开发通过N	(67)
PCA降维算法总结以及m	(64)
AdaBoost中利用Haar特	(62)
AdaBoost中利用Haar特	(45)
在Android中使用JNI调用	(28)
ASM(Active Shape Mod	(17)
C# Socket编程基础入门	(11)
opencv 目录文件遍历	(10)
应用程序部署总结	(10)

最新评论	
AdaBoost中利用Haar特征进行人	gg-123: 你这个基本上也是看别人的抄的吧? 我没有发现你自己的理解在里边
PCA 降维算法详解 以及代码示例	丑小鸭的路途: 写的很好, 谢谢博主分享, 正在学习PCA, 很受用
ASM(Active Shape Model) 主动	getbetter飞飞: 楼主, 很佩服您在这方面的学习, 想像您请教一下人脸对齐中的公式(5)是怎么来的? 期待您的回复~
Android(安卓)开发通过NDK调用	a999888huang: 博主, 请问你是怎么解决怎么问题的? 而且当手动把文件添加到android项目的libs目录后NDK在 ...
matlab实现gabor filter (4)	zbxzc: 你找的这个伊朗人写的代码有问题, lamda是波长, U、V分别是x、y方向的中心频率, 要多尺度检测的话...
gabor滤波器的几种实现方式	mxllv: 你好 我想问一下 我想用这个Gabor小波做视网膜的血管提取 能不能给出一些资料或思路 本...
C# Socket编程基础入门	songhuangong123: 学习了~~
C#设计模式之21——状态模式	qsh_zh: if-else或者switch过多就是坏味道, 状态模式很好地解决了这个问题。
Android(安卓)开发通过NDK调用	af913337456: @lixue20141529: 我放弃不搞了
PCA降维算法总结以及matlab实	LeijianHP: 我有X: 16000*3000 (3000为样本), 使用 = princomp(X); X=score...

http://www.ilovematlab.cn/thread-54493-1-1.html
http://www.ilovematlab.cn/forum.php?mod=viewthread&tid=146626
http://www.ilovematlab.cn/forum.php?mod=viewthread&tid=204069
http://www.ilovematlab.cn/forum.php?mod=viewthread&tid=54600
http://search.discuz.qq.com/s/aa8585553/princomp+%E9%99%8D%E7%BB%B4.html
http://www.ilovematlab.cn/thread-68796-1-1.html
http://www.ilovematlab.cn/thread-209229-1-1.html
http://www.ilovematlab.cn/thread-209229-1-1.html
http://blog.sina.com.cn/s/blog_61c0518f0100f4mi.html
http://blog.csdn.net/haitao111313/article/details/7875392
http://media.cs.tsinghua.edu.cn/~ahz/digitalimageprocess/chapter11/chapt11_ahz.htm
http://hi.baidu.com/845777018/item/7438e555df1138404fff2011
http://en.wikipedia.org/wiki/Principal_component_analysis
http://baike.baidu.com/view/852194.htm
http://wenku.baidu.com/view/bd9284fcfab069dc51220107.html
http://wenku.baidu.com/view/c0bde56da98271fe910ef9b8.html
http://wenku.baidu.com/view/9f69930790c69ec3d5bb75d3.html
http://www.ilovematlab.cn/thread-54600-1-1.html
http://www.cnblogs.com/sunwufan/archive/2011/08/31/2159952.html
http://zhidao.baidu.com/question/416895922.html
上面的网址都是一些pca原理啊, 实现什么的介绍。

具体的PCA的算法的理论基础呢, 我这里就不详细说了, 因为我也没有看具体详细, 所以如果想要彻底的弄明白PCA的工作原来, 还是请到wiki上看吧, 写的非常清晰, 我因为临时用一下, 就写个大概的原理就可以了。

P C A 原理:

PCA的原理就是将原来的样本数据投影到一个新的空间中, 相当于我们在矩阵分析里面学习的将一组矩阵映射到另外的坐标系下。通过一个转换坐标, 也可以理解成把一组坐标转换到另外一组坐标系下, 但是新的坐标系下, 表示原来的原本不需要那么多的变量, 只需要原来样本的最大的一个线性无关组的特征值对应的空间的坐标即可。

比如, 原来的样本是30*1000000的维数, 就是说我们有30个样本, 每个样本有1000000个特征点, 这个特征点太多了, 我们需要对这些样本的特征点进行降维。那么在降维的时候会计算一个原来样本矩阵的协方差矩阵, 这里就是1000000*1000000, 当然, 这个矩阵太大了, 计算的时候有其他的方式进行处理, 这里只是讲解基本的原理, 然后通过这个1000000*1000000的协方差矩阵计算它的特征值和特征向量, 最后获得具有最大特征值的特征向量构成转换矩阵。比如我们的前29个特征值已经能够占到所有特征值的99%以上, 那么我们只需要提取前29个特征值对应的特征向量即可。这样就构成了一个1000000*29的转换矩阵, 然后用原来的样本乘以这个转换矩阵, 就可以得到原来的样本数据在新的特征空间的对应的坐标。30*1000000 * 1000000*29 = 30 *29, 这样原来的训练样本每个样本的特征值的个数就降到了29个。

一般来说, PCA降维后的每个样本的特征的维数, 不会超过训练样本的个数, 因为超出的特征是没有意义的。

下面是百度百科中对pca降维的一段解释, 还是挺清晰的:

“对于一个训练集, 100个对象模板, 特征是10维, 那么它可以建立一个100*10的矩阵, 作为样本。求这个样本的协方差矩阵, 得到一个10*10的协方差矩阵, 然后求出这个协方差矩阵的特征值和特征向量, 应该有10个特征值和特征向量, 我们根据特征值的大小, 取前四个特征值所对应的特征向量, 构成一个10*4的矩阵, 这个矩阵就是我们要求的特征矩阵, 100*10的样本矩阵乘以这个10*4的特征矩阵, 就得到了一个100*4的新的降维之后的样本矩阵, 每个特征的维数下降了。

当给定一个测试的特征集之后, 比如1*10维的特征, 乘以上面得到的10*4的特征矩阵, 便可以得到一个1*4的特征, 用这个特征去分类。”

我对 P C A 的一些了解
我的pca迷惑
迷惑一

刚开始接触PCA的时候，咨询了一个浙大的博士朋友，这朋友告诉我，如果对训练样本进行降维，那么样本的数量必须大于特征的维数，然后我当时就迷惑了，那我怎么办啊，我的人脸表情图像顶多有几百张就算多了，但是每个图像提取的特征的维数将近有几十万，我不可能找那么多样本去啊。当时有这个迷惑也是因为matlab给出的一个实现在pca降维的函数的说明，就是princomp，这个函数的说明也是用的样本的个数多余特征的维数。后来经过试验是证实，证实了那个浙大的博士的认识是错误的，pca降维肯定不需要样本的个数大于特征的维数，要不然还降维个什么意思。比如我有30*1000000的特征矩阵，那么降维后肯定是每个样本在新的空间中的表示的特征维数不超过30。

迷惑二

另外一个迷惑，在最初刚开始做的时候，就是为什么这么大的数据，比如30*1000000直接就降到了30*29，这不是减少的数据有点太多了么，会不会对性能造成影响。之所以有这个迷惑，是因为最初并不了解pca的工作方式。pca并不是直接对原来的数据进行删减，而是把原来的数据映射到新的一个特征空间中继续表示，所有新的特征空间如果有29维，那么这29维足以能够表示非常非常多的数据，并没有对原来的数据进行删减，只是把原来的数据映射到新的空间中进行表示，所以你的测试样本也要同样的映射到这个空间中进行表示，这样就要求你保存住这个空间坐标转换矩阵，把测试样本同样的转换到相同的坐标空间中。

有些同学在网上发帖子问对训练样本降维以后，怎么对测试样本降维，是不是还是使用princomp这个函数进行降维，这个是错误的。如果你要保证程序运行正常，就要保证训练样本和测试样本被映射到同一个特征空间，这样才能保证数据的一致性。

迷惑三

网上有不同的pca降维的代码，每个代码也实现的不一样，那么对于同一个数据是否是pca降维以后都是获得相同的数据呢，也就是说不管你用哪种方式进行pca降维，不管你是从哪里下载到的或者自己根据算法实现的pca降维，同样的矩阵降维以后的数据是否一致？这个我个人认为，不同的算法最后导致的pca降维的数据肯定不一致。因为pca降维以后，只是把原来的数据映射到新的特征空间，所以如果你的算法不同，那么选择的协方差矩阵肯定就不同，最后获得的转换矩阵肯定也不一样。那么训练样本和测试样本和不同的转换矩阵相乘以后最终肯定会获得不同的降维坐标。所以使用不同的算法应该最后不会有相同的坐标结果，这个也是我一直实验的结果，我也使用了matlab自带的princomp降维，并且使用相同的数据使用网上下载的一些降维方法进行降维，得到的数据都不一致。

比如说princomp这个matlab自带的函数，在降维之前就将每一个样本减去了一个所有样本的平均值，也可能有很多样本没有减去平均值。princomp这里使用一行表示一个样本，每行包括这个样本的所有的特征值。而网上大部分都是每一列表示一个样本，这样这一列的所有行都表示这个样本的特征值。网上的程序使用列表示样本是有一定好处的，比如我的样本是1000000*30，总共有30个训练样本，每个样本的特征值个数是1000000，那么这个矩阵获得的协方差矩阵是30*30，计算起来非常的方便，不想30*1000000这样的矩阵获得到的协方差矩阵式1000000*1000000，直接就内存溢出了，不过matlab有自己的实现方式，巧妙的解决了这个问题。

pca的实现（matlab）

我在网上看了很多pca降维的例子，都大同小异，原理差不多，都是活的原来矩阵的协方差矩阵，然后计算协方差矩阵的特征值和特征向量，最后通过特征向量的根据特征值由大到小的排序进行KL变换神马的获得一个转换矩阵。

1. matlab自带的实现方式

PCA在matlab中的实现举例

以下资料来自matlab的help，翻译和注解部分由笔者添加：(重点部分添加了翻译！)

princomp-----函数名称

Principal component analysis (PCA) on data

Syntax-----函数调用语法

[COEFF,SCORE] = princomp(X)

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

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

[...] = princomp(X,'econ')

Description -----函数描述

COEFF = princomp(X) performs principal components analysis (PCA) on the n-by-p data matrix X, and returns the principal component coefficients, also known as loadings. Rows of X correspond to observations, columns to variables. COEFF is a p-by-p matrix, each column containing coefficients for one principal component. The columns are in order of decreasing component variance.

在n行p列的数据集X上做主成分分析。返回主成分系数。X的每行表示一个样本的观测值，每一列表示特征变量。COEFF是一个p行p列的矩阵，每一列包含一个主成分的系数，列是按主成分变量递减顺序排列。(按照这个翻译很难理解，其实COEFF是X矩阵所对应的协方差阵V的所有特征向量组成的矩阵，即变换矩阵或称投影矩阵，COEFF每列对应一个特征值的特征向量，列的排列顺序是按特征值的大小递减排序，后面有具体例子解释，见说明1)

princomp centers X by subtracting off column means, but does not rescale the columns of X. To perform principal components analysis with standardized variables, that is, based on correlations, use princomp(zscore(X)). To perform principal components analysis directly on a covariance or correlation matrix, use pcacov.

计算PCA的时候，MATLAB自动对列进行了去均值的操作，但是并不对数据进行规格化，如果要规格化的话，用princomp(zscore(X))。另外，如果直接有现成的协方差阵，用函数pcacov来计算。

[COEFF,SCORE] = princomp(X) returns SCORE, the principal component scores; that is, the representation of X in the principal component space. Rows of SCORE correspond to observations, columns to components.

返回的SCORE是对主分的打分，也就是说原X矩阵在主成分空间的表示。SCORE每行对应样本观测值，每列对应一个主成份(变量)，它的行和列的数目和X的行列数目相同。

[COEFF,SCORE,latent] = princomp(X) returns latent, a vector containing the eigenvalues of the covariance matrix of X.

返回的latent是一个向量，它是X所对应的协方差矩阵的特征值向量。

[COEFF,SCORE,latent,tsquare] = princomp(X) returns tsquare, which contains Hotelling's T2 statistic for each data point.

返回的tsquare，是表示对每个样本点Hotelling的T方统计量(我也不很清楚是什么东东)。

The scores are the data formed by transforming the original data into the space of the principal components. The values of the vector latent are the variance of the columns of SCORE. Hotelling's T2 is a measure of the multivariate distance of each observation from the center of the data set.

所得的分(scores)表示由原数据X转变到主成分空间所得到的数据。latent向量的值表示SCORE矩阵每列的方差(见说明2)。Hotelling的T方是用来衡量多变量间的距离，这个距离是指样本观测值到数据集中心的距离。

When $n \leq p$, SCORE(:,n:p) and latent(n:p) are necessarily zero, and the columns of COEFF(:,n:p) define directions that are orthogonal to X.

[...] = princomp(X,'econ') returns only the elements of latent that are not necessarily zero, and the corresponding columns of COEFF and SCORE, that is, when $n \leq p$, only the first n-1. This can be significantly faster when p is much larger than n.

当维数p超过样本个数n的时候，用[...] = princomp(X,'econ')来计算，这样会显著提高计算速度

Examples--举例

(上面说了那么多废话，看了还不一定懂，还不如举例容易理解，下面样本数据集为ingredients，这个数据集是matlab自带的)

Compute principal components for the ingredients data in the Hald data set, and the variance accounted for by each component.

```
load hald; %载入matlab内部数据

[pc,score,latent,tsquare] = princomp(ingredients); %调用pca分析函数

ingredients,score,pc,latent,tsquare %显示得到的结果

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

score =

36.8218 -6.8709 -4.5909 0.3967
29.6073 4.6109 -2.2476 -0.3958
-12.9818 -4.2049 0.9022 -1.1261
23.7147 -6.6341 1.8547 -0.3786
-0.5532 -4.4617 -6.0874 0.1424
-10.8125 -3.6466 0.9130 -0.1350
-32.5882 8.9798 -1.6063 0.0818
22.6064 10.7259 3.2365 0.3243
-9.2626 8.9854 -0.0169 -0.5437
-3.2840 -14.1573 7.0465 0.3405
9.2200 12.3861 3.4283 0.4352
-25.5849 -2.7817      8
-26.9032 -2.9310      6

pc =

-0.0678 -0.6460 0
-0.6785 -0.0200 -1
0.0290 0.7553 0.4036 0.5156
0.7309 -0.1085 -0.4684 0.4844
```

latent =

517.7969

关闭



2.6198

3.3681

0.5668

3.4818

3.9794

2.6086

7.4818

4.1830

2.2327

2.7216

%下面我们来做一个验证

%下面为计算ingredients协方差矩阵:

cov_ingredients=cov(ingredients)

cov_ingredients =

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

%下面为计算ingredients所对应的协方差矩阵(也就是cov_ingredients矩阵)的特征值和特征

%向量, 下面的矩阵V为特征向量, D为特征值(对比上面的latent)组成的对角线矩阵

[V,D] = eig(cov_ingredients)

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
```

%说明1: 对比一下矩阵V和矩阵pc, 现在很容易明白为什么COEFF是按列递减顺序排列的

%了! (V中第三列与pc中倒数第三列差个负号, 学过线性代数的人都知道这没问题)

%下面再验证一下说明2

```
diag(cov(score))
```

```
ans =
```

```
517.7969
```

```
67.4964
```

```
12.4054
```

```
0.2372
```

%说明2: 以上结果显示latent确实表示SCORE矩阵每列的方差, 517.7969表示第一列方差

下面做图表示结果:

上面说了半天还没有达到我们终极想要的, 其实我们要的是由函数[pc,score,latent,tsquare] = princomp(ingredients)它所产生的pc和latent。由latent可以算出降维后的空间所能表示原空间的程度, 只要这个累积的值大于95%就行了。

The following command and plot show that two components account for 98% of the variance:

```
cumsum(latent)./sum(latent)
```

```
ans =
```

```
0.86597
```

```
0.97886
```

```
0.9996
```

```
1
```

%由以上ans值可以看出前两个主成分就能表示原空间的97.886%,所以取pc中的前两列可

%做主成分变换矩阵tranMatrix = pc(:,1:2)。则从原来的4维空间降到2维空间。对任意一个

%原空间样本,例如a=(7,26,6,60)变到低维空间的表达式为a1 = a*tranMatrix。(当然你也可

%以取pc中的前三列, 由原来的4维空间变到3维空间)

```
biplot(pc(:,1:2),'Scores',score(:,1:2),'VarLabels',...
```

```
{'X1' 'X2' 'X3' 'X4'})
```

上面这个matlab函数的说明呢, 只是引用百度百科, 也可以看看matlab的函数说明, 但是多少还是有点难懂。我把我的理解简单的说说。

[COEFF, SCORE, LATENT, TSQUARED] = PRINCOMP(X)

上面这个函数, coeff矩阵是返回的转换矩阵, 也就是把样本转换到新的空间中的准换矩阵, 这个准换矩阵式比较大的, 比如你的降维矩阵式30*100000, 那么这个准换矩阵一般都是10000*29的维数。

score是原来的样本矩阵在新的坐标系中的表示, 也就是原来的样本乘上转换矩阵, 但是还不是直接乘, 要减去一个样本的均值。将原来的数据转换到新的样本空间中的算法是这样实现的:

```
x0 = bsxfun(@minus,x,mean(x,1));
```

```
score = x0 * coeff;
```

然后就会得到和[COEFF, SCORE, LATENT, TSQUARED] = PRINCOMP(X) 输出一样的score数据。同时这个也是原来的样本矩阵降维后的结果，如果使用降维后的数据就使用这个数据。一般情况下，如果你的每个样本的特征维数远远大于样本数，比如30*1000000的维数,princomp要加上'econ', 就是princomp(x,'econ')这样使用，可以很大程度的加快计算速度，而且不会内存溢出，否则会经常报内存溢出。

[...] = PRINCOMP(X,'econ') returns only the elements of LATENT that are not necessarily zero, i.e., when N <= P, only the first N-1, and the corresponding columns of COEFF and SCORE. This can be significantly faster when P >> N.

latent是返回的按降序排列的特征值，根据这个你可以手动的选择降维以后的数据要选择前多少列。

```
cumsum(latent)./sum(latent)
```

，通过这样计算特征值的累计贡献率，一般来说都选择前95%的特征值对应的特征向量，还是原来的矩阵30*1000000，如果你计算得到前25个特征值的累计贡献率已经超过99.9%，那么就完全可以只要降维后的数据的前25列。

tsquared是个什么东西我也不知道。。。不过貌似很少有人能用到，网络上也没有神马资料，各位如果需要用的再查阅吧，一般情况下也用不到。

如果你需要对测试样本降维，一般情况下，使用matlab自带的方式，肯定需要对测试样本减去一个训练样本均值，因为你在给训练样本降维的时候减去了均值，所以测试样本也要减去均值，然后乘以coeff这个矩阵，就获得了测试样本降维后的数据。比如说你的测试样本是1*1000000，那么乘上一个1000000*29的降维矩阵，就获得了1*29的降维后的测试样本的降维数据。

princomp(x)使用的行表示一个样本，每行的所有的列数据都是这个样本的特征值。降维以后比如是30*29，那么每一行就是降维以后的数据。每个样本有29个特征值。

2. 一个自实现的pca降维方式

下面是来自mpb同学的一个自实现的例子，很牛的一个人，我们本科同学。

原文地址：<http://blog.csdn.net/mpbchina/article/details/7384425>

下面引用原文内容：

```
01. %训练
02. %Lx=X'*X
03. clear;
04. clc;
05. train_path='..\Data\TrainingSet\';
06. phi=zeros(64*64,20);
07. for i=1:20
08. path=strcat(train_path,num2str(i),'.bmp');
09. Image=imread(path);
10. Image=imresize(Image,[64,64]);
11. phi(:,i)=double(reshape(Image,1,[]));
12. end;
13. %mean
14. mean_phi=mean(phi,2);
15. mean_face=reshape(mean_phi,64,64);
16. Image_mean=mat2gray(mean_face);
17. imwrite(Image_mean,'meanface.bmp','bmp');
18. %demean
19. for i=1:19
20. X(:,i)=phi(:,i)-mean_phi;
21. end
22. Lx=X'*X;
23. tic;
24. [eigenvector,eigenvalue]=eigs(Lx,19);
25. toc;
26. %normalization
```



```

27. for i=1:19
28.     %K-L变换
29.     UL(:,i)=X*eigenvector(:,i)/sqrt(eigenvalue(i,i));
30. end
31. %display Eigenface
32. for i=1:19
33.     Eigenface=reshape(UL(:,i),[64,64]);
34.     figure(i);
35.     imshow(mat2gray(Eigenface));
36. end

```

得到的均值图像mean_face:



前19个最大主元对应的“特征脸”:



测试:

测试用样本:



[plain]

```

01. %使用测试样本进行测试
02. clc;
03. test_path='../Data\TestingSet\';
04. error=zeros([1,4]);
05. for i=1:4
06.     path=strcat(test_path,num2str(i),'.bmp');
07.     Image=imread(path);
08.     Image=double(imresize(Image,[64,64]));
09.     phi_test=zeros(64*64,1);
10.     phi_test(:,1)=double(reshape(Image,1,[]));
11.     X_test=phi_test-mean_phi;
12.     Y_test=UL'*X_test;
13.     X_test_re=UL*Y_test;
14.     Face_re=X_test_re+mean_phi;
15.     calculate error rate
16.     e=Face_re-phi_test;
17.
18.
19. %%display figure
20.     Face_re_2=reshape(Face_re(:,1),[64,64]);
21.     figure(i);
22.
23.     imshow(mat2gray(Image));
24.     title('Original');
25.     figure(10+i);
26.     imshow(mat2gray(Face_re_2));
27.     title('Reconstruct');
28.     error(1,i)=norm(e);
29.

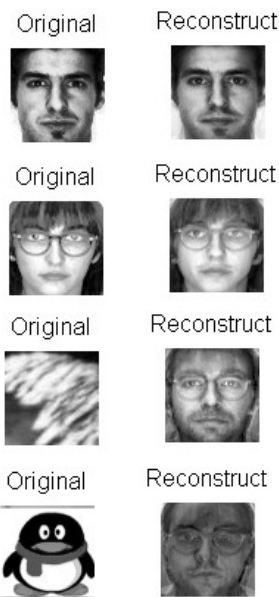
```

```

30. %dispaly error rate
31. error_rate=error(1,i);
32. display(error_rate);
33. end

```

重建出的测试样本与原样本的对比:



四副测试样本的重建误差分别为:

```

1.4195e+003
1.9564e+003
4.7337e+003
7.0103e+003

```

可见测试样本为人脸的样本的重建误差显然小于非人脸的重建误差。

上面的降维的例子中, 每一列表示一个样本, 这样就一共有 4096×20 的待降维矩阵, 然后对这个矩阵降维, 请注意, 如果采用列表示一个样本, 那么获得的降维矩阵, 是一个 4096×19 的矩阵, 然后用这个降维矩阵对测试样本和训练样本降维, 我们的测试样本是 4096×1 的矩阵, 降维的时候这样:

```
Y_test=UL'*X_test;
```

UL是计算获得降维矩阵, UL' (对UL进行转至) 获得的 19×4096 的矩阵, $19 \times 4096 \times 4096 \times 1$, 就获得了 19×1 的降维后的数据。

如果是使用matlab自带的princomp进行降维, 那么得到的coeff就是降维矩阵, 使用测试样本, 这里的训练样本和测试样本都要转换成行向量, 每一行表示一个样本, 测试样本是 1×4096 , 降维矩阵是 4096×29 , 那么就是 用待降维的样本 x 乘上降维矩阵, $x \times \text{coeff}$, 注意这两种不同的样本表示方法中降维的使用, 降维矩阵的不同位置。这样降维后获得 $1 \times 4096 \times 4096 \times 29 = 1 \times 29$ 的降维后的数据。

通过 上面的自己实现的pca降维的代码, 还可以对降维后的数据进行重建, 获得重建后的图像, 上面的程序中已经给出了。下面给出一个通过princomp降维后再对降维后的数据进行重建的程序。

通过 princomp降维后的数据进行重建

```

[csharp]
01. clear;
02. clc;
03. train_path='E:\TrainingSet\angry\positive\';
04. images = dir('E:\TrainingSet\angry\positive\*.bmp');
05. phi=zeros(30,64*64);
06.
07. % 加载样本图像到 30*(64*64)的矩阵中, 每一行代表一幅图像
08. for i=1:30
09.     path=strcat(train_path,images(i).name);
10.     Image=imread(path);

```

```

11.     Image=imresize(Image,[64,64]);
12.     phi(i,:)=double(reshape(Image,1,[]));
13. end;
14.
15. % 计算平均脸，并保存用以查看
16. mean_phi=mean(phi,1);
17. mean_face=reshape(mean_phi,64,64);
18. Image_mean=mat2gray(mean_face);
19. imwrite(Image_mean,'meanface2.bmp','bmp');
20.
21. % 使用matlab自带的pca进行降维
22. [coeff, score, latent, TSQUARED] = princomp(phi,'econ');
23.
24. %display Eigenface
25. for i=1:29
26.     Eigenface=reshape(coeff(:,i),[64,64]);
27.     figure(i);
28.     imshow(mat2gray(Eigenface));
29. end
30.
31. % 进行测试
32. %使用测试样本进行测试
33. clc;
34. test_path='E:\BIT\code\FER\meanface.bmp';
35. error=zeros([1,4]);
36.
37. Image=imread(test_path);
38. Image=double(imresize(Image,[64,64]));
39. phi_test=zeros(1,64*64);
40. phi_test(1,:)=double(reshape(Image,1,[])); % 读入的测试图像保存为一行，行向量
41. X_test=phi_test-mean_phi; % 检测训练样本的平均脸
42. Y_test=X_test*coeff; % 进行降维<span style="background-color: rgb(248, 248, 248);"></span>

[csharp] C {
01. X_test_re=Y_test*coeff'; % 重构
02. Face_re=X_test_re+mean_phi;
03. %calculate error rate
04. e=Face_re-phi_test;
05.
06. %%display figure
07. Face_re_2=reshape(Face_re(1,:),[64,64]);
08. figure(i);
09.
10. imshow(mat2gray(Image));
11. title('Original');
12. figure(10+i);
13. imshow(mat2gray(Face_re_2));
14. title('Reconstruct');
15. error(1,i)=norm(e);
16.
17. %dispaly error rate
18. error_rate=error(1,i);
19. display(error_rate);

```

上面的程序关键处都有注释，应该挺好理解的。

关于网络上的一些解释个人理解（仅供大家参考理解）

1.

原文地址：<http://www.cnblogs.com/sunwufan/archive/2011/08/31/2159952.html>

原文：

最近看了些主成分分析，混迹Matlab论坛，翻了n多帖子，对princomp函数有了些了解。

在此只讲一些个人理解，并没有用术语，只求通俗。

贡献率：每一维数据对于区分整个数据的贡献，贡献率最大的显然是主成分，第二大的是次成分.....

[coef, score, latent, t2] = princomp(x);（个人观点）：

x: 为要输入的n维原始数据。带入这个matlab自带函数，将会生成新的n维加工后的数据（即score）。此数据与之前的n维原始数据一一对应。

score: 生成的n维加工后的数据存在score里。它是对原始数据进行的分析，进而在新的坐标系下获得的数据。他将这n维数据按贡献率由大到小排列。（即在改变坐标系的情况下，又对n维数据排序）

latent: 是一维列向量，每一个数据是对应score里相应维的贡献率，因为数据有n维所以列向量有n个数据。由大到小排列（因为score也是按贡献率由大到小排列）。

coef: 是系数矩阵。通过cofe可以知道x是怎样转换成score的。

则模型为从原始数据出发:

score= bsxfun(@minus, x, mean(x, 1))*coef; (作用: 可以把测试数据通过此方法转变为新的坐标系)

逆变换:

x= bsxfun(@plus, score*inv(coef), mean(x, 1))

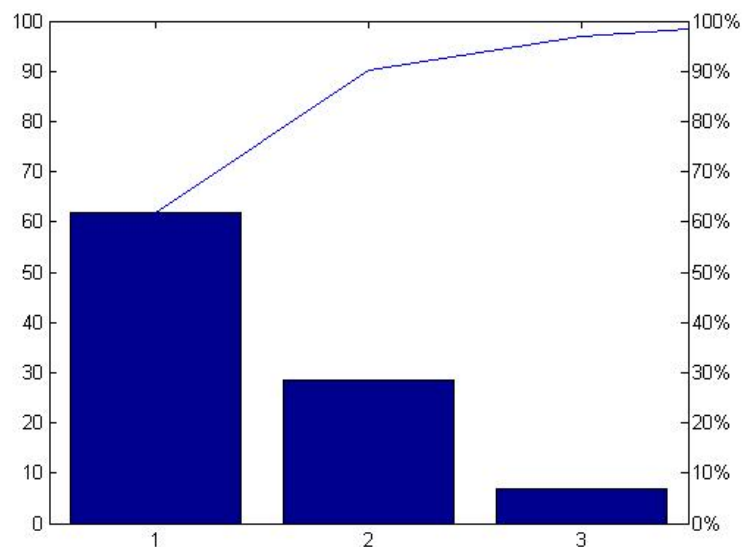
例子:

[View Code](#)

```
%%
%清屏
clear
%%
%初始化数据
a=
[-14.8271317103068, -3.00108550936016, 1.52090778549498, 3.95534842970601; -16.2288612441648, -2.801874
33749996, -0.410815700402130, 1.47546694457079; -15.1242838039605, -2.59871263957451, -0.35996567444673
7, 1.34583763509479; -15.7031424565913, -2.53005662064257, 0.255003254103276, -0.179334985754377; -17.78
92158910100, -3.32842422986555, 0.255791146332054, 1.65118282449042; -17.8126324036279, -4.097195279534
07, -0.879821957489877, -0.196675865428539; -14.9958877514765, -3.90753364293621, -0.418298866141441, -0
.278063876667954; -15.5246706309866, -2.08905845264568, -1.16425848541704, -1.16976057326753];

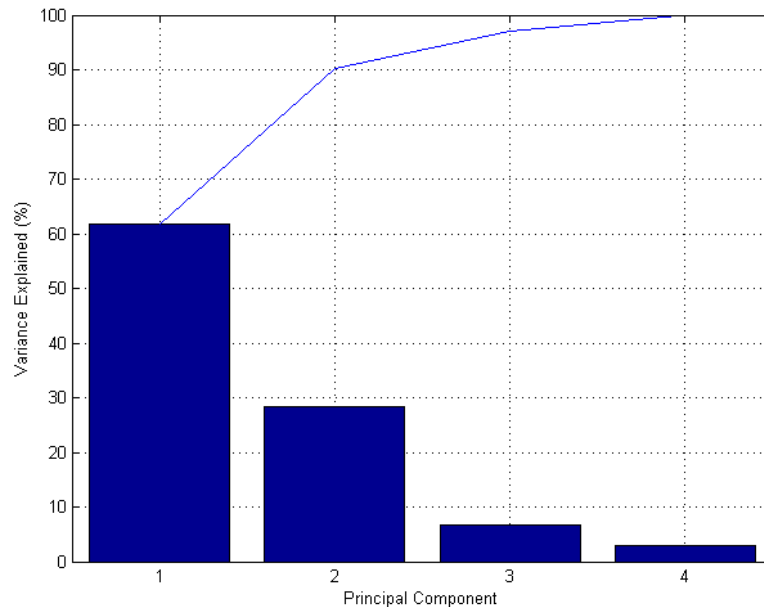
x=a;
%%
%调用princomp函数
[coef, score, latent, t2] = princomp(x);
score
%测试score是否和score_test一样
score_test=bsxfun(@minus, x, mean(x, 1))*coef;
score_test

latent=100*latent/sum(latent)%将latent总和统一为100，便于观察贡献率
pareto(latent);%调用matla画图
```



上图是通过自带函数绘制，当贡献率累加至95%，以后的维数会不在显示，最多只显示10维。

下面用自己编写的表示：



之前的**错误**认识：

1. 认为主成分分析中latent显示的贡献值是原始数据的，其实是加工后的数据的。解释：对原始数据既然选择PCA方法，那么计算机认为原始数据每维之间可能存在关联，你想去掉关联、降低维数。所以采用这种方法的。所以计算机并不关心原始数据的贡献值，因为你不会去用了，用的是加工后的数据（这也是为什么当把输入数据每一维的顺序改变后，score、latent不受影响的原因）。
2. 认为PCA分析后自动降维，不对。PCA后会有贡献值，是输入者根据自己想要的贡献值进行维数的改变，进而生成数据。（一般大家会取贡献值在85%以上，要求高一点95%）。
3. PCA分析，只根据输入数据的特征进行主成分分析，与输出有多少类型，每个数据对应哪个类型无关。如果样本已经分好类型，那PCA后势必对结果的准确性有一定影响，我认为对于此类数据的PCA，就是在降维与准确性间找一个平衡点的问题，让数据即不会维数多而使运算复杂，又有较高的分辨率。

我的个人见解：这篇文章中的解释挺靠谱的，可以用来参考。第二点其实matlab的输出结果score这个数据已经是降维后的数据，不过大家可以根据自己的需要取前多少列的数据。

2。

原文地址：<http://www.ilovematlab.cn/thread-54600-1-1.html>

部分原文：

回复 8# 5342245 的帖子

设原始数据为X，先不做任何预处理。

```
[coef,score,latent,t2] = princomp(X);
```

则那些参数的底层算法大体过程如下：

```
x0 = bsxfun(@minus,X,mean(X,1)); %x0为将X去均值后的数据。
```

```
[coef,ignore] = eig(x0'*x0); 这就是coef的由来。 【当然最终的还有排序什么乱七八糟的。。】
```

```
score = x0*coef % 这就是score的由来，就是一个简单的线性变换，将原来的X的坐标转换到主成分空间中的坐标。仅此而已
```

则模型为从原始数据出发：

```
score = bsxfun(@minus,X,mean(X,1))*coef;
```

逆变换：

```
X = bsxfun(@plus,score*inv(coef),mean(X,1))
```

以上这些你可以自己验证，看是否正确。

关于你的第三问。对于每一个主成分，就看coef的相应的列就能知道原始的变量那个对该主成分贡献大了啊。。

上面是没有预处理的。如果加了可逆的预处理。则原始数据亦可从预处理后的数据表示出。进而 bla bla....
=====这回够通俗易懂吧。。 $O(n_n)O$

PS：pca算法流程，你熟悉吗？只要知道那个算法过程。这些都不难理解啊。。
建议您看看书把pca算法流程再过一遍。。否则别人再怎么说明也没用。。。

我的个人见解：

这里我想说的是，在对测试样本进行降维的时候，一定要减去训练样本的均值，使用训练样本得到的转换矩阵，保证训练样本和测试样本转换到相同的样本空间中，这样才有意义。大家有时间可以去看看英文的资料，说的都比较详细。再用测试样本减去均值以后，就可以进行转换了。

很多同学可能在开始的时候和我一样，都是不知道如果对测试样本进行降维，很多人就选择了还是使用princomp这个函数处理测试样本，那么这样测试样本被映射到一个新的空间中，和原来的训练样本完全不是在一个空间，一点意义都没有，还是要使用测试样本减去均值，然后乘上训练样本降维的时候获得降维矩阵，转换到相同的空间中。

基本的对pca的认识就都说完了，比较乱，没有条理，不过如果认真看下来的话，应该还是可以理解的。目前网上没有关于pca的综合的介绍个注意事项，说以我就把我的经验和大家分享一下，还望文明转载，转载声明出处。我也没有对pca进行详细的学习，肯定有不正确的地方，还请大家多多指教，共同探讨。

上一篇 [使用matlab对训练样本图像降维，并对测试图像使用变换矩阵降维并重构](#)

下一篇 [整理一些网络上的PCA算法，供大家参考](#)

顶 49 踩 0

我的同类文章

Algorithms (46)

- 编写程序 控制cpu占用率
- 编写程序 控制cpu占用率-3
- 写程序 控制cpu占用率-4 绘制CPU使用率的正弦...
- 同时控制CPU的两个内核，显示两个曲线
- 编程之美- 烙饼排序问题
- 编写程序 控制cpu占用率2
- 编写程序 控制cpu占用率-4 绘制CPU使用率的正...
- 写程序+控制cpu占用率-4+绘制CPU使用率的正...
- 编程之美 中国象棋将帅问题 位操作
- 一字棋游戏设计-极大极小搜索

更多

主题推荐

matlab

算法

猜你在找

有趣的算法（数据结构）
数据结构基础系列（1）：数据结构和算法
iOS自动化测试第一季
数据结构和算法
Python自动化测试MonkeyRunner

PCA降维算法总结以及matlab实现PCA
关于matlab的PCA降维的一些资料
基于MATLAB运用PCA+SVM的特征脸方法人脸识别
PCA降维简介
基于PCA的人脸识别02 - 实现一个最基本的系统原

Miss Diāne
黛丝恩

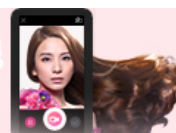
参与活动

送



HEBE亲笔签名海报

卡西欧自拍神器



飘逸女神
大咖秀

了解详情 >

查看评论

40楼 [LejianHP](#) 2015-10-20 19:52发表



我有X: 16000*3000 (3000为样本), 使用
`[coeff,score,latent,tsquare]=princomp(X);`
`X=score(:,1:1000);`变为1000*3000的特征作为识别, 为什么识别率和用没降维的相比下降很多啊? ?

39楼 [jacobian_t](#) 2015-06-04 14:31发表



"而网上大部分都是每一列表示一个样本, 这样这一列的所有行都表示这个样本的特征值。网上的程序使用列表示样本是有一定好处的, 比如我的样本是1000000*30, 总共有30个训练样本, 每个样本的特征值个数是1000000, 那么这个矩阵获得的协方差矩阵是30*30, 计算起来非常的方便, 不想30*1000000这样的矩阵获得到的协方差矩阵式1000000*1000000, 直接就内存溢出了, 不过matlab有自己的实现方式, 巧妙的解决了这个问题"

30*30根本就不是特征的协方差矩阵, 这是最基本的。

38楼 [tmac5655](#) 2015-05-25 09:42发表



请问一下你的重建误差是如何计算的?

37楼 [zyy19900402](#) 2015-04-27 10:44发表



十分感谢博主, 写的很好, 也都是好东西, 感谢

36楼 [Bicelove](#) 2015-03-30 16:39发表



楼主你好! 最近在使用PCA进行降维, 有幸读到了您的相关博客! 我是要处理100*25137的数据, 100是样本数, 25137是维度, 我现在需要把维度降下来, 但求得特征值与特征向量均是complex double数据, 不知是什么原因, 请问楼主遇见过这种情况嘛?

Re: [jsjipipi](#) 2015-04-07 14:42发表



回复u010555688: 你好, 我也是得到的复数, 不知道你解决了吗, 望一起学习啊, 求回复

35楼 [paladinosment](#) 2014-12-08 23:16发表



神贴。楼主太强了~~~

34楼 [数据小冰](#) 2014-11-22 15:58发表



最近正在看pac资料, 对我很有帮助

33楼 [sinat_21760611](#) 2014-10-09 16:23发表



同问~~~这两句是什么意思?

32楼 [gaocheng007](#) 2014-09-12 15:47发表



你好, 可以加我的qq吗?和你交流下296800938

Re: [watkins](#) 2014-09-15 09:40发表



回复gaocheng007: 我不怎么用QQ, 可以邮件交流: watkinsong@163.com或者 wei.x.song@oracle.com

31楼 [BadNull](#) 2014-09-05 08:50发表



"很多同学可能在开始的时候和我一样, 都是不知道如果对测试样本进行降维, 很多人就选择了还是使用princomp这个函数处理测试样本, 那么这样测试样本被映射到一个新的空间中, 和原来的训练样本完全不是在一个空间, 一点意义都没有, 还是要使用测试样本减去均值, 然后乘上训练样本降维的时候获得降维矩阵, 转换到相同的空间中。"希望能把这句话举个例子详细说一下, 意思是样本X, princomp(X)之后, 得到COEFF, 比如要得到前两个主成分, 不能直接X*COEFF? 而要先X0=bsxfun(@minus,X,mean(X,1));, 然后X0*COEFF(:,1:2)才是样本的前两个主成分?

Re: [小村长](#) 2014-11-27 13:54发表



回复BadNull: 这个直接使用SCORE就可以了哈哈

30楼 [susan810290107](#) 2014-08-28 17:27发表



请问一下协方差矩阵的特征值是不是应该非负啊? 我的特征值有负数是什么原因呢?

Re: [小村长](#) 2014-11-27 13:57发表



回复susan810290107: 你的数据最好先归一化下。

29楼 [watkins](#) 2014-08-13 20:48发表

全新的算法总结: <http://blog.csdn.net/watkinsong/article/details/38536463>



28楼 [dodojiang](#) 2014-08-01 10:08发表



找来找去终于找到最全的了，而且介绍也很容易理解，非常感谢！

27楼 [hondef](#) 2014-07-01 17:42发表



时隔很久才看到您的文章，不知道会不会回复，我是想问一下，上面的代码是可以完成“不同人脸”的识别，还是可以完成“人脸”和“非人脸”的识别？

Re: [watkins](#) 2014-07-03 16:55发表



回复[hondef](#): 一般情况下，我们做的都是同类事物的识别，本文所给出的都是同类人脸的识别，在学术领域，我暂时还没有看到过人脸和非人脸的区分，做这个难度太大，非人脸包含太多的种类，不过有一种简单的办法就是在图像上做基于Haar特征的人脸检测，检测到人脸就划分为人脸，否则就划分为非人脸。

26楼 [hellofuturecj](#) 2014-06-13 16:43发表



写的很赞啊，迷惑部分是亮点，怒顶！

Re: [watkins](#) 2014-07-03 17:09发表



回复[hellofuturecj](#): 哈哈，谢谢。大家共同学习

25楼 [ganxiewusideren](#) 2014-05-18 01:03发表



感谢无私的楼主，感谢你，你帮助我学习了好多东西，谢谢你

Re: [watkins](#) 2014-07-03 17:09发表



回复[ganxiewusideren](#): 谢谢，共同学习，共同进步～

24楼 [huacuo2013](#) 2014-05-08 19:38发表



一般来说，PCA降维后的每个样本的特征的维数，不会超过训练样本的个数，因为超出的特征是没有意义的。

楼主，我不是很理解这句话。为什么没意义？

30*10000的数据，其中30为样本数，10000为特征数，出于某种需要，我利用PCA降维到30*80，其中80为降维后的特征数。这样不行吗？至于80是我需要的维度，是我人为定的。

希望楼主能解答。

Re: [watkins](#) 2014-07-03 17:08发表



回复[huacuo2013](#): 我个人的理解是：因为降维其实就是转换到其他的样本空间重新表示特征，对于30个样本，29个维度已经足以表示所有的30个样本，这样的话用29个维度和80个维度，理论上来说没有区别，因为后边的超过30的维度都是浪费的。

Re: [小村长](#) 2014-11-27 13:59发表



回复[weixingstudio](#): 不是很理解。。。。

23楼 [fangth](#) 2014-03-17 10:18发表



写得真好，谢谢露珠！

Re: [watkins](#) 2014-07-03 17:10发表



回复[fangth](#): 谢谢，共同学习

22楼 [sulanqing](#) 2014-03-11 15:55发表



博士写的有彩，受益了！

21楼 [sulanqing](#) 2014-03-11 15:53发表



博主写的有彩，很是受益！

20楼 [光谱师](#) 2014-03-10 14:59发表



这篇文章对我的帮助太大了。请问博主什么时候能写一篇关于PLS的类似文章。不尽感激。

19楼 [weitaogudan](#) 2013-12-26 00:39发表



我或者使用score这个矩阵，但是降维后还是不能匹配。

```
[coeff,score,latent,tsquare] = princomp(descr1,'econ');
descr1=score(:,1:24);
```



```
x0 = bsxfun(@minus,descr2,mean(descr1,1));
descr2=x0*coeff(:,1:24);
```

18楼 [weitaogudan](#) 2013-12-25 23:33发表



亲们，我看了一下你的这篇文章。有点问题想请教一下哦！我是要做pca-sift实现模板和样本的匹配。主要是：sift变换后，想要对得到的特征描述子descr进行降维处理。但是发现，对矩阵进行降维之后，不能用来进行匹配了。很是费解，难道是我的降维程序写错了吗？麻烦您指点一下。

%进行sift变换

```
[frames1,descr1,gss1,dogss1] = do_sift( I1 );
```

```
[frames2,descr2,gss2,dogss2] = do_sift( I2 );
```

得到两个矩阵descr1 (346*128)和descr2 (145*128)。128是特征向量的维数。我直接自己定义降维后是24维（已经验证过是可以取得90%以上的信息的）

%对特征描述子进行降维（默认降到24维。具体维数可自己定义）

```
[coeff,score,latent,tsquare] = princomp(descr1,'econ');
```

```
descr1=descr1*coeff(:,1:24); %我没有使用score这个矩阵
```

```
descr2=descr2*coeff(:,1:24);
```

降维后的descr1和 descr2失去了后续匹配的意义了……没有办法进行匹配了。（没有降维前，是可以进行匹配的）

Re: [weitaogudan](#) 2013-12-25 23:46发表



回复weitaogudan：我或者使用score这个矩阵，但是降维后还是不能匹配。

```
[coeff,score,latent,tsquare] = princomp(descr1,'econ');
```

```
descr1=score(:,1:24);
```

```
x0 = bsxfun(@minus,descr2,mean(descr2,1));
```

```
descr2=x0*coeff(:,1:24);
```

17楼 [ice110956](#) 2013-11-06 09:22发表



PCA计算的协方差是一样的，同样的降维维数，得到的结果也是一样的。

16楼 [silence1214](#) 2013-10-06 11:22发表



上面发了几个评论，在说一句吧：pca求出来的协方差的特征向量是没问题的，和作者的协方差是不一样的，但是作者在pca的那个地方没有画出来分别对应的投影，只是画出来了特征脸，应该这个地方的特征连和上面实验中的投影出来的我觉得不一样的。在下面重构的地方注意到，作者分了2个步骤连起来其实就是 $UL*UL*xxx$ 的，pca中直接是用coeff来的，这个可能是等价的关系，但是没具体看。

希望作者能够解答我上面一些疑问，也是朋友推荐过来看的这篇博文，谢谢。

15楼 [silence1214](#) 2013-10-06 11:03发表



```
Y_test=UL*X_test;
```

```
X_test_re=UL*Y_test;
```

在重构中，这2句话的作用是什么呢？谢谢楼主

14楼 [silence1214](#) 2013-10-06 10:37发表



其实你这个地方不是降的维度你是降低的数据样本数，因为在你的第一个测试例子中，样本维度是大小是4096(64*64)*20的，后面取了19个计算出来的是4096*19。。减少的是样本数，不是维度(64*64)。。可能在实际需求中是存在这种的计算方法的，但是这里面有一些概念性的东西我个人觉得说的不到位哈，包括协方差的计算不太正确。。可能是你实际需要这么来的吧

13楼 [silence1214](#) 2013-10-06 10:11发表



协方差的定义是不同样本之间同纬度数据之间的运算，如果一组样本10000维度吗，一共30个样本的话，只有 $1/30*A*A'$ 才是协方差矩阵，A是一个10000*30的矩阵，不是说30*10000的矩阵，你这样计算出来的30*30的就不是协方差矩阵哟，但是你后面的结果怎么看着好像也不错啊

12楼 [anakin320](#) 2013-09-06 16:55发表



请问楼主，我对数据进行pca降维后分类，涉及到正负样本集，正样本集里有训练样本和测试样本，负样本集里也有训练样本和测试样本。我运行时用正样本的训练集调用了princom函数，得到coeff1和score1，用coeff1来转换正样本的测试集得到score1'（已减去训练样本均值）；然后又用负样本的训练集调用princom函数，同样得到coeff2和score2，用coeff2转换负样本得到score2'（已减去训练样本均值）。这样得到了4个降了维的特征数据集，然后去分类，发现分类结果全部判定为负样本，想请教楼主是怎么回事？

Re: [watkins](#) 2013-09-08 10:52发表



回复anakin320：很明显的一个错误，正样本和负样本需要在同一个样本空间中践行降维，你的正样本和负样本单独降维，这样降维后的正样本与负样本根本不在一个特征空间中，所以肯定是有问题的。应该将正样本与负样本放在一起同时降维。

Re: [anakin320](#) 2013-09-09 08:24发表



回复weixingstudio：你的意思是把正负样本合在一起？我本来是训练用正样本200个，训练用负样本200个，你意思合在一起400个样本输入到princom函数里，然后再用转换矩阵coeff去转换测试的样本（也合在一起转换）？那这样的话得到的降维矩阵每个样本对应的样本标签还一致么？

Re: [watkins](#) 2013-09-09 09:07发表

回复anakin320：对，就是这个意思，只有这样做才能保证正样本和负样本是在一个特征



空间中，训练分类器才会有意思。降维之后的顺序和原来的顺序一致的，没有区别。至于测试样本，使用降维矩阵单个的降维和一起降维没有区别的，这个可以随意。

Re: [anakin320](#) 2013-09-09 10:31发表



回复weixingstudio: 好的，非常感谢！经你这样一解释才终于弄明白了。

11楼 [wangyingnihao](#) 2013-08-03 10:25发表



你好，我想问一下降维后样本的顺序与降维前样本顺序是一致的吗？

Re: [watkins](#) 2013-08-05 09:24发表



回复wangyingnihao: 降维后的样本顺序肯定和降维之前的是一致的，要不降维之后怎么使用呢。

10楼 [byxiangfei](#) 2013-07-03 14:02发表



不好意思，看到您说100000*30 和30*1000000 那块了。

是这样有个概念，协方差矩阵都是对维度做协方差的，所以按照定义来想，必须是1000000*1000000的协方差矩阵的。协方差矩阵并不是对样本30个，做协方差。您想，他是在降维对吧，又不是降样本。。所以要把1000000个维度，降下来才行。

参考文章是: <http://pinkyje.com/2010/08/31/covariance/>

Re: [watkins](#) 2013-07-05 17:26发表



回复byxiangfei: 恩，你说的对，呵呵，我就是低级理解。以后还要多多向你请教。

Re: [大愚若智_](#) 2014-07-21 09:18发表



回复weixingstudio: 我也是这么理解的。不过30*30的样本协方差矩阵也是有应用的。比如同一个信号，测量了30次，使用30*30的协方差进行PCA是可以提高信噪比的。

9楼 [u010914456](#) 2013-06-01 15:42发表



太谢谢你的回复了，你说的降维矩阵就是13*4转换矩阵的逆吧 但它是奇异的 只能是用广义逆矩阵了吧

8楼 [u010914456](#) 2013-06-01 10:47发表



你好，我想问下降维之后的数据能不能通过乘以转换矩阵的转置得到原来未降维时的数据呢？就是我现在将10*13的数据降到了10*4，用这些数据去训练神经网络，测试的时候神经网络输出了10*4的数据，想把这些数据通过PCA逆变换变换成10*13的矩阵，有什么办法吗？

Re: [watkins](#) 2013-06-01 13:25发表



回复u010914456: 可以将降维后的数据乘以降维矩阵逆矩阵，然后加上降维之前的平均值，就可以获得原来的数据。但是存在一定的误差。因为降维的时候就减去了平均值，所以最后还要加上。

7楼 [alexsun01](#) 2013-05-27 15:50发表



很有帮助，我想问下有没有关于PCA-SIFT方面的例子。

Re: [weitaogudan](#) 2013-12-26 00:37发表



回复alexsun01: 亲，你好！请问你pca-sift实现了吗？我现在正遇到问题，想请教一下您哦！！我的问题就在评论里面
用户: weitaogudan

Re: [watkins](#) 2013-05-27 16:08发表



回复alexsun01: 我没有做过SIFT，不过opencv有SIFT的实现吧，你理解了pca的原理，降维使用就可以了。

6楼 [流小淘](#) 2013-05-08 19:46发表



实在太感谢了，受益很多

5楼 [wolf3077956](#) 2013-04-30 13:28发表



我想把一个71*328的矩阵降到71*3.怎么写Matlab代码吗？71是样本个数。能不能帮我一下啊。加我QQ491805841

Re: [watkins](#) 2013-04-30 19:00发表



回复wolf3077956: 这个实现很简单啊，你得到的降维矩阵只要取前三列就可以了，然后用原来的矩阵乘以降维矩阵的前三列，再加上均值，就可以了。

4楼 [lvgyx](#) 2013-04-26 15:59发表



Too many thanks ! 正好解决了我的疑惑，让我有了一个基本的了解。

3楼 [cuiyongxu1109](#) 2013-04-08 11:12发表



我也在做这方面，可以加你qq吗，我的318473933

2楼 [jbr1025](#) 2013-03-18 12:44发表



我有一个500*15360的大矩阵（500是样本数），然后我就调用matlab自带的princomp()函数，报错如下
Error using svd
Out of memory. Type HELP MEMORY for your options.
Error in pca>localSVD (line 357)
[U,sigma,coeff] = svd(x, 0);
Error in pca (line 249)
[U,sigma,coeff,wasNaN] = localSVD(x, n,...
Error in princomp (line 29) [varargout{1:nargout}]=pca(varargin{1},'Algorithm','svd','Economy',fEconomy);
还是有out of memory的问题，不知道楼主是怎么避免内存问题的？谢谢！

Re: [y1f13](#) 2014-03-27 11:17发表



回复jbr1025: 样本个数对内存影响不是很大，因为你的特征数已经很大了，因为降维是要降特征，所以你肯定要生成特征数*特征数 大的协方差矩阵，所以你的例子会生成15360*15360 大的矩阵，按照每个数据8个字节存储（有1.74G大的内存，如果是实验室电脑还能承受，个人电脑就难咯），楼主所说5、6万的特征做降维，能做到降维，这个还请楼主指教如何做到的？如果使用参数econ,不是对样本降维了么

Re: [watkins](#) 2013-03-21 16:13发表



回复jbr1025: 这个我也不太清楚，我降维的时候每个样本的特征值也很多，大概有五六万，但是样本的个数没有你这么多，我样本只有30个。所以没有出现内存溢出的情况。你可以先对原有特征进行一个手动删选什么的试试。

1楼 [LuLuLangLang](#) 2013-02-25 11:20发表



用得到，这段时间会常来。。。

您还没有登录,请[\[登录\]](#)或[\[注册\]](#)

* 以上用户言论只代表其个人观点，不代表CSDN网站的观点或立场

核心技术类目

全部主题 Hadoop AWS 移动游戏 Java Android iOS Swift 智能硬件 Docker
OpenStack VPN Spark ERP IE10 Eclipse CRM JavaScript 数据库 Ubuntu NFC
WAP jQuery BI HTML5 Spring Apache .NET API HTML SDK IIS Fedora XML
LBS Unity Splashtop UML components Windows Mobile Rails QEMU KDE Cassandra
CloudStack FTC coremail OPhone CouchBase 云计算 iOS6 Rackspace Web App
SpringSide Maemo Compuware 大数据 aptech Perl Tornado Ruby Hibernate ThinkPHP
HBase Pure Solr Angular Cloud Foundry Redis Scala Django Bootstrap

[公司简介](#) | [招贤纳士](#) | [广告服务](#) | [银行汇款帐号](#) | [联系方式](#) | [版权声明](#) | [法律顾问](#) | [问题报告](#) | [合作伙伴](#) | [论坛反馈](#)

网站客服 杂志客服 微博客服 webmaster@csdn.net 400-600-2320 | 北京创新乐知信息技术有限公司 版权所有 | 江苏乐知网络技术有限公司 提供商务支持
京 ICP 证 09002463 号 | Copyright © 1999-2014, CSDN.NET, All Rights Reserved