

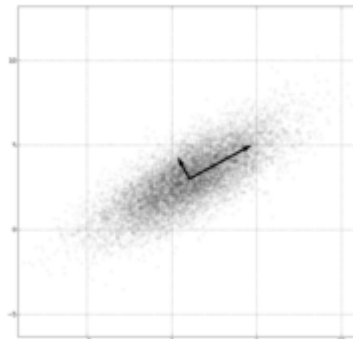
主成分分析

維基百科，自由的百科全书

在多元统计分析中，**主成分分析**（英語：**Principal components analysis, PCA**）是一種统计分析、簡化數據集的方法。它利用正交变换来对一系列可能相关的变量的观测值进行线性变换，从而投影为一系列线性不相关变量的值，这些不相关变量称为主成分（Principal Components）。具体地，主成分可以看做一个线性方程，其包含一系列线性系数来指示投影方向。PCA对原始数据的正则化或预处理敏感（相对缩放）。

基本思想：

- 将坐标轴中心移到数据的中心，然后旋转坐标轴，使得数据在C1轴上的方差最大，即全部n个数据个体在该方向上的投影最为分散。意味着更多的信息被保留下来。C1成为**第一主成分**。
- C2第二主成分**：找一个C2，使得C2与C1的协方差（相关系数）为0，以免与C1信息重叠，并且使数据在该方向的方差尽量最大。
- 以此类推，找到第三主成分，第四主成分……第p个主成分。p个随机变量可以有p个主成分^[1]。



一個高斯分布，平均值為(1, 3)，標準差在(0.878, 0.478)方向上為3、在其正交方向上為1的主成分分析。黑色的兩個向量是此分布的共變異數矩陣的特征向量，其長度為對應的特征值之平方根，並以分布的平均值為原點。

主成分分析经常用于减少数据集的维数，同时保留数据集當中对方差贡献最大的特征。这是通过保留低維主成分，忽略高維主成分做到的。这样低維成分往往能够保留住数据的最重要部分。但是，这也不是一定的，要视具体应用而定。由于主成分分析依赖所给数据，所以数据的准确性对分析结果影响很大。

主成分分析由卡尔·皮尔逊於1901年發明^[2]，用於分析數據及建立數理模型，在原理上与主轴定理相似。之后在1930年左右由哈罗德·霍特林独立发展并命名。依据应用领域的不同，在信号处理中它也叫做离散K-L 转换（discrete Karhunen–Loève transform (KLT)）。其方法主要是通過對共變異數矩陣進行特征分解^[3]，以得出數據的主成分（即特征向量）與它們的權值（即特征值^[4]）。PCA是最簡單的以特征量分析多元統計分布的方法。其結果可以理解為對原數據中的方差做出解釋：哪一個方向上的數據值對方差的影響最大？換而言之，PCA提供了一種降低數據維度的有效辦法；如果分析者在原數據中除掉最小的特征值所對應的成分，那麼所得的低維度數據必定是最優化的（也即，這樣降低維度必定是失去訊息最少的方法）。主成分分析在分析複雜數據時尤為有用，比如人臉識別。

PCA是最简单的以特征量分析多元统计分布的方法。通常，这种运算可以被看作是揭露数据的内部结构，從而更好地展現数据的變異度。如果一个多元数据集是用高维数据空间之坐标系來表示的，那么PCA能提供一幅较低维度的图像，相當於数据集在讯息量最多之角度上的一個投影。这样就可以利用少量的主成分讓数据的维度降低了。

PCA 跟因子分析密切相关。因子分析通常包含更多特定領域底層結構的假設，並且求解稍微不同矩陣的特徵向量。

PCA 也跟典型相關分析(CCA)有關。CCA定義的坐標系可以最佳地描述兩個數據集之間的互協方差，而PCA定義了新的正交坐標系，能最佳地描述單個數據集當中的變異數。

目录

[数学定义](#)

[讨论](#)

[符号和缩写表](#)

[主成分分析的属性和限制](#)

[主成分分析和信息理论](#)

[使用统计方法计算PCA](#)

[组织数据集](#)

[计算经验均值](#)

[计算平均偏差](#)

[求协方差矩阵](#)

[查找协方差矩阵的特征值和特征向量](#)

[参见](#)

[注释](#)

[参考](#)

数学定义

PCA的数学定义是：一个正交化线性变换，把数据变换到一个新的坐标系统中，使得这一数据的任何投影的第一大方差在第一个坐标（称为第一主成分）上，第二大方差在第二个坐标（第二主成分）上，依次类推^[5]。

定义一个 $n \times m$ 的矩阵 \mathbf{X} 为去平均值（以平均值为中心移动至原点）的数据，其行为数据样本，列为数据类别（注意，这里定义的是 \mathbf{X}^T 而不是 \mathbf{X} ）。则 \mathbf{X} 的奇异值分解为 $\mathbf{X} = \mathbf{W}\mathbf{\Sigma}\mathbf{V}^T$ ，其中 $m \times m$ 矩阵 \mathbf{W} 是 $\mathbf{X}\mathbf{X}^T$ 的特征向量矩阵， $\mathbf{\Sigma}$ 是 $m \times n$ 的非负矩形对角矩阵， \mathbf{V} 是 $n \times n$ 的 $\mathbf{X}^T\mathbf{X}$ 的特征向量矩阵。据此，

$$\begin{aligned}\mathbf{Y}^T &= \mathbf{X}^T \mathbf{W} \\ &= \mathbf{V} \mathbf{\Sigma}^T \mathbf{W}^T \mathbf{W} \\ &= \mathbf{V} \mathbf{\Sigma}^T\end{aligned}$$

当 $m < n - 1$ 时， \mathbf{V} 在通常情况下不是唯一定义的，而 \mathbf{Y} 则是唯一定义的。 \mathbf{W} 是一个正交矩阵， $\mathbf{Y}^T \mathbf{W}^T = \mathbf{X}^T$ ，且 \mathbf{Y}^T 的第一列由第一主成分组成，第二列由第二主成分组成，依此类推。

为了得到一种降低数据维度的有效办法，我们可以利用 \mathbf{W}_L 把 \mathbf{X} 映射到一个只应用前面 L 个向量的低维空间中去：

$$\mathbf{Y} = \mathbf{W}_L^\top \mathbf{X} = \mathbf{\Sigma}_L \mathbf{V}^\top$$

其中 $\mathbf{\Sigma}_L = \mathbf{I}_{L \times m} \mathbf{\Sigma}$ ，且 $\mathbf{I}_{L \times m}$ 為 $L \times m$ 的單位矩陣。

\mathbf{X} 的单向量矩阵 \mathbf{w} 相当于协方差矩阵的特征向量 $\mathbf{C} = \mathbf{X} \mathbf{X}^\top$,

$$\mathbf{X} \mathbf{X}^\top = \mathbf{W} \mathbf{\Sigma} \mathbf{\Sigma}^\top \mathbf{W}^\top$$

在欧几里得空间给定一组点数，第一主成分对应于通过多维空间平均点的一条线，同时保证各个点到这条直线距离的平方和最小。去除掉第一主成分后，用同样的方法得到第二主成分。依此类推。在 $\mathbf{\Sigma}$ 中的奇异值均为矩阵 $\mathbf{X} \mathbf{X}^\top$ 的特征值的平方根。每一个特征值都与跟它们相关的方差是成正比的，而且所有特征值的总和等于所有点到它们的多维空间平均点距离的平方和。PCA 提供了一种降低维度的有效办法，本质上，它利用正交变换将围绕平均点的点集中尽可能多的变量投影到第一维中去，因此，降低维度必定是失去讯息最少的方法。PCA 具有保持子空间拥有最大方差的最优正交变换的特性。然而，当与离散余弦变换相比时，它需要更大的计算需求代价。非线性降维技术相对于 PCA 来说则需要更高的计算要求。

PCA 对变量的缩放很敏感。如果我们只有两个变量，而且它们具有相同的样本方差，并且成正相关，那么 PCA 将涉及两个变量的主成分的旋转。但是，如果把第一个变量的所有值都乘以 100，那么第一主成分就几乎和这个变量一样，另一个变量只提供了很小的贡献，第二主成分也将和第二个原始变量几乎一致。这就意味着当不同的变量代表不同的单位（如温度和质量）时，PCA 是一种比较武断的分析方法。但是在 Pearson 的题为 "On Lines and Planes of Closest Fit to Systems of Points in Space" 的原始文件里，是假设在欧几里得空间里不考虑这些。一种使 PCA 不那么武断的方法是使用变量缩放以得到单位方差。

讨论

通常，为了确保第一主成分描述的是最大方差的方向，我们会使用平均减法进行主成分分析。如果不执行平均减法，第一主成分有可能或多或少的对应于数据的平均值。另外，为了找到近似数据的最小均方误差，我们必须选取一个零均值^[6]。

假设零经验均值，数据集 \mathbf{X} 的主成分 w_1 可以被定义为：

$$\mathbf{w}_1 = \arg \max_{\|\mathbf{w}\|=1} \text{Var}\{\mathbf{w}^\top \mathbf{X}\} = \arg \max_{\|\mathbf{w}\|=1} E\left\{\left(\mathbf{w}^\top \mathbf{X}\right)^2\right\}$$

为了得到第 k 个主成分，必须先从 \mathbf{X} 中减去前面的 $k - 1$ 个主成分：

$$\hat{\mathbf{X}}_{k-1} = \mathbf{X} - \sum_{i=1}^{k-1} \mathbf{w}_i \mathbf{w}_i^\top \mathbf{X}$$

然后把求得的第 k 个主成分带入数据集，得到新的数据集，继续寻找主成分。

$$\mathbf{w}_k = \arg \max_{\|\mathbf{w}\|=1} E\left\{\left(\mathbf{w}^\top \hat{\mathbf{X}}_{k-1}\right)^2\right\}.$$

PCA 相当于在气象学中使用的经验正交函数 (EOF)，同时也类似于一个线性隐层神经网络。隐含层 K 个神经元的权重向量收敛后，将形成一个由前 K 个主成分跨越空间的基础。但是与 PCA 不同的是，这种技术并不一定会产生正交向量。

PCA是一种很流行且主要的的模式识别技术。然而，它并不能最优化类别可分离性^[7]。另一种不考虑这一点的方法是线性判别分析。

符号和缩写表

Symbol符号	Meaning意义	Dimensions尺寸	Indices指数
$\mathbf{X} = \{X[m, n]\}$	由所有数据向量集组成的数据矩阵，一列代表一个向量	$M \times N$	$m = 1 \dots M$ $n = 1 \dots N$
N	数据集中列向量的个数	1×1	标量
M	每个列向量的元素个数	1×1	标量
L	子空间的维数, $1 \leq L \leq M$	1×1	标量
$\mathbf{u} = \{u[m]\}$	经验均值向量	$M \times 1$	$m = 1 \dots M$
$\mathbf{s} = \{s[m]\}$	经验标准方差向量	$M \times 1$	$m = 1 \dots M$
$\mathbf{h} = \{h[n]\}$	所有的单位向量	$1 \times N$	$n = 1 \dots N$
$\mathbf{B} = \{B[m, n]\}$	对均值的偏离向量	$M \times N$	$m = 1 \dots M$ $n = 1 \dots N$
$\mathbf{Z} = \{Z[m, n]\}$	Z-分数，利用均值和标准差计算得到	$M \times N$	$m = 1 \dots M$ $n = 1 \dots N$
$\mathbf{C} = \{C[p, q]\}$	协方差矩阵	$M \times M$	$p = 1 \dots M$ $q = 1 \dots M$
$\mathbf{R} = \{R[p, q]\}$	相关矩阵	$M \times M$	$p = 1 \dots M$ $q = 1 \dots M$
$\mathbf{V} = \{V[p, q]\}$	\mathbf{C} 的所有特征向量集	$M \times M$	$p = 1 \dots M$ $q = 1 \dots M$
$\mathbf{D} = \{D[p, q]\}$	主对角线为特征值的对角矩阵	$M \times M$	$p = 1 \dots M$ $q = 1 \dots M$
$\mathbf{W} = \{W[p, q]\}$	基向量矩阵	$M \times L$	$p = 1 \dots M$ $q = 1 \dots L$
$\mathbf{Y} = \{Y[m, n]\}$	\mathbf{X} 和 \mathbf{W} 矩阵的投影矩阵	$L \times N$	$m = 1 \dots L$ $n = 1 \dots N$

主成分分析的属性和限制

如上所述，主成分分析的结果依赖于变量的缩放。

主成分分析的适用性受到由它的派生物产生的某些假设^[8]的限制。

主成分分析和信息理论

通过使用降维来保存大部分数据信息的主成分分析的观点是不正确的。确实如此，当没有任何假设信息的信号模型时，主成分分析在降维的同时并不能保证信息的不丢失，其中信息是由香农熵^[9]来衡量的。基于假设得 $\mathbf{x} = \mathbf{s} + \mathbf{n}$ 也就是说，向量 x 是含有信息的目标信号 s 和噪声信号 n 之和，从信息论角度考虑主成分分析在降维上是最优的。

特别地，Linsker证明了如果 s 是高斯分布，且 n 是与密度矩阵相应的协方差矩阵的高斯噪声，

使用统计方法计算PCA

以下是使用统计方法计算PCA的详细说明。但是请注意，如果利用奇异值分解（使用标准的软件）效果会更好。

我们的目标是把一个给定的具有 M 维的数据集 \mathbf{X} 变换成具有较小维度 L 的数据集 \mathbf{Y} 。现在要求的就是矩阵 \mathbf{Y} ， \mathbf{Y} 是矩阵 \mathbf{X} Karhunen–Loève 变换。： $\mathbf{Y} = \mathbf{KLT}\{\mathbf{X}\}$

组织数据集

假设有一组 M 个变量的观察数据，我们的目的是减少数据，使得能够用 L 个向量来描述每个观察值， $L < M$ 。进一步假设，该数据被整理成一组具有 N 个向量的数据集，其中每个向量都代表 M 个变量的单一观察数据。

- $\mathbf{x}_1 \dots \mathbf{x}_N$ 为列向量，其中每个列向量有 M 行。
- 将列向量放入 $M \times N$ 的矩阵 \mathbf{X} 裡。

计算经验均值

- 对每一维 $m = 1, \dots, M$ 计算经验均值
- 将计算得到的均值放入一个 $M \times 1$ 维的经验均值向量 \mathbf{u} 中

$$u[m] = \frac{1}{N} \sum_{n=1}^N X[m, n]$$

计算平均偏差

对于在最大限度地减少近似数据的均方误差的基础上找到一个主成分来说，均值减去法是该解决方案的不可或缺的组成部分^[10]。因此，我们继续如下步骤：

- 从数据矩阵 \mathbf{X} 的每一列中减去经验均值向量 \mathbf{u}
- 将平均减去过的数据存储在 $M \times N$ 矩阵 \mathbf{B} 中

$$\mathbf{B} = \mathbf{X} - \mathbf{u}\mathbf{h}$$

where \mathbf{h} is a $1 \times N$ row vector of all 1s:

其中 \mathbf{h} 是一个全 1s 的 $1 \times N$ 的行向量

$$h[n] = 1 \quad \text{for } n = 1, \dots, N$$

求协方差矩阵

- 从矩阵 \mathbf{B} 中找到 $M \times M$ 的经验协方差矩阵 \mathbf{C}

$$\mathbf{C} = \mathbb{E}[\mathbf{B} \otimes \mathbf{B}] = \mathbb{E}[\mathbf{B} \cdot \mathbf{B}^*] = \frac{1}{N-1} \sum \mathbf{B} \cdot \mathbf{B}^*$$

其中 \mathbb{E} 为期望值

\otimes 是最外层运算符

$*$ 是共轭转置运算符。

请注意，如果 \mathbf{B} 完全由实数组成，那么共轭转置与正常的转置一样。

- 为什么是 $N-1$, 而不是 N , [Bessel's correction](#)^[11] 给出了解释

查找协方差矩阵的特征值和特征向量

- 计算矩阵 \mathbf{C} 的特征向量

$$\mathbf{V}^{-1} \mathbf{C} \mathbf{V} = \mathbf{D}$$

其中， \mathbf{D} 是 \mathbf{C} 的特征值对角矩阵，这一步通常会涉及到使用基于计算机的计算特征值和特征向量的算法。在很多矩阵代数系统中这些算法都是现成可用的，如 \mathbf{R} 语言，[MATLAB](#),^{[12][13]} [Mathematica](#),^[14] [SciPy](#), [IDL](#)(交互式数据语言), 或者 [GNU Octave](#) 以及 [OpenCV](#)。

- 矩阵 \mathbf{D} 为 $M \times M$ 的对角矩阵
- 各个特征值和特征向量都是配对的， m 个特征值对应 m 个特征向量。

参见

- | | |
|--|---|
| ▪ Correspondence analysis | ▪ 矩阵分解 |
| ▪ Canonical correlation | ▪ Nonlinear dimensionality reduction |
| ▪ CUR matrix approximation (can replace of low-rank SVD approximation) | ▪ Oja's rule |
| ▪ Detrended correspondence analysis | ▪ Point distribution model (PCA applied to morphometry and computer vision) |
| ▪ Dynamic mode decomposition | ▪ Principal component regression |
| ▪ 特征脸(Eigenface) | ▪ Singular spectrum analysis |
| ▪ 多线性主成分分析(Multilinear PCA) | ▪ 奇异值分解 |
| ▪ Geometric data analysis | ▪ Sparse PCA |
| ▪ Factorial code | ▪ 变换编码 |
| ▪ 独立成分分析 | ▪ 最小二乘法 |
| ▪ 核主成分分析 | ▪ Low-rank approximation |
| ▪ 圖模式 | |
| ▪ 马尔可夫链 | |
| ▪ 马尔可夫逻辑网络 | |

注释

1. 主成分分析 (principal components analysis, PCA) —— 无监督学习. (原始内容存档于 2020-08-19) .
2. Pearson, K. On Lines and Planes of Closest Fit to Systems of Points in Space (PDF). Philosophical Magazine. 1901, **2** (6): 559–572 [2012-01-24]. (原始内容 (PDF)存档于 2013-10-20) .
3. Abdi, H., & Williams, L.J. Principal component analysis.. Wiley Interdisciplinary Reviews: Computational Statistics,. 2010, **2**: 433–459.
4. Shaw P.J.A. (2003) *Multivariate statistics for the Environmental Sciences*, Hodder-Arnold. ISBN 978-0-340-80763-7.
5. Jolliffe I.T. Principal Component Analysis (<http://www.springer.com/west/home/new+%26+forthcoming+titles+%28default%29?SGWID=4-40356-22-2285433-0>) (页面存档备份 (<https://web.archive.org/web/20191016161904/http://www.springer.com/west/home/new+%26+forthcoming+titles+%28default%29?SGWID=4-40356-22-2285433-0>), 存于互联网档案馆), Series: Springer Series in Statistics (<http://www.springer.com/west/home/statistics/statistical+theory+and+methods?SGWID=4-10129-69-173621571-0>) (页面存档备份 (<https://web.archive.org/web/20191016161926/http://www.springer.com/west/home/statistics/statistical+theory+and+methods?SGWID=4-10129-69-173621571-0>), 存于互联网档案馆) , 2nd ed., Springer, NY, 2002, XXIX, 487 p. 28 illus. ISBN 978-0-387-95442-4
6. A. A. Miranda, Y. A. Le Borgne, and G. Bontempi. *New Routes from Minimal Approximation Error to Principal Components* (http://www.ulb.ac.be/di/map/yleborgn/pub/NPL_PCA_07.pdf) (页面存档备份 (https://web.archive.org/web/20180721043245/http://www.ulb.ac.be/di/map/yleborgn/pub/NPL_PCA_07.pdf), 存于互联网档案馆) , Volume 27, Number 3 / June, 2008, Neural Processing Letters, Springer
7. Fukunaga, Keinosuke. *Introduction to Statistical Pattern Recognition*. Elsevier. 1990. ISBN 0122698517.
8. Jonathon Shlens, A Tutorial on Principal Component Analysis. (<http://www.sn1.salk.edu/~shlens/pca.pdf>) (页面存档备份 (<https://web.archive.org/web/20100215191953/http://www.sn1.salk.edu/~shlens/pca.pdf>), 存于互联网档案馆)
9. Geiger, Bernhard; Kubin, Gernot (Sep 2012), Relative Information Loss in the PCA (<http://arxiv.org/abs/1204.0429>) (页面存档备份 (<https://web.archive.org/web/20190514224655/http://arxiv.org/abs/1204.0429>), 存于互联网档案馆)
10. A.A. Miranda, Y.-A. Le Borgne, and G. Bontempi. *New Routes from Minimal Approximation Error to Principal Components* (http://www.ulb.ac.be/di/map/yleborgn/pub/NPL_PCA_07.pdf) (页面存档备份 (https://web.archive.org/web/20180721043245/http://www.ulb.ac.be/di/map/yleborgn/pub/NPL_PCA_07.pdf), 存于互联网档案馆) , Volume 27, Number 3 / June, 2008, Neural Processing Letters, Springer
11. Bessel's correction (https://en.wikipedia.org/wiki/Bessel%27s_correction) (页面存档备份 (https://web.archive.org/web/20201203023311/https://en.wikipedia.org/wiki/Bessel%27s_correction), 存于互联网档案馆) Bessel's correction
12. eig function (<http://www.mathworks.com/access/helpdesk/help/techdoc/ref/eig.html#998306>) (页面存档备份 (<https://web.archive.org/web/20100430173026/http://www.mathworks.com/access/helpdesk/help/techdoc/ref/eig.html#998306>), 存于互联网档案馆) Matlab documentation
13. MATLAB PCA-based Face recognition software. [2012-04-30]. (原始内容存档于 2012-03-09) .
14. Eigenvalues function (<http://reference.wolfram.com/mathematica/ref/Eigenvalues.html>) (页面存档备份 (<https://web.archive.org/web/20140414023625/http://reference.wolfram.com/mathematica/ref/Eigenvalues.html>), 存于互联网档案馆) Mathematica documentation

参考

- Jolliffe, I. T. Principal Component Analysis. Springer-Verlag. 1986: 487 [2012-01-24]. ISBN 978-0-387-95442-4. doi:10.1007/b98835. （原始内容存档于2019-10-16）.

取自“<https://zh.wikipedia.org/w/index.php?title=主成分分析&oldid=64285017>”

本页面最后修订于2021年2月14日 (星期日) 12:15。

本站的全部文字在知识共享 署名-相同方式共享 3.0协议之条款下提供，附加条款亦可能应用。（请参阅使用条款）
Wikipedia®和维基百科标志是维基媒体基金会的注册商标；维基™是维基媒体基金会的商标。
维基媒体基金会是按美国国内税收法501(c)(3)登记的非营利慈善机构。