A Principal Components Decomposition Algorithm is a matrix decomposition algorithm that can be applied by a PCA system (to solve a PCA task to return principal components of a matrix).

- **AKA:** PCA.
- **Context:**
  - In a large data set most of the data are spread around its mean. If we can shift our parameter axes along with this mean point which is of maximum variance then there is a chance that all the hidden nature of the data will be revealed and can be measured through these parameter axes.The PCA provides the algorithm to find the axes around which most of the data are spread.

    [math]\longrightarrow[/math] [math]\longrightarrow[/math]

    Here in the first graph the data are plotted in a 2-dimensional space.Then the coordinate axes [math]X[/math] and [math]Y[/math] are shifted to the eigen axes [math]e_1[/math] and [math]e_2[/math].Again most of the data are spread around the eigen axis [math]e_1[/math] then [math]e_2[/math], so the data can be studied with respect to one axis [math]e_1[/math]. This is called dimensionality reduction.PCA provides the algorithm to perform all such task.
  - It can be used as a Matrix Dimensionality Compression Algorithm.
- **Example(s):**
  - a Covariance-method PCA Algorithm.
  - an EM-based PCA Algorithm (EM algorithm for PCA).
- **Counter-Example(s):**
  - Projection Pursuit.
  - Multidimensional Scaling Algorithm.
  - Singular Value Decomposition Algorithm.
  - Karhunen-Loeve Transform.
- **See:** Linear Combination, Covariance Matrix, Linear Model, Linear Combination, PCA Score.

References

2014

- Sebastian Rashka. (2014). “Implementing a Principal Component Analysis (PCA) in Python step by step." Blog post
  - QUOTE: Both Multiple Discriminant Analysis (MDA) and Principal Component Analysis (PCA) are linear transformation methods and closely related to each other. In PCA, we are interested to find the directions (components) that maximize the variance in our dataset, where in MDA, we are additionally interested to find the directions that maximize the separation (or discrimination) between different classes (for example, in pattern classification problems where our dataset consists of multiple classes. In contrast two PCA, which ignores the class labels). In other words, via PCA, we are projecting the entire set of data (without class labels) onto a different subspace, and in MDA, we are trying to determine a suitable subspace to distinguish between patterns

that belong to different classes. Or, roughly speaking in PCA we are trying to find the axes with maximum variances where the data is most spread (within a class, since PCA treats the whole data set as one class), and in MDA we are additionally maximizing the spread between classes. In typical pattern recognition problems, a PCA is often followed by an MDA.

2012

- http://en.wikipedia.org/wiki/Principal_component_analysis
  - QUOTE: **Principal component analysis (PCA)** is a mathematical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. The number of principal components is less than or equal to the number of original variables. This transformation is defined in such a way that the first principal component has the largest possible variance (that is, accounts for as much of the variability in the data as possible), and each succeeding component in turn has the highest variance possible under the constraint that it be orthogonal to (i.e., uncorrelated with) the preceding components. Principal components are guaranteed to be independent only if the data set is jointly normally distributed. PCA is sensitive to the relative scaling of the original variables. Depending on the field of application, it is also named the discrete **Karhunen–Loève transform** (KLT), the **Hotelling transform** or proper orthogonal decomposition (**POD**).

    PCA was invented in 1901 by Karl Pearson.[1] Now it is mostly used as a tool in exploratory data analysis and for making predictive models. PCA can be done by eigenvalue decomposition of a data covariance (or correlation) matrix or singular value decomposition of a data matrix, usually after mean centering (and normalizing or using Z-scores) the data matrix for each attribute.[2] The results of a PCA are usually discussed in terms of component scores, sometimes called factor scores (the transformed variable values corresponding to a particular data point), and loadings (the weight by which each standardized original variable should be multiplied to get the component score).[3]

    PCA is the simplest of the true eigenvector-based multivariate analyses. Often, its operation can be thought of as revealing the internal structure of the data in a way that best explains the variance in the data. If a multivariate dataset is visualised as a set of coordinates in a high-dimensional data space (1 axis per variable), PCA can supply the user with a lower-dimensional picture, a "shadow" of this object when viewed from its (in some sense) most informative viewpoint. This is done by using only the first few principal components so that the dimensionality of the transformed data is reduced.

    PCA is closely related to factor analysis. Factor analysis typically incorporates more domain specific assumptions about the underlying structure and solves eigenvectors of a slightly different matrix.

1. Pearson, K. (1901). "On Lines and Planes of Closest Fit to Systems of Points in Space" (PDF). *Philosophical Magazine* **2** (6): 559–572. http://stat.smmu.edu.cn/history/pearson1901.pdf.

2. Abdi. H., & Williams, L.J. (2010). "Principal component analysis.". *Wiley Interdisciplinary Reviews: Computational Statistics,* **2**: 433–459.

3. Shaw P.J.A. (2003) *Multivariate statistics for the Environmental Sciences*, Hodder-Arnold. ISBN 0-340-80763-6.

2011

- (Sammut & Webb, 2011) ⇒ Claude Sammut, and Geoffrey I. Webb. (2011). “Principal Component Analysis.” In: (Sammut & Webb, 2011) p.795

2009

- http://www.statistics.com/resources/glossary/p/pca.php
  - QUOTE: The purpose of principal component analysis is to derive a small number of linear combinations (principal components) of a set of variables that retain as much of the information in the original variables as possible. This technique is often used when there are large numbers of variables, and you wish to reduce them to a smaller number of variable combinations by combining similar variables (ones that contain much the same information).

    Principal components are linear combinations of variables that retain maximal amount of information about the variables. The term "maximal amount of information" here means the best least-square fit, or, in other words, maximal ability to explain variance of the original data.

    In technical terms, a principal component for a given set of N-dimensional data, is a linear combination of the original variables with coefficients equal to the **components** of an eigenvector of the correlation or covariance matrix. Principal components are usually sorted by descending order of the eigenvalues - i.e. the first principal component corresponds to the eigenvector with the maximal eigenvalue.

- (Johnstone & Lu, 2009) ⇒ Iain M Johnstone, and Arthur Yu Lu. (2009). "On Consistency and Sparsity for Principal Components Analysis in High Dimensions." doi:10.1198/jasa.2009.0121
  - QUOTE: Suppose [math]{x_i, i=1,…,n}[/math] is a dataset of [math]n[/math] observations on [math]p[/math] variables. Standard principal components analysis (PCA) looks for vectors ξ that maximize :[math]var(ξTxℓ.)/‖ξ‖2. (1)[/math] If [math]ξ_1, …, ξ_k[/math] have already been found by this optimization, then the maximum defining <math>ξ_{k+1}<math> is taken over vectors ξ orthogonal to <math>ξ1, …, ξk<math>.

2006

- (Hinton & Salakhutdinov, 2006) ⇒ Geoffrey E. Hinton, and Ruslan R. Salakhutdinov. (2006). "Reducing the Dimensionality of Data with Neural Networks." In: Science, 313(5786). doi:10.1126/science.1127647
  - QUOTE: … A simple and widely used method is principal components analysis (PCA), which finds the directions of greatest variance in the data set and represents each data point by its coordinates along each of these directions. We describe a nonlinear generalization of PCA that uses an adaptive, multilayer "encoder" network.

2002

- (Jollife, 2002) ⇒ Ian T. Jolliffe. (2002). "Principal Component Analysis, 2nd ed." Springer. ISBN:0-387-95442-2
  - QUOTE: Principal component analysis is central to the study of multivariate data. Although one of the earliest multivariate techniques it continues to be the subject of much research, ranging from new model-based approaches to algorithmic ideas from neural networks. It is extremely versatile with applications in many disciplines. …

- (Fodor, 2002) ⇒ Imola K. Fodor. (2002). "A Survey of Dimension Reduction Techniques." LLNL technical report, UCRL ID-148494
  - QUOTE: Principal component analysis (PCA) is the best, in the mean-square error sense, linear dimension reduction technique [25, 28]. Being based on the covariance matrix of the variables, it is a second-order method. In various fields, it is also known as the singular value decomposition (SVD), the Karhunen-Loμeve transform, the Hotelling transform, and the empirical orthogonal function (EOF) method.

In essence, PCA seeks to reduce the dimension of the data by finding a few orthogonal linear combinations (the PCs) of the original variables with the largest variance. The first PC, s1, is the linear combination with the largest variance.

1999

- (Tipping & Bishop, 1999) ⇒ Michael E. Tipping, and Christopher M. Bishop. (1999). “Probabilistic Principal Component Analysis].” In: Journal of the Royal Statistical Society, 61(3).

1901

- (Pearson, 1901) ⇒ K. Pearson. (1901). “On Lines and Planes of Closest Fit to Systems of Points in Space" In: Philosophical Magazine, 2(11). doi:10.1080/14786440109462720.
