Principal components analysis research paper examples

Science, Statistics

Principal concept analysis may be defined as the statistical analysis tool relevant for the transformation of detailed or complex data into simple and easy to understand data. The method focuses on the use of principal components of statistics for possible interpretations. In most cases, the method does not include statistical assumptions. However, the method will always give extremely significant results for data that contains multivariate normal distribution. The data is extensively significant in determining the hypothetical variables or components that are extensively useful in the multidimensional data. In other words, the method PCA is a method significant in projecting one-dimensional variable onto a two-dimensional plane with minimum loss of variance (MORPHOMETRICS 86).

The main purpose for PCA is to reorganize extensively garbled data set. The method aims at filtering the problem with the data set and tries to come up with the hidden dynamics. For example, the x-axis may possess some dynamics, which may only be relevant if the significance in them is brought up. The determination is extensively significant in defining whether the dynamics are significant or are redundant (Shlens 05).

PCA may be defined in terms of standardization. If the PCA does not have some standardization, it may be referred to as PCA on the variance-covariance matrix. If it has some standardization, it may be referred to as the PCA on the correlation matrix. While carrying out the PCA, it may be advisable to remove the effect of the overall size of every specimen. In that case, one of the preferable techniques is the application of the row normalization with the sum of squares of the variants for the specimen forced to one. A special form of PCA is highly essential in the analysis of

compositional data.

Every component works with its latent root of the eigenvalue, which shows the relative value of the overall variance that the component describes. In this case, the principal components are provided in order of diminishing eigenvalues. The main idea of PCA is to discover the tendency for the data to be concentrated in a low-dimensional space, showing the availability of some low-dimensional space that shows the level of correlation between variables. At some point, the PCA employs some statistical properties of the data set like multivariate normality. Violation of the principles would degrade the explanation ability or potential of the axes. Like other indirect ordination methods, PCA is a descriptive as well as explorative method that has no statistical importance.

The reliability of the PCA, highly depends on the understanding for the extensions that make it work accordingly. The first element is linearity. In statistics, linearity describes the problem as a change of basis. Nonlinearity consideration defines nonperformance of the PCA thus extending the algorithm and the element have been described as Kernel PCA (MORPHOMETRICS 89).

The second element that is equally fundamental is that mean and variance are sufficient statistics. The formality of statistics shows that sufficient statistics defines the notion that the mean and the variance are the complete definitions of a probability distribution. The only statistics that is fully defined by variance alone is the Gaussian distribution (Shlens 13).

The third element for the definition of the PCA is that large variances have important dynamics. The assumption entails the assumption that the data

has an extensively high signal to noise ratio (SNR). Therefore, principal components with more related variances stand for interesting dynamics, with the ones with lower variances standing for the noise (Shlens 13).

The fifth assumption is that the principal components are orthogonal. The assumption is extensively simple through a technique that shows PCA as a necessary problem that is solvable with the decomposition of linear algebra techniques (Shlens 13).

The PCA solution mainly works with the decomposition of the linear algebra. This mainly works with the simple consideration of the eigenvector as the element for decomposition. The person dealing with the decomposition has to consider the data set X, and an m × n matrix, with m being the number of measurement types and n number of data trials. The algebraic equation works with the consideration of all the principles necessary for an outstanding equation. The necessary principles for the matrix are defined accordingly (Shlens 02).

PCA is exclusively useful in computational biology through the use of high-dimensional data sets. In most cases, three-dimensional visualizations are relevant for similar explorations with the samples plotted according to the correlation with the components. Since two or three dimensional visualizations may lose a lot of information, it is essential to try various combinations when visualizing data set. Since the principal components are uncorrelated, they may stand for different aspects of the samples. PCA can serve a crucial role before the classification of samples is carried out. The SVD is extensively applicable in when gene tests are being carried out. The application process includes the identification of patterns that correlate with

the experimental artifacts and filtering the out to estimate the missing data, associating genes, as well as expression patterns with set of regulators as well as helping in uncovering the dynamic architecture of cellular phenotypes (Ringnér 304).

Works Cited

" MORPHOMETRICS." MORPHOMETRICS. 4 Mar. 2010. Web. 9 Dec. 2014. .

Top of Form

Ringnér, Markus. " What Is Principal Component Analysis?" Computational Biology 26. 3 (2008): 303-04. Print.

Shlens, Jon. " A TUTORIAL ON PRINCIPAL COMPONENT ANALYSIS Derivation, Discussion and Singular Value Decomposition." JonShelns 1. 1 (2003): 1-15. Print.