

UCLA Department of Statistics

History of Nonlinear Principal Component Analysis

Jan de Leeuw

February 6, 2011



Abstract

- *Multiple Correspondence Analysis (MCA)* is discussed as a form of *Nonlinear Principal Component Analysis (NLPCA)*.
- It is compared with other forms of NLPCA that have been proposed over the years: Shepard-Kruskal- Breiman-Friedman-Gifi PCA with optimal scaling, aspect analysis of correlations, Guttman's MSA, Logit and Probit PCA of binary data, and Logistic Homogeneity Analysis.



Linear PCA

History

- (Linear) Principal Components Analysis (PCA) is sometimes attributed to Hotelling (1933), but that is surely incorrect.
- The equations for the principal axes of quadratic forms and surfaces, in various forms, were known from classical analytic geometry.
- There are some modest beginnings in Galton's *Natural Inheritance* of 1889, where the principal axes are connected for the first time with the "correlation ellipsoid".
- There is a full-fledged (although tedious) discussion of the technique in Pearson (1901), and there is a complete application (7 physical traits of 3000 criminals) in MacDonell (1902), by a Pearson co-worker.
- There is proper attribution in: Burt, C., *Alternative Methods of Factor Analysis and their Relations to Pearson's Method of "Principle Axes"*, Br. J. Psych., Stat. Sec., 2 (1949), pp. 98-121.



Linear PCA

How To

- Hotelling's introduction of PCA follows the now familiar route of making successive orthogonal linear combinations with maximum variance. He does this by using Von Mises (Power) iterations, discussed in 1929 by Von Mises and Pollaczek-Geiringer.
- Pearson, following Galton, used the correlation ellipsoid throughout. This seems to me the more basic approach.
- He cast the problem in terms of finding low-dimensional subspaces (lines and planes) of best (least squares) fit to a cloud of points, and connects the solution to the principal axes of the correlation ellipsoid.
- In modern notation, this means minimizing $\mathbf{SSQ}(Y - XB')$ over $n \times r$ matrices X and $m \times r$ matrices B . For $r = 1$ this is the best line, etc.



Correspondence Analysis

History

- Simple Correspondence Analysis (CA) of a bivariate frequency table was first discussed, in fairly rudimentary form, by Pearson (1905), by looking at transformations linearizing regressions. See De Leeuw, *On the Prehistory of Correspondence Analysis*, *Statistica Neerlandica*, 37, 1983, 161–164.
- This was taken up by Hirshfeld (Hartley) in 1935, where the technique was presented in a fairly complete form (to maximize correlation and decompose contingency). This approach was later adopted by Gebelein, and by Renyi and his students in their study of maximal correlation.



Correspondence Analysis

History

- In the 1938 edition of *Statistical Methods for Research Workers* Fisher scores a categorical variable to maximize a ratio of variances (quadratic forms). This is not quite CA, because it is presented in an (asymmetric) regression context.
- Symmetric CA and the reciprocal averaging algorithm are discussed, however, in Fisher (1940) and applied by his co-worker Maung (1941a,b).
- Then in the early sixties the chi-square distance based form of CA, relating CA to metric multidimensional scaling (MDS), with an emphasis on geometry and plotting, was introduced by Benzécri (thesis of Cordier, 1965).



Multiple Correspondence Analysis

History

- Different weighting schemes to combine quantitative variables to an index that optimizes some variance-based discrimination or homogeneity criterion were proposed in the late thirties by Horst (1936), by Edgerton and Kolbe (1936), and by Wilks (1938).
- The same idea was applied to quantitative variables in a seminal paper by Guttman (1941), that presents, for the first time, the equations defining *Multiple Correspondence Analysis (MCA)*.
- The equations are presented in the form of a row-eigen (scores), a column-eigen (weights), and a singular value (joint) problem.
- The paper introduces the “codage disjonctif complet”, the “Tableau de Burt”, and points out the connections with the chi-square metric.
- There is no geometry, and the emphasis is on constructing a single scale. In fact Guttman warns against extracting and using additional eigen-pairs.



Multiple Correspondence Analysis

Further History

- In Guttman (1946) scale or index construction was extended to paired comparisons and ranks. In Guttman (1950) it was extended to scalable binary items.
- In the fifties and sixties Hayashi introduced the quantification techniques of Guttman in Japan, where they were widely disseminated through the work of Nishisato. Extensions and variations were added
- Starting in 1968, MCA was studied as a simple form of metric MDS by De Leeuw.
- Although the equations defining MCA were the same as those defining PCA, the relationship between the two remained problematic.
- These problems are compounded by “horse shoes” or the “effect Guttman”, i.e. artificial curvilinear relationships between successive dimensions (eigenvectors).



Nonlinear PCA

What ?

PCA can be made non-linear in various ways.

- 1 First, we could seek indices which discriminate maximally and are non-linear combinations of variables. This generalizes the weighting approach (Hotelling).
- 2 Second, we could find nonlinear combinations of components that are close to the observed variables. This generalizes the reduced rank approach (Pearson).
- 3 Third, we could look for transformations of the variables that optimize the linear PCA fit. This is known (term of Darrell Bock) as the *optimal scaling (OS)* approach.



Nonlinear PCA

Forms

- The first approach has not been studied much, although there are some relations with Item Response Theory.
- The second approach is currently popular in Computer Science, as “nonlinear dimension reduction”. I am currently working on a polynomial version, but there is not unified theory, and the papers are usually of the “*well, we could also do this*” type familiar from cluster analysis.
- The third approach preserves many of the properties of linear PCA and can be connected with MCA as well. We shall follow its history and discuss the main results.



Nonlinear PCA

PCA with OS

- Guttman observed in 1959 that if we require that the regression between monotonically transformed variables are linear, then the transformations are uniquely defined. In general, however, we need approximations.
- The loss function for PCA-OS is $\mathbf{SSQ}(Y - XB')$, as before, but now we minimize over components X , loadings B , and transformations Y .
- Transformations are defined column-wise (over variables) and belong to some restricted class (monotone, step, polynomial, spline).
- Algorithms often are of the *alternating least squares* type, where optimal transformation and low-rank matrix approximation are alternated until convergence.



PCA-OS

History of programs

- Shepard and Kruskal used the monotone regression machinery of non-metric MDS to construct the first PCA-OS programs around 1962. The paper describing the technique was not published until 1975.
- Around 1970 versions of PCA-OS (sometimes based on Guttman's rank image principle) were developed by Lingoes and Roskam.
- In 1973 De Leeuw, Young, and Takane started the ALSOS project, which resulted in PRINCIPALS (published in 1978), and PRINQUAL in SAS.
- In 1980 De Leeuw (with Heiser, Meulman, Van Rijckevorsel, and many others) started the Gifi project, which resulted in PRINCALS, in SPSS CATPCA, and in the R package homals by De Leeuw and Mair (2009).
- In 1983 Winsberg and Ramsay published a PCA-OS version using monotone spline transformations.
- In 1987 Koyak, using the ACE smoothing methodology of Breiman and Friedman (1985), introduced mdrace.



PCA/MCA

The Gifi Project

