QUASI IMAGE TRANSFORMATION OF A SET OF STANDARDIZED QUANTITATIVE VARIABLES

Konstantin Momirovic
Institute for criminological and sociological research
Belgrade

Quasi image transformation of a set of standardized quantitative variables is defined as such a transformation that maximizes the covariances between original and so transformed variables. Quasi antiimage variables are defined as the differences between original and quasi image variables. Relations among quasi image, quasi antiimage, image, antiimage and variables in the universal metric space are defined, as well as the relations of quasi image variables with principal components and variables transformed to Mahalanobis form. Some interesting properties of quasi image transformation are discussed and their application to distance measures are mentioned.

KEY WORDS

image variables/antiimage variables/quasi image variables/
quasi antiimage variables/principal components/
univesal metrics/Mahalanobis form

0. INTRODUCTION

Image transformation of a set of quantitative variables (Guttman, 1953) can be defined as the transformation of each variable from the set so that the transformed variable, obtained as linear combination of other variables in the set, is as much similar, in the least squares sense, to the original variable. This is, of course, equivalent to the maximization of (multiple) correlation of original variable and transformed variable, usually refered to as image variable.

Instead of maximization of correlation, Stalec and Momirovic (1983) proposed the maximization of covariance between a variable and linear combination of other variables in the set. This leads to a robust method for regression analysis, somethimes called "stupid regression analysis", with some very desirable properties (Dobric, Stalec and Momirovic, 1984; Dobric, 1986; Dobric and Momirovic, 1991). Among the most important are little sensitivity to outliers, insensitivity to singularity or near singularity of the correlation matrix of variables in the set, and higher generalizability of so obtained regression estimates.

The aim of this paper is to propose a similar transformation of the whole set of variables, and to demonstrate some properties of this transformation, including the relationships with other transformations, such as image and antiimage transformation, transformation to principal components, and transformation of variables to Mahalanobis form.

1. DEFINITIONS

Let $\mathbf{Z} = (z_{ij})$, $i = 1,...,n$; $j = 1,...,m$; $m<n$ be a data matrix, in standard normal form, obtained by the description of a set of objects $E = (e_i, i = 1,...,n)$ on a set of quantitative variables $V = (v_j, j = 1,...,m)$. Then $\mathbf{R} = \mathbf{Z'Z}$ will be the intercorrelation matrix of variables from V obtained on E, due to the fact that vector variables of $\mathbf{Z}$ are centered and normed to 1.

In accordance with the derivations in Stalec and Momirovic (1983) and Dobric, Stalec and Momirovic (1984) define

$$\mathbf{Q} = \mathbf{R} - \mathbf{I},$$

$\mathbf{I}$ denoting identity matrix of order m, as the matrix with vectors proportional to vectors which transform a set of variables to variables with maximized covariances with original variables.

Brutto quasi image transformation (QUIT) is defined by

$$\mathbf{P} = \mathbf{ZQ} = \mathbf{Z}(\mathbf{R} - \mathbf{I}) = \mathbf{ZR} - \mathbf{Z}$$

with the obvious properties

$$\mathbf{Z'P} = \mathbf{R}(\mathbf{R} - \mathbf{I}) = \mathbf{R}^2 - \mathbf{R}$$

and

$$\text{trace}(\mathbf{R}^2 - \mathbf{R}) = \text{maximum}.$$

Brutto quasi antiimage transformation is defined by

$$\mathbf{F} = \mathbf{Z} - \mathbf{P} = 2\mathbf{Z} - \mathbf{ZR} = \mathbf{Z}(2\mathbf{I} - \mathbf{R}) = \mathbf{ZD}$$

with

$$\mathbf{D} = (2\mathbf{I} - \mathbf{R}),$$

with property

$$Z'F = R(2I - R) = RD.$$

The covariance matrix of quasi image variables is

$$P'P = QRQ = (R - I)R(R - I) = (R^2 - R)(R - I)$$

and the covariance matrix of quasi antiimage variables is

$$F'F = DRD = (2I - R)R(2I - R) = (2R - R^2)(2I - R).$$

Finally, the crosscovariance matrix between quasi image and quasi antiimage variables is

$$P'F = QRD = (R - I)R(2I - R) = (R^2 - R)(2I - R).$$

2. RELATIONS OF QUASI IMAGE AND QUASI ANTIIMAGE VARIABLES WITH IMAGE, ANTIIMAGE AND VARIABLES IN UNIVERSAL METRIC SPACE

Image variables are defined by (Guttman, 1953; Kaiser, 1963)

$$G = Z(I - R^{-1}U^2)$$

with

$$U^{-2} = \mathrm{diag}\ R^{-1}$$

and antiimage variables are defined by

$$A = Z - G = ZR^{-1}U^2.$$

Variables in universal metric space are defined by (Harris, 1962)

$$H = ZU^{-1}.$$

Relations of these variables with quasi image and quasi antiimage are presented:

$$P'G = (R - I)(R - U^2),$$

$$P'A = (R - I)U^2,$$

$$P'H = (R^2 - R)U^{-1},$$

$$F'G = (2I - R)(R - U^2),$$

$$F'A = (2I - R)U^2,$$

$$F'H = (2R - R^2)U^{-1}.$$

Note the following important properties:

(1) $\text{diag}\,((R - I)(R - U^2)) = \text{diag}\,R^2 - I$

(2) $\text{diag}\,(R - I)U^2 = 0$

(3) $\text{diag}\,((R^2 - R)U^{-1}) = \text{diag}\,(R^2 - I)U^{-1}$

(4) $\text{diag}\,((2I - R)(R - U^2) = \text{diag}\,R^2 - U^2$

(5) $\text{diag}\,(2I - R)U^2 = U^2$

(6) $\text{diag}\,(2R - R)U^{-1} = \text{diag}\,(2I - R^2)U^{-1}$.

3. RELATIONS OF QUASI IMAGE VARIABLES WITH PRINCIPAL COMPONENTS AND VARIABLES TRANSFORMED TO MAHALANOBIS FORM

Let be

$$Z = YLX'$$

the basic structure of original variables in standard normal form, with Y, $Y'Y = I$ the matrix of left eigenvectors, X, $X'X = XX' = I$ the matrix of right eigenvectors, and L the diagonal matrix of singular values.

Then quasi image variables have the structure

$$P = YLX'(XLX' - I),$$

the (nonstandardized) principal components have the structure

$$K = ZX = YL$$

and the variables transformed to Mahalanobis form have the structure

$$M = ZR^{-1/2} = YX'.$$

The relations of quasi image variables with principal components and Mahalanobis variables are

$$P'K = X(L^4 - L^2)$$

and

$$P'M = X(L^3 - L)X'.$$

4. NOTES AND COMMENTS

(1) Both quasi image and quasi anti image variables lie in the same space as the standardized variables because Z'Z, P'P and F'F can be diagonalized in the same base, defined by the right eigenvectors of Z.

(2) Crosscovariance matrix of quasi image and quasi anti image variables is a symmetric matrix.

(3) Crosscorrelation matrices of quasi image variables with original variables and variables transformed to Mahalanobis form are symmetric matrices.

(4) Relation among original, quasi image, principal components and Mahalanobis variables are the simple functions of singular values and right eigenvectors of original data matrix.

Therefore, the Minkowski distances between objects in quasi image space are metric distances, and can be used for any classification algorithm.

REFERENCES

(1) Dobric, V.; Stalec, J.; Momirovic, K. (1984):
Note on some relationships between least squares and robust linear regression estimates.
Proc. 6th International Symposium "Computer at the University", 507: 1-7.

(2) Dobric, V. (1986):
On a class of robust methods for multivariate data analysis.
COMPSTAT 86, Physica-Verlag, Heidelberg, 211-216.

(3) Dobric, V.; Momirovic, K. (1991):
Some relations between least squares and robust linear regression estimates.
Information Technology Interfaces, 5: 1-9.

(4) Guttman, L. (1953):
Image theory for the structure of quantitative variates.
Psychometrika, 18: 227-296.

(5) Harris, C.W. (1962):
Some Rao-Guttman relationships.
Psychometrika, 27: 247-263.

(6) Kaiser, H.F. (1963):
Image analysis. In C.W. Harris, "Problems in measuring change".
University of Wisconsin Press, Madison.

(7) Stalec, J.; Momirovic, K. (1983):
Some properties of a very simple model for robust regression analysis.
Proc. 5th International Symposium "Computer at the University", 453-461.