Journal of Statistical
Theory & Methods
1998, Vol. 9, No. 1, pp. 63 ~ 70

A Study of Singular Value Decomposition in Data Reduction techniques

Yang Kyu Shin [1]

Abstract

The singular value decomposition is a tool which is used to find a linear structure of reduced dimension and to give interpretation of the lower dimensional structure about multivariate data. In this paper the singular value decomposition is reviewed from both algebraic and geometric point of view and, is illustrated the way which the tool is used in the multivariate techniques finding a simpler geometric structure for the data.

Key Words and Phrases: Dimension reduction, Multidimensional scaling, Principal component analysis, Singular value decomposition, Spectral decomposition.

1. Introduction

A vast majority of data in social science is multivariate and a wide variety of multivariate techniques - principal component analysis, factor analysis, multidimensional scaling, etc. - are available to analyze such multivariate data. The aim of these techniques is to find a simple geometric structure among data points which would either reduce the dimensionality or suggest a possible internal relationships among units or variables(Murtagh and Heck(1987)).

The singular value decomposition is the tool being used in these multivariate techniques for those purposes(Gnanadesikan(1977)). Nishisato and Shizuhiko(1996) used the dual space property of the singular value decomposition for gleaning in the field of dual scaling. Choi and Huh(1996) derived the resistant version of the singular value decomposition for principal component analysis. Moreover the singular value decomposition of the data matrix is computationally far more efficient than the spectral decomposition of the sample covariance matrix when the number

[1]Dept. of Information Statistics, Kyungsan University, Kyungpook, 712-240, Korea.

of variable is large. Section 2 reviews the singular value decomposition form both algebraic and geometric point of view. In section 3 we illustrate the way the singular value decomposition is used in the multivariate techniques, especially in the principal component analysis and the mutidimensional scaling. Finally we conclude the properties of the singular value decomposition in multivariate techniques.

2. Singular value decomposition of a data matrix

2.1 Algebraic analysis about singular value decomposition

Let X be a $p \times n$ data matrix and consider the grammians XX^t and X^tX. Then the eigenvalues of XX^t and X^tX are nonnegative. Moreover the eigenvalues of XX^t are same as those of X^tX. Suppose $a_1, \cdots, a_k$ are the orthonormal eigenvectors of XX^t and that $a_1, \cdots, a_k$ correspond to the positive eigenvalues of ${\lambda_1}^2, \cdots, {\lambda_k}^2$ of XX^t. Then $\{a_1, \cdots, a_k\}$ is an orthonormal basis of the sample space. Hence data matrix X can be written as

$$\begin{aligned} X &= (X^t a_1)a_1^t + \cdots + (X^t a_k)a_k^t + (X^t a_{k+1})a_{k+1}^t + \cdots + (X^t a_p)a_p^t \qquad (1) \\ &= \Big[a_1(a_1^t X) + \cdots + a_k(a_k^t X)\Big] + \Big[a_{k+1}(a_{k+1}^t X) + \cdots + a_p(a_p^t X)\Big] \\ &= (a_1 a_1^t + \cdots + a_k a_k^t)X + (a_{k+1}a_{k+1}^t + \cdots + a_p a_p^t)X \\ &= (a_1 a_1^t + \cdots + a_k a_k^t)X \end{aligned}$$

because for $j = k+1, \cdots, p, \quad XX^t a_j = 0$

For $j = 1, \cdots, k$

$$\begin{aligned} XX^t a_j &= \lambda_j^2 a_j \\ a_j^t XX^t a_j &= \lambda_j^2 a_j^t a_j \\ \|a_j^t X\| &= \lambda_j > 0 \end{aligned}$$

Define $\{c_1, \cdots, c_k\}$ such that

$$\lambda_j c_j = X^t a_j \ \ \textbf{for} \ \ j = 1, \cdots, k \qquad (2)$$

From (2)

$$\lambda_i \lambda_j c_i^t c_j = \ a_i^t XX^t a_j = \lambda_j^2 a_i^t a_j$$

Thus

$$c_i^t c_j = \begin{cases} 1 \ \text{if} \ i{=}j \\ 0 \ \text{if} \ i \neq j \end{cases} \qquad 1 \leq i, \ j \leq k$$

On the other hand

$$\lambda_j X^t X c_j = X^t(XX^t a_j) = X^t(\lambda_j^2) = \lambda_j^2 \lambda_j c_j$$

Therefore $X^t X c_j = \lambda_j^2 c_j$
Hence $c_1, \cdots, c_k$ are the orthonormal eigenvectiors of $X^t X$ correspond to the k positive eigenvalues $\lambda_1^2, \cdots, \lambda_k^2$ *of* $X^t X$.

By (2), (1) is written

$$X = \lambda_1 a_1 c_1^t + \cdots + \lambda_k a_k c_k^t \quad (3)$$

(3) is called the singular value decompostion of X. The numbers $\lambda_1, \cdots, \lambda_k$ are the singular values of X and the vectors $a_1, \cdots, a_k$ and $c_1, \cdots, c_k$ are the right and left singular vectors of X. Shin(1982) compared the singular value decomposition of the data matrix and the spectral decomposition of the sample covariance matrix. The singular value decomposition is more powerful than commonly used spectral decomposition.

2.2 Geometric analysis about singular value decomposition

The singular value decomposition of X provides an immediate analysis of the effect of X regarded as a linear transformation acting on the vectors of Euclidean n-space E_n and p-space E_p resp.. For any such vector a is of the form

$$a = \beta_1 a_1 + \cdots + \beta_k a_k + \beta_{k+1} a_{k+1} + \cdots + \beta_p a_p$$

where $a_{k+1}, \cdots, a_p$ completes the orthonormal basis of E_p, and

$$X^t a = \beta_1 \lambda_1 c_1 + \cdots + \beta_k \lambda_k c_k$$

Thus a is first projected into the k-manifold spanned by $a_1, \cdots a_k$, then the k co-ordinate are scaled by factors $\lambda_1, \cdots, \lambda_k$, and finally the resulting vector is pictured with the same co-ordinates in the k-manifold spanned by $c_1, \cdots, c_k$.

If q is a positive integer,

$$(XX^t)^q = \lambda_1^{2q} a_1 a_1^t + \cdots + \lambda_k^{2q} a_k a_k^t$$
$$(X^t X)^q = \lambda_1^{2q} c_1 c_1^t + \cdots + \lambda_k^{2q} c_k c_k^t$$
$$(XX^t)^q X = \lambda_1^{2q+1} a_1 c_1^t + \cdots + \lambda_k^{2q+1} a_k c_k^t$$
$$(X^t X)^q X^t = \lambda_1^{2q+1} c_1 a_1^t + \cdots + \lambda_k^{2q+1} c_k a_k^t$$

Then $(XX^t)^q$ and $(X^tX)^q$ are matrices representing the projection of p and n space into the k-manifolds spanned by $(a_1, \cdots, a_k)$ and $(c_1, \cdots, c_k)$ resp.. By means of generalized matrix inverses, we can obtain the equations for negative integers q. The iterative procedure for calculating the singular value decomposition of data matrix is easy to program for a computer and will usually be numerically stable when only the first few terms of the singular value decomposition are required. If all the terms are wanted then analogues of many of the comments and techniques in Wilkinson(1965), chapter 9, would be relevant. The iterative method is especially pertinent for large sparse matrices.

3. Singular value decomposition in data reduction techniques

3.1 Singular value decomposition in principal component analysis

Main concern of principal component analysis is the recognition of lower dimensional linear subspaces which the multi-responses observations may, statistically, lie. The basic idea of principal component analysis is to describe the dispersion of an array of n point in p demensional space by introductory a new set of orthogonal linear co-ordinates so that the sample variances of the given points which respect to these derived co-cordinates are in decreasing order of magnitude(Jackson and Hearne(1975)). The first principal component is such that the projections of the given points onto it have maximum variance among all possible linear co-dordinates ; the second principal component has maximum variance subject to being orthogonal to the first ; and so on. We hope that the first principal component will be counted for most of the variation in the original data so that the effective dimensionality of the data can be reduced. By centring of the original data, we can have the origin as mean of data. Suppose $X = (X_1, \cdots, X_p)$ is a p-dimensional random variable with mean 0 and covariance matrix $\sum(= XX^t)$. A $(p \times n)$ dimensional random matrix X represent a multivariate statistical sample of n observations on p variable. Then the jth principal component is the eigenvector associated with the jth largest eigenvalue λ_j of covariance matrix $\sum$. If $\lambda_i \neq \lambda_j$ the elements can be chosen to be orthogonal, although an infinity of such orthogonal vectors exists. Let us denote the $(p \times p)$ matrix of eigenvectors by A, where $A = (a_1, \cdots, a_p)$ and the $(p \times 1)$ vector of principal components by Y, then $Y = A^tX$. The $(p \times p)$ covariance matrix of Y be denote by $\bigwedge$ and is clearly given by

$$\wedge = \begin{pmatrix} \lambda_1 & \cdot & \cdot & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot \\ \cdot & & \cdot & \cdot \\ 0 & \cdot & \cdot & \lambda_p \end{pmatrix}$$

Hence

$$\sum = \sum_{i=1}^{p} \lambda_i a_i a_i^t$$

From above results principal component analysis is related the spectral decomposition of the covariance matrix. Since the spectral decomposition is the special case of the singular value decomposition, we can say that principal component is related to the singular value decompositon of the data matrix X.

The equations, $X^t \sum^{-1} X = c$, for arrange of non-negative value of c define a family of concentric ellipsoides in the p dimensional space of X. The principal component transformation of the data is just the projections of the observations onto the principal axes of this family. Although the eigenvector corresponding to the largest eigenvalue provides the projecton of each point onto the first principal component, the equation of the first principal component co-ordinate is given by the conjunction of the equations of planes defined by the remaining eigenvectors. More generally, most of the variability of a p-dimensional sample is confined $(p-q)$ dimensional linear subspace, that subspace is described by the $(p-q)$ eigenvectors which correspond to the $(p-q)$ small eigenvalues.

By eigenanalysis, we judge the relative magnitudes of the eigenvalues, both for isloating "negligibly small" once and for inferring groupings, if any, among the others. Thus by the singular value decomposition of the data matrix, we can have the variational interpretation of the data.

3.2 Singular value decomposition in multidimensional scaling

Given a set of observed measures of similarity or dissimilarity between every pair of n objects, find a representation of the objects as points in Euclidean space such that the inter-point distances in some sense "match" the observed similarity or dissimilarity. Multidimensional scaling aims to find a configuration in a much smaller number of dimensions which approximately reproduce the given dissimilarities. Suppose the data matrix X have the exact co-ordinate of n points in p-dimensional Euclidean space. Let $B_{n\times n} = X^t X$, where (r, s) term of B is given by

$$b_{rs} = \sum_{j=1}^{p} X_{rj} X_{sj} \tag{4}$$

The $(n \times n)$ matrix of squared Euclidean distance, $D = (d_{rs})$, is then such that

$$\begin{aligned} d_{rs} &= \textstyle\sum_{j=1}^{p} (x_{rj} - x_{sj})^2 \\ &= b_{rr} + b_{ss} - b_{rs} \end{aligned} \tag{5}$$

In mutidimensional scaling, consider the inverse problem. Suppose we know the distance but not the co-ordinate. So we want to estimate the co-ordinate. First we find the B matrix. From (4) and (5)

$$b_{rs} = -\frac{1}{2}[d_{rs}^2 - d_{r.}^2 - d_{.s}^2 + d_{..}^2]$$

where
$$d_{r.}^2 = \text{ average term in } r\text{th row}$$
$$d_{.s}^2 = \text{average term in } s\text{th column}$$
$$d_{..}^2 = \text{overall average squared distances.}$$

Since D consists of the squares of exact Euclidean distances, B is a positive symmetric matrix. Suppose rank of B is k where $k \le n$. Then B will have k nonzero eigenvalues which we arrange in the order of magnitude so that $\lambda_1 \ge \cdots \ge \lambda_k > 0$. The corresponding eigenvectors of unit length will be denoted by $\{c_i\}$. To scale the eigenvectors so that their sum of squares is equal to λ_i, we set

$$e_i = \sqrt{\lambda_i} c_i$$

By the Young-Hausehold factorization theorem, a positive semidefinite matrix B can be factorized into the form X^tX. Let $X^t = (e_1, \cdots, e_k)$, then

$$(X^tX)c_i = X^t \begin{pmatrix} e_1^t \\ \cdot \\ \cdot \\ \cdot \\ e_i^t \\ \cdot \\ \cdot \\ \cdot \\ e_k^t \end{pmatrix} c_i = X \begin{pmatrix} 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \\ \sqrt{\lambda_i} \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{pmatrix} = \sqrt{\lambda_i} e_i = \lambda_i c_i = Bc_i$$

Since the $\{c_i\}$ forms an orthonormal basis, $B = X^tX$. Hence the co-ordinate of the rth point or individual are given by the rth component of the $\{e_i\}$. Let we denote the $(k \times k)$ matrix of eigenvectors of B by C, where $C = (c_1, \cdots, c_k)$.

Then

$$B = X^tX$$

$$= \begin{pmatrix} e_1, & \cdots & ,e_k \end{pmatrix} \begin{pmatrix} e_1 \\ \cdot \\ \cdot \\ \cdot \\ c_k \end{pmatrix}$$

$$= \begin{pmatrix} c_1, & \cdots & ,c_k \end{pmatrix} \begin{pmatrix} \lambda_1 & \cdots & 0 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 0 & \cdots & \lambda_k \end{pmatrix} \begin{pmatrix} c_1 \\ \cdot \\ \cdot \\ \cdot \\ c_k \end{pmatrix}$$

Therefore, multidimensional scaling is related to the singular value decomposition of the co-ordinate matrix B.

4. Conclusion

By the sngular value decomposition of a data matrix X, we can find the linear structure in data reduction techniques such as principal component analysis and multidimensional scaling.

The grammian of the (mean-adjusted) data matrix is a real symmetric matrix. Thus the linear structure in principal component analysis is an ellipsoide concerned on the origin and using as co-ordinate axes as its principal axes. Also the structure in multidimensional scaling is an ellipsoide concerned the origin and using as co-ordinate axes as the principal axes resulting from a principal component analysis. The above results are related to the spectral decomposition of XX^t(*or* X^tX). But the square of singular value of X is the eigenvalue of XX^t(*or* X^tX). The right singular vector is the eigenvector of XX^t and the left singular vector is the eigenvector of X^tX. Then the stationary point of above ellipsoide is the right(or left) singular vectors of data matrix X. Therefore, through the singular value decomposition of X, we can find the linear structure of the reduced dimension and give interpretation of that subspace.

Moreover there are certain advantages in using singular value decomposition over spectral decomposition as the tool for multivariate data reduction. The singular value decomposition is computationally far more efficient than the spectral decomposition of the sample covariance matrix when the number of variable is large. And working directly with data matrix X, we can also maintain first-hand feeling for the data which would have been at best diminished if XX^t were used as instead.

References

1. Choi, Y. S. and Huh, M. H. (1996). Resistant Singular Value decomposition and its statistical applications, *Journal of the Korea Statistical Society,* Vol 25, No 1 , 49-66.

2. Gnanadesikan, R. (1977). *Method for statistical Data Analysis of Multivariate Observations,* John Wiley and Sons, Inc. New York.

3. Jackson, J. E. and Hearne, F. J. (1975). Relationships among coefficients of vectors used in principal components, *Technometrics,* vol 15, 601-610.

4. Murtagh, F. and Heck, A. (1987). *Multivariate data Analysis,* D. Reidel Publishing Company. Holland.

5. Nishisato and Shizuhiko. (1996). Gleaning in the field of dual scaling. *Psycometrika,* Vol 61, 559-599.

6. Shin, Y.K. (1982). The Singular value decomposition in data analytic multivariate analysis, *MSc Thesis,* Korea University.