IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
Vol. 07, Issue 07 (July. 2017), ||V1|| PP 66-70

A Study On Dimension Reduction By Using Singular Value Decomposition

[1]Jajimogga Raghavendar[*], [2] V.Dharmaiah
[1]Department of Mathematics, Nalla Malla Reddy Engineering College
Divyanagar, Ghatkesar Mandal, Ranga Reddy District- 500088. Telangana
[2]Department of Mathematics, Retd. Professor, Osmania University, Hyderabad, Telangana-500007
Corresponding Author: Jajimogga Raghavendar

Abstract: Singular value decomposition(SVD) plays a vital role in matrix transformation. It is the basis on which number of vector- based methods like Principal Component Analysis (PCA), Independent Analysis. As we work with big data, there is a chance of facing these methods rather frequently.by transforming the matrix in multiple ways using SVD, it makes us to check the meaning of words from various angles. Usually a matrix can be decomposed into 3 matrices. A is the original matrix, and U and V are the new basis. And D stands in between U and Metrics only has diagonal elements are similar values of A. U rows are document vectors but now with different coefficient V: columns of V^T are word vectors.A sparse matrix solver is topic that people have been working on for 50 years in HPC (High Performance Computing). There are several math solver libraries like Petsc, Trilinos, and Python or Mat lab in order to compute these 3 matrices.The Matrix can get large when we go to the internet scale but SVD can handle the computation of a few hundred million cells in a large sparse matrix with the help of dimension reduction by shrinking the size of the matrix.

Keywords: *SVD; PCA; Big Data;*

Date of Submission: 26-07-2017 Date of acceptance: 08-08-2017

I. INTRODUCTION

This paper define what is PCA and SVD , and to give some example of PCA and SVD . then discuss about how to reduce the dimension of data (in matix) by using SVD: the basic idea of this method putting zero in smallest singular value in SVD of A. then find product of changed decomposed matrices , we get a new matrix as a result. This resultant matrix dimension is reduced compare to the given matrix.

1.1 Principal Component Analysis (PCA)

PCA is used as a technique for collection a dataset consisting of a set of tuples point in a high dimension resembling space and finding directions along which the tuple line up fast. This is to formulate the set of tuple as a matrix A and find eigen vector of AA^T and A^TA. These eigen vector can be considered as affixed rotation in a high dimensional space. If we apply this transformation to the original data the axis corresponding to the principle eigen vector is the axis along which the point are most speed out.

1.2 Singular Value Decomposition

We now look at a process in which a higher dimensional matrix converts into lower dimensions matrix by using singular value decomposing.
Let A be a $m \times n$ matrix and rank is r then

$$A_{m\times n} = U_{m\times r} D_{r\times r} V^T_{r\times n}$$

1. U is $m \times r$ column orthogonal matrix
2. V is $n \times r$ column orthogonal matrix
3. $D$is a diagonal matrix the elements are called the singular value of A.

II. METHOD AND ANALYSIS

The data is two dimensional a number of dimensions that is too small to make PCA really useful. The data shown in fig 1.1 has only four points.

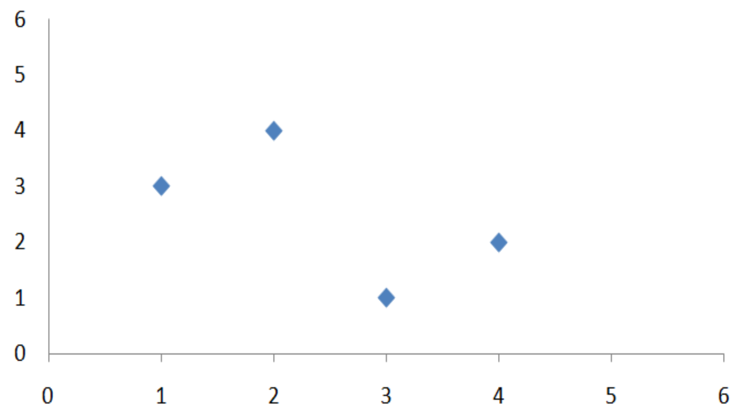

Fig 2.1 four points in two dimensional space

Let us represent the points by a matrix A with four rows one for each point and two columns, corresponding to the X- axis and Y- axis. This matrix is

$$A = \begin{bmatrix} 1 & 3 \\ 2 & 4 \\ 3 & 1 \\ 4 & 2 \end{bmatrix}$$

$$A^T A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 2 & 4 \\ 3 & 1 \\ 4 & 2 \end{bmatrix} = \begin{bmatrix} 30 & 22 \\ 22 & 30 \end{bmatrix}$$

We may find the eigen value for the matrix $A^T A$ and corresponding eigen vectors $\lambda = 52{,}8$

Therefore the unit eigen vector corresponding to the principle eigen vakue 52 is $E_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$

For the second eigen value in 8 the unit eigen vector corresponding to the principle eigen value 2 is

$$E_2 = \begin{bmatrix} \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

Now, let us construct the matrix of eigen vector for the matrix $A^T A$ placing the principle eigen vector first,

$$\therefore E = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

Any matrix of orthogonal vectors represent a rotation of the axes of a Euclidean space.
The matrix above can be viewed as a rotation 45 degrees anti clock wise.
For example
Let us multiply A by E the product is

$$AE = \begin{bmatrix} 1 & 3 \\ 2 & 4 \\ 3 & 1 \\ 4 & 2 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} \frac{4}{\sqrt{2}} & \frac{2}{\sqrt{2}} \\ \frac{6}{\sqrt{2}} & \frac{2}{\sqrt{2}} \\ \frac{4}{\sqrt{2}} & -\frac{2}{\sqrt{2}} \\ \frac{6}{\sqrt{2}} & -\frac{2}{\sqrt{2}} \end{bmatrix}$$

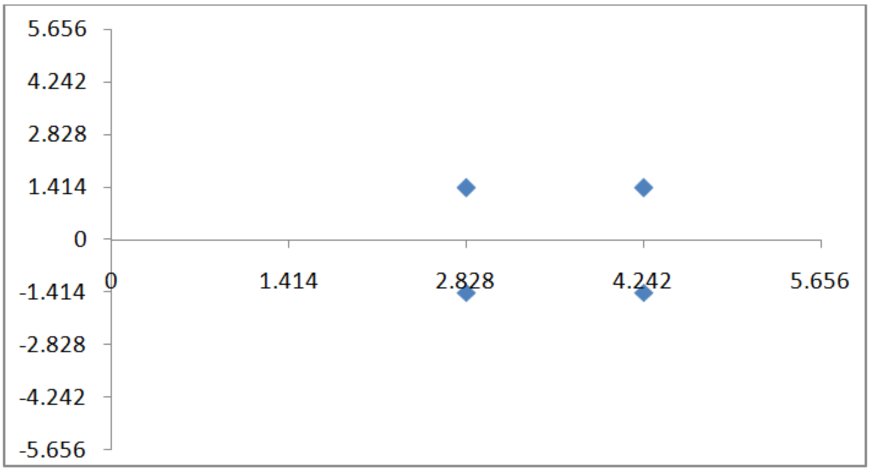

Fig 2.2 the points of Fig 2.1 in the new coordinate system

We observed that AE is the point of A transformed into a new coordinate space. In this case the X-axis is most significant. If we want to transform A to a space with fewer dimensions than the choice.that preserved the most significant is the one that uses the eigen vector associate with the largest eigen values and ignore the other eigen value. Let A be the matrix from fig 1.1. this data has only two dimensions, project the data onto a one dimension space that is we compute.

$$AE_1 = \begin{bmatrix} \frac{4}{\sqrt{2}} \\ \frac{6}{\sqrt{2}} \\ \frac{4}{\sqrt{2}} \\ \frac{6}{\sqrt{2}} \end{bmatrix}$$

The point of A by this projection onto the axis of Fig 2.2 while the first and third point project to the same point, as the second and fourth. This representation makes the best possible one dimensional and districtions among the point.

Example

Table 2.1 gives a rank 2 matrix representations rating of subject by the student. In this contrived example there are two concepts understanding the subjects: "Science & Engineering" .

	M	**P**	**C**	**E**	**D**
B_1	1	1	1	0	0
B_2	2	2	2	0	0
B_3	3	3	3	0	0
B_4	5	5	5	0	0
G_1	0	0	0	4	4
G_2	0	0	0	5	5
G_3	0	0	0	2	2

Table 2.1

M – Maths, **P** – Physics , **C** –Chemsitry, **E** – Engineering Mechanics, **D** – Drawing,

B_1 , B_2,B_3,, B_4 are boys G_1,G_2,G_3 are girls

All the boys rate only science subject and all the girls rate only Engineering. It is this existence of two strictly adherel to concepts. That gives the matrix a rank of 2.

The decomposition of the matrix A from table 2.1 into U , D andV Since the rank of A is 2.

$$A = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 4 & 4 \\ 0 & 0 & 0 & 5 & 5 \\ 0 & 0 & 0 & 2 & 2 \end{bmatrix}$$

$$U = \begin{bmatrix} -0.1601 & 0 & 0.9784 & -0.0151 & -0.0774 & -0.0968 & -0.0387 \\ -0.3203 & 0 & 0.0519 & -0.4406 & 0.4991 & 0.6239 & 0.2496 \\ -0.4804 & 0 & -0.0313 & 0.8374 & 0.1544 & 0.1930 & 0.0772 \\ -0.8006 & 0 & -0.1976 & -0.3232 & -0.2768 & -0.3460 & -0.1384 \\ 0 & -0.5963 & 0 & 0 & 0.6444 & -0.4444 & -0.1778 \\ 0 & -0.7454 & 0 & 0 & -0.4444 & 0.4444 & -0.2222 \\ 0 & -0.2981 & 0 & 0 & -0.1778 & -0.2222 & 0.9111 \end{bmatrix}$$

$$D = \begin{pmatrix} 10.8167 & 0 & 0 & 0 & 0 \\ 0 & 9.4868 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$V = \begin{pmatrix} -0.5774 & 0 & -0.8165 & 0 & 0 \\ -0.5774 & 0 & 0.4082 & -0.7071 & 0 \\ -0.5774 & 0 & 0.4082 & 0.7071 & 0 \\ 0 & -0.7071 & 0 & 0 & -0.7071 \\ 0 & -0.7071 & 0 & 0 & 0.7071 \end{pmatrix}$$

Which is singular value decomposition of matrix A by deleting the zeros of D cooresopoding columns of U and V^T then we get

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 4 & 4 \\ 0 & 0 & 0 & 5 & 5 \\ 0 & 0 & 0 & 2 & 2 \end{bmatrix} =$$

$$\begin{bmatrix} -0.1601 & 0 \\ -0.3203 & 0 \\ -0.4804 & 0 \\ -0.8006 & 0 \\ 0 & -0.5963 \\ 0 & -0.7454 \\ 0 & -0.2981 \end{bmatrix} \begin{bmatrix} 10.8167 & 0 \\ 0 & 9.4868 \end{bmatrix} \begin{bmatrix} -0.5774 & -0.5774 & -0.5774 & 0 & 0 \\ 0 & 0 & 0 & -0.7071 & -0.7071 \end{bmatrix}$$

III. INTERPRETATION OF SVD

Let us think of the rows of A as students and the column of A are subjects then matrix U connects students. For example the B_1 who cooorespond to row one of A like only subject of Science the second columnof first row of U is zero because B_1 doesnot like Engineering subject at all.

The matrix V relates to the subjects. The value -0.5774 in each of the first three columns of the first rows of V^T indicate first three subjects.while the zeros in last two columns of the first row indicates Engineering subjects are not opt by boys.the second row of V^T tells the subjects E and M are exculsivly are Engineeering subjects. The matrix D gives strengh of each the concepts. In our example the strengh of Science concepts is 10.8167while the strengh of Engineering concepts is 9.4868. Therefore the Science concepts are stronger because the data provides more information about the subject of that genereand the students who like them. The best way to reduce the dimension of three matrices is to set the smallest of the singular value to zero. If we set the smallest singular value to zero, then we can also eliminate the cooresponding column of U and V.

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 4 & 4 \\ 0 & 0 & 0 & 5 & 5 \\ 0 & 0 & 0 & 2 & 2 \end{bmatrix} = \begin{bmatrix} -0.1601 \\ -0.3203 \\ -0.4804 \\ -0.8006 \\ 0 \\ 0 \\ 0 \end{bmatrix} [10.8167][-0.5774 \quad -0.5774 \quad -0.5774 \quad 0 \quad 0]$$

$$\approx \begin{pmatrix} 0.9999 & 0.9999 & 0.9999 & 0 & 0 \\ 2.0005 & 2.0005 & 2.0005 & 0 & 0 \\ 3.0004 & 3.0004 & 3.0004 & 0 & 0 \\ 5.0002 & 5.0002 & 5.0002 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

IV. CONCLUSION

This paper is proven thatto reduce the dimension of the large matrix into two or more other matrices whose sizes are much smaller than the original by using singular values decompositionThe original matrix can be appoximately reconstructed by taking three matrices product. Then we can enhance this process to the large data.

ACKNOWLEDGEMENT

Author is thankful to Dr. DivyaNalla , Principal for her encouragement and support and also thanks to Mr. S. Karunakar Reddy, Mr. M. Karunakar , J. BalaAnasuyafor their technical support.

REFERENCES

[1] D.S. Dodson, R.G. Grimes, and J.G. Lewis, Sparse Extensions to the Fortran Basic Linear Algebra Subprograms, ACM Transactions on Mathematical Software, June 1991, Vol 17, No. 2.

[2] George and J. W-H. Liu, Computer Solution of Large Sparse Positive Definite Systems, Prentice-Hall Inc., Englewood Cliffs, New Jersey, 1981.

[3] 3.E. Ng and B. W. Peyton, Block Sparse Cholesky Algorithms on Advanced Uniprocessor Computers, SIAM M. Sci Comput., 14:1034-1056, 1993.

[4] Ian S. Duff, Roger G. Grimes and John G. Lewis, User's Guide for the Harwell-Boeing Sparse Matrix Collection (Release I), Technical Report TR/PA/92/86, CERFACS, Lyon, France, October 1992.

[5] Developments and ApplicationsofNonlinear Principal omponent Analysis a Review we Kruger, Junping Zhang, Lei XieHujun Yin Pages 68-95.