INTERNATIONAL JOURNAL OF RESEARCH IN COMPUTER APPLICATION & MANAGEMENT

IJRCM

A Monthly Double-Blind Peer Reviewed (Refereed/Juried) Open Access International e-Journal - Included in the International Serial Directories

Indexed & Listed at:

Ulrich's Periodicals Directory ©, ProQuest, U.S.A., EBSCO Publishing, U.S.A., Cabell's Directories of Publishing Opportunities, U.S.A.,

Open J-Gage, India [link of the same is duly available at Inflibnet of University Grants Commission (U.G.C.)],

Index Copernicus Publishers Panel, Polandwith IC Value of 5.09 &number of libraries all around the world.

Circulated all over the world & Google has verified that scholars of more than 2718 Cities in 161 countries/territories are visiting our journal on regular basis.

Ground Floor, Building No. 1041-C-1, Devi Bhawan Bazar, JAGADHRI – 135 003, Yamunanagar, Haryana, INDIA

http://ijrcm.org.in/

CONTENTS

CALL FOR MANUSCRIPTS

We invite unpublished novel, original, empirical and high quality research work pertaining to recent developments & practices in the areas of Computer Science & Applications; Commerce; Business; Finance; Marketing; Human Resource Management; General Management; Banking; Economics; Tourism Administration & Management; Education; Law; Library & Information Science; Defence & Strategic Studies; Electronic Science; Corporate Governance; Industrial Relations; and emerging paradigms in allied subjects like Accounting; Accounting Information Systems; Accounting Theory & Practice; Auditing; Behavioral Accounting; Behavioral Economics; Corporate Finance; Cost Accounting; Econometrics; Economic Development; Economic History; Financial Institutions & Markets; Financial Services; Fiscal Policy; Government & Non Profit Accounting; Industrial Organization; International Economics & Trade; International Finance; Macro Economics; Micro Economics; Rural Economics; Co-operation; Demography: Development Planning; Development Studies; Applied Economics; Development Economics; Business Economics; Monetary Policy; Public Policy Economics; Real Estate; Regional Economics; Political Science; Continuing Education; Labour Welfare; Philosophy; Psychology; Sociology; Tax Accounting; Advertising & Promotion Management; Management Information Systems (MIS); Business Law; Public Responsibility & Ethics; Communication; Direct Marketing; E-Commerce; Global Business; Health Care Administration; Labour Relations & Human Resource Management; Marketing Research; Marketing Theory & Applications; Non-Profit Organizations; Office Administration/Management; Operations Research/Statistics; Organizational Behavior & Theory; Organizational Development; Production/Operations; International Relations; Human Rights & Duties; Public Administration; Population Studies; Purchasing/Materials Management; Retailing; Sales/Selling; Services; Small Business Entrepreneurship; Strategic Management Policy; Technology/Innovation; Tourism & Hospitality; Transportation Distribution; Algorithms; Artificial Intelligence; Compilers & Translation; Computer Aided Design (CAD); Computer Aided Manufacturing; Computer Graphics; Computer Organization & Architecture; Database Structures & Systems; Discrete Structures; Internet; Management Information Systems; Modeling & Simulation; Neural Systems/Neural Networks; Numerical Analysis/Scientific Computing; Object Oriented Programming; Operating Systems; Programming Languages; Robotics; Symbolic & Formal Logic; Web Design and emerging paradigms in allied subjects.

Anybody can submit the **soft copy** of unpublished novel; original; empirical and high quality **research work/manuscript** ***anytime*** in ***M.S. Word format*** after preparing the same as per our **GUIDELINES FOR SUBMISSION**; at our email address i.e. infoijrcm@gmail.com or online by clicking the link **online submission** as given on our website (***FOR ONLINE SUBMISSION, CLICK HERE***).

GUIDELINES FOR SUBMISSION OF MANUSCRIPT

1. **COVERING LETTER FOR SUBMISSION:**

DATED: _____

THE EDITOR
IJRCM

Subject: **SUBMISSION OF MANUSCRIPT IN THE AREA OF .**

(e.g. Finance/Marketing/HRM/General Management/Economics/Psychology/Law/Computer/IT/Engineering/Mathematics/other, please specify)

DEAR SIR/MADAM

Please find my submission of manuscript entitled '_____' for possible publication in your journals.

I hereby affirm that the contents of this manuscript are original. Furthermore, it has neither been published elsewhere in any language fully or partly, nor is it under review for publication elsewhere.

I affirm that all the author (s) have seen and agreed to the submitted version of the manuscript and their inclusion of name (s) as co-author (s).

Also, if my/our manuscript is accepted, I/We agree to comply with the formalities as given on the website of the journal & you are free to publish our contribution in any of your journals.

NAME OF CORRESPONDING AUTHOR:
Designation:
Affiliation with full address, contact numbers & Pin Code:
Residential address with Pin Code:
Mobile Number (s):
Landline Number (s):
E-mail Address:
Alternate E-mail Address:

NOTES:

a) The whole manuscript is required to be in ***ONE MS WORD FILE*** only (pdf. version is liable to be rejected without any consideration), which will start from the covering letter, inside the manuscript.
b) The sender is required to mentionthe following in the **SUBJECT COLUMN** of the mail:
New Manuscript for Review in the area of (Finance/Marketing/HRM/General Management/Economics/Psychology/Law/Computer/IT/ Engineering/Mathematics/other, please specify)
c) There is no need to give any text in the body of mail, except the cases where the author wishes to give any specific message w.r.t. to the manuscript.
d) The total size of the file containing the manuscript is required to be below **500 KB**.
e) Abstract alone will not be considered for review, and the author is required to submit the complete manuscript in the first instance.
f) The journal gives acknowledgement w.r.t. the receipt of every email and in case of non-receipt of acknowledgment from the journal, w.r.t. the submission of manuscript, within two days of submission, the corresponding author is required to demand for the same by sending separate mail to the journal.

2. **MANUSCRIPT TITLE:** The title of the paper should be in a 12 point Calibri Font. It should be bold typed, centered and fully capitalised.

3. **AUTHOR NAME (S) & AFFILIATIONS:** The author (s) **full name**, **designation**, **affiliation** (s), **address**, **mobile/landline numbers**, and **email/alternate email address** should be in italic & 11-point Calibri Font. It must be centered underneath the title.

4. **ABSTRACT:** Abstract should be in fully italicized text, not exceeding 250 words. The abstract must be informative and explain the background, aims, methods, results & conclusion in a single para. Abbreviations must be mentioned in full.

5. **KEYWORDS:** Abstract must be followed by a list of keywords, subject to the maximum of five. These should be arranged in alphabetic order separated by commas and full stops at the end.

6. **MANUSCRIPT:** Manuscript must be in ***BRITISH ENGLISH*** prepared on a standard A4 size ***PORTRAIT SETTING PAPER***. It must be prepared on a single space and single column with 1" margin set for top, bottom, left and right. It should be typed in 8 point Calibri Font with page numbers at the bottom and centre of every page. It should be free from grammatical, spelling and punctuation errors and must be thoroughly edited.

7. **HEADINGS:** All the headings should be in a 10 point Calibri Font. These must be bold-faced, aligned left and fully capitalised. Leave a blank line before each heading.

8. **SUB-HEADINGS:** All the sub-headings should be in a 8 point Calibri Font. These must be bold-faced, aligned left and fully capitalised.

9. **MAIN TEXT:** The main text should follow the following sequence:

INTRODUCTION

REVIEW OF LITERATURE

NEED/IMPORTANCE OF THE STUDY

STATEMENT OF THE PROBLEM

OBJECTIVES

HYPOTHESES

RESEARCH METHODOLOGY

RESULTS & DISCUSSION

FINDINGS

RECOMMENDATIONS/SUGGESTIONS

CONCLUSIONS

SCOPE FOR FURTHER RESEARCH

ACKNOWLEDGMENTS

REFERENCES

APPENDIX/ANNEXURE

It should be in a 8 point Calibri Font, single spaced and justified. The manuscript should preferably not exceed ***5000 WORDS***.

10. **FIGURES &TABLES:** These should be simple, crystal clear, centered, separately numbered &self explained, and **titles must be above the table/figure**. **Sources of data should be mentioned below the table/figure**. It should be ensured that the tables/figures are referred to from the main text.

11. **EQUATIONS:**These should be consecutively numbered in parentheses, horizontally centered with equation number placed at the right.

12. **REFERENCES:** The list of all references should be alphabetically arranged. The author (s) should mention only the actually utilised references in the preparation of manuscript and they are supposed to follow **Harvard Style of Referencing**. The author (s) are supposed to follow the references as per the following:

- All works cited in the text (including sources for tables and figures) should be listed alphabetically.
- Use (**ed.**) for one editor, and (**ed.s)** for multiple editors.
- When listing two or more works by one author, use --- (20xx), such as after Kohl (1997), use --- (2001), etc, in chronologically ascending order.
- Indicate (opening and closing) page numbers for articles in journals and for chapters in books.
- The title of books and journals should be in italics. Double quotation marks are used for titles of journal articles, book chapters, dissertations, reports, working papers, unpublished material, etc.
- For titles in a language other than English, provide an English translation in parentheses.
- The location of endnotes within the text should be indicated by superscript numbers.

PLEASE USE THE FOLLOWING FOR STYLE AND PUNCTUATION IN REFERENCES:

BOOKS

- Bowersox, Donald J., Closs, David J., (1996), "Logistical Management." Tata McGraw, Hill, New Delhi.
- Hunker, H.L. and A.J. Wright (1963), "Factors of Industrial Location in Ohio" Ohio State University, Nigeria.

CONTRIBUTIONS TO BOOKS

- Sharma T., Kwatra, G. (2008) Effectiveness of Social Advertising: A Study of Selected Campaigns, Corporate Social Responsibility, Edited by David Crowther & Nicholas Capaldi, Ashgate Research Companion to Corporate Social Responsibility, Chapter 15, pp 287-303.

JOURNAL AND OTHER ARTICLES

- Schemenner, R.W., Huber, J.C. and Cook, R.L. (1987), "Geographic Differences and the Location of New Manufacturing Facilities," Journal of Urban Economics, Vol. 21, No. 1, pp. 83-104.

CONFERENCE PAPERS

- Garg, Sambhav (2011): "Business Ethics" Paper presented at the Annual International Conference for the All India Management Association, New Delhi, India, 19–22 June.

UNPUBLISHED DISSERTATIONS AND THESES

- Kumar S. (2011): "Customer Value: A Comparative Study of Rural and Urban Customers," Thesis, KurukshetraUniversity, Kurukshetra.

ONLINE RESOURCES

- Always indicate the date that the source was accessed, as online resources are frequently updated or removed.

WEBSITES

- Garg, Bhavet (2011): Towards a New Natural Gas Policy, Political Weekly, Viewed on January 01, 2012 http://epw.in/user/viewabstract.jsp

IMPLEMENTATION OF PCA WITH SVD TO REDUCE PRECISION LOSS

AMITPREET KOUR
STUDENT
DEPARTMENT OF COMPUTER SCIENCE
GURU NANAK DEV UNIVERSITY
AMRITSAR

RAMANDEEP KAUR
ASST. PROFESSOR
DEPARTMENT OF COMPUTER SCIENCE
GURU NANAK DEV UNIVERSITY
AMRITSAR

ABSTRACT

The objective of data mining is to identify valid novel, potentially useful, and understandable correlations and patterns in existing data. Finding useful patterns in data is known by different names (including data mining) in different communities (e.g., knowledge extraction, information discovery, information harvesting, data archeology, and data pattern processing). The volume of electronically accessible data in warehouses and on the Internet is growing faster than the speedup in processing times predicted by Moore's Law. Consequently, classical data mining algorithms that require one or more computationally intensive passes over the entire database are becoming prohibitively slow, and this problem will only become worse in the future. The scalability of mining algorithms has therefore become a major research topic. One approach to the scalability problem is to run mining algorithms on a small subset of the data. That is, data reduction is very important issue.Data reduction techniques can be applied to obtain a red-uced representation of the data set that is much smaller in volume, yet closely maintains the integrity of the original data. Various data reduction techniques are used like PCA,SVD. It is observed that PCA causes precision loss of data. So in this paper, an algorithm is proposed which leads to minimal loss in the precision by applying SVD on the eigen vectors. Then these eigen vectors are used to obtain Principal Components.

KEYWORDS
Data Mining, Eigen vectors, PCA, precision loss, SVD.

INTRODUCTION

Data mining is the process of discovering potentially valuable patterns, associations, trends, sequences and dependencies in data. It is the overall process of discovering new pattern or building models from a given dataset. There are many steps involved in Knowledge Discovery in Databases (KDD) which include data selection, data cleaning and preprocessing, data transformation and reduction, data mining task and algorithm selection, and finally post-processing and interpretation of discovered knowledge.The primary goals of data mining in practice are prediction and description. Prediction involves using some variables or fields in the database to predict unknown or future values of other variable of interest. Description focuses on finding interpretable patterns that describes the data.The sheer volume of data today and its expected increase in the next few days are some of the key challenges in data mining and knowledge discovery applications[1]. Besides a huge number of data samples that are collected and processed, the high dimensional nature of data arising in many applications causes a need to develop effective and efficient techniques that are able to deal with massive amount of data[9]. As the dimensionality of data increases, many types of data analysis and classification problems become significantly harder.
Dimensionality reduction, feature selection, numerosity reduction, and data compression are some of the techniques to reduce the attribute space.
PCA is one the data reduction technique, it reduces the high dimensional data by calculating the eigen values and eigen vectors. But the main drawback of this method is that it leads to the precision loss.

DATA REDUCTION TECHNIQUES

Dimensionality reduction is the process of reducing the number of random variables or attributes under consideration by producing the linear combination of the original data. Dimensionality reduction methods include transforms and principal components analysis which transform or project the original data onto a smaller space. The main advantage of this technique is that the size of attribute space is reduced without loosing a lot of information of the original attribute space. The disadvantage is the fact that the linear combinations of the original features are usually not interpretable and the information about how much an original attribute contributes is often lost.
Feature Selection works by removing redundant or irrelevant features from the data set as they can lead to a reduction of the classification accuracy or clustering quality and to unnecessary increase in the computational cost. The advantage of feature selection is that no information about the importance of single feature is lost. On the other hand, if a small set of features is required and the original features are diverse, information may be lost as some of the features must be omitted.
Numerosity reduction techniques replace the original data volume by alternative smaller forms of data representation.
Data compression, transformations are applied so as to obtain a reduced or "compressed" representation of original data.
In this paper we will mainly deal with dimensionality reduction technique of data reduction. Data reduction can be lossless or lossy. In dimensionality reduction, reduction of the dimensionality is done by using low rank approximation techniques to create new attributes that are combinations of the old, original variables. Dimensionality reduction is also referred to as feature transformation, feature extraction or feature construction. Some of the techniques covered under this are Principal Component Analysis(PCA), Singular Value Decomposition(SVD), Pseudoinverse(also called as Moore Penrose Pseudoinverse), etc.

A. PRINCIPAL COMPONENT ANALYSIS

Principal component analysis (PCA)[2] is a standard tool in modern data analysis - in diverse fields from neuroscience to computer graphics - because it is a simple, non-parametric method for extracting relevant information from confusing data sets.
Often, the variables under study are highly correlated and as such they are effectively "saying the same thing". It may be useful to transform the original set of variables to a new set of uncorrelated variables called principal components. These new variables are linear combinations of original variables and are derived in decreasing order of importance so that the first principal component accounts for as much as possible of the variation in the original data. These new attributes (called principle components (PCs)) have to meet the following criteria: The PCs are (i) linear combinations of the original attributes, (ii) orthogonal to each other, and (iii) capture the maximum amount of variation in the data. Also, PCA is a linear dimensionality reduction technique, which identifies orthogonal directions of maximum variance in the original data, and projects the data into a lower-dimensionality space formed of a sub set of the highest variance components.

The goal of PCA is to reduce the dimensionality of the data while retaining asmuch as possible of the variation present in the original dataset.

Advantages and disadvantages of PCA:

PCA is a relatively simple, non-parametric, generic method that is useful for finding new, more informative, uncorrelated features and it can be used to reduce dimensionality by rejecting low variance features. Since the principal components are orthogonal[7][8] to each other, every principal component is uncorrelated to every other principal component (i. e., they do not contain any redundant information). The principal components are designed to account for the highest percentage of the variation among the variables with as few PCs as possible. Thus, often the first few PCs account for some large percentage of the total variance, allowing for a compact representation of the data with only low dimensions.

However, PCA is limited to re-expressing the data as combinations of its basis vectors. A main drawback of PCA[5] is that each PC is a linear combination of all original variables, thus leading to a potentially difficult interpretation of the PCs. On the contrary, in a system with many variables PCA may be used to project the dimension down to a reasonable number of plots, and the principal components could be rotated towards a more meaningful representation. Moreover, PCA[6] is sensitive with respect to the units of measurement. If the units and the variances of attributes vary a lot, then variables with high variance tend to dominate the first few principal components. In this case, the data need to be normalized prior to the PCA transformation. Computational complexity**:** From a computational point of view, the eigen value decomposition for solving the PCA transformation is rather expensive in terms of runtime, especially for a large number of attributes. There are several algorithms for solving symmetric eigen problems, but all of them are of order O(n3). So the complexity of solving PCA also revolves around the same.

B. SINGULAR VALUE DECOMPOSITION

Singular Value Decomposition[3] can be viewed from three points of view:

1. It is a method for transforming correlated variables into a set of uncorrelated ones that better expose the various relationships among the original data items.
2. SVD is a method for identifying and ordering the dimensions along which data points exhibit the most variation.
3. Once we have identified where the most variation is, it is possible to find the best

approximation of the original data point usingfewer dimensions. Hence SVD can be seen as a method of data reduction.

SVD[4] is based on a theorem of linear algebra which says that a rectangular matrix can be broken down into a product of three matrices:

i. An orthogonal matrix U [i.e. $U^TU=I$] and the columns of U are orthonormal eigen vectors of $A.A^T$.
ii. A diagonal matrix S, containing the square roots of eigen values from U or V in descending order.
iii. Transpose of an orthogonal matrix V^T [i.e. $V^TV=I$] and the columns of V are orthonormal eigen vectors of $A^T.A$.

i.e. $A_{mn}=U_{mn}S_{mn}V^T_{nn}$

The truncated SVD (or reduced rank SVD) to A can be found by setting all but the first k largest singular values equal to zero and using only the first k columns of U and V .

This is usually denoted like $A_k=U_kS_kV^T_k$

or more explicitly**,**

$$A_k \simeq (\mathbf{u}_1, \ldots, \mathbf{u}_k) \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_k \end{pmatrix} \begin{pmatrix} \mathbf{v}_1^\top \\ \vdots \\ \mathbf{v}_k^\top \end{pmatrix},$$

Computational complexity

In general, the computational complexity of the SVD algorithm for computing all three matrices U, S and V is $O(4m^2n+8mn^2+9n^3)$, for computing only the matrices S and V it is $O(4mn^2 + 8n^3)$, making the decomposition unfeasible for large datasets. The RSVD algorithm has a complexity of O(4m2n+22n3) for computing all three matrices, and a complexity of $O(2mn^2 + 11n^3)$ for computing only S and V .

Relation to PCA

Calculating the SVD is equivalent of finding the eigen values and eigen vectors of AA^T and A^TA. The eigenvectors of AA^T make up the columns of U, the eigenvectors of A^TA make up the columns of V. Moreover, the singular values in S are the square roots of eigen values from AA^T and A^TA. The SVD-PCA connection stems from the straightforward linear algebra calculations

$AA^T=(USV^T)(VS^TU^T)=US^2U^T$, *and*

$A^TA=(VS^TU^T)(USV^T)=VS^2V^T$

An often mentioned difference between these two approaches is that PCA removes the mean of each variable whereas SVD uses the original data (however, the mean could also be removed before computing the SVD). Especially for sparse data it is not always preferable to remove the mean of the data.

Methods

There are various methods to solve matrix equations that are specialised to particular problems. We can select between these based on the type of data. Various methods used for the same are as follows:

1. *LAPACK*: It is the default method for solving dense numerical matrices. When the matrix is square and non-singular the routines dgesv, dlange, and dgecon are used for real matrices and zgesv, zlange, and zgecon for complex matrices. When the matrix is non-square or singular dgelss is used for real matrices and zgelss for complex matrices. If the input matrix uses arbitrary-precision numbers, then LAPACK algorithms extended for arbitrary-precision computation are used.
2. *Multifrontal*: The Multifrontal method is a direct solver used by default if the input matrix is sparse. If the input matrix to the Multifrontal method is dense, it is converted to a sparse matrix. The implementation of the Multifronal method uses the UMFPACK library.
3. *Krylov*: The Krylov method is an iterative solver that is suitable for large sparse linear systems, such as those arising from numerical solving of PDEs. Basically two Krylov methods are implemented: Conjugate Gradient (for symmetric positive definite matrices) and BiCGSTAB (for non-symmetric systems). The default method for Krylov, BiCGSTAB, is more expensive but more generally applicable. The ConjugateGradient method is suitable for symmetric positive definite systems, always converging to a solution (though the convergence may be slow). If the matrix is not symmetric positive definite the ConjugateGradient may not converge to a solution.
4. *Cholesky*: The Cholesky method is suitable for solving symmetric positive definite systems. This method is far more fast and stable. For dense matrices the Cholesky method uses LAPACK functions such as dpotrf and dpotrs for real matrices and zpotrf and zpotrs for complex matrices. For sparse matrices the Cholesky method uses the TAUCS library.
5. *Symbolic Methods*: There are a number of methods that are specific to symbolic and

Exact computation: CofactorExpansion, DivisionFreeRowReduction, and OneStepRowReduction.

RELATED WORK

Data mining (sometimes called data or knowledge discovery)[1] is the process of analyzing data from different perspectives and summarizing it into useful information. With this increase has come the need to be able to store, transmit, and query large volumes of image data efficiently. To overcome problems associated with high dimensionality, such as high storage and retrieval times, a dimension reduction step is usually applied to the vectors to concentrate relevant information in a small number of dimensions. Principal Component Analysis (PCA) is a well-known dimension reduction scheme. PCA is sensitive to the

relative scaling of the original variables. Depending on the field of application, it is also named the discrete Karhunen–Loève transform (KLT), the Hotelling transform or proper orthogonal decomposition (POD).PCA[7] was invented in 1901 by Karl Pearson.

PCA can be done by eigen value decomposition of a data covariance matrix or singular value decomposition of a data matrix, usually after mean centering the data for each attribute.PCA finds its application in number of fields ,but the major drawback of PCA is that it leads to precision loss,i.e, the reduced data when converted back to the actual data, the difference is large.

PROPOSED ALGORITHM

As PCA suffers from the drawback of precision loss. So in this section we are proposing an algorithm that will work on this limitation. In this algorithm we are combining another data reduction technique SVD[4] with PCA .PCA is a Technique to convert data set in a form that can regenerate or found in High Dimension. Its help in compression. After compress data, reduced data is formed , also we can regenerate original data with different techniques.

Steps :

Step 1: Find Original Data in X and Y co-ordinates
Step 2: Find Mean XI and YI for X and Y Co-ordinates.
Step 3: Calculate X-XI and Y-YI and XI-X,YI-Y
Step 4: Calculate VAR (X,X)
Step 5: Calculate VAR (X,Y)
Step 6: Calculate VAR (Y,X)
Step 7: Calculate VAR (Y,Y)
Step 8: Calculate CO VAR Matrix
Step 9: Calculate EigenVector
Step 10 : Apply SVD on EigenVector As $X=U\Sigma V^T$
*Step 11 :*Then Eigen Value Will be $XX^T=(U\Sigma V^T)(U\Sigma V^T)^T=(U\Sigma V^T)(V\Sigma U^T)=U\Sigma^2U^T$
Step 12 : Finally we get a a Eigen Value from Vector a Singular Values.
Step 13 : Calculate Feature Matrix Using Eigen Vector And Eigen Value
Step 14 : Finally We get a PCA data with multiply Different Feature Matrix with Original Matrix.

Since the formation of Eigen Vector with Co Var Matrix causes loss of precision . In this thesis, this loss is avoided by using SVD on the Eigen vector.

EXPERIMENTAL RESULTS

The platform as used by us for the implementation of the proposed algorithm is java. The application of the above algorithm has shown that the data reduction is more accurate by using the above technique and the results we have got are more satisfactory.

Actual Data

1.5000 1.4000
1.5000 1.7000
2.3000 2.6000
1.9000 1.2000
1.1000 1.0000
2.3000 2.7000
2.3000 2.6000
1.0000 1.1000
1.6000 1.4000
1.3000 1.9000

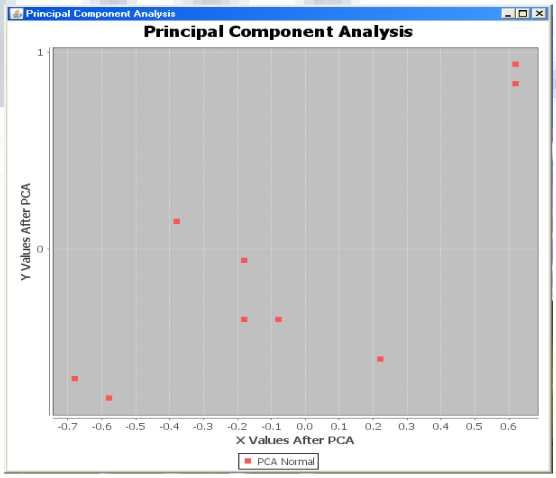
FINAL OUTPUT

Reduced data with PCA only	Reduced data when PCA is applies with SVD
-0.3976	-2.1503
-0.1533	-0.8290
1.0439	5.6449
-0.3284	-1.7761
-0.9555	-5.1671
1.1253	6.0854
1.0439	5.6449
-0.9321	-5.0404
-0.3396	-1.8365
-0.1065	-0.5757

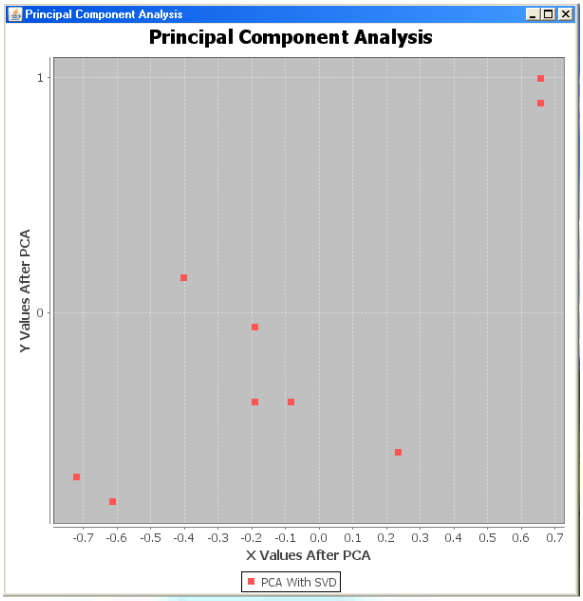
From the table we can deduce that the second method i.e., PCA with SVD is better than the first one. As we get the data which is more closer to the actual data, thus reducing precision loss which is more in the first case

These results can be shown graphically,

CASE 1: WHEN PCA IS USED ONLY



CASE 2: WHEN PCA IS APPLIED WITH SVD



From the above graphs we can see that in the second we are getting more clear data.

CONCLUSIONS

In this paper we have discussed about data reduction techniques and from them we conclude various advantages of data reduction. Principal component analysis is a powerful tool for reducing a number of observed variables into a smaller number of artificial variables that account for most of the variance in the data set.PCA finds its application in image recognization and various other fields. But the precision loss is high in PCA. Which can be reduced by applying Singular value computation technique for calculating eigen values in the PCA method. This will help to obtain more clear data i.e., more accurate data near to the actual data.

REFERENCES

1. Alter O., Brown P.O., Botstein D. *Singular value decomposition for genome-wide expression data processing and modeling.* Proc Natl Acad Sci USA 2000; 97:10101-06.
2. Fayyad, U., Piatetsky-Shapiro, G., and Smyth "From Data Mining to Knowledge Discovery: An Overview" *Advances in Knowledge Discovery and Data Mining, MIT Press*, 1-36, Cambridge, 1996.
3. Jolliffe, I. T. (2002). *Principal component analysis* (Second ed.): Springer.
4. Larose, D.T., *An Introduction to Data Mining. Discovering Knowledge in Data.* 2005: Willi & Sons,Inc.
5. M.A. Carreira-Perpinan. *A review of dimension reduction techniques. Technical report CS-96-09,Department of CSE* , University of Sheffield, 1997
6. Mark Richardson tutorial on "*Principal Component Analysis*",2009.
7. Michael E. Wall "*Singular value decomposition and principal component analysis*",2003.
8. Principal components analysis for correlated curves and seasonal commodities: The case of the Petroleum markets (Tolmasky & Hindanov, 2002).
9. Shlens, J., *A Tutorial on Principal Component Analysis*. 2009.
10. Smith, L., *A Tutorial on Principal Component Analysis*. 2002.

REQUEST FOR FEEDBACK

Dear Readers

At the very outset, International Journal of Research in Computer Application and Management (IJRCM) acknowledges & appreciates your efforts in showing interest in our present issue under your kind perusal.

I would like to request you tosupply your critical comments and suggestions about the material published in this issue as well as on the journal as a whole, on our E-mail**infoijrcm@gmail.com** for further improvements in the interest of research.

If youhave any queries please feel free to contact us on our E-mail **infoijrcm@gmail.com**.

I am sure that your feedback and deliberations would make future issues better – a result of our joint effort.

Looking forward an appropriate consideration.

With sincere regards

Thanking you profoundly

Academically yours

Sd/-

Co-ordinator

ABOUT THE JOURNAL

In this age of Commerce, Economics, Computer, I.T. & Management and cut throat competition, a group of intellectuals felt the need to have some platform, where young and budding managers and academicians could express their views and discuss the problems among their peers. This journal was conceived with this noble intention in view. This journal has been introduced to give an opportunity for expressing refined and innovative ideas in this field. It is our humble endeavour to provide a springboard to the upcoming specialists and give a chance to know about the latest in the sphere of research and knowledge. We have taken a small step and we hope that with the active co-operation of like-minded scholars, we shall be able to serve the society with our humble efforts.

Our Other Journals