Vol 04, Special Issue01; 2013
http://ijpaper.com/
PUBLICATIONS OF PROBLEMS & APPLICATION IN ENGINEERING RESEARCH - PAPER
CSEA2012
ISSN: 2230-8547; e-ISSN: 2230-8555

STUDY ON DIMENSIONALITY REDUCTION TECHNIQUES AND APPLICATIONS

G.N.RAMADEVI [1] K.USHARANI 2

[1] Research Scholar, Department of Computer Science, S.P.M.V.V, Tirupati.
2 Assoc.Prof., Department of Computer science , S.P.M.V.V, Tirupati.

ABSTRACT

Data is not collected only for data mining. Data accumulates in an unprecedented speed. Data preprocessing is an important part for effective machine learning and data mining. Data mining is discovering interesting knowledge from large amounts of data, which is the integral part of the KDD (Knowledge Discovery in Databases), which is the overall process of converting raw data into useful information. Dimension reduction can be beneficial not only for reason of computational efficiency but also it can improve the accuracy of the analysis. The set of techniques that can be employed for dimension reduction can be partitioned into two important ways i.e., they are feature selection and feature extraction [1]. They can be applied for supervised or unsupervised learning. In this paper we are presented the study of the dimension reduction techniques and its applications in real life. The PCA (Principal Component Analysis) is one of the dimensionality reduction techniques of feature reduction algorithm to reduce the dimensionality of the dataset without losing the data. PCA can be applied on data before clustering will results more accurate and reduce the time substantially. PCA is used for data visualization and noise reduction.

Keywords: Dimensionality Reduction, Feature selection, Feature Extraction, PCA(Principal Component Analysis), EM (Expectation Maximization), SVM (Support Vector Machines) ,RFE (Recursive Feature Elimination), LDA (Linear Discriminant Analysis), CCA (Canonical Correlation Analysis). PLS (Partial Least Squares).

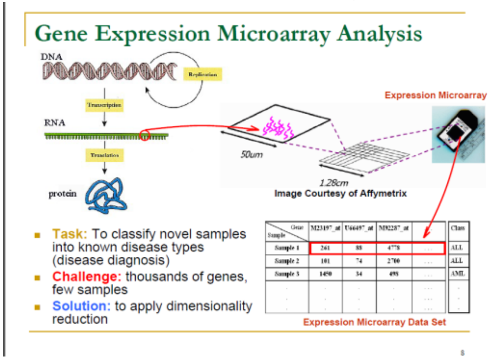
1. INTRODUCTION

1.1. Dimensionality Reduction:

Dimensionality reduction is an effective approach to downsizing the data [1]. It is a methodology that attempts to project a set of high dimensional vectors to a lower dimensionality space while retaining metrics among them. The machine learning and data mining techniques may not be effective for high-dimensional data because of the curse of dimensionality and query accuracy and efficiency will degrade rapidly as the dimension increases [1]. The Dimension reduction is used for 1) Visualization: To projection of high-dimensional data onto 2D or 3D. 2) Data Compression: Efficient storage and retrieval. 3) Noise removal: Positive effect on query accuracy. Dimensionality techniques are used for handling of high dimensional data as in gene expression microarray analysis, text categorization, with hundreds to tens of thousands of features, with many irrelevant and redundant features and recent research results leads to redundancy based feature selection. The motivation for dimension reduction can be summarized as follows [2] 1) the identification of a reduced set of features that are predictive of outcomes can be very useful from a knowledge discovery perspective. 2) For many learning algorithms, the training and/or classification time increases directly with the number of features. 3) Noisy or irrelevant features can have the same influence on classification as predictive features so they will impact negatively on accuracy [4].

1.2. Real-World Applications: The dimensionality reduction real world applications are **1)** Customer relationship management. 2) Text categorization. **3)** Image retrieval. **4)** Gene expression microarray data analysis [10] (see in Fig (c))**.** 5) Intrusion detection. 6).Protein Classification. 7). Face recognition (see in Fig (d)). 8) Handwritten digit recognition (see in Fig (d)).

I. Applications for Dimensionality Reduction are shown in figures [A] and [B].



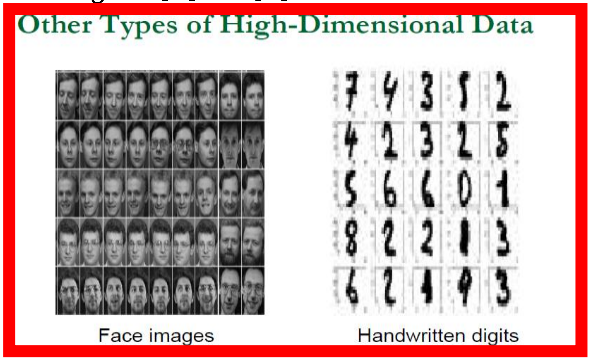
Gene Expression Microarray Analysis

- **Task:** To classify novel samples into known disease types (disease diagnosis)
- **Challenge:** thousands of genes, few samples
- **Solution:** to apply dimensionality reduction

Gene / Sample	M23197_at	U66497_at	M92287_at	...	Class
Sample 1	261	88	4778	...	ALL
Sample 2	101	74	2700	...	ALL
Sample 3	1450	34	498	...	AML
...	...	...	...	...	...

Expression Microarray Data Set

Fig # A : Gene Expression Microarray Analysis



Other Types of High-Dimensional Data

Fig # B: Face recognition and Hand written digits

Vol 04, Special Issue01; 2013
http://ijpaper.com/
PUBLICATIONS OF PROBLEMS & APPLICATION IN ENGINEERING RESEARCH - PAPER
CSEA2012
ISSN: 2230-8547; e-ISSN: 2230-8555

2. Major Techniques of Dimensionality Reduction [1]: The major techniques of dimensionality reduction are 1**)** Feature selection. 2) Feature Extraction (reduction). Feature selection is a process that chooses an optimal subset of features according to objective function. Its objectives are 1)To reduce dimensionality and remove noise 2)To improve mining performance 3)Speed of learning 4)Predictive accuracy 5)Simplicity and comprehensibility of mined results. Feature extraction or Feature reduction refers to the mapping of the original high-dimensional data onto a lower-dimensional space. For a given set of data points of p variables compute their low-dimensional representation. The major differences between feature reduction and feature selection are In feature reduction all original features are used and the transformed features are linear combinations of the original features and in feature selection, only a subset of the original features are selected and they are Continuous versus discrete. The major goals of Techniques of dimensionality reduction are: **1)**High effectiveness, 2)Able to handle both irrelevant and redundant features, 3)Not pure individual feature evaluation, 4)High efficiency, 5)Less costly than existing subset evaluation methods, 6)Not traditional heuristic search methods. Limitations of Existing Methods in Feature Selection [1] are *a) Individual feature evaluation:* Focusing on identifying relevant features without handling feature redundancy, Time complexity: *O* (*N*). *b) Feature subset evaluation:* Relying on minimum feature subset heuristics to implicitly handling redundancy while pursuing relevant features and Time complexity: at least *O* (*N2*).

3. Learning Strategies for dimensionality reduction [2][1]: The learning process is supervised or unsupervised. *A) Unsupervised learning (Clustering):* The class labels of training data are unknown. Given a set of measurements, observations, etc. establish the existence of clusters in the data. B*) Supervised Learning (Classification):* The training data (observation, measurements, etc.) are accomplished by labels indicating the class of the observations. The new data is classified based on the training set. C*) Semi-supervised clustering:* Learning approaches that use user input (i.e. constraints or labeled data). Clusters are defined so that user-constraints are satisfied. The Dominant strategies used in practice are Principal Components Analysis (PCA) which is an unsupervised feature transformation technique and supervised feature selection strategies such as the use of Information Gain for feature ranking/selection. Based on different problem settings there are different criterion for Feature Reduction are 1*) Unsupervised Setting:* To minimize the information loss, *2) Supervised Setting:* To maximize the class discrimination.

4. Models of Feature Selection [1]:

1) Filter model: a) Separating feature selection from classifier learning. b) Relying on general characteristics of data (information, distance, dependence, consistency). c) No bias towards any learning algorithm. d) Fast. *2) Wrapper model:* a) Relying on a Predetermined Classification Algorithm. b) Using Predictive Accuracy as goodness measure. c) High Accuracy and d) Computationally Expensive. *3) Validate Selection Results:*1) *Direct Evaluation* (if we know a priori): It is suitable for artificial data sets. It is based on prior knowledge about data. 2) *Indirect Evaluation* (if we don't know): It is suitable for real-world data seta. It is based on a) Number of features selected, b) Performance on selected features (predictive accuracy, goodness of resulting clusters). c) Speed.

II. FILTER AND WRAPPER MODELS OF DATA FLOW DIAGRAMS ARE SHOWN BELOW IN FIGURES [C] AND [D].

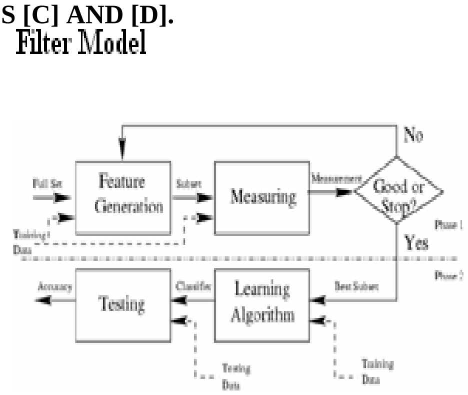


Figure # C

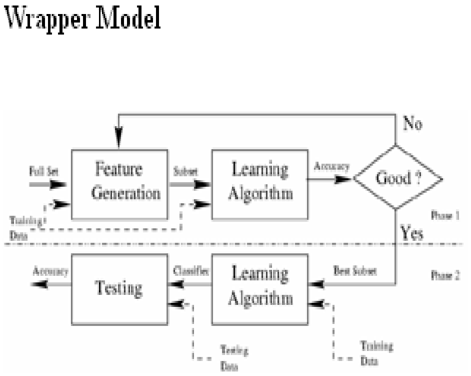
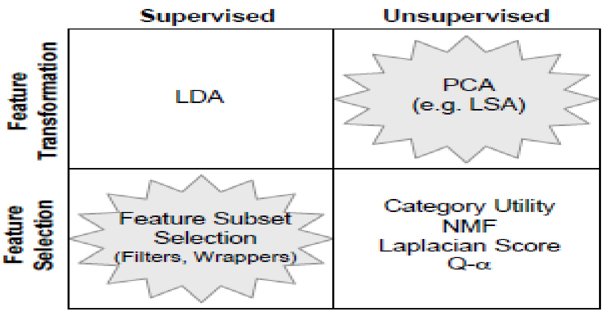


Figure # D

5. Types of Dimensionality Reduction Algorithms [2][7]: *a) Linear dimensionality reduction algorithms:* All data lay in a globally linear space [2]. Examples: a) LSI (Latent Semantic Indexing). b) PCA (Principal Component Analysis). c) LDA (Linear Discriminant Analysis). d) CCA (Canonical Correlation Analysis)**. (**e) PLS (Partial Least Squares). b) *Non linear dimensionality reduction algorithms:* All data lay in a locally linear subspace [2]. Examples: a) Non linear feature reduction using kernels (non linear PCA with kernels). b) Manifold learning.

6. Feature Transformation [2]: Feature transformation refers to a family of data pre-processing techniques that transform the original features of a data set to an alternative, more compact set of dimensions. These techniques can be sub-divided into two categories. *a) Feature extraction* involves the production of a new set of features from the original features in the data, through the application of some mapping. A well-known unsupervised feature extraction method includes Principal Component Analysis (PCA) [2] and spectral clustering the important corresponding supervised approach is Linear Discriminant Analysis (LDA) [1]. *b) Feature generation* involves the discovery of missing information between features in the original dataset, and the augmentation of that space through the construction of additional features that emphasize the newly discovered information. In feature extraction, Where the number of extracted dimensions will generally be significantly less than the original number of features. In contrast, feature generation often expands the dimensionality of the data, though feature selection techniques can subsequently be applied to select a smaller subset of useful features.



	Supervised	Unsupervised
Feature Transformation	LDA	PCA (e.g. LSA)
Feature Selection	Feature Subset Selection (Filters, Wrappers)	Category Utility NMF Laplacian Score Q-α

Figure # E

III. In figure [E] shows the distinguish between supervised and unsupervised techniques and differences between feature transformation and feature extraction techniques .The dominant techniques are feature subset selection and principal component analysis.

7. Feature Reduction Algorithms[1]:

7.1. Unsupervised [1]: 1) Latent Semantic Indexing (LSI): SVD (singular value decomposition) on matrix for the latent structure of data [4]. 2) Independent Component Analysis (ICA). 3) Principal Component Analysais (PCA). 4) Manifold learning algorithms.

7.1.1. Singular Value Decomposition (SVD) [6]: A technique for matrix decomposition. It transforms a single matrix in a product of three matrixes. Latent semantic indexing (LSI): Apply SVD on matrix for the latent structure of data. Decomposition into Eigen values and Eigenvectors is applied to square matrices. Data tables are usually non square in these case we apply SVD. *Eigen Value Decomposition [4]***:** It is a specialization of SVD and in PCA uses an Eigen value decomposition application on the data covariance matrix. *Landmark Multi-Dimensional Scaling (LMDS) [4]: It is an* alternative to classic MDS for large data sets. *Multidimensional Scaling (MD) [2]:* In which all data are randomly projected to a lower dimensionality space. IsoMap & C-IsoMap [1]: It is uses in special cases where Euclidean metric does not apply.

7.2.2. *Principal Components Analysis (PCA):* It is the dominant feature transformation technique is that transforms the data into a reduced space that captures most of the variance in the data. PCA is an unsupervised technique in that it does not take class labels into account. In contrast Linear Discriminant Analysis (LDA) seeks a transformation that maximizes between-class separation.

7.1.3. Manifold learning [1]: Discover low dimensional representations (smooth manifold) for data in high dimension. A manifold is a topological space which is locally Euclidean.

7.2. Supervised [1]: 1) Linear Discriminate Analysis (LDA). 2) Canonical Correlation Analysis (CCA). 3) Partial Least Squares (PLS).

7.2.1. *Linear discriminant analysis (LDA)*[4]: LDA is also closely related to PCA and factor analysis in that they both look for linear combinations of variables which best explain the data. LDA explicitly attempts to model the difference between the classes of data. PCA on the other hand does not take into account any difference in class, and factor analysis builds the feature combinations based on differences rather than similarities. Discriminant analysis is also different from factor analysis in that it is not an interdependence technique: a distinction between independent variables and dependent variables (also called criterion variables) must be made.LDA works when the measurements made on independent variables for each observation are continuous quantities. When dealing with categorical independent variables, the equivalent technique is discriminant correspondence analysis. Linear discriminant analysis (LDA) and the related Fisher's linear discriminant are methods used in statistics, pattern recognition and machine learning to find a linear combination of features which characterizes or separates two or more classes of objects or events. The resulting combination may be used as a linear classifier or more commonly, for dimensionality reduction before later classification.

7.2.2. Canonical Correlation Analysis (CCA)[1]: It measures the linear relationship between two multidimensional variables. It finds two bases, one for each variable, that are optimal with respect to correlations. CCA applications are in economics, medical studies, and bioinformatics.
7.2.3. Partial least squares regression (PLS regression)[1]: It is a statistical method that bears some relation to principal component regression; instead of finding hyper planes of minimum variance between the response and independent variables, it finds a linear regression model by projecting the predicted variables and the observable variables to a new space.
7.3. Semi-supervised: Learning approaches that uses user input (i.e. constraints or labeled data) [2].
8. Methods for Result Evaluation: *1) Learning curves:* For results in the form of a ranked list of features. *2) Before-and-after comparison:* For results in the form of a minimum subset. *3) Comparison using different classifiers:* To avoid learning bias of a particular classifier. *4) Repeating experimental results:* For non-deterministic results [1].

IV. Guideline for Feature Selecting Algorithms are shown in the following figure # F[1]

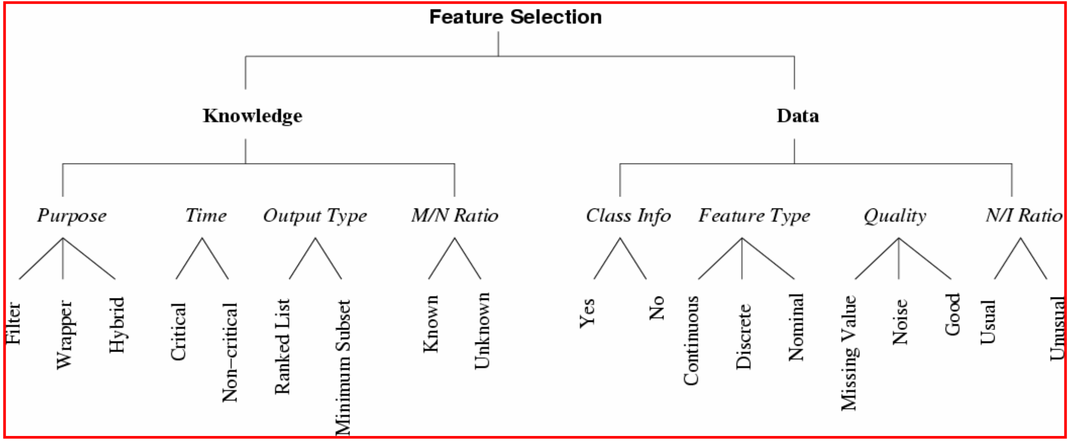


Figure # F

9. Feature selection [2]: Feature selection techniques have become an apparent need in many applications for which datasets with tens or hundreds of thousands of variables are studied. It has better predictive performance. It is computationally efficiency from working with fewer inputs. Cost savings from having to measure fewer features and simpler, more intelligible models. The following Figure [G] shows the feature selection process with validation [1].

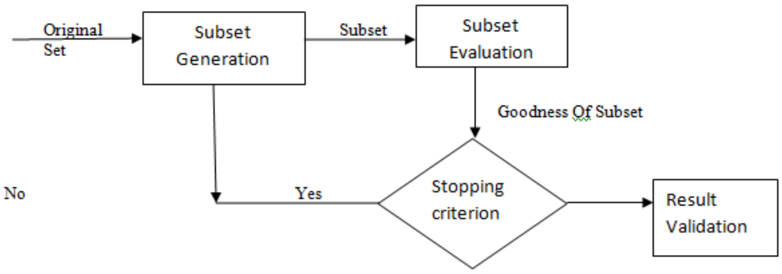
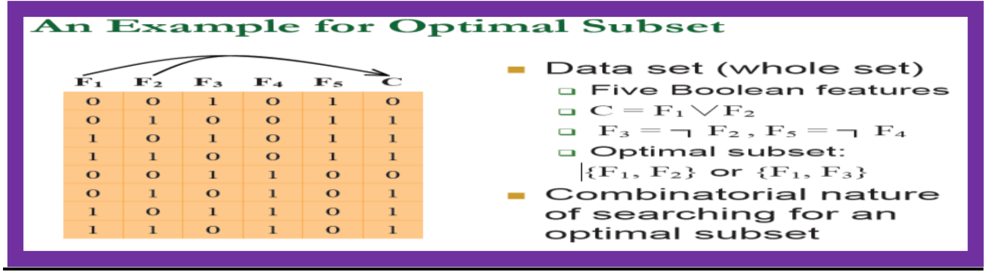


Figure # G: Feature Selection Process with Validation

9.1. Feature Selection Real-World Applications [1]: a) Customer Relationship management. b) Text Categorization. c) Image retrieval: CBIR (Content-Based Image Retrieval) . d) Gene Expression microarray data analysis. e)Intrusion detection.

9.2. Basic Methods for Feature Selection [1]: ***1) Subset Optimality for Classification***. ***2)Search Directions:*** a)Sequential Forward Selection, b)Sequential Backward Elimination, c)Bidirectional Generation, d)Random Generation. ***3) Search Strategies:*** Exhaustive/complete Search, Heuristic search, Nondeterministic search. Example DFS and BFS ***4) Evaluation measures for ranking and Selecting Features:*** a) Information Measures:- Entropy of variable X, Entropy of X after observing Y, Information Gain IG (X / Y) = H(X) - H(X / Y).
b) b) Accuracy Measures, c) Consistency Measures.
9.3. Models of Feature Selection:
a) Filter Model. b) Wrapper Model.
9.4 Representative Algorithms for Classification [1]:

1) *Filter algorithms*: a) Feature ranking algorithms example: Relief [1]. b) Subset search algorithms [13] Example: consistency-based algorithms. c) Focus [13].
2**)** ***Wrapper algorithms:*** a) Feature ranking algorithms Example: SVM (Support Vector Machines) [6]. b) Subset search algorithms, Example: RFE [3] (Recursive Feature Elimination). With reference to the paper [1] the following figure [1] shows an example of searching for an optimal subset.



An Example for Optimal Subset

F_1	F_2	F_3	F_4	F_5	C
0	0	1	0	1	0
0	1	0	0	1	1
1	0	1	0	1	1
1	1	0	0	1	1
0	0	1	1	0	0
0	1	0	1	0	1
1	0	1	1	0	1
1	1	0	1	0	1

- Data set (whole set)
  - □ Five Boolean features
  - □ $C = F_1 \vee F_2$
  - □ $F_3 = \neg F_2$, $F_5 = \neg F_4$
  - □ Optimal subset: $\{F_1, F_2\}$ or $\{F_1, F_3\}$
- Combinatorial nature of searching for an optimal subset

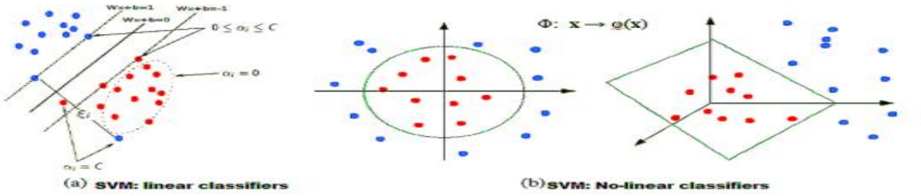
Figure # 1: An Example for Optimal Subset

9.4.1. Filter Algorithm[13]: Focus Algorithm:

Focus:

```
        Input: K - all features x in data D.
                P - Inconsistency rate as evaluation measure.
Initializes: S = { }
For i = 1 to N
   For each subset S of size i
       IF CalP(S,D) = 0   /* CalP(S,D) returns inconsistency*/
       return S .
        Output: S – A Minimum subset that satisfies P
```

9.4.2. Wrapper Algorithms: SVM (Support Vector Machines) [3]:-The Support Vector Machines are a set of supervised learning methods used for classification. They belong to a family of generalized linear classifiers. Their main aim is to simultaneously minimize the empirical classification error and maximize the geometric margin.



(a) SVM: linear classifiers (b) SVM: No-linear classifiers

Figure 2. SVM-linear (Figure 2a) & SVM-nonlinear (Figure 2b)

With reference to the paper [3], the above Figure [2] shows the SVM-linear in Figure (2a) and SVM-nonlinear in Figure (2b).

9.4.3. RFE-SVM: The well-studied RFE-SVM algorithm is a wrapper feature selection method. This generates the ranking of features using backward feature elimination. It was originally proposes to perform gene selection for cancer classification [3]. Its basic idea is to eliminate redundant genes and yields better and more compact gene subsets.

9.4.4. The RFE-SVM algorithm [3]: can be broken into four steps:
1. Train an SVM on the training set;
2. Order features using the weights of the resulting classifier;
3. Eliminate features with the smallest weight;
4. Repeat the process with the training set restricted to the remaining features.

9.5. Representative Algorithms for Clustering: 1) ***Filter algorithms***: Example: a filter algorithm based on entropy measure [13]. 2) ***Wrapper algorithms***: Example: FSSEM –a wrapper algorithm based on EM (Expectation Maximization) clustering algorithm [12].

9.5.1. FSSEM Algorithm [12] (Feature subset selection wrapped around EM clustering)

1) EM Clustering (Expectation Maximization)

a) To estimate the maximum likelihood mixture model parameters and the cluster probabilities of each data point.
b) Each data point belongs to every cluster with some probability.

2) Feature selection for EM

a) Searching through feature subsets.
b) Applying EM on each candidate subset.
c) Evaluating goodness of each candidate subset based on the goodness of resulting clusters.

9.5.2. Effect of Features on Clustering: Example from (*Dash et al., ICDM, 2002*) Synthetic data in (3,2,1)-dimensional spaces,1)75 points in three dimensions in Fig(3) , 2)Three clusters in F1-F2 dimensions in Fig(4), 3) Each cluster having 25 points in Fig(5) .

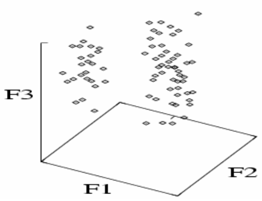


Figure # 3

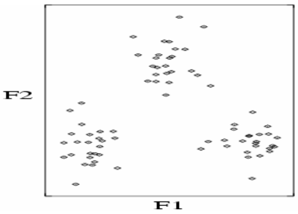
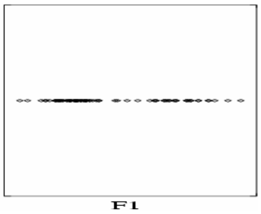


Figure #4



Figures #5

10. Feature Reduction Algorithms [1]:

a) Linear Models: 1) Principal Component Analysis: Principal component analysis (PCA) is to reduce the Dimensionality of a data set by finding a new set of variables, smaller than the original set of variables and retains most of the sample's information. By information we mean the variation present in the sample, given by the correlations between the original variables. The new variables, called principal components (PCs), are uncorrelated, and are ordered by the fraction of the total information each retains. **Applications of PCA**: 1) Eigen faces for recognition, 2) Principal Component Analysis for clustering gene expression data, 3) Probabilistic Disease Classification of Expression-Dependent Proteomic Data from Mass Spectrometry of Human Serum. **PCA Disadvantages**: **1.** not efficient for non linear data.

10.1. Algebraic Derivation Of PCs (Principal Components)

Main steps for computing PCS

- Form The Covariance matrix S
- Compute its Eigen vectors:
- The first p eigenvectors $\{a_i\}_{i=1}^{p}$ form the p PCs $G \leftarrow [a_1, a_2, \cdots, a_p]$
- The transformation G consists of the p PCs.

A test point $x \in \Re^d \rightarrow G^T x \in \Re^p$.

11) Feature Reduction Algorithms [1]: b)Non Linear Model : Isomap:: In statistics, **Isomap** is one of several widely used low-dimensional embedding methods, where geodesic distances on a weighted graph are incorporated with the classical scaling (metric multidimensional scaling). Isomap is used for computing a quasi-isometric, low-dimensional embedding of a set of high-dimensional data points. The algorithm provides a simple method for estimating the intrinsic geometry of a data manifold based on a rough estimate of each data point's neighbors on the manifold. Isomap is highly efficient and generally applicable to a broad range of data sources and dimensionalities.

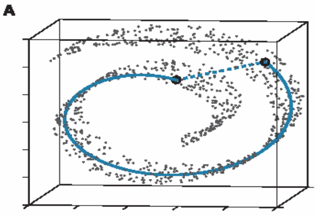


Figure (A)

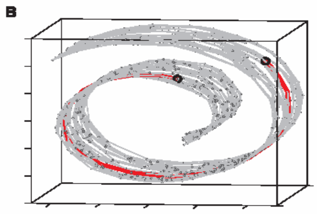


Figure (B)

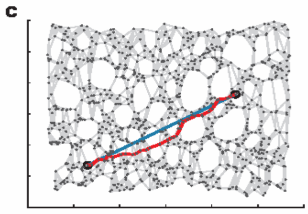


Figure (C)

With reference to the paper [1] we are presented the example 1) In Figure (A) Constructing neighborhood graph G, **k**-nearest neighborhood (k=7). D_G is 1000 by 1000(Euclidean) distance matrix of two neighbors. 2) In Figure (B) for each pair of points in G, Computing shortest path distances----geodesic distances. 3) In Figure(C) use Classical MDS (Multidimensional scaling) with geodesic distances. Euclidean distance-→Geodesic distance (see Figure C).

11.1. The Isomap Algorithm:

Step1: Construct Neighborhood Graph. (See Figure A)
Step2: Compute Shortest Paths. (See Figure B)
Step3: Construct D-dimensional Embedding. (See Figure C)

11.2. IsomapAdvantages: 1) Nonlinear. 2) Globally optimal: Still produces globally optimal low-dimensional Euclidean representation even though input space is highly folded, twisted, or curved. 3) Guarantee asymptotically to recover the true dimensionality.

11.3. Isomap Disadvantages: 1) May not be stable, dependent on topology of data. 2) Guaranteed asymptotically to recover geometric structure of nonlinear manifolds–As N increases, pair wise distances provide better approximations to geodesics, but cost more computation–If N is small, geodesic distances will be very in accurate.

12. Laplacian Eigen maps [11]: Laplacian Eigen maps for Dimensionality Reduction and Data Representation. Key steps: 1) Build the adjacency graph. 2) Choose the weights for edges in the graph (similarity). 3) Eigen-decomposition of the graph laplacian. 4) Form the low-dimensional embedding

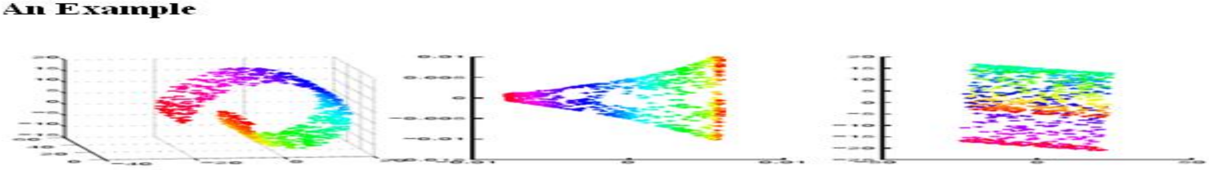


Figure (1):"Swiss Roll" Figure(2): Laplacian Representation Figure(3): PCA representation

With reference to the paper [6] we compare a spectral 2-D representation of the "Swiss Roll" with "PCA". As a result we concluded that PCA is limited to projections and cannot produce a good representation of non-linear data.

13. Trends in Dimensionality Reduction [1]: *1) Dimensionality reduction for complex data* :a) Biological data. b) Streaming data. *2) Incorporating prior knowledge:* a) Semi-supervised dimensionality reduction. *3) Combining feature selection with extraction:* a) Develop new methods which achieve feature "selection" while efficiently considering feature interaction among all original features.

14. Conclusion:

The objective with this paper was to provide the study of the variety of strategies that can be employed for dimensionality reduction when processing high dimension data. Dimension reduction techniques are often applied as a data pre-processing step or as part of the data analysis to simplify the data model [2]. It involves in the identification of suitable low-dimensional representation for the original high-dimensional data set. By working with this reduced representation, tasks such as classification or clustering can often yield more accurate and readily interpretable results, while computational costs may also be significantly reduced. PCA is the dominant technique if the data is not labelled. If the data is labelled, the LDA can be applied to discover a projection of data that separates the classes [1]. When feature selection is required and the data is labelled then the problem is well posed. A variety of filter and wrapper-based techniques for feature selection were discussed. In this paper we conclude that unsupervised feature selection is more difficult problem than the supervised. In this paper we are discussed about the variety of feature extraction algorithms for unsupervised learning, supervised learning and dimensionality reduction linear and non-linear applications.

15. References:

1. Lei Yu Binghamton University, Jieping Ye,Huan Liu ,Arizona State University, "Dimensionality Reduction for datamining-Techniques, Applications and Trends".
2. " Dimension Reduction" P_adraig Cunningham University College Dublin Technical Report UCD-CSI-2007-7 August 8th, 2007
3. "A Novel RFE-SVM-based Feature Selection Approach for Classification" by Mouhamadou Lamine Samb, Fodé Camara, Samba Ndiaye, Yahya .
4. "A Survey of Dimension Reduction Techniques" I.K. Fodor, US Department of Energy, 2002.
5. "Sparse Multidimensional scaling using landmark points", vin de silva, Joshua B.tenendaum, 2004.
6. "Singular Value Decomposition (SVD)" M. Vazirgiannis, M. Halkidi, D. Gunopulos-PKDD 2006.
7. "Global versus local methods in nonlinear dimensionality reduction". vin de silva, Joshua B.tenendaum,2004.
8. "Novel Aspects in Unsupervised Learning: Semi-Supervised and Distributed Algorithms". Mariahalkidi,Michalis Vazirgiannis, Dimitrios Gunopulos.
9. "Gene expression microarray data analysis", Golubet al., 1999 (MIT), Xing et al., 2001 (UC Berkeley).
10. J.G. Dy and C.E.Broadley. Feature selection for unsupervised Learning. The journal of Machine Learning Research, 5:845-889, 2004.
11. "Laplacian Eigenmaps for dimensionality reduction and data representation" M. Belkin, P.Niyogi.
12. "FSSEM –a wrapper algorithm based on EM (Expectation Maximization) clustering algorithm", Dyand Brodley, ICML, 2000.
13. "Subset search algorithms" Example: consistency-based algorithms. "Focus" (Almuallim & Dietterich, 1994).
14. "Feature ranking algorithms" example: Relief, Kira& Rendell 1992.