Journal of Physics: Conference Series **490** (2014) 012081 doi:10.1088/1742-6596/490/1/012081

Multiscale principal component analysis

A A Akinduko and A N Gorban

Mathematics Department, University of Leicester, Leicestershire, LE1 7RH, UK

E-mail: aaa78@le.ac.uk and ag153@le.ac.uk

Abstract. Principal component analysis (PCA) is an important tool in exploring data. The conventional approach to PCA leads to a solution which favours the structures with large variances. This is sensitive to outliers and could obfuscate interesting underlying structures. One of the equivalent definitions of PCA is that it seeks the subspaces that maximize the sum of squared pairwise distances between data projections. This definition opens up more flexibility in the analysis of principal components which is useful in enhancing PCA. In this paper we introduce scales into PCA by maximizing only the sum of pairwise distances between projections for pairs of datapoints with distances within a chosen interval of values $[l,u]$. The resulting principal component decompositions in Multiscale PCA depend on point (l,u) on the plane and for each point we define projectors onto principal components. Cluster analysis of these projectors reveals the structures in the data at various scales. Each structure is described by the eigenvectors at the medoid point of the cluster which represent the structure. We also use the distortion of projections as a criterion for choosing an appropriate scale especially for data with outliers. This method was tested on both artificial distribution of data and real data. For data with multiscale structures, the method was able to reveal the different structures of the data and also to reduce the effect of outliers in the principal component analysis.

1. Introduction

In 1901, Pearson proposed approximating high dimensional data with lines and planes and hence invented the Principal Component Analysis (PCA). At present time, PCA is a powerful analysis tool with many interesting applications which include: dimension reduction, blind source separation, data visualization, image compression, and with relevance in many applied disciplines such as quantitative finance, biology, pharmaceutics, taxonomy, healthcare and many more [1-3].

However, despite the many applications of PCA, it is not without its drawbacks. An example of such drawbacks is that PCA is based on the covariance matrix which is sensitive to outliers. In this paper, outliers are defined as data elements with large distance from the other data elements in a data sample. Even though outliers can be filtered before performing PCA on the dataset, however in some contexts, identifying outliers could be cumbersome. In addition to the above, datasets are usually noisy (here we define noises as data elements with rather small variance) and the presence of noise in data analysis can further obfuscate the underlying structure(s) of the data being investigated [5].

One of the definitions of PCA is that PCA finds subspaces (lines, planes or higher dimensional subspaces) that maximize the sum of point-to-point squared distances between the orthogonal projections of data points to them. In this paper, scale was introduced to enhance the performance of PCA on datasets. That is, we will use in the definition of multiscale PCA maximization of the sum of point-to-point squared distances between the orthogonal projections of data points for the pairs of points with distances in some intervals. The result of this is PCA decomposition of the data which

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

depend on the scale chosen. A further study of these PCA decompositions reveals some underlying structures which could have been obfuscated by other structures such has the presence of outliers or repeated patterns as shown later. We also proposed a criterion for determining the appropriate scale for computing the principal components for data with outliers.

2. Definitions and Mathematical Background

2.1. Weighted PCA

Consider the problem of finding the principal component using weighted pairwise distances of projected data. This problem is stated below.

$$\begin{aligned} &\sum_{i<j} w_{ij}[dist^2(P_L\mathbf{x}_i, P_L\mathbf{x}_j)] \\ D_X &= \sum_{i<j} w_{ij} \| P_L(\mathbf{x}_i - \mathbf{x}_j)\|^2 \to \max \\ &\text{Subject to: } (\mathbf{v}_\alpha, \mathbf{v}_\beta) = \delta_{\alpha\beta}. \end{aligned} \tag{1}$$

Where $w_{ij} = w_{ji}$ is the non-negative weight assigned to the distance between element i and j and $w_{ij} = 0,$ for $i = j$.

Let $L^w = [L^w{}_{ij}] = \left(\delta_{ij}(\sum_{j=1}^{n} w_{ij}) - w_{ij} \right)$, where $w_{ij} = 0,$ for $i = j$. Let us introduce the matrix $\tilde{M}_{\alpha\beta} = (X_\alpha{}^T L^w X_\beta)$. In these notations, the problem given by (15) is reduced to

$$\begin{aligned} &\max_{v_1,\ldots,v_k} \sum_{\alpha=1}^{k} \mathbf{v}_\alpha^T \tilde{M} \mathbf{v}_\alpha \\ &\text{Subject to } (\mathbf{v}_\alpha, \mathbf{v}_\beta) = \delta_{\alpha\beta} \;\; \alpha, \beta = 1,2,..,k, \end{aligned} \tag{2}$$

where $\tilde{M}$ is a symmetric positive semi-definite matrix, and the eigenvectors corresponding to the sorted eigenvalues of the matrix $\tilde{M}$ is a maximizer of the constrained maximization problem (2). In the case of degenerated eigenvalues, the set $\mathbf{e}_1,\ldots,\mathbf{e}_m$ is not uniquely defined.

3. Multiscale PCA (MPCA)

In this section, we introduce the Multiscale PCA (MPCA) algorithm to enhance the robustness of the PCA especially in revealing hidden structure(s) that may be present in dataset but which the conventional approach might not reveal. MPCA compute principal components by maximizing the sum of pairwise distances between data projection for only pairs of datapoints for which the distance is within the chosen scale:

$$\begin{cases} w_{ij} = 1 & l \le \| \mathbf{x}_i - \mathbf{x}_j \|_2 \le u \\ w_{ij} = 0 & otherwise. \end{cases} \tag{3}$$

In the scale interval, (l,u), l is the lower limit of the scale and u is the upper limit. Let $d^{\min}$ be the minimum pairwise distance greater than zero and $d^{\max}$ be the maximum pairwise distance in the data. We select the pairs (l,u) from a triangle $\Delta=\{(l,u): d^{\min}\leq l<u\leq d^{\max}\}$.

With this control over the pairwise distances, we are able compute PCA at various scales and the outcome of this is scale dependent PCA which can reveal interesting underlying structure(s) that may be present in data. For example, reducing the upper limit of the scale while keeping the lower limit at 0 translate to computing PCA by considering smaller distances and excluding very large distances. This has the effect of minimizing without explicit exclusion the contribution of certain influential data elements in the analysis of the principal components.

3.1. The Multiscale PCA Algorithm

Here we discuss the Multiscale PCA Algorithm.

1. Given the data sample.
2. Centralize the data by subtracting the mean of the variables from each observation.
3. Find the dissimilarity matrix by computing the Euclidean distance.
4. Choose an appropriate scale between 0 and the maximum distance. For easy analysis, a scale between 0 and 1 could be chosen and then multiplied by the maximum distance. For this paper when using scale between 0 and 1 we call it standard scale.
5. Calculated the binary weight as given in equation (27)
6. Calculate the matrix L^w as given below
7. Calculate the matrix $A=Y^TL^wY$, where Y is the centralized data matrix.
8. Find the sorted eigenvalues of the matrix A in descending order of magnitude and project the data onto their corresponding eigenvectors. This will be the principal components at the selected scale.

4. Clustering Analysis on the Interval of Scales

To further study these structures, we consider clustering analysis on the interval of scales and we introduce the Ratio of Distortion in this section. We represent the PCA structure of the data at any point (l,u) by the sum of the projectors corresponding to MPCA at that point. This will be denoted by $\rho_k=\sum_{i=1}^{k}\mathbf{e}_i\otimes\mathbf{e}_i$, $k=1,2,,,,m-1$. The full description of the principal components decompositions of data X is given by an ordered set ("cortege") of projectors $\rho_1,\rho_2,...,\rho_{m-1},\rho_m=1$ (compare to [4]) If we arrange the e_i $i=1,2,...,k$ as columns of matrix E, then $\rho_kX=EE^TX$ and $\rho_k=EE^T$. For $k=m,\ EE^T=\mathbf{I}$.

MPCA lead to scale dependent PCA structures and with these PCA structures represented as defined above, we can study the structures in our data further by analyzing these projectors.

The PCA structures associated with two different points on the plane is said to be similar if their corresponding projectors ρ_k are similar. We guess that in some cases there are clear internal structures in the data which depend on scales. Performing MPCA on the data leads to a continuum of PCA structures depending on scales used and to reveal the structures in the data, we join scales with similar PCA structures and separate scales with dissimilar PCA structures. This leads to the idea of clustering of scales. We represent the distance between two points on the scale by the distance between their corresponding PCA structures. Clustering analysis of the scales group similar PCA structures together and this reveals some structures in the data. We describe each cluster by the projector corresponding to the medoid point of the cluster. We denote the projector ρ_k at a point χ_p by ρ_{χ_p}. For any pair of points χ_p,χ_q in the space of scales we can compute the distance between the associated projectors

$\rho_{\chi_p}, \rho_{\chi_q}$ for a given k using the matrix norm. This clustering is illustrated on a simple example of data with hierarchical microstructures (Fig. 1-3).

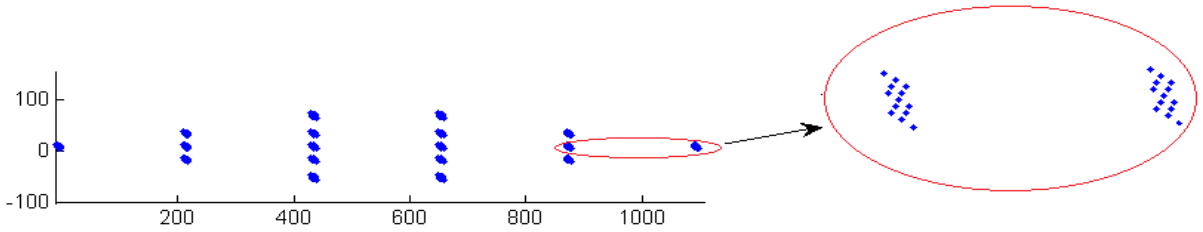


Figure 1. Scatter plot of data with hierarchical microstructure.

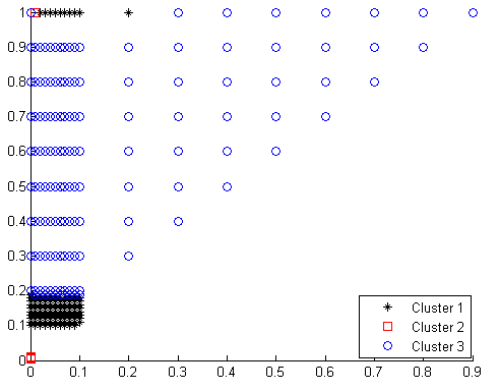


Figure 2. Cluster of scales on the plane for data from Fig. 1.

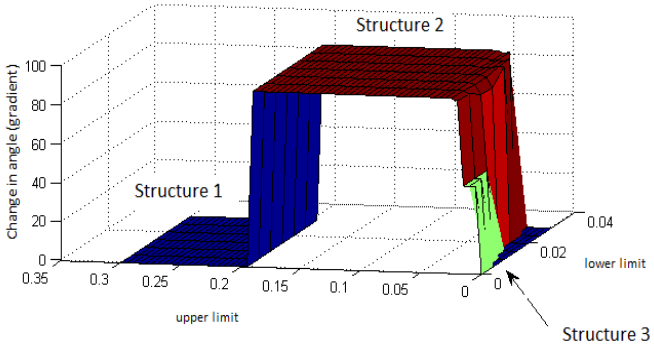


Figure 3. The angle between the first principal component using PCA and the first principal component using MPCA at a given scale.

5. Conclusion

MPCA is developed to solve the problem of revealing hidden geometric structures in data. The result of MPCA on data leads to a continuum of PCA structures of the data which is dependent on the intervals chosen. For data with clear multiscale structures, the cluster analysis of the scales reveals some underlying structures in the data which conventional PCA cannot reveal due to the fact that such structures are obfuscated by other structures of higher variance.

References

[1] Jolliffe I T 2002 Principal Component Analysis Second Edition (New York: Springer)
[2] Gorban A N and Zinovyev A Y 2009. Principal Graphs and Manifolds *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods and Techniques* (Hershey, PA: IGI Global) pp 28-59.
[3] Burges C J C 2010 Geometric Methods for Feature Extraction and Dimensional Reduction - A Guided Tour *Data Mining and Knowledge Discovery Handbook* (New York: Springer) ed O Maimon and L Rokach . 2nd Edition ISBN 978-0-387-09822-7 pp 53-82.
[4] Arnold R and Jupp P E 2013 Statistics of orthogonal axial frames, *Biometrika* pp 1-16
[5] Koren Y and Carmel L 2004 Robust linear dimensionality reduction, Visualization and Computer Graphics, *IEEE Transactions* vol.10, no.4, pp.459-70.
[6] Akinduko A A, Gorban A N 2013 arXiv:1307.8339 [stat.ME]