COMSTA: 6664 Model 3G pp. 1–13 (col. fig: NIL)

ARTICLE IN PRESS

Computational Statistics and Data Analysis xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Generalized biplots for stress-based multidimensionally scaled projections

J.T. Fry [a,*], Matt Slifko [b], Scotland Leman [b]

[a] Department of Mathematics, Bucknell University, United States
[b] Department of Statistics, Virginia Tech, United States

ARTICLE INFO

Article history:
Received 29 November 2017
Received in revised form 5 August 2018
Accepted 5 August 2018
Available online xxxx

Keywords:
Biplots
Multidimensional scaling
Principal component analysis
Classical multidimensional scaling
Stress function
Low dimensional projection

ABSTRACT

Dimension reduction and visualization are staples of data analytics. Methods such as Principal Component Analysis (PCA) and Multidimensional Scaling (MDS) provide low dimensional (LD) projections of high dimensional (HD) data while preserving an HD relationship between observations. Traditional biplots assign meaning to the LD space of a PCA projection by displaying LD axes for the attributes. These axes, however, are specific to the linear projection used in PCA. Stress-based MDS (s-MDS) projections, which allow for arbitrary stress and dissimilarity functions, require special care when labeling the LD space. An iterative scheme is developed to plot an LD axis for each attribute based on the user-specified stress and dissimilarity metrics. The resulting plot, which contains both the LD projection of observations and attributes, is referred to as the Generalized s-MDS Biplot. The details of the Generalized s-MDS Biplot methodology, its relationship with PCA-derived biplots, and an application to a real dataset are provided.

1. Introduction

Dimension reduction and data visualization are staples of any good analysis, whether as an exploratory or disseminating tool. Visualizations provide an opportunity for the analyst to discover underlying structures and gain insights not easily gleaned by examining the raw data itself (Keim, 2002). Early techniques, such as Chernoff's faces (Chernoff, 1973) and star plots (Chambers, 1983) simultaneously display all attributes for a given observation at once. These techniques, however, lose utility when the data becomes larger and the dimensionality increases. Consequently, many exploratory approaches for modern data involve dimension reduction. We briefly discuss two popular data reduction techniques: Principal Component Analysis (PCA) and Multidimensional Scaling (MDS).

PCA and MDS are among the most commonly used methods for reducing dimensionality. PCA provides a new set of orthogonal axes in the directions that maximize the variance of the reduced dimensional, projected data. To produce the low dimensional (LD) projection, the user removes the axes that capture the lowest amounts of variance. MDS is a more general framework for finding an LD projection that most closely (based on some criterion) matches high dimensional (HD) dissimilarities. Stress-based MDS (s-MDS) provides additional flexibilities by allowing users to select dissimilarity metrics in both the HD attribute space and the LD projected space, as well as the stress function used to related HD and LD dissimilarities. PCA and MDS will be more thoroughly discussed in Sections 3 and 4. It should be mentioned that PCA is a specific case of an MDS algorithm, which we provide details in Section 4.2.

* Correspondence to: Bucknell University, Department of Mathematics, Olin Science Building, 570 Vaughan Lit Drive, Lewisburg, PA 17837, United States.
E-mail address: jt.fry@bucknell.edu (J.T. Fry).

https://doi.org/10.1016/j.csda.2018.08.003
0167-9473/

Please cite this article in press as: Fry J.T., et al., Generalized biplots for stress-based multidimensionally scaled projections. Computational Statistics and Data Analysis (2018), https://doi.org/10.1016/j.csda.2018.08.003.

Although s-MDS preserves the average dissimilarity between observations based on the chosen stress function, we lose a sense of how the original attributes affect positioning. For example, the PCA axes are principal components representing a linear combination of the attributes. To rectify this, researchers have developed ways of labeling the LD space. Gabriel (1971) developed the original biplot, a PCA-specific technique that adds vectors to the PCA projection to represent a projection of the HD axes. It is important to emphasize that the *bi* in biplot refers to the simultaneous display of information about (1) observations and (2) attributes in the same space, rather than referring to the dimensionality of the LD space. Gower (1992) expanded upon the PCA Biplot by allowing other distance metrics. Using an approximation based on Euclidean distance, axes are linearly projected based on the specified distance function. Referred to as the Nonlinear Biplot, these projections often create highly curved LD axes. Cheng and Mueller (2016) propose the Data Context Map, which also displays both the observations and attributes as points in the same space. This is achieved by creating a large composite matrix with observations and attributes that are treated as observations. As a consequence, the projection of the observations is affected by the treatment of the attributes, instead of simply labeling the already created projection.

Our proposed solution for providing context to the LD space of an s-MDS projection is the Generalized s-MDS Biplot. The remainder of the manuscript is organized as followed. First, we review PCA and introduce the original biplot. We then discuss s-MDS and Classical MDS, while establishing the connection between them. Next, we detail our method for the Generalized s-MDS Biplot, and show its connection with the PCA Biplot. Finally, we apply our Generalized s-MDS Biplot to a real dataset and discuss the generated projections.

2. Notation

For clarity and ease of reading, we define some notation that will be used throughout the manuscript. We let $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)'$ denote an $n \times p$ matrix of HD data containing n observations of p continuous attributes. We assume that $\boldsymbol{X}$ is full column rank and has unitless columns with column means of 0. Utilizing the singular value decomposition (SVD), we can write $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Lambda}^{1/2}\boldsymbol{V}'$, where $\boldsymbol{\Lambda} = diag(\lambda_1, \ldots, \lambda_p)$ ($\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$) is a diagonal matrix with the p positive eigenvalues of $\boldsymbol{X}'\boldsymbol{X}$ and $\boldsymbol{X}\boldsymbol{X}'$ in descending order, and $\boldsymbol{U} = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p)$ and $\boldsymbol{V} = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p)$ are $n \times p$ and $p \times p$ orthonormal matrices whose columns contain the eigenvectors of $\boldsymbol{X}\boldsymbol{X}'$ and $\boldsymbol{X}'\boldsymbol{X}$, respectively. We can further partition the SVD as $\boldsymbol{X} = (\boldsymbol{U}_1, \boldsymbol{U}_2)diag(\boldsymbol{\Lambda}_1, \boldsymbol{\Lambda}_2)^{1/2}(\boldsymbol{V}_1, \boldsymbol{V}_2)'$ where $\boldsymbol{\Lambda}_1$ contains the first m eigenvalues and $\boldsymbol{U}_1$ and $\boldsymbol{V}_1$ contain the corresponding eigenvectors. We let $\boldsymbol{Z} = (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n)'$ be an $n \times m$, $m < p$ matrix of LD coordinates corresponding to $\boldsymbol{X}$. Similarly to $\boldsymbol{X}$, $\boldsymbol{Z}$ can be decomposed into $\tilde{\boldsymbol{U}}\tilde{\boldsymbol{\Lambda}}^{1/2}\tilde{\boldsymbol{V}}'$.

3. Principal Component Analysis

We begin with data matrix $\boldsymbol{X}$, which consists of n observations of p attributes. PCA finds a new orthogonal basis, whose elements are linear combinations of the original attributes, that maximizes the total variance in the projected space. To find new basis vectors $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_p$, we sequentially solve the following constrained optimization:

$$\begin{aligned} &\underset{\boldsymbol{e}_j}{ArgMax} && Var(\boldsymbol{X}\boldsymbol{e}_j), \\ &\text{subject to:} && \boldsymbol{e}_j'\boldsymbol{e}_j = 1, \\ & && \boldsymbol{e}_j'\boldsymbol{e}_k = 0,\ j \neq k. \end{aligned}$$

Solving for $\boldsymbol{e}_1$ provides the principal direction that captures the most variance. Given $\boldsymbol{e}_1$, $\boldsymbol{e}_2$ is the principal direction that captures the second-most variance while being orthogonal to $\boldsymbol{e}_1$; we continue in this manner until we solve for all p basis vectors. The constraints ensure that we do not simply make $\boldsymbol{e}_j$ extremely large to achieve the maximization and also that the basis vectors are orthogonal. PCA has a simple, closed-form solution: $\boldsymbol{e}_j = \boldsymbol{v}_j$, the eigenvector associated with the jth largest eigenvalue of $\boldsymbol{X}'\boldsymbol{X}$. We can then obtain orthogonal, HD coordinates $\tilde{\boldsymbol{X}}$ via the linear projection $\boldsymbol{X}\boldsymbol{V}$.

While the original goal of PCA is to find a new basis for the data, it is convenient to use PCA for dimension reduction as well. To reduce the data to m dimensions, we need only to keep the first m columns of $\tilde{\boldsymbol{X}}$, which provides the linear projection of $\boldsymbol{X}$ along the first m principal components. The quality of the projection can easily be quantified by the proportion of total variance preserved, given by $(\sum_{j=1}^{m}\lambda_j)/(\sum_{j=1}^{p}\lambda_j)$. When the proportion of variance captured is higher, the projection more accurately reflects the HD structure. When $m \in \{2, 3\}$, PCA is a common tool for visually exploring underlying structures.

3.1. Principal Component Analysis Biplot

Gabriel's PCA Biplot (Gabriel, 1971) is an extension of the PCA projection that labels the projection space in terms of the HD attributes. Consider the SVD of the HD data $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Lambda}^{1/2}\boldsymbol{V}'$. $\boldsymbol{X}$ can be further decomposed into $b\boldsymbol{U}\boldsymbol{\Lambda}^{\alpha/2}\boldsymbol{\Lambda}^{(1-\alpha)/2}\boldsymbol{V}'/b$, where $\alpha \in [0, 1]$ and b is a scalar. Gabriel shows that we can consider $b\boldsymbol{U}\boldsymbol{\Lambda}^{\alpha/2}$ as information about the observations and $\boldsymbol{V}\boldsymbol{\Lambda}^{(1-\alpha)/2}/b$ as information about the attributes embedded in the raw data.

As in PCA, for dimension reduction we extract the first m columns of each matrix, $\boldsymbol{U}_1$, $\boldsymbol{\Lambda}_1$, and $\boldsymbol{V}_1$. The matrix product $\tilde{\boldsymbol{X}} = \boldsymbol{U}_1\boldsymbol{\Lambda}_1^{1/2}\boldsymbol{V}_1'$ is a rank deficient approximation of $\boldsymbol{X}$. To obtain an LD projection of the observations, we plot the n rows of $\boldsymbol{Z} = b\boldsymbol{U}_1\boldsymbol{\Lambda}_1^{\alpha/2}$. Similarly, we plot the p rows of $\boldsymbol{\Lambda}_1^{(1-\alpha)/2}\boldsymbol{V}_1/b$ as arrow-vectors (axes) from the origin, indicating the direction of the projection in terms of each attribute. Longer arrows represent the important variables driving the projection.

Please cite this article in press as: Fry J.T., et al., Generalized biplots for stress-based multidimensionally scaled projections. Computational Statistics and Data Analysis (2018), https://doi.org/10.1016/j.csda.2018.08.003.