Journal of Multivariate Analysis 140 (2015) 47–59

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Predictive nonlinear biplots: Maps and trajectories

S.K. Vines

Department of Mathematics and Statistics, The Open University, Walton Hall, Milton Keynes, UK

ARTICLE INFO

Article history:
Received 12 September 2014
Available online 25 April 2015

AMS subject classifications:
62H99
62H25
62-09

Keywords:
Euclidean-embeddable dissimilarity function
Nonlinear biplot
Normal projection
Prediction
Prediction region
Predictive trajectory

ABSTRACT

When the difference between samples is measured using a Euclidean-embeddable dissimilarity function, observations and the associated variables can be displayed on a nonlinear biplot. Furthermore, a nonlinear biplot is predictive if information on variables is added in such a way that it allows the values of the variables to be estimated for points in the biplot. In this paper an r dimensional biplot which maps the predicted value of a variable for every point in the plot, is introduced. Using such maps it is shown that even with continuous data, predicted values do not always vary continuously across the biplot plane. Prediction trajectories that are appropriate for summarising such non-continuous prediction maps are also introduced. These prediction trajectories allow information about two or more variables to be estimated even when the underlying predicted values do not vary continuously.

1. Introduction

A biplot is a plot in which information about samples and variables is simultaneously displayed. The term 'biplot' was first coined in [5], with the prefix 'bi-' intended to reflect that two different modes are displayed rather than the number of dimensions used for the display. Typically biplots are 2-dimensional, making them easy to display on paper or on computer screens, although this does not have to be the case.

This paper will deal with biplots in which samples are represented by points on the plot whose positions are determined using classical scaling. Thus this method falls within the class of multidimensional techniques that produce low-dimensional representations of points based on their dissimilarities.

Furthermore, it will be assumed that the functional form of the underlying dissimilarity function is known, and is in the class of Euclidean-embeddable functions. This class of dissimilarity functions includes the familiar Pythagorean distance, along with other dissimilarity functions such as the square-root of the City Block (Manhattan) distance and Clark's distance [10]. This means that the low-dimensional representation of the points is a projection of a high-dimensional configuration that exactly represents the dissimilarities between points instead of rather an low-dimensional approximate representation of dissimilarities obtained directly by, for example, minimising Stress or S-Stress.

The aim will be to add information about variables to the plot in such a way that values of the variables can be associated with the configuration of the points. This will primarily be done by adding trajectories to the plot, one for each variable. The trajectories will in general be nonlinear, and hence such biplots are known as nonlinear biplots [9]. These trajectories will be calculated by adding points that correspond to positions along an axis.

E-mail address: s.k.vines@open.ac.uk.

http://dx.doi.org/10.1016/j.jmva.2015.04.010
0047-259X/

In multidimensional scaling, different approaches are available to add points to an existing configuration of points (see for example [2]). In this paper, an approach which matches the construction of the existing configuration will be followed. That is, knowledge of the form of the dissimilarity function will be used to calculate the exact position of the extra points in a sufficiently high-dimensional space. Projection is then used to place these points on the biplot.

In linear (PCA) biplots (biplots that are produced when Pythagorean distance is chosen to be the dissimilarity function), the position of marker points on the trajectories representing variables depends on whether trajectories are to be used for interpolation or prediction [8, p. 15]. That is, on whether the trajectory is going to be used to placing a new observation in the most appropriate place on the biplot (interpolation) or to be used to determine what values of the original variables are best associated with a point on the biplot, usually one of those already plotted (prediction). In nonlinear biplots the trajectories themselves also generally depend on whether they are going to be used for interpolation or prediction (see for example [11]). Here the focus will be on prediction trajectories. That is, trajectories complete with marker points, suitable for prediction purposes.

On nonlinear biplots, prediction trajectories also depend on the method by which points in the biplot are to be projected on to the trajectory. Here the focus will be on normal projection prediction trajectories. With normal projection prediction trajectories, a projection P^* of any point P in the biplot on the trajectory is where the line PP^* intersects the trajectory orthogonally. The position of P^* along the trajectory then indicates the predicted value to be associated with P.

In the next section, existing methodology that has been used to calculate normal projection prediction trajectories for nonlinear biplots will be described. This existing methodology relies on the assumption that the dissimilarity function is smooth. This assumption is not always appropriate as there are Euclidean-embeddable dissimilarity functions that are not smooth everywhere. Hence the existing methodology cannot be applied to all such dissimilarity functions. So in Section 3 an alternative approach to prediction in nonlinear biplots is introduced so that normal projection prediction trajectories can be calculated regardless of whether the dissimilarity function is smooth.

As Section 3 will also show, the alternative approach to prediction introduced in this paper will allow prediction maps for each variable to be created — that is, plots where every point is coloured according to the value it predicts. Such maps can be used to explore how predicted values vary across the biplot plane. For example in Section 3.3 prediction maps will used to illustrate a new observation about nonlinear biplots: the dimension of prediction regions (regions on the biplot that all predict the same value of a variable) depends on whether the dissimilarity function is smooth. A mathematical explanation for this observation will be given in Section 3.4.

Prediction maps, by colouring every point in the biplot, effectively preclude the depiction of more than one variable on the same biplot. So, in Section 4 for the special case of 2-dimensional nonlinear biplots with 2-dimensional prediction regions, a new method of calculating a prediction trajectory through the biplot space to approximate the prediction regions is proposed. Then, by superimposing the prediction trajectories for the different variables on the same plot, the ability to compare different variables on the same biplot is restored.

2. Displaying variables in nonlinear biplots

2.1. Preliminaries

Let $\boldsymbol{X}$ represent an $n \times p$ data matrix of n samples and p variables, with its ith row vector $\boldsymbol{x}_i' = (x_{i1}, \ldots, x_{ip})$ representing the ith sample. Further suppose that the dissimilarity between two samples $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ is measured by the dissimilarity function, $d(\boldsymbol{x}_i, \boldsymbol{x}_j)$ which is Euclidean-embeddable. That is, it is possible to find a configuration of n points in m-dimensional space such that the Euclidean distance between the points representing samples i and j is $d(\boldsymbol{x}_i, \boldsymbol{x}_j)$.

Let $\boldsymbol{\Delta}$ be the doubly-centred matrix of dissimilarities multiplied by $-\frac{1}{2}$,

$$\boldsymbol{\Delta} = -\frac{1}{2}\left(\boldsymbol{I} - \frac{1}{n}\mathbf{1}\mathbf{1}'\right)\boldsymbol{D}\left(\boldsymbol{I} - \frac{1}{n}\mathbf{1}\mathbf{1}'\right)$$

where the (i, j)th element of $\boldsymbol{D}$ is $d^2(\boldsymbol{x}_i, \boldsymbol{x}_j)$, $\boldsymbol{I}$ is the $n \times n$ identity matrix and $\mathbf{1}$ is the $n \times 1$ vector of ones. A consequence of Euclidean-embeddability is that $\boldsymbol{\Delta}$ is a positive semi-definite matrix [13]. So, via the spectral decomposition of $\boldsymbol{\Delta}$, it is possible to find a $n \times m$ real matrix $\boldsymbol{Y}$ such that $\boldsymbol{Y}\boldsymbol{Y}' = \boldsymbol{\Delta}$ and that $\boldsymbol{Y}'\boldsymbol{Y} = \boldsymbol{\Lambda}$ where $\boldsymbol{\Lambda}$ is a diagonal matrix with entries $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m > 0$ in the main diagonal. Let $\boldsymbol{y}_i'$ be the ith row of $\boldsymbol{Y}$. Then $\boldsymbol{y}_i$ can be regarded as the location of the ith sample in m-dimensional space such that for $j = 1, \ldots, n$, the distance between $\boldsymbol{y}_i$ and $\boldsymbol{y}_j$ matches $d(\boldsymbol{x}_i, \boldsymbol{x}_j)$.

Usually $m = (n - 1)$ meaning that the exact correspondence between inter-point distances and dissimilarities cannot normally be directly plotted on a low dimensional plot. However, as a result of least squares properties of spectral decompositions, the best rank r approximation of $\boldsymbol{\Delta}$ is obtained by simply using the first r columns of $\boldsymbol{Y}$ as the positions of the samples (see, for example, [9]).

Suppose now that we are interested in a new point $\boldsymbol{\mu}' = (\mu_1, \ldots, \mu_p)$. Let $\boldsymbol{d}(\boldsymbol{\mu})$ be the $n \times 1$ vector of squared dissimilarities between $\boldsymbol{\mu}$ and the samples $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$. That is, $\boldsymbol{d}'(\boldsymbol{\mu}) = (d^2(\boldsymbol{x}_1, \boldsymbol{\mu}), \ldots, d^2(\boldsymbol{x}_n, \boldsymbol{\mu}))$. Then, setting $\boldsymbol{z}'(\boldsymbol{\mu}) = (z_1(\boldsymbol{\mu}), \ldots, z_m(\boldsymbol{\mu}), z_{m+1}(\boldsymbol{\mu}))'$ where

$$(z_1(\boldsymbol{\mu}), \ldots, z_m(\boldsymbol{\mu}))' = \frac{1}{2}\boldsymbol{\Lambda}^{-1}\boldsymbol{Y}'\left(\frac{1}{n}\boldsymbol{D}\mathbf{1} - \boldsymbol{d}(\boldsymbol{\mu})\right) \tag{1}$$