B

Barycentric Discriminant Analysis

Hervé Abdi[1], Lynne J. Williams[2] and
Michel Béra[3]
[1]School of Behavioral and Brain Sciences, The
University of Texas at Dallas, Richardson,
TX, USA
[2]BC Children's Hospital MRI Research Facility,
Vancouver, BC, Canada
[3]Centre d'Étude et de Recherche en Informatique
et Communications, Conservatoire National des
Arts et Métiers, Paris, France

Synonyms

Intraclass analysis; Mean-centered partial least
square correlation

Glossary

Barycenter	The mean of the observations from a given category (also called center of gravity, center of mass, mean vector, or centroid)
Confidence interval	An interval encompassing a given proportion (e.g., 95%) of an estimate of a parameter (e.g., a mean)
Discriminant analysis	A technique whose goal is to assign observations to some predetermined categories
Discriminant factor scores	A linear combination of the variables of a data matrix. Used to assign observations to categories
Design matrix (aka group matrix)	In a group matrix, the rows represent observations and the columns represent a set of exclusive groups (i.e., an observation belongs to one and only one group). A value of 1 at the intersection of a row and a column indicates that the observation represented by the row belongs to the group represented by the column. A value of 0 at the intersection of a row and a column indicates that the observation represented by the row does not belong to the group represented by the column
Estimation bias	The difference between the computed value of a barycenter and the mean of the bootstrapped estimates of this barycenter
Fixed effect model	Analysis in which the observations that are predicted were used to compute the predictive model

Inertia	A weighted sum of squared distances of a set of points to a given point (often the barycenter of the points). The concept of inertia generalizes the concept of the variance of a set of points.
Learning set	The observations that are used to compute the predictive model
Multiblock Barycentric Analysis	Also called multitable or multiple subjects barycentric analysis. A barycentric analysis in which the data table is made of a concatenation of subdata-tables. Acronyms are MUBADA, MUSUBADA, and MUDICA
Prediction interval	An interval encompassing a given proportion (e.g., 95%) of the predicted observations from a given category
Random effect model	Analysis in which the observations that are predicted were *not* used to compute the predictive model
Testing set	A set of observations that were not used to compute the predictive model but are used to evaluate the quality of the prediction
Tolerance interval	An interval encompassing a given proportion (e.g., 95%) of the observations from a given category

Definition

Barycentric discriminant analysis (BADA) is a robust version of discriminant analysis that is used to assign, to predefined groups (also called categories), observations described by multiple variables. By contrast with traditional discriminant analysis, BADA can be used even when the number of observations is smaller than the number of variables – this makes BADA particularly suited for the analysis of Big Data.

Introduction

Barycentric discriminant analysis (BADA (Abdi 2007a; Bastin et al. 1982; Abdi and Williams 2010c; Bergougnan and Couraud 1982; Beaton et al. 2014)) is a robust version of discriminant analysis that is used – like discriminant analysis (Abdi 2003) – when multiple measurements describe a set of observations in which each observation belongs to one category (i.e., group) from a set of a priori defined categories. The goal of BADA is to combine the measurements to create new variables (called components or discriminant variables) that best separate the categories. These discriminant variables are also used to assign the original observations or "new" observations to the a-priori defined categories.

For example, BADA can be used (1) to assign subjects to a given diagnostic group (i.e., Alzheimer's disease, other dementia, normal aging (Abdi et al. 2012a)) on the basis on brain imaging data or psychological tests (here the a-priori categories are the clinical groups), (2) to assign wines to a region of production on the basis of several physical and chemical measurements (here the a-priori categories are the regions of production (Abdi 2007b)), (3) to use brain scans taken on a given participant to determine what type of object (e.g., a face, a cat, a chair) was watched by the participant when the scans were taken (here the a-priori categories are the types of object (Abdi et al. 2012b; St. Laurent et al. 2011)), (4) to use DNA measurements to predict if a person is at risk for a given health problem (here the a-priori categories are the types of health problem (El Behi et al. 2017; Cioli et al. 2014)). BADA is a very general discriminant technique that can also be used in cases for which standard discriminant analysis cannot be used. This is the case, for example, when there are more variables than observations (a case often called the "$N << P$ problem") or when the measurements are qualitative (instead of quantitative as required by standard discriminant analysis).

Key Points

Barycentric discriminant analysis is a robust version of discriminant analysis that is used when multiple measurements describe a set of observations in which each observation belongs to one category (i.e., group) from a set of a priori defined categories. BADA combines the original variables to create new variables that best separate the groups and that can also be used to optimally assign old or new observations to these categories. The quality of the performance is evaluated by cross-validation techniques that estimate the performance of the classification model for new observations.

BADA is a very versatile technique that can be declined in several different varieties that can handle, for example, qualitative data and data structured in blocks. This versatility makes BADA particularly suited for the analysis of multimodal and Big data.

Historical Background

BADA can be seen as a particular version of the two-table method of Tucker's interbattery analysis (Tucker 1958) – a technique developed in the 1950s as a robust version of canonical correlation analysis. BADA, however, was first developed as such in the 1970s by the French school of data analysis (Benzécri 1977; Nakache et al. 1977; Leclerc 1976) for the specific case of classifying qualitative data (under the name of discriminant correspondence analysis). Subsequent developments occurred within the framework of co-inertia analysis (an approach generalizing Tucker's method that is particularly popular in Ecological studies, see Doledec and Chessel 1994) and also within the framework of partial least square correlation methods (another approach generalizing Tucker's method that is particularly popular in brain imaging and genetic studies, see Krishnan et al. 2010).

Barycentric Discriminant Analysis (BADA)

BADA is, in fact, a *class* of methods which all rely on the same principle: Each category of interest is represented by the *barycenter* of its observations (i.e., the weighted average; the barycenter is also called the *center of gravity* of the observations of a given category), and a generalized principal component analysis (GPCA) is performed on the category by variable matrix. This analysis gives a set of discriminant factor scores for the categories and another set of factor scores for the variables. The original observations are then projected onto the category factor space, and this operation provides a set of factor scores for the observations. The distance of each observation to all the categories is computed from the factor scores and each observation is assigned to its closest category. The comparison between the a-priori and a-posteriori category assignments is used to assess the quality of the discriminant procedure. The prediction for the observations which were used to compute the barycenters is called the *fixed effect* prediction. Fixed effect performance is evaluated by counting the number of correct and incorrect assignments and storing these numbers in a confusion matrix. Another index of the performance of the fixed effect model – equivalent to a squared coefficient of correlation – is the ratio

$$R^2 = \frac{\text{category variance}}{\text{category variance } + \text{ variance of the observations within category}}.$$

(1)

This coefficient R^2 is interpreted as the proportion of variance of the observations explained by the categories or as the proportion of the variance explained by the discriminant model. The performance of the fixed effect model can also be represented graphically as a *tolerance* ellipsoid that encompasses a given proportion (say 95%) of the observations. The overlap between the tolerance

ellipsoids of two categories is roughly proportional to the number of misclassifications between these two categories.

New observations can also be projected onto the discriminant factor space, and they can be assigned to the closest category. When the actual assignment of these observations is not known, the model can be used to *predict* category membership. The model is then called a *random* effect model (as opposed to the *fixed* effect model). An obvious problem, then, is to evaluate the quality of the prediction for new observations. Ideally, the performance of the random effect model is evaluated by counting the number of correct and incorrect classifications for new observations and computing a confusion matrix on these *new* observations. However, it is not always practical or even feasible to obtain new observations, and therefore the random effect performance is, in general, evaluated using computational cross-validation techniques such as the *jackknife* (Abdi and Williams 2010a) or the *bootstrap* (Efron and Tibshirani 1993). For example, a jackknife approach (also, better called, "leave one out" LOO) can be used by which each observation is taken out of the set, in turn, and predicted from the model built on all the other observations. The predicted observations are then projected onto the space of the fixed effect discriminant scores. The quality of the LOO prediction can also be represented graphically as a *prediction* ellipsoid. A prediction ellipsoid encompasses a given proportion (say 95%) of the *new* observations. The overlap between the prediction ellipsoids of two categories is roughly proportional to the number of misclassifications of *new* observations between these two categories.

The stability of the discriminant model can be assessed by a cross-validation model such as the bootstrap (Efron and Tibshirani 1993; Diaconis and Efron 1983). In this procedure, multiple sets of observations (called "Bootstrapped samples") are generated by sampling with replacement from the original set of observations and the category barycenters are computed from each of these sets.

These barycenters are then projected onto the discriminant factor scores. The variability of the barycenters can be represented graphically as a *confidence* ellipsoid that encompasses a given proportion (say 95%) of the barycenters. When the confidence intervals of two categories do not overlap, this indicates that these two categories are *significantly* different.

In summary, BADA is a GPCA performed on the category barycenters. Recall that GPCA encompasses different techniques such as, for example, correspondence analysis, biplot, Hellinger distance analysis, discriminant analysis, and canonical variate analysis (Abdi 2007a; Gittins 1980; Greenacre 1984). For each specific type of GPCA, we have a corresponding version of BADA. For example, when the GPCA used is correspondence analysis, this gives the most well-known version of BADA: discriminant correspondence analysis (DICA, also called correspondence discriminant analysis, see, e.g., Abdi 2007b, Celeux and Nakache 1994, Leclerc 1976, Saporta and Niang 2006, Bastin et al. 1982, Bergougnan and Couraud 1982, Doledec and Chessel 1994).

Notations

The original data matrix is an I observations by J variables matrix denoted $\mathbf{X}$. Prior to the analysis, the matrix $\mathbf{X}$ can be preprocessed by centering (i.e., subtracting the column mean from each column), by transforming each column into a Z-score, or even by normalizing each row such that the sum of its elements or the sum of its squared elements is equal to one. The observations in $\mathbf{X}$ are partitioned into N a-priori categories of interest, with I_n being the number of observations of the nth category (and so $\sum_n^N I_n = I$). The general structure of the data matrix can be illustrated as follows (note that for convenience, the numbering of the rows is done within category):

$$\begin{array}{c}
\text{Category 1} \left\{ \begin{array}{cccccc}
1 & x_{1,1} & \cdots & x_{1,j} & \cdots & x_{1,J} \\
\vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
I_1 & x_{I_1,1} & \cdots & x_{I_1,j} & \cdots & x_{I_1,J}
\end{array} \right. \\
\vdots \\
\text{Category } n \left\{ \begin{array}{cccccc}
1 & x_{1,1} & \cdots & x_{1,j} & \cdots & x_{1,J} \\
\vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
I_n & x_{I_n,1} & \cdots & x_{I_n,j} & \cdots & x_{I_n,J}
\end{array} \right. \\
\vdots \\
\text{Category } N \left\{ \begin{array}{cccccc}
1 & x_{1,1} & \cdots & x_{1,j} & \cdots & x_{1,J} \\
\vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
I_N & x_{I_N,1} & \cdots & x_{I_N,j} & \cdots & x_{I_N,J}
\end{array} \right.
\end{array} \quad (2)$$

Notations for the Categories (Rows)

We denote by $\mathbf{Y}$ the I by N design matrix (also called a group matrix) for the categories (i.e., groups) describing the rows of $\mathbf{X}$: $y_{i,n} = 1$ if row i belongs to category n, $y_{i,n} = 0$ otherwise. We denote by $\mathbf{m}$ the I by 1 vector of *masses* for the rows of $\mathbf{X}$ and by $\mathbf{M}$ the I by I diagonal matrix whose diagonal elements are the elements of $\mathbf{m}$ (i.e., using the diag operator which transforms a vector into a diagonal matrix, we have $\mathbf{M} = \mathrm{diag}\{\mathbf{m}\}$). We denote by $\mathbf{b}$ the N by 1 vector of masses for the categories describing the rows of $\mathbf{X}$ and by $\mathbf{B}$ the N by N diagonal matrix whose diagonal elements are the elements of $\mathbf{b}$. Masses are positive numbers; it is convenient (but not necessary) to have the sum of the masses equal to one.

Barycentric Discriminant Analysis: Computation

The first step of BADA is to compute the barycenter of each of the N categories describing the rows. The barycenter of a category is the weighted average of the rows where the weights are the masses rescaled such that the sum of the weights for one category is equal to one. The N by J matrix of barycenters is computed as

$$\mathbf{R} = \mathrm{diag}\{\mathbf{Y}^{\top}\mathbf{M}\mathbf{1}\}^{-1}\mathbf{Y}^{\top}\mathbf{M}\mathbf{X}, \quad (3)$$

where $\mathbf{1}$ is a conformable vector of 1s (the diagonal matrix $\mathrm{diag}\{\mathbf{Y}\mathbf{M}\mathbf{1}\}^{-1}$ serves to rescale the

masses of the rows such that their sum is equal to one for each category).

Masses, Weights, and GPCA

The type of preprocessing of the data and the choice of $\mathbf{B}$ (matrix of masses for the categories) and $\mathbf{W}$ (matrix of weights for the variables) is crucial, because this choice determines the specific type of GPCA used. For example, discriminant correspondence analysis is obtained by transforming the rows of $\mathbf{R}$ into relative frequencies, by using the relative frequencies of the barycenters as their masses and by using the inverse of the column frequencies for the weights of the variables. As another example, standard discriminant analysis is obtained when $\mathbf{W}$ is equal to the inverse of the within group variance-covariance matrix.

GPCA of the Barycenter Matrix
The $\mathbf{R}$ matrix is then analyzed using a GPCA under the constraints provided by the matrices $\mathbf{B}$ (for the N categories) and $\mathbf{W}$ (for the columns). Specifically, the matrix $\mathbf{R}$ is analyzed with the generalized singular value decomposition (see, e.g., Abdi 2007a; Greenacre 1984) as

$$\mathbf{R} = \mathbf{P}\mathbf{\Delta}\mathbf{Q}^{\top} \quad \text{with} \quad \mathbf{P}^{\top}\mathbf{B}\mathbf{P} = \mathbf{Q}^{\top}\mathbf{W}\mathbf{Q} = \mathbf{I}, \quad (4)$$

where $\mathbf{\Delta}$ is the L by L diagonal matrix of the singular values (with L being the number of non-zero singular values), and $\mathbf{P}$ (respectively $\mathbf{Q}$) being the N by L (respectively J by L) matrix of the left (respectively right) generalized singular vectors of $\mathbf{R}$.

Factor Scores

The N by L matrix of factor scores for the categories is obtained as

$$\mathbf{F} = \mathbf{P}\mathbf{\Delta} = \mathbf{R}\mathbf{W}\mathbf{Q}. \quad (5)$$

The variance of the columns of $\mathbf{F}$ is given by

the square of the corresponding singular values (i.e., the "eigenvalue" denoted λ; these are stored in the diagonal matrix $\mathbf{\Lambda}$). This can be shown by combining Eqs. 4 and 5 to give:

$$\mathbf{F}^{\top}\mathbf{BF} = \mathbf{\Delta P}^{\top}\mathbf{BP\Delta} = \mathbf{\Delta}^2 = \mathbf{\Lambda}. \qquad (6)$$

The I rows of matrix $\mathbf{X}$ can be projected onto the space defined by the factor scores of the barycenters (this procedure is called a projection as "supplementary," or "illustrative," or even "passive" elements; for more details see Abdi and Williams 2010b). Note that the matrix $\mathbf{WQ}$ from Eq. 5 is a projection matrix. Therefore, the I by L matrix $\mathbf{H}$ of the factor scores for the rows of $\mathbf{X}$ can be computed as

$$\mathbf{H} = \mathbf{XWQ}. \qquad (7)$$

These projections are barycentric, because the weighted average of the factor scores of the rows of a category gives the factors scores of the category. This can be shown by first computing the barycenters of the row factor scores as (cf. Eq. 3) as

$$\overline{\mathbf{H}} = \operatorname{diag}\{\mathbf{YM1}\}^{-1}\ \mathbf{YMH}, \qquad (8)$$

then plugging in Eq. 7 (see also Eq. 5) and developing. Taking this into account, Eq. 8 becomes

$$\begin{aligned}\overline{\mathbf{H}} &= \operatorname{diag}\{\mathbf{YM1}\}^{-1}\ \mathbf{YMXWQ}\\ &= \mathbf{RWQ}\ = \mathbf{F}.\end{aligned} \qquad (9)$$

Loadings

The loadings describe the variables of the barycentric data matrix and are used to identify the variables important for the separation between the groups. As for standard PCA, there are several ways of defining the loadings. The loadings can be defined as the correlation between the columns of matrix $\mathbf{R}$ and the factor scores. Alternatively, the loadings can also be defined as the matrix $\mathbf{Q}$ or (as we did in our example), even, as

$$\mathbf{G} = \mathbf{Q\Delta}. \qquad (10)$$

Quality of the Prediction

The performance, or quality of the prediction of a discriminant analysis, is assessed by predicting the category membership of the observations and by comparing the predicted with the actual category membership. The pattern of correct and incorrect classifications can be stored in a confusion matrix in which the columns represent the actual categories and the rows the predicted categories. At the intersection of a row and a column is the number of observations from the column category assigned to the row category.

The performance of the model can be assessed for the observations used to compute the categories: this is the *fixed effect* model. In addition, the performance of the model can be estimated for *new* observations (i.e., observations not used to compute the model): This is the *random effect* model.

Fixed Effect: Old Observations

The *fixed effect* model predicts the category assignment for the observations that were used to compute the barycenters of the categories. In order to assign an observation to a category, the first step is to compute the distance between this observation and all N categories. Then, the observation is assigned to the closest category. Several possible distances can be chosen, but a natural choice is the Euclidean distance computed in the factor space. If we denote by $\mathbf{h}_i$ the vector of factor scores for the ith observation, and by $\mathbf{f}_n$ the vector of factor scores for the nth category, then the squared Euclidean distance between the ith observation and the nth category is computed as

$$d^2(\mathbf{h}_i,\ \mathbf{f}_n) = (\mathbf{h}_i - \mathbf{f}_n)^{\top}(\mathbf{h}_i - \mathbf{f}_n). \qquad (11)$$

Obviously, other distances are possible (e.g., Mahalanobis distance), but the Euclidean distance has the advantage of being "directly read" on the map.

Tolerance Intervals

The quality of the category assignment of the actual observations can be displayed using *tolerance* intervals (Abdi et al. 2009; Krzanowski and Radley 1989). A tolerance interval encompasses a given proportion of a sample or a population. When displayed in two dimensions, these intervals have the shape of an ellipse and are called *tolerance ellipsoids*. For BADA, a category tolerance ellipsoid is plotted on the category factor score map. This ellipsoid is obtained by fitting an ellipse which includes a given percentage (e.g., 95%) of the observations. Tolerance ellipsoids are centered on their categories, and the overlap of the tolerance ellipsoids of two categories reflects the proportion of misclassifications between these two categories.

Random Effect: New Observations

The *random effect* model evaluates the quality of the assignment of *new* observations to categories. This estimation is obtained, in general, by using cross-validation techniques that partition the data into a *learning set* (used to create the model) and a *testing set* (used to evaluate the model). A convenient variation of this approach is the jackknife (aka the "leave one out" or LOO) approach: Each observation is taken out from the data set, in turn, and is then projected onto the factor space of the remaining observations in order to predict its category membership. For the estimation to be unbiased, the left-out observation should not be used in any way in the analysis. In particular, if the data matrix is preprocessed, the left-out observation should not be used in the preprocessing. So, for example, if the columns of the data matrix are transformed into Z scores, the left-out observation should *not* be used to compute the means and standard deviations of the columns of the matrix to be analyzed, but these means and standard deviations will be used to compute the Z-score for the left-out observation.

The assignment of an observation to a category follows the same procedure as for a fixed effect model: the observation is projected onto the category factor scores, and the observation is assigned to the closest category. Specifically, we denote by $\mathbf{X}_{-i}$ the data matrix without the ith observation, and by $\mathbf{x}_i$ the ith observation. If $\mathbf{X}_{-i}$ is preprocessed (e.g., centered and normalized), the preprocessing parameters will be estimated without $\mathbf{x}_i$ (e.g., the mean and standard deviation of $\mathbf{X}_{-i}$ are computed *without* $\mathbf{x}_i$), and $\mathbf{x}_i$ will be preprocessed with the parameters estimated for $\mathbf{X}_{-i}$ (e.g., $\mathbf{x}_i$ will be centered and normalized using the means and standard deviations of the columns of $\mathbf{X}_{-i}$). Then the matrix of barycenters (denoted $\mathbf{R}_{-i}$) is computed and its generalized eigendecomposition is obtained as (cf. Eq. 4):

$$\mathbf{R}_{-i} = \mathbf{P}_{-i}\mathbf{\Delta}_{-i}\mathbf{Q}_{-i}^{\top} \quad \text{with} \quad \mathbf{P}_{-i}^{\top}\mathbf{W}_{-i}\mathbf{P}_{-i}$$
$$= \mathbf{Q}_{-i}^{\top}\mathbf{B}_{-i}\mathbf{Q}_{-i} = \mathbf{I} \tag{12}$$

(with $\mathbf{B}_{-i}$ and $\mathbf{W}_{-i}$ being the mass and weight matrices for $\mathbf{R}_{-i}$). The matrix of factor scores denoted $\mathbf{F}_{-i}$ is obtained as (cf. Eq. 5)

$$\mathbf{F}_{-i} = \mathbf{P}_{-i}\mathbf{\Delta}_{-i} = \mathbf{R}_{-i}\mathbf{W}_{-i}\mathbf{Q}_{-i}. \tag{13}$$

The jackknifed projection of the ith observation, denoted $\tilde{\mathbf{h}}_i$, is obtained (cf. Eq. 7) as

$$\tilde{\mathbf{h}}_i = \mathbf{x}_i\mathbf{W}_{-i}\mathbf{Q}_{-i}. \tag{14}$$

Distances between the ith observation and the N categories can be computed (cf. Eq. 11) with the factor scores. The observation is then assigned to the closest category.

Prediction Intervals

In order to display the quality of the prediction for *new* observations we use *prediction* intervals. In order to compute these intervals, the first step is to project the jackknifed observations onto the original complete factor space. There are several ways to project a jackknifed observation onto the factor score space. Here we describe a two-step procedure. First, the observation is projected onto the jackknifed space and is reconstructed from its projections. Then, the reconstituted observation

is projected onto the full factor score solution. Specifically, a jackknifed observation is reconstituted from its factor scores as (cf. Eqs. 4 and 14):

$$\tilde{\mathbf{x}}_i = \tilde{\mathbf{h}}_i \mathbf{Q}_{-i}^\top. \tag{15}$$

The projection of the jackknifed observation is denoted $\hat{\mathbf{h}}_i$ and is obtained by projecting $\tilde{\mathbf{x}}_i$ as a supplementary element in the original solution. Specifically, $\hat{\mathbf{h}}_i$ is computed as

$$\begin{aligned}
\hat{\mathbf{h}}_i &= \tilde{\mathbf{x}}_i \mathbf{WQ} && (cf. \text{ Eq. 5}) \\
&= \tilde{\mathbf{h}}_i \mathbf{Q}_{-i}^\mathbf{T} \mathbf{WQ} && (cf. \text{ Eq. 15}) \\
&= \mathbf{x}_i \mathbf{W}_{-i} \mathbf{Q}_{-i} \mathbf{Q}_{-i}^\mathbf{T} \mathbf{WQ} && (cf. \text{ Eq. 14}). \tag{16}
\end{aligned}$$

Prediction ellipsoids are not necessarily centered on their categories (the distance between the center of the ellipse and the category represents the estimation *bias*). Overlap of two predictions intervals directly reflects the proportion of misclassifications for the "new" observations.

Quality of the Category Separation

R^2 and Permutation Test

In order to evaluate the quality of the discriminant model, we use a coefficient inspired by the coefficient of correlation. Because BADA is a barycentric technique, the total *inertia* (i.e., the "variance") of the observations to the grand barycenter (i.e., the barycenter of all categories) can be decomposed into two additive quantities: (1) the inertia of the observations relative to the barycenter of their own category and (2) the inertia of the category barycenters to the grand barycenter.

Specifically, if we denote by $\overline{\mathbf{f}}$ the vector of the coordinates of the grand barycenter (i.e., each component of this vector is the average of the corresponding components of the barycenters), the total inertia, denoted $\mathcal{I}_{\text{Total}}$, is computed as the sum of the squared distances of the observations to the grand barycenter (11):

$$\begin{aligned}
\mathcal{I}_{\text{Total}} &= \sum_i^I m_i d^2 (\mathbf{h}_i, \overline{\mathbf{f}}) \\
&= \sum_i^I m_i (\mathbf{h}_i - \overline{\mathbf{f}})^\top (\mathbf{h}_i - \overline{\mathbf{f}}). \tag{17}
\end{aligned}$$

The inertia of the observations relative to the barycenter of their own category is abbreviated as the "inertia within." It is denoted $\mathcal{I}_{\text{Within}}$ and computed as

$$\begin{aligned}
\mathcal{I}_{\text{Within}} &= \sum_n^N \sum_{\substack{i \text{ in} \\ \text{category } n}} m_i d^2 (\mathbf{h}_i, \mathbf{f}_n) \\
&= \sum_n^N \sum_{\substack{i \text{ in} \\ \text{category } n}} m_i (\mathbf{h}_i - \mathbf{f}_n)^\top (\mathbf{h}_i - \mathbf{f}_n) \tag{18}
\end{aligned}$$

The inertia of the barycenters to the grand barycenter is abbreviated as the "inertia between." It is denoted $\mathcal{I}_{\text{Betwen}}$ and computed as

$$\begin{aligned}
\mathcal{I}_{\text{Between}} &= \sum_i^I b_i \times d^2 (\mathbf{f}_i, \overline{\mathbf{f}}) = \sum_n^N b_n \times d^2 (\mathbf{f}_n, \overline{\mathbf{f}}) \\
&= \sum_n^N b_n \times (\mathbf{f}_n - \overline{\mathbf{f}})^\top (\mathbf{f}_n - \overline{\mathbf{f}}). \tag{19}
\end{aligned}$$

So the additive decomposition of the inertia can be expressed as

$$\mathcal{I}_{\text{Total}} = \mathcal{I}_{\text{Within}} + \mathcal{I}_{\text{Between}}. \tag{20}$$

This decomposition is similar to the familiar decomposition of the sum of squares in the analysis of variance. This suggests that the intensity of the discriminant model can be tested by the ratio of between inertia to the total inertia, as is done in analysis of variance and regression. This ratio – denoted R^2 (see also Eq. 1) – is computed as:

$$R^2 = \frac{\mathcal{I}_{\text{Between}}}{\mathcal{I}_{\text{Total}}} = \frac{\mathcal{I}_{\text{Between}}}{\mathcal{I}_{\text{Between}} + \mathcal{I}_{\text{Within}}} . \quad (21)$$

The R^2 ratio takes values between 0 and 1, the closer to one the better the model. The significance of R^2 can be assessed by permutation tests, and confidence intervals can be computed using cross-validation techniques such as the jackknife and the bootstrap.

Confidence Intervals

The stability of the position of the categories can be displayed using *confidence* intervals. A confidence interval reflects the variability of a population *parameter* or its estimate. In two dimensions, this interval becomes a confidence ellipsoid. The problem of estimating the variability of the position of the categories cannot, in general, be solved analytically and cross-validation techniques need to be used. Specifically, the variability of the position of the categories is estimated by generating *bootstraped* samples from the sample of observations (Manly 1997). A bootstraped sample is obtained by sampling *with replacement* from the observations (recall that when sampling with replacement some observations may be absent and some others maybe repeated). The "bootstraped barycenters" obtained from these samples are then projected onto the discriminant factor space and, finally, an ellipse is plotted such that it comprises a given percentage (e.g., 95%) of these bootstraped barycenters. When the confidence intervals of two categories do not overlap, these two categories are "significantly different" at the corresponding alpha level (e.g., $\alpha = .05$).

An Example: Four Assessors Taste 12 Wines from three Regions

It is a common belief that the taste of a wine depends upon its place of origin (aka appellation or region). As an (fictitious) illustration, we have sampled 12 wines coming from three different French wine regions and asked four professional assessors (unaware of the origin of the wines) to rate these wines. Each assessor generated a list of descriptors (up to six) and rated (with a 7-point scale) each wine on these descriptors. The assessors also evaluated an additional French wine from an unknown region with the goal of predicting the origin of the wine from the assessors' ratings. The data are given in Table 1.

Tables 2 and 3 give the factor scores for the analysis and Fig. 1 displays them. We can see that the wine regions appear to be well differentiated, with wines from the Loire being characterized as "fruity" and "pleasant," wines from the Rhône being characterized as having a high "alcohol" content, and the wines from Beaujolais being characterized as "tangy" and "tannic."

Reliability and Stability of the Analysis

Proportion of Explained Variance: R^2

The reliability and the stability of the analysis is evaluated by computing R^2 (see Eq. 21). Its large value: $R^2 = .94$ confirms that the wine regions are well identified. In addition, the *p*-value of $p < .001$ (obtained from a permutation test using 10,000 permutations) indicates that the discrimination of the three wine regions is reliable.

Confidence Intervals

The original $I \times J$ matrix was resampled 1000 times with replacement, and the barycenters for each wine region for each of the 1000 samples were projected onto the original factor space. The 950 closest barycenters were kept, and we generated the 95% confidence ellipsoid which comprised these 950 barycenters. These confidence ellipsoids are displayed in Fig. 2a. Because there is no overlap between the three wines regions, they are significantly different at the $p < .05$ significance level.

Barycentric Discriminant Analysis, Table 1 Data from the three wine regions example: 12 wines from three different wine regions are rated on a 7-point scale on descriptors assigned by each of four assessors

Wine	Région	Assessor 1					Assessor 2				Assessor 3						Assessor 4		
		Wood	Fruit	Sugar	Alcohol	Hedonic	Tannic	Fruity	Phenol	Hedonic	Tangy	Sour	Fruity	Easy	Smooth	Pleasant	Alcohol	Hedonic	Tannic
1	Loire	1	3	2	1	1	2	5	2	2	1	1	3	4	5	4	1	2	1
2	Loire	2	3	3	2	3	2	6	1	3	1	2	4	2	4	5	2	2	2
3	Loire	1	2	2	1	2	2	4	2	4	2	1	5	3	5	6	1	1	3
4	Loire	1	3	3	2	4	1	6	1	3	1	1	4	4	5	5	1	2	1
$\overline{X}$	Loire	1.25	2.75	2.50	1.50	2.50	1.75	5.25	1.50	3.00	1.25	1.25	4.00	3.25	4.75	5.00	1.25	1.75	1.75
1	Rhône	1	2	1	3	3	3	3	4	5	3	1	3	1	1	3	3	5	1
2	Rhône	2	1	1	3	2	5	2	3	4	4	1	4	2	2	2	3	4	2
3	Rhône	3	2	2	2	1	4	4	4	6	3	2	2	1	1	1	2	3	3
4	Rhône	2	3	3	3	4	5	6	5	5	3	1	1	1	2	3	3	3	4
$\overline{X}$	Rhône	2.00	2.00	1.75	2.75	2.50	4.25	3.75	4.00	5.00	3.25	1.25	2.50	1.25	1.50	2.25	2.75	3.75	2.50
1	Beaujolais	3	1	3	1	1	5	4	1	1	4	4	1	2	2	2	3	5	1
2	Beaujolais	2	1	3	1	2	6	4	2	1	4	3	3	3	3	4	1	2	4
3	Beaujolais	3	2	2	2	4	7	4	1	2	5	5	1	2	2	1	1	1	4
4	Beaujolais	3	1	1	1	3	6	1	1	1	4	5	2	1	3	2	1	2	4
$\overline{X}$	Beaujolais	2.75	1.25	2.25	1.25	2.50	6.00	3.25	1.25	1.25	4.25	4.25	1.75	2.00	2.50	2.25	1.50	2.50	3.25
	Mystery	1	3	2	2	1	2	6	1	3	1	1	4	4	5	5	1	2	2

Barycentric Discriminant Analysis, Table 2 Factor scores for the observations, categories, and blocks. Jackknifed factor scores for the observations

Axis	λ	%	Loire wines				Region Loire	Rhône wines				Region Rhône	Beaujolais wines				Region Mystery Beaujolais	Wine
			1	2	3	4		1	2	3	4		1	2	3	4		
Factor scores for observations and regions																		
1	0.67	62.91	1.03	0.94	1.10	1.40	1.11	−0.07	−0.44	−0.47	−0.22	−0.30	−0.76	−0.34	−1.13	−1.03	−0.82	1.23
2	0.39	37.09	0.31	0.19	0.16	0.25	0.23	−1.09	−0.62	−0.85	−0.86	−0.86	0.38	0.68	0.66	0.79	0.63	0.27
Jackknifed factor scores for observations																		
1	–	–	1.02	0.97	1.09	1.45	–	−0.10	−0.44	−0.44	−0.22	–	−0.70	−0.37	−1.09	−0.96	–	–
2	–	–	0.33	0.15	0.09	0.32	–	−1.06	−0.53	−0.81	−0.65	–	0.12	0.63	0.67	0.82	–	–

Barycentric Discriminant Analysis, Table 3 Factor scores for the variables

Axis	λ	%	Wood	Fruit	Sugar	Alcohol	Hedonic	Tannic	Fruity	Phenol	Hedonic	Tangy	Sour	Fruity	Easy	Smooth	Pleasant	Alcohol	Hedonic	Tannic
1	0.67	62.91	−0.59	0.59	0.18	−0.07	0	−1.72	0.85	−0.22	0.34	−1.24	−1.00	0.93	0.66	1.15	1.25	−0.27	−0.49	−0.59
2	0.39	37.09	0.16	−0.16	0.26	−0.65	0	0.28	0.02	−1.22	−1.49	0.09	1.00	−0.07	0.49	0.73	0.33	−0.60	−0.66	0.16

Barycentric Discriminant Analysis,
Fig. 1 Barycentric
discriminant analysis.
Projections on the first two
dimensions. (**a**) The *N* set.
Rows of **R**: wine regions
barycenters. (**b**) The *J* set.
Columns (i.e., descriptors).
The wine categories have
also been projected for ease
of interpretation. Both
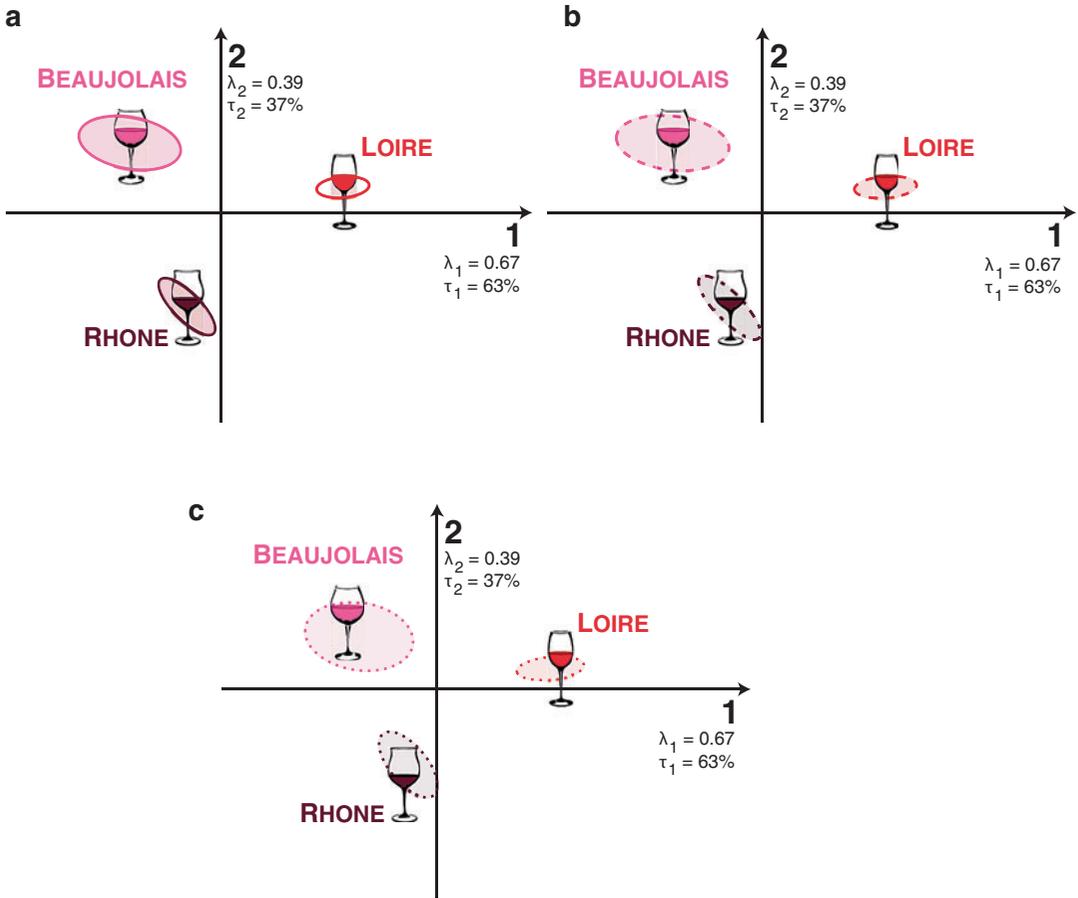figures have the same scale
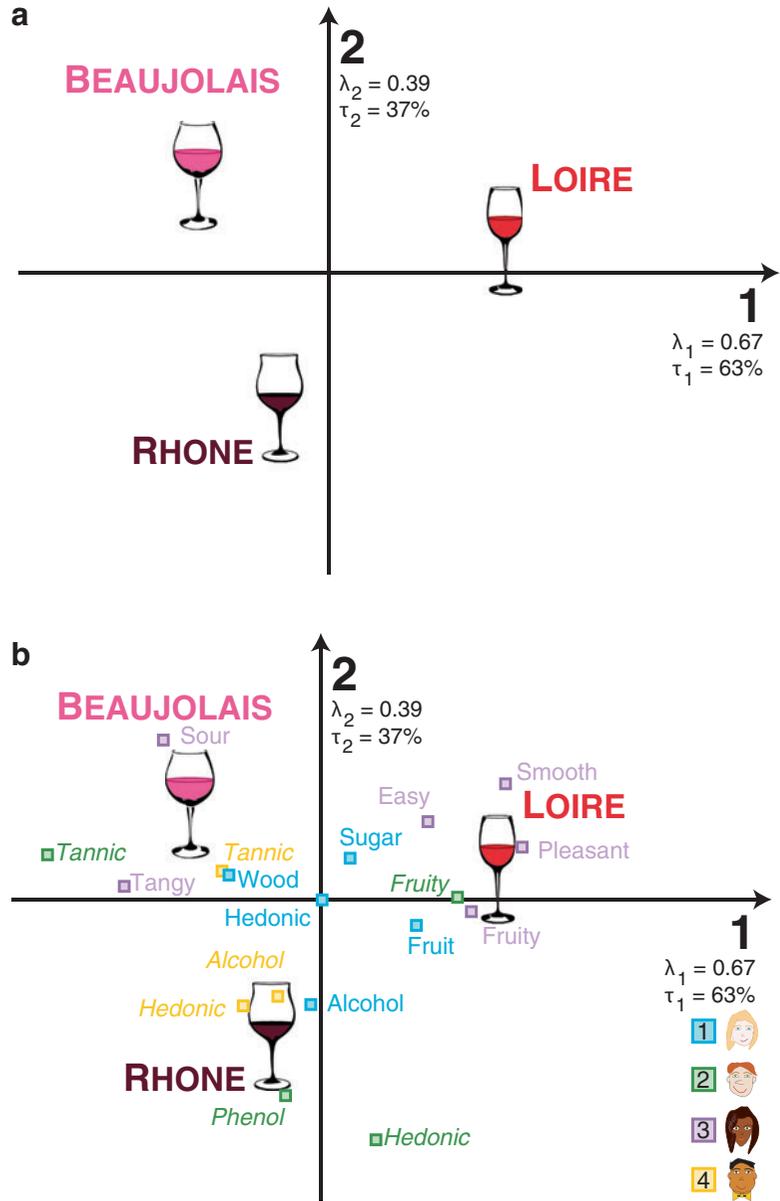
Reliability of the Prediction

Fixed Effect Model

Confusion Matrix
The fixed effect quality of the model is evaluated
by the following confusion matrix:

$$\begin{bmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{bmatrix}. \qquad (22)$$

In this matrix, the rows are the assigned cate-
gories and the columns are the real categories. All
wines within the sample were correctly classified.
The overall quality of the fixed effect model can
be computed from the diagonal of the matrix.

Barycentric Discriminant Analysis, Fig. 2 Barycentric discriminant analysis. Projections on the first two dimensions. (**a**) Confidence ellipses. (**b**) Tolerance ellipses. (**c**) Prediction ellipses. All figures are shown on the same scale
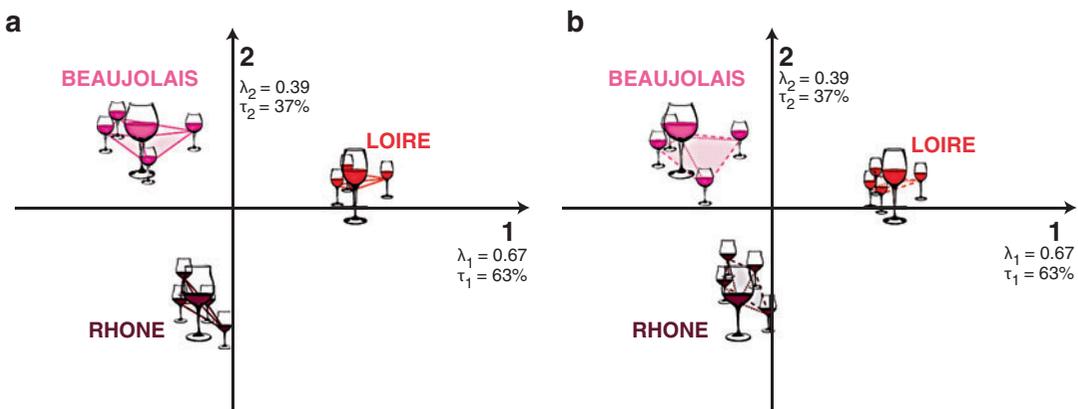
Barycentric Discriminant Analysis, Fig. 3 Barycentric discriminant analysis. Projections on the first two dimensions. (**a**) Fixed effect model. The three wine regions and the convex hulls for the wines. (**b**) Random effect model.

The jackknifed wines have been projected back onto the fixed effect solution. The convex hulls show that the random effect categories have a larger variability and have moved. All figures are shown on the same scale

Here we find that all 12 wines were correctly classified.

The projections of the wines within the sample into the original GPCA space are shown in Fig. 3a. The quality of the model can be evaluated by drawing the convex hull of each category. For the fixed effect model, the center of gravity of the convex hulls are the category barycenters.

Tolerance Intervals

The reliability of the prediction for the fixed effect model can also be displayed graphically as tolerance ellipsoids. These are shown in Fig. 2b. Overlap, between the tolerance ellipsoids, represents the proportion of misclassification of observations within the sample. Because there is no overlap, there were no misclassified wines within the sample.

Random Effect Model

Confusion Matrix

A jackknife procedure was used in order to evaluate the generalization capacity of the analysis to *new* wines. Each wine was taken out of the sample, in turn, a GPCA was performed on the remaining sample of 11 wines, and the left-out wine was then projected onto the discriminant factor space (see Eq. 14) and was assigned to its closest category. This gave the following random effect confusion matrix:

$$\begin{bmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{bmatrix}. \tag{23}$$

The random effect performance is perfect, with all 12 wines correctly assigned.

Prediction Intervals

The projections of the jackknifed wines onto the original GPCA space (computed according to Eqs. 15 and 16) are given in Table 2 and displayed in Fig. 3b. The quality of the model can be illustrated by drawing the convex hull for these

observations. All the wines were correctly classified, but note that, compared to the fixed effect (cf. Fig. 3a), the placement of the convex hull expands and shifts for each of the three wine categories.

The reliability of the predictions for the random effect model can also be displayed as prediction ellipsoids. Overlap between ellipsoids represents the proportion of misclassifications of *new* observations. This is shown in Fig. 2c. Because there is no overlap between the prediction ellipsoids, the *new* observations were all correctly classified.

Extensions of BADA

Discriminant Correspondence Analysis   When the observations are described by qualitative variables, the standard version of BADA cannot be used because, BADA – being based on the principal component analysis of the barycentric matrix – requires quantitative variables. Correspondence analysis and multiple correspondence analysis – extensions of principal component analysis for qualitative data – can, in this case, be substituted to PCA. BADA then becomes Discriminant Correspondence Analysis (DICA, see Abdi et al. 2012a, b; Abdi 2007b; Benzécri 1977; Nakache et al. 1977; Williams et al. 2010) –also known as Correspondence Discriminant Analysis (see Doledec and Chessel 1994; Perriere et al. 1996).

Multiblock Analysis (MUBADA, MUSUBADA)

Because BADA is based on GPCA, it can also analyze data tables obtained by the concatenation of data-tables or data matrices that we call here blocks (aka subtables). In this case, BADA becomes multitable (or multiblock) barycentric discriminant analysis (MUBADA Abdi et al. 2012a) and when the blocks are subjects (e.g.,
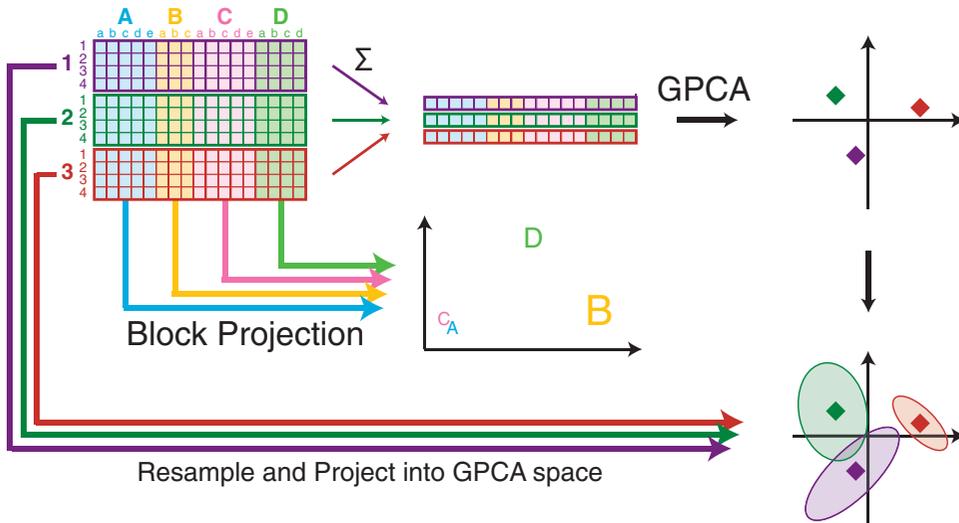
Barycentric Discriminant Analysis, Fig. 4 The different steps of a multiblock version of BADA (e.g., MUBADA)

here, the assessors): multisubject barycentric discriminant analysis (MUSUBADA (Abdi et al. 2012b)).

For this type of data, the importance (often called the *contribution*) of each block to the overall discrimination can also be evaluated and represented as a graph. For example, in the wine data set, each assessor can be seen as a block of data, and it will be of interest to evaluate the contribution of each assessor to the overall discrimination between the wine regions. A graphical sketch of MUBADA is given in Fig. 4.

Multiblock Analysis: Notations for Blocks (Columns)

In a multiblock analysis, the columns of matrix $\mathbf{X}$ can be arranged in K a priori blocks (or subtables). The number of columns of the kth block are denoted J_k (and so $\sum_k^K J_k = J$). The data matrix corresponding to the nth category and the k block is denoted $\mathbf{X}_{n,k}$. With this notation, the matrix $\mathbf{X}$ can decomposed into N by K blocks as

$$
\mathbf{X} = \begin{array}{c} \\ 1 \\ \vdots \\ n \\ \vdots \\ N \end{array}
\begin{array}{cccccc}
1 & \cdots & k & \cdots & K \\
\left[\begin{array}{ccccc}
\mathbf{X}_{1,1} & \cdots & \mathbf{X}_{1,k} & \cdots & \mathbf{X}_{1,K} \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
\mathbf{X}_{n,1} & \cdots & \mathbf{X}_{n,k} & \cdots & \mathbf{X}_{n,K} \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
\mathbf{X}_{N,1} & \cdots & \mathbf{X}_{N,k} & \cdots & \mathbf{X}_{N,K}
\end{array}\right].
\end{array}
$$

$$(24)$$

For the wine example, the matrix $\mathbf{X}_{n,k}$ contains the ratings of the kth assessor for all the wines belonging to the nth region (cf. Table 1).

We denote by $\mathbf{Z}$ the J by K design matrix for the blocks from the columns of $\mathbf{X}$: $z_{j,k} = 1$ if column j belongs to block k, $z_{j,k} = 0$ otherwise. We denote by $\mathbf{w}$ the J by 1 vector of *weights* for the columns of $\mathbf{X}$ and by $\mathbf{W}$ the J by J diagonal matrix whose diagonal elements are the elements of $\mathbf{w}$. We denote by $\mathbf{c}$ the K by 1 vector of weights for the blocks of $\mathbf{X}$ and by $\mathbf{C}$ the K by K diagonal matrix whose diagonal elements are the elements of $\mathbf{c}$. Weights are positive numbers, and it is convenient (but not necessary) to have the sum of the weights equal to one.

Partial Projections

Each block can be projected in the common solution. The procedure starts by rewriting Eq. 4 in order to show the blocks:

		Regions		
Axis	Assessor	Loire	Rhône	Beaujolais
1	Assessor 1	0.26	−0.04	−0.23
2	Assessor 1	0.02	−0.25	0.23
1	Assessor 2	1.36	−0.12	−1.24
2	Assessor 2	0.15	−1.75	1.60
1	Assessor 3	2.55	−0.82	−1.72
2	Assessor 3	0.44	−0.97	0.53
1	Assessor 4	0.29	−0.21	−0.07
2	Assessor 4	0.30	−0.45	0.15

Barycentric Discriminant Analysis, Table 4 Partial factor scores for the blocks (i.e., assessors)

Barycentric Discriminant Analysis, Fig. 5 Multiblock barycentric discriminant analysis. Projections on the first two dimensions. Position of the assessors relative to the barycenters of the wine regions. The assessors are projected as supplementary elements

$$\mathbf{R} = \mathbf{P\Delta Q}^{\top} = \mathbf{P\Delta}[\mathbf{Q}_1, \ldots, \ \mathbf{Q}_k, \ldots, \ \mathbf{Q}_K]^{\top}, \quad (25)$$
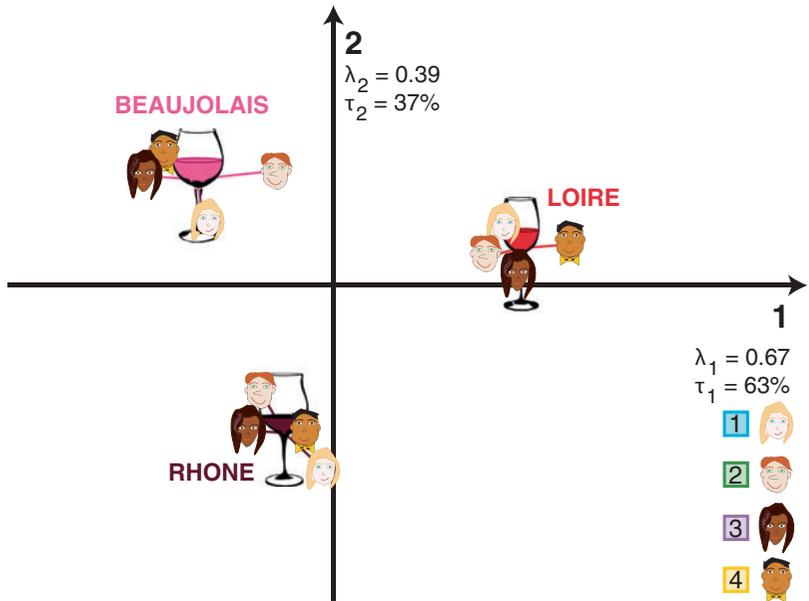
where $\mathbf{Q}_k$ is the kth block (comprising the J_k columns of $\mathbf{Q}$ corresponding to the J_k columns of the kth block). Then, Eq. 5 is rewritten to get the projection for the k-th block as

$$\mathbf{F}_k = K\mathbf{X}_k\mathbf{W}_k\mathbf{Q}_k \quad (26)$$

(where $\mathbf{W}_k$ is the weight matrix for the J_k columns of the k-th block).

Inertia of a Block

Recall from Eq. 6 that, for a given dimension, the variance of the factor scores of all the J columns of matrix $\mathbf{R}$ is equal to the eigenvalue of this dimension. Because each block comprises a set of columns, the contribution of a block to a dimension can be expressed as the sum of this dimension squared factor scores of the columns of this block. Precisely, the inertia for the kth table and the ℓth dimension is computed as:

Barycentric Discriminant Analysis, Table 5 Partial inertias of the blocks

Axis	Assessor 1	Assessor 2	Assessor 3	Assessor 4	$\sum = \lambda$
1	0.0408	0.2139	0.3743	0.0370	0.6660
2	0.0302	0.2113	0.1056	0.0456	0.3927
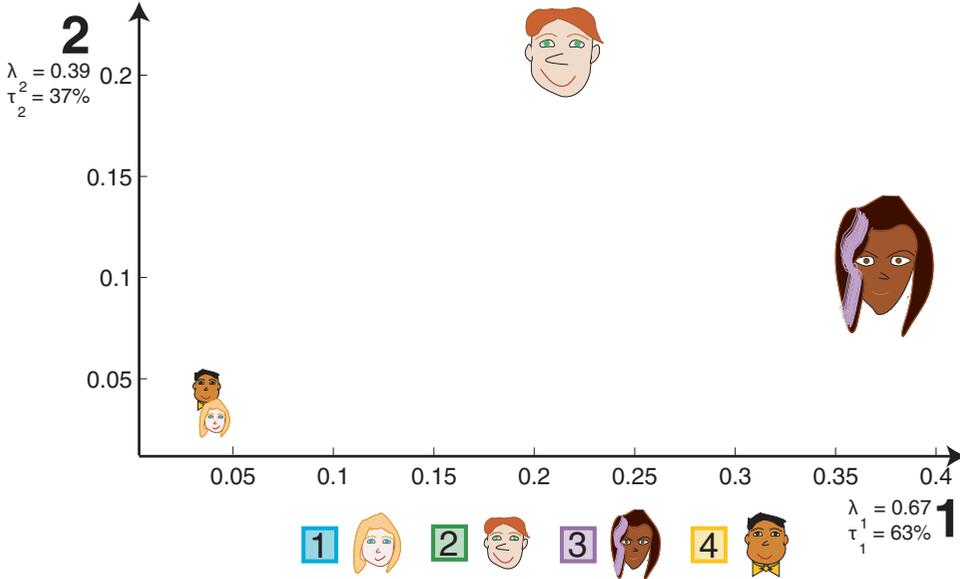
Barycentric Discriminant Analysis, Fig. 6 Multiblock barycentric discriminant analysis. Partial inertia of the blocks (i.e., assessors) on the first two dimensions

$$\mathcal{I}_{\ell,k} = \sum_{j \in J_k} w_j f_{\ell,j}^2. \tag{27}$$

Note that the sum of the inertia of the blocks gives back the total inertia:

$$\lambda_\ell = \sum_k \mathcal{I}_{\ell,k}. \tag{28}$$

Back to the example

To see how each assessor separated the wine regions, we project – as supplementary elements – each assessor's factor scores for the three wine regions (cf. Eq. 26). These factor scores are given in Table 4 and are displayed in Fig. 5.

Inertia of the Blocks

The respective importance of each assessor (i.e., block) is quantified by its associated partial inertia (see Eq. 27). The partial inertias are given in

Table 5 and are displayed in Fig. 6. We can see that Assessors 2 and 3 account for most of the inertia in the analysis with Assessor 3 determining most of Dimension 1 and Assessor 2 determining most of Dimension 2.

Special Cases of Multiblock Barycentric Analysis

When the variables measured are qualitative data (e.g., such as counts, 0/1 data, or ranks), the principal component approach cannot be used because it requires quantitative data. In such cases, the principal component approach is replaced by correspondence analysis and the multiblock version becomes multiblock correspondence analysis (MUDICA, see, e.g., Williams et al. 2010).

Related Methods

In the same way that linear discriminant analysis can be considered as a particular case of canonical correlation analysis when one of the matrices is a "design matrix" (i.e., a matrix that codes the group to which each observation belongs, also called a "group matrix"), BADA can be considered as a particular case of Tucker's interbattery analysis ((Tucker 1958) also known as co-inertia analysis (Chessel and Mercier 1993; Doledec and Chessel 1994), partial-least square correlation, PLSC (Krishnan et al. 2010)) when one of the matrices used is a design matrix. In the framework of PLSC, BADA is equivalent to "mean-centered-PLSC" (Krishnan et al. 2010). Along similar lines, BADA is also a particular case of constrained principal component analysis (Takane 2013) when one of the data matrices used by the method is a group matrix.

The multiblock or multisubject versions of BADA (i.e., MUBDADA and MUSUBADA) are also closely related to other multitables techniques such as multiple factor analysis (Abdi et al. 2013) and the STATIS family of techniques (Abdi et al. 2012c).

Implementation

BADA and several related methods such as MUBADA, MUSUBADA, DICA, and MUDICA are available from several R packages including, among others, TExPosition (Beaton et al. 2014) and ade4 (but often under different names (Dray and Dufour 2007)).

Key Applications

BADA and its derivatives – or variations thereof – are used when the analytic problem is to assign observation to predefined categories, and this makes these techniques ubiquitous in almost any domains of inquiry from marketing to brain imaging genetics (and other "omics") and network analysis.

Future Directions

BADA is still a domain of intense research with future developments likely to be concerned with multitable extensions (e.g., (Horst 1961)), "robustification," and sparsification (Witten et al. 2009). All these approaches will make BADA and its related techniques even more suitable for the analysis of very large data sets that are becoming prevalent in analytics.

Cross-References

▶ Canonical Correlation Analysis, Correspondence Analysis
▶ Eigenvalues, Singular Value Decomposition
▶ Iterative Methods for Eigenvalues/Eigenvectors
▶ Least Squares
▶ Matrix Algebra, Basics of
▶ Matrix Decomposition
▶ Principal Component Analysis
▶ Regression Analysis
▶ Spectral Analysis
▶ Visualization of Large Networks

References

Abdi H (2003) Multivariate analysis. In: Lewis-Beck M, Bryman A, Futing T (eds) Encyclopedia for research methods for the social sciences. Sage, Thousand Oaks, pp 699–702

Abdi H (2007a) Singular value decomposition (SVD) and generalized singular value decomposition (GSVD). In: Salkind NJ (ed) Encyclopedia of measurement and statistics. Sage, Thousand Oaks, pp 907–912

Abdi H (2007b) Discriminant correspondence analysis (DICA). In: Salkind NJ (ed) Encyclopedia of measurement and statistics. Sage, Thousand Oaks, pp 270–275

Abdi H, Williams LJ (2010a) Jackknife. In: Salkind NJ (ed) Encyclopedia of research design. Sage, Thousand Oaks

Abdi H, Williams LJ (2010b) Principal component analysis. Wiley Interdiscip Rev: Comput Stat 2:433–459

Abdi H, Williams LJ (2010c) Barycentric discriminant analysis (BADIA). In: Salkind NJ (ed) Encyclopedia of measurement and statistics. Sage, Thousand Oaks, pp 64–65

Abdi H, Dunlop JP, Williams LJ (2009) How to compute reliability estimates and display confidence and

tolerance intervals for pattern classifiers using the bootstrap and 3-way multidimensional scaling (DISTATIS). NeuroImage 45:89–95

Abdi H, Williams LJ, Beaton D, Posamentier M, Harris TS, Krishnan A, Devous MD (2012a) Analysis of regional cerebral blood flow data to discriminate among Alzheimer's disease, fronto-temporal dementia, and elderly controls: a multi-block barycentric discriminant analysis (MUBADA) methodology. J Alzheimer Dis 31:s189–s201

Abdi H, Williams LJ, Connolly AC, Gobbini MI, Dunlop JP, Haxby JV (2012b) Multiple subject Barycentric discriminant analysis (MUSUBADA): how to assign scans to categories without using spatial normalization. Comput Math Methods Med 2012:1–15. https://doi.org/10.1155/2012/634165

Abdi H, Williams LJ, Valentin D, Bennani-Dosse M (2012c) STATIS and DISTATIS: optimum multi-table principal component analysis and three way metric multidimensional scaling. Wiley Interdiscip Rev: Comput Stat 4:124–167

Abdi H, Williams LJ, Valentin D (2013) Multiple factor analysis: principal component analysis for multi-table and multi-block data sets. Wiley Interdiscip Rev: Comput Stat 5:149–179

Bastin C, Benzécri JP, Bourgarit C, Caze P (1982) Pratique de l'Analyse des Données. Dunod, Paris, pp 102–104

Beaton D, Chin Fatt CR, Abdi H (2014) An ExPosition of multivariate analysis with the singular value decomposition in R. Comput Stat & Data Anal 72:176–189

Benzécri J-P (1977) Analyse discriminante et analyse factorielle. Les Cahiers de l'Analyse des Données 2:369–406

Bergougnan D, Couraud C (1982) Pratique de la discrimination barycentrique. Les Cahiers de l'Analyse des Données 7:341–354

Celeux P, Nakache JP (1994) Analyse discriminante sur variables qualitatives. Polytechnica, Paris

Chessel D, Mercier P (1993) Couplage de triplet statistiques et liaisons espèce-environnement. In: Lebreton JD, Asselain B (eds) Biométrie et Environnement. Dunod, Paris, pp 15–43

Cioli C, Abdi H, Beaton D, Burnod Y, Mesmoudi S (2014) Human cortical gene expression and properties of functional networks. PLoS One 9(12):1–28

Diaconis P, Efron B (1983) Computer-intensive methods in statistics. Scientific American 248:116–130

Doledec S, Chessel D (1994) Co-inertia analysis: an alternative method for studying species- environment relationships. Freshw Biol 31:277–294

Dray S, Dufour AB (2007) The ade4 package: implementing the duality diagram for ecologists. J Stat Softw 22(4):1–20

Efron B, Tibshirani RJ (1993) An introduction to the bootstrap. Chapman & Hall, New York

El Behi M, Sanson C, Bachelin C, Guillot-Noel L, Fransson J, Stankoff B, Maillart E, Sarrazin N, Guillemot V, Abdi H, Rebeix I, Fontaine B, Zujovic V (2017) Adaptive human immunity drives remyelination in a mouse model of demyelination. Brain 140(4):967–980

Gittins R (1980) Canonical analysis: a review with applications in ecology. Springer Verlag, New York

Greenacre MJ (1984) Theory and applications of correspondence analysis. Academic Press, London

Horst P (1961) Relations among m sets of measures. Psychometrika 26:129–149

Krishnan A, Williams LJ, McIntosh AR, Abdi H (2010) Partial least squares (PLS) methods for neuroimaging: a tutorial and review. NeuroImage 56:455–475

Krzanowski WJ, Radley D (1989) Nonparametric confidence and tolerance regions in canonical variate analysis. Biometrics 45:1163–1173

Leclerc A (1976) Une etude de la relation entre une variable qualitative et un groupe de variables qualitatives. Int Stat Rev 44:241–248

Manly BFJ (1997) Randomization, bootstrap, and Monte Carlo methods in biology, 2nd edn. Chapman & Hall, New York

Nakache J-P, Lorente P, Benzcri J-P, Chastang J-F (1977) Aspects pronostiques et therapeutiques de l'infarctus myocardique aigu compliqu d'une dfaillance sévère de la pompe cardiaque. Application des methodes de discrimination Les Cahiers de l'Analyse des Données 2:415–434

Perriere G, Lobry JR, Thioulouse J (1996) Correspondence discriminant analysis: a multivariate method for comparing classes of protein and nucleic acid sequences. CABIOS 12:519–524

Saporta G, Niang N (2006) Correspondence analysis and classification. In: Greenacre M, Blasius J (eds) Multiple correspondence analysis and related methods. Boca Raton, Chapman & Hall/CRC, pp 371–392

St. Laurent M, Abdi H, Burianová H, Grady GL (2011) Influence of aging on the neural correlates of autobiographical, episodic, and semantic memory retrieval. J Cogn Neurosci 23:4150–4163

Takane Y (2013) Constrained principal component analysis and related techniques. CRC Press, Boca Raton

Tucker LR (1958) An inter-battery method of factor analysis. Psychometrika 23:111–136

Williams LJ, Abdi H, French R, Orange JB (2010) A tutorial on multi-block discriminant correspondence analysis (MUDICA): a new method for analyzing discourse data from clinical populations. J Speech Lang Hear Res 53:1372–1393

Witten DM, Tibshirani R, Hastie T (2009) A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. Biostatistics 10:515–534