Discriminant Analysis for Multi-way Data

Gisela Lechuga, Laurent Le Brusquet, Vincent Perlbarg, Louis Puybasset, Damien Galanaud, Arthur Tenenhaus

Discriminant Analysis for Multi-way Data

Lechuga G.,[a*] Le Brusquet L.,[a] Perlbarg V.,[b] Puybasset L.,[c] Galanaud G.[d] and Tenenhaus A.[a]

[a]Supélec Sciences des Systèmes - EA4454 (E3S), Gif-sur-Yvette, France
[b]Sorbonne Universités, UPMC Univ Paris 06 UMCR 2, INSERM UMR_S 1146, CNRS UMR 7371, Laboratoire d'Imagerie Biomédicale, F-75005, Paris, France
[c]AP-HP, Pitié-Salpêtrière Hospital, Surgical Neuro-Intensive Care Unit, Paris, France
[d]AP-HP, Pitié-Salpêtrière Hospital, Department of Neuroradiology, Paris, France

Keywords: Discriminant analysis, multi-way data, medical imaging, high dimension

Introduction

In standard multivariate data analysis, individuals $\times$ variables data table is usually considered (two-way data table). However, from a practical view point this simple data structure appears to be somehow limitated. It is the case for instance when individuals are charaterized by the temperature at different locations sampled over different times, leading to a three-way data structure. Such multi-way structure can be viewed as a stack of matrices $\underline{\mathbf{X}} = \left\{X_{ijk}\right\}_{1\leq i\leq I,\, 1\leq j\leq J,\, 1\leq k\leq K}$ from which the I horizontal slices describe the individuals $i = 1,...,I$, the J lateral slices describe the variables (temperature) $j = 1,...,J$ and the K frontal slices describe the different time points $k = 1,...,K$. Many two-way data analysis methods have been extended to the multi-way configuration. For instance, a multi-way formulation of Partial Least Squares Regression (N-PLS) has been proposed in [1]. N-PLS relies on the maximization of a covariance criterion but explicitely takes into acount the multi-way structure of the input data. In this paper, we present a Multi-way formulation of Fisher Discriminant Analysis (MFDA) in an attempt to improve the interpretability of the resulting model compared with the results obtained with unfolded methods. MFDA is illustrated on a real multi-modal Magnetic Resonance Brain Imaging (MRI) dataset [1].

1 Multi-way FDA analysis

FDA is defined by the optimization problem: $\mathbf{w}^* = \arg\max_w \frac{\mathbf{w}^t\mathbf{S}_B\mathbf{w}}{\mathbf{w}^t\mathbf{S}_T\mathbf{w}+\lambda\mathbf{w}^t\mathbf{w}}$, where $\mathbf{S}_B$ is the between covariance matrix, $\mathbf{S}_T$ is the total covariance matrix and $\lambda\mathbf{w}^t\mathbf{w}$ is an additive regularization term required in the high dimensional setting. MFDA seeks to maintain the natural tensor structure of the input data by constraining $\mathbf{w}$ to be of the form $\mathbf{w} = \mathbf{w}^K \otimes \mathbf{w}^J$. $\mathbf{w}^K$ is a weight vector associated with the K modalities while $\mathbf{w}^J$ is the weight vector related to the J variables. From the following equalities an alternating algorithm is developed to maximize the FDA optimization problem subject to the structural constraint that $\mathbf{w} = \mathbf{w}^K \otimes \mathbf{w}^J$.

$$\begin{aligned}
\mathbf{w}^t\mathbf{S}_{B/T}\mathbf{w} &= (\mathbf{w}^K \otimes \mathbf{w}^J)^t(\mathbf{X}^u)^t\mathbf{M}_{B/T}\mathbf{X}^u(\mathbf{w}^K \otimes \mathbf{w}^J). \\
&= (\mathbf{w}^J)^t(\mathbf{X}^u(\mathbf{w}^K \otimes \mathbb{1}_J))^t\mathbf{M}_{B/T}(\mathbf{X}^u(\mathbf{w}^K \otimes \mathbb{1}_J))\mathbf{w}^J. \quad (1) \\
&= (\mathbf{w}^K)^t(\mathbf{X}^u(\mathbb{1}_K \otimes \mathbf{w}^J))^t\mathbf{M}_{B/T}(\mathbf{X}^u(\mathbb{1}_K \otimes \mathbf{w}^J))\mathbf{w}^K. \quad (2)
\end{aligned}$$

*Corresponding author. E-mail: gisela.lechuga@supelec.fr.

[1]This study was funded by a grant from the French Ministry of Health (Projet Hospitalier de Recherche Clinique registration #P051061 [2005]) and from departmental funds from the Assistance Publique-Hôpitaux de Paris. The research leading to these results has also received funding from the program "Investissements d'avenir" ANR-10-IAIHU-06.

$\mathbf{M}_{B/T}$ are positive semi-definite matrices that depend only of the label vector $\mathbf{y}$ and $\mathbf{X}^u$ is the unfolded matrix. The alternating algorithm is described in Algorithm 1, which starts by assigning random initial values for $\mathbf{w}^J$ or $\mathbf{w}^K$ and iterates a sequence of FDA problems. More specifically, each update boils down to perform FDA between $\mathbf{y}$ and either $\mathbf{X}^J$ or $\mathbf{X}^K$ where $\mathbf{X}^J = \mathbf{X}^u(\mathbf{w}^K \otimes 1\!\!1_J) = \sum_{k=1}^K (\mathbf{w}^K)_k X_{1,\ldots k}$ and $\mathbf{X}^K = \mathbf{X}^u(1\!\!1_K \otimes \mathbf{w}^J) = \sum_{j=1}^J (\mathbf{w}^J)_j X_{1,\ldots J}$. We can note that $\mathbf{X}^J$ (resp. $\mathbf{X}^K$) is a $I \times J$ (resp. $I \times K$) matrix. Algorithm 1 allows to calculate $(\mathbf{w}_1^J, \mathbf{w}_1^K)$, which

Algorithm 1 Algorithm to calculate the first axis of Multi-way FDA analysis

Require: $\varepsilon > 0$, $\mathbf{w}^{K(0)}$
  $q \leftarrow 0$
  repeat
    $\mathbf{w}^{J(q+1)} \leftarrow \arg\max_{\mathbf{w}^J} \frac{(\mathbf{w}^J)^t(\mathbf{X}^u(\mathbf{w}^{K(q)} \otimes 1\!\!1_J))^t \mathbf{S}_B (\mathbf{X}^u(\mathbf{w}^{K(q)} \otimes 1\!\!1_J))\mathbf{w}^J}{(\mathbf{w}^J)^t(\mathbf{X}^u(\mathbf{w}^{K(q)} \otimes 1\!\!1_J))^t \mathbf{S}_W (\mathbf{X}^u(\mathbf{w}^{K(q)} \otimes 1\!\!1_J))\mathbf{w}^J} \leftarrow \mathrm{FDA}(\mathbf{y}, \mathbf{X}^u(w^{K(q)} \otimes 1\!\!1_J))$
    $\mathbf{w}^{K(q+1)} \leftarrow \arg\max_{\mathbf{w}^K} \frac{(\mathbf{w}^K)^t(\mathbf{X}^u(1\!\!1_K \otimes \mathbf{w}^{J(q+1)}))^t \mathbf{S}_B (\mathbf{X}^u(1\!\!1_K \otimes \mathbf{w}^{J(q+1)}))\mathbf{w}^K}{(\mathbf{w}^K)^t(\mathbf{X}^u(1\!\!1_K \otimes \mathbf{w}^{J(q+1)}))^t \mathbf{S}_W (\mathbf{X}^u(1\!\!1_K \otimes \mathbf{w}^{J(q+1)}))\mathbf{w}^K} \leftarrow \mathrm{FDA}(\mathbf{y}, \mathbf{X}^u(1\!\!1_K \otimes \mathbf{w}^{J(q+1)}))$
  until $\|\mathbf{w}^{K(q)} - \mathbf{w}^{K(q+1)}\| < \varepsilon$
  return $(\mathbf{w}^{K(q)}, \mathbf{w}^{J(q)})$

corresponds to the first discriminant axis. The following $C-1$ discriminant axes, $\mathbf{w}_s^J$, $\mathbf{w}_s^K$, $s = 2, \ldots, C-1$ are obtained subject to the additional orthogonality constraint between $\mathbf{w}_{s+1}^K$ and $\mathbf{w}_1^K, \ldots \mathbf{w}_s^K$ [2].

2 Results

MFDA is applied on multi-modal diffusion images acquired on individuals divided into 3 classes: 39 controls, 65 coma patients with a positive outcome and 39 coma patients with a negative outcome ($I = 143$). 4 diffusion images namely fractional anisotropy (FA), mean diffusivity, axial diffusivity and radial diffusivity were acquired from the entire brain of the patients and controls ($K = 4$). Each image has a size of $91 \times 109 \times 91$ voxels, reshaped into a 1×902629 vector ($J = 902629$). We mention that due to the dimensionality of the dataset, a kernel version of FDA is used. The leave-one-out test error rate obtained with MFDA is equal to 71% whereas for the unfolded method the accuracy was of 76%. This slight loss in accuracy is compensated by an improvement in the interpretability of the obtained classifier as seen in Figures 1 and 2 which shows an axial cut at a central slice. Such improvement is partly due to the chosen structure for modeling $\mathbf{w}$; MFDA clearly separates the influence of spatial positions and the influence of the modalities. FDA applied to $\mathbf{X}^u$ results in 8 weight matrices (4 for each eigenvector), which complicate the interpretability, opposed to only 2 weight matrices obtained with MFDA which integrate all the modalities. Interestingly in our application, we have exhibited from MFDA that the discriminating voxels are located, as expected, within the main white matter bundles. Indeed, traumatic brain injury is characterized by the presence of diffuse axonal injury mainly located within deep and axial white matter bundle.

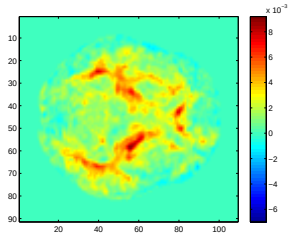

Figure 1. MFDA obtained weights.

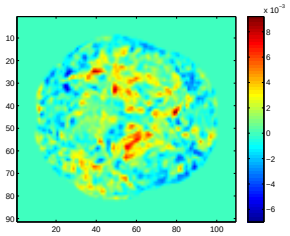

Figure 2. FDA FA obtained weights.

References

[1] R. Bro, *Multi-way Analysis in the food industry: models, algorithms and applications*. PhD thesis, Royal Veterinary and Agricultural University, 1998.

[2] L. L. Brusquet and A. Tenenhaus, "Analyse factorielle discriminante multi-voie," *Journees de Statistique* , 2013.