Machine learning methods to help classify liver tumors from radiomics data

Alexandre SELVESTREL[1]

[1]Laboratoire des systèmes, Orsay France
[2]Centrale-Supélec, Gif-sur-Yvette France

Abstract

In this report, we investigate whether tensor-based machine learning algorithms outperform more traditional machine learning on tensor-structured data. Specifically, we benchmark these methods on a radiomic and clinical dataset of liver tumors. The radiomic variables for each tumor are categorized into three main groups: first-order, shape, and texture, and are observed at four time points corresponding to different phases after contrast medium injection: arterial, portal, venous, and late. This organization gives the radiomic data a tensorial structure, and we aim to determine if models that leverage this structure lead to better predictions.

More precisely, we aim to distinguish patients whose tumor type is "CCK" from those whose tumor type is "CHC". This is a complex task, since only microscopy of cancerous tissue can characterize without hesitation the tumor type. Indeed, experts do not always agree with each other when they analyze X-rays alone. We begin by proposing a logistic regression penalized by lasso as a baseline. We then compare its results with those of algorithms exploiting the tensorial structure of the data: group lasso [1], multiway logistic regression [2, 3] and finally multiway and multiblock logistic regression.

Keywords: Machine learning, Multibloc, Multiway, Liver tumor .

1   Introduction

We speak of tensorial data as soon as the explanatory variables are structured along several dimensions. In our case, we have two dimensions for our explanatory variables: the quantity of interest and the time at which it is measured (arterial, portal, venous or late). So, instead of having a matrix of explanatory variables, we have a tensor of explanatory variables $\underline{\mathbf{X}} = (x_{ijk})_{i\in[\![1,n]\!], j\in[\![1,J]\!], k\in[\![1,K]\!]}$, where i is the individual, j is the quantity of interest and k is the observation time. To avoid confusion with the notion of dimension of a vector space, and in line with the conventions promoted in *Kolda and Bader* (2009) [4], the three dimensions of the tensor $\underline{\mathbf{X}}$ will be referred to as "mode" in the following.

In order to train models that are not specific to tensor data (like a penalized logistic regression), we unfold the tensor $\underline{\mathbf{X}}$ and transform it into a matrix by placing each slice at a fixed k next to the other. This gives the following matrix: $\mathbf{X}_{(1)} = [\mathbf{X}_{::1}, \ldots, \mathbf{X}_{::K}]$. However, in doing so, the model is no longer informed that $\underline{\mathbf{X}}$ is made up of the same quantities observed at different times. One of the reason for which this is important is that there is a non negligeable correlation between variables describing the same quantity of interest at

different times. We'll still use the group lasso to restore this information to the model by grouping together variables belonging to the same time, but this is not enough to recover the full tensor aspect of the initial data. Furthermore, as pointed out by *Le Brusquet et al.* (2014) [2], unfolding the matrix greatly increases the number of predictors compared to tensor methods, which can increase computation times. However, given that in our dataset, K is small ($K = 4$), this increase in computation time is not observed with our data.

Introduire ici les CHC et CCK et l'interet de l'etude. Les mixtes notamment.

2 Notations

In this section, we specify the notations used in this paper

- The concatenation of two matrices $\mathbf{A}$ and $\mathbf{B}$ by juxtaposing their columns side by side is denoted $[\mathbf{A} \ \mathbf{B}]$.
- To avoid overuse of the symbol ^{T}, we also define a notation to designate the juxtaposition of two matrices one below the other. Thus, the matrix defined by block with $\mathbf{A}$ above $\mathbf{B}$ is denoted $[\mathbf{A}; \mathbf{B}]$. It can also be written $[\mathbf{A}^T \ \mathbf{B}^T]^T$ but this multiplies the ^{T} symbols, which impairs legibility.
- Since vectors are column matrices, using the same notation, we write the concatenation of two vectors $\mathbf{u}$ and $\mathbf{v}$ as follows: $[\mathbf{u}; \mathbf{v}]$.
- The vector (column) whose elements are $(u_i)_{i \in [\![1,I]\!]}$ is also denoted $(u_1, u_2, \cdots u_I)$.
- If $\mathbf{X}$ is a matrix of explanatory variables, $\mathbf{x}_i$ is the vector (column) composed of the i-th row of $\mathbf{X}$.
- The vector of length I filled with 1 is denoted by $\mathbb{1}_I$.
- We denote $\mathrm{Diag}(\mathbf{u})$ the diagonal matrix whose diagonal is the vector $\mathbf{u}$.
- For other notations, we use the conventions employed by *Kolda and Bader* (2009) [4]. In particular, when we transform a tensor $\underline{\mathbf{X}}$ into a matrix (by unfolding it) using the first mode to form the rows, we do so as follows: $[\mathbf{X}_{::1} \ \ldots \ \mathbf{X}_{::K}]$ and we denote $\mathbf{X}_{(1)}$ the matrix thus obtained.

3 State of the art

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

4 Detailed presentation of the data

Détailler le jeu de données et expliquer qu'on a enlevé la phase veineuse. Indiquer aussi qu'on a des données cliniques en plus... Justifier le refus de l'inputation de données (qui risquerait de biaiser les résultats vu le nombre de données manquantes sur certains individus). Expliquer la séparation en blocs.

5   Methodology

In this section, we present in detail the different models that we benchmarked on our dataset. We describe briefly the non tensorial methods so we can focus more on the tensorial ones.

5.1   Non tensorial methods

For these methods, we start by unfolding the tensorial data $\underline{\mathbf{X}}$ into the matrix $\mathbf{X}_{(1)} = \begin{bmatrix} \mathbf{X}_{::1} & \ldots & \mathbf{X}_{::K} \end{bmatrix}$. We then complete this matrix by concatenating (along the columns) the matrix of non tensorial data $\mathbf{X}_{\text{tab}}$ (where "tab" stands for "tabular"). By doing so we obtain $\mathbf{X}_{\text{tot}} = \begin{bmatrix} \mathbf{X}_{(1)} & \mathbf{X}_{\text{tab}} \end{bmatrix}$.

We first train a penalized logistic regression lasso on $\mathbf{X}_{\text{tot}}$. Then, still based on the matrix $\mathbf{X}_{\text{tot}}$, we train a group lasso [1]. The idea of this second model is to group together all the explanatory variables corresponding to a same slice of $\underline{\mathbf{X}}$. We do this first for slices at given j and then for slices at given k. The goal is to give the model a vague description of the tensorial structure of the data. Finally, because we also identified blocs of variables in our tensorial data, we try a last group lasso by grouping the variables along these blocs.

5.2   Multiway logistic regression with lasso

We now turn to tensor approaches. We start by studying a multiway logistic regression penalized by lasso. This model is described for rank 1 in *Le Brusquet et al.* (2014) [2] and in *Girka et al.* (2021) [3] for its extension to rank $R \in \mathbb{N}^*$. In this report, we directly describe the generalization to rank $R \in \mathbb{N}^*$, rank 1 being a special case of this model.

The fundamental idea of the model is to decompose the parameter $\boldsymbol{\beta}_{\text{tens}} \in \mathbb{R}^{JK}$ associated with the tensor explanatory variables of the logistic regression as:

$$\boldsymbol{\beta}_{\text{tens}} = \sum_{r=1}^{R} \boldsymbol{\beta}_r^K \otimes \boldsymbol{\beta}_r^J \tag{1}$$

with for all $r \in [\![1, R]\!]$, $\boldsymbol{\beta}_r^J \in \mathbb{R}^J$ and $\boldsymbol{\beta}_r^K \in \mathbb{R}^K$. To take account of the M clinical variables, which are not tensorial, we associate them with a coefficient $\boldsymbol{\beta}_{\text{tab}} \in \mathbb{R}^M$. In this way, the parameter $\boldsymbol{\beta}$ of the logistic regression is written: $[\boldsymbol{\beta}_{\text{tens}}; \boldsymbol{\beta}_{\text{tab}}]$.

As usual with logistic regressions, we consider that each realization of the explained variable y_i ($i \in [\![1, n]\!]$) follows an independent Bernoulli law conditionally on $\mathbf{x_i}$. This law is defined by:

$$\mathbb{P}(y_i = 1 \,|\, \mathbf{x}_i) = \frac{1}{1 + \exp(-\mathbf{x}_i^T \boldsymbol{\beta} - \beta_0)} \tag{2}$$

where $\beta_0 \in \mathbb{R}$ is the intercept

We set $\boldsymbol{\beta}^J = \begin{bmatrix} \boldsymbol{\beta}_1^J; & \cdots & ; \boldsymbol{\beta}_R^J \end{bmatrix}$ and $\boldsymbol{\beta}^K = \begin{bmatrix} \boldsymbol{\beta}_1^K; & \cdots & ; \boldsymbol{\beta}_R^K \end{bmatrix}$. In order to simplify the calculations, while ensuring that the penalty continues to promote sparse models, we adapt the definition of the lasso penalty. The new penalty defines the following optimization problem:

$$\beta_0, \boldsymbol{\beta}^J, \boldsymbol{\beta}^K, \boldsymbol{\beta}_{\text{tab}} = \underset{\beta_0, \boldsymbol{\beta}^J, \boldsymbol{\beta}^K, \boldsymbol{\beta}_{\text{tab}}}{\operatorname{argmin}} \left(-\sum_{i=1}^{N} \log(\mathbb{P}(y_i = 1 \,|\, \mathbf{x}_i)) + \sum_{r=1}^{R} \|\boldsymbol{\beta}_r^K \otimes \boldsymbol{\beta}_r^J\|_1 + \|\boldsymbol{\beta}_{\text{tab}}\|_1 \right) \tag{3}$$

Optimization is performed by alternating directions between $[\beta_0; \boldsymbol{\beta}^J; \boldsymbol{\beta}_{\text{uni}}]$ and $[\beta_0; \boldsymbol{\beta}^K; \boldsymbol{\beta}_{\text{uni}}]$. The stopping criterion is defined by the relative difference between the value of the objective function after optimization in the first direction and the value of the same function after optimization in the second direction. We note that optimizing the loss function in each of these directions is tantamount to performing a simple logistic regression with a lasso penalty. Indeed, if we denote C the loss function of classical logistic regression penalized by lasso (for any $K_0 \in \mathbb{N}^\star$):

$$C : \begin{cases} \mathbb{R} \times \mathbb{R}^{K_0} \times \mathbb{R}^{N \times K_0} \times \mathbb{R}^N \times \mathbb{R} \longrightarrow \mathbb{R} \\ \\ (\beta_0, \boldsymbol{\beta}, \mathbf{X}, \mathbf{y}, \lambda) \longmapsto -\sum_{i=1}^{N} [y_i(\beta_0 + \mathbf{x}_i^T \boldsymbol{\beta}) - \log(1 + \exp(\beta_0 + \mathbf{x}_i^T \boldsymbol{\beta}))] + \lambda \|\boldsymbol{\beta}\|_1 \end{cases} \tag{4}$$

optimizing the overall loss function with respect to $[\beta_0; \boldsymbol{\beta}^J; \boldsymbol{\beta}_{\text{uni}}]$ amounts to solve

$$\underset{(\beta_0, \boldsymbol{\beta}) \in \mathbb{R} \times \mathbb{R}^{JR+M}}{\operatorname{argmin}} C(\beta_0, (\mathbf{Q}^J)^{-1}\boldsymbol{\beta}, \mathbf{Z}^J \mathbf{Q}^J, \mathbf{y}, \lambda) \tag{5}$$

Where $\mathbf{Q}^J$ and $\mathbf{Z}^J$ are defined as follows:

$$\mathbf{Z}^J = [\mathbf{Z}_1^J \ \cdots \ \mathbf{Z}_R^J \ \mathbf{X}_{\text{tab}}] \tag{6}$$

$$\text{where } \forall r \in [\![1, R]\!], \quad \mathbf{Z}_{\mathbf{r}}^{\mathbf{J}} = \sum_{k=1}^{K} \mathbf{X}_{::k}(\beta_r^K)_k \quad (\mathbf{Z}_r^J \in \mathbb{R}^{N \times J}) \tag{7}$$

$$\mathbf{Q}^J = \operatorname{Diag}([\|\boldsymbol{\beta}_1^K\|_1^{-1} \mathbb{1}_J; \ \cdots \ ; \|\boldsymbol{\beta}_R^K\|_1^{-1} \mathbb{1}_J; \mathbb{1}_M]) \tag{8}$$

Girka et al. (2021) [3] demonstrate this result by noting that for $i \in [\![1, n]\!]$,

$$\mathbf{x}_{(1)_i}^T \left(\sum_{r=1}^{R} \boldsymbol{\beta}_r^K \otimes \boldsymbol{\beta}_r^J \right) = \sum_{r=1}^{R} \left[\left(\mathbf{x}_{(1)_i}^T \left(\boldsymbol{\beta}_r^K \otimes \mathbf{I}_J \right) \right] \boldsymbol{\beta}_r^J \tag{9}$$

$$= \sum_{r=1}^{R} (\mathbf{z}_r^J)_i^T \boldsymbol{\beta}_r^J \tag{10}$$

and that

$$\sum_{r=1}^{R} \|\boldsymbol{\beta}_r^K \otimes \boldsymbol{\beta}_r^J\|_1 = \|\mathbf{R}_{\text{tens}}^J \boldsymbol{\beta}^J\|_1 \tag{11}$$

$$\text{with} \quad \mathbf{R}_{\text{tens}}^J = \operatorname{Diag}([\|\boldsymbol{\beta}_1^K\|_1 \mathbb{1}_J; \ \cdots \ ; \|\boldsymbol{\beta}_R^K\|_1 \mathbb{1}_J]) \tag{12}$$

Thus,

$$(\mathbf{x}_{\text{tot}})_i^T \boldsymbol{\beta} = (\mathbf{z}_i^J)^T [\boldsymbol{\beta}^J; \boldsymbol{\beta}_{\text{tab}}] \tag{13}$$

$$\text{and} \quad \sum_{i=1}^{N} \|\boldsymbol{\beta}_r^K \otimes \boldsymbol{\beta}_r^J\|_1 + \|\boldsymbol{\beta}_{\text{tab}}\|_1 = \|(\mathbf{Q}^J)^{-1}\boldsymbol{\beta}\|_1 \tag{14}$$

This justifies the previous results

For optimization with respect to $\left[\beta_0; \boldsymbol{\beta}^K; \boldsymbol{\beta}_{\text{tab}}\right]$, the method is analogous. The only difference concerns the definition of $\mathbf{Z}^K$. It is:

$$\mathbf{Z}^K = \begin{bmatrix} \mathbf{Z}_1^K & \cdots & \mathbf{Z}_R^K & \mathbf{X}_{\text{tab}} \end{bmatrix} \tag{15}$$

$$\text{with} \forall r \in [\![1, R]\!] \quad \mathbf{Z}_r^K = \sum_{j=1}^J (\beta_r^J)_j \mathbf{X}_{:j:} \tag{16}$$

This is justified by:

$$\mathbf{x}_{(1)_i}^T \left(\sum_{r=1}^R \boldsymbol{\beta}_r^K \otimes \boldsymbol{\beta}_r^J \right) = \sum_{r=1}^R \left[\left(\mathbf{x}_{(1)_i}^T \left(I_K \otimes \boldsymbol{\beta}_r^J \right) \right] \boldsymbol{\beta}_r^K \tag{17}$$

$$= \sum_{r=1}^R (\mathbf{z}_r^K)_i^T \boldsymbol{\beta}_r^K \tag{18}$$

5.3   Multiway and multibloc logistic regression with lasso

We now present the lasso-penalized multiway and multiblock logistic regression. This model draws heavily on the multiway logistic regression we have just presented, while also taking into account a block structure of tensor data. More precisely, each of these blocs will have its own independent coefficient $\boldsymbol{\beta}_l$, which was not the case in the previous model. As tabular quantities are not measured at different times, they are not placed in any particular block. They will be included in the model in the same way as in the multiway case. Mathematically, we define the model as follows:

Let $L \in \mathbb{N}^*$ denote the number of blocks of variables. For any $l \in [\![1, L]\!]$, let d_l be the number of tensorial quantities of interest in block l. Thus we have :

$$\sum_{l=1}^L d_l = J$$

We reorganize $\underline{\mathbf{X}}$ by grouping together slices $\mathbf{X}_{:j:}$ associated with quantities from the same block. More precisely, for all $l \in [\![1, L]\!]$, we call $\underline{\mathbf{X}}^l$ the tensor constituted by the slices $\mathbf{X}_{:j:}$ associated with the l-th bloc. We then concatenate these tensors along their second mode to obtain the new tensor of explanatory variables: $\underline{\mathbf{X}}'$.

The new $\boldsymbol{\beta}$ structure is defined by blocks. It is:

$$\boldsymbol{\beta} = \left[\sum_{r=1}^R \boldsymbol{\beta}_{(1,r)}^K \otimes \boldsymbol{\beta}_{(1,r)}^J; \ \cdots \ ; \ \sum_{r=1}^R \boldsymbol{\beta}_{(L,r)}^K \otimes \boldsymbol{\beta}_{(L,r)}^J \ ; \ \boldsymbol{\beta}_{\text{tab}} \right] \tag{19}$$

With for all $(l, r) \in [\![1, L]\!] \times [\![1, R]\!]$, $\boldsymbol{\beta}_{(l,r)}^J \in \mathbb{R}^{d_l}$ and $\boldsymbol{\beta}_{(l,r)}^K \in \mathbb{R}^K$

We call $\boldsymbol{\beta}^J$ and $\boldsymbol{\beta}^K$ the vectors

$$\boldsymbol{\beta}^J = \left[\boldsymbol{\beta}_{(1,1)}^J; \ \cdots \ ; \boldsymbol{\beta}_{(1,R)}^J; \ \cdots \ \cdots \ ; \boldsymbol{\beta}_{(L,1)}^J \ \cdots \ ; \boldsymbol{\beta}_{(L,R)}^J \right] \tag{20}$$

$$\boldsymbol{\beta}^K = \left[\boldsymbol{\beta}_{(1,1)}^K; \ \cdots \ ; \boldsymbol{\beta}_{(1,R)}^K; \ \cdots \ \cdots \ ; \boldsymbol{\beta}_{(L,1)}^K \ \cdots \ ; \boldsymbol{\beta}_{(L,R)}^K \right] \tag{21}$$

In a similar way to what is done in the multiway model, we adapt the lasso penalty, so that the new optimization problem becomes:

$$\beta_0, \boldsymbol{\beta}^J, \boldsymbol{\beta}^K, \boldsymbol{\beta}_{\text{tab}} = \underset{\beta_0, \boldsymbol{\beta}^J, \boldsymbol{\beta}^K, \boldsymbol{\beta}_{\text{tab}}}{\text{argmin}} \left(-\sum_{i=1}^N \log(\mathbb{P}(y_i = 1 | \mathbf{x}_i)) + \sum_{l=1}^L \sum_{r=1}^R \|\boldsymbol{\beta}_{(l,r)}^K \otimes \boldsymbol{\beta}_{(l,r)}^J\|_1 + \|\boldsymbol{\beta}_{\text{tab}}\|_1 \right) \tag{22}$$

Once again, this problem is solved by alternating optimization directions $\left[\beta_0; \boldsymbol{\beta}^J; \boldsymbol{\beta}_{\text{tab}} \right]$ and $\left[\beta_0; \boldsymbol{\beta}^K; \boldsymbol{\beta}_{\text{tab}} \right]$. Each of these two problems can be reduced to a lasso-penalized classical logistic regression.

Indeed, optimizing according to $\left[\beta_0; \boldsymbol{\beta}^J; \boldsymbol{\beta}_{\text{tab}} \right]$ is equivalent to searching

$$\underset{(\beta_0, \boldsymbol{\beta}) \in \mathbb{R} \times \mathbb{R}^{RJ+M}}{\text{argmin}} C(\beta_0, (\mathbf{Q}^J)^{-1} \boldsymbol{\beta}, \mathbf{Z}^J \mathbf{Q}^J, \mathbf{y}, \lambda) \tag{23}$$

Where $\mathbf{Q}^J$ and $\mathbf{Z}^J$ are defined as follows:

$$\mathbf{Z}^J = \left[\mathbf{Z}_{(1,1)}^J \; \cdots \; \mathbf{Z}_{(1,R)}^J \; \cdots \; \cdots \; \mathbf{Z}_{(L,1)}^J \; \cdots \; \mathbf{Z}_{(L,R)}^J \; \mathbf{X}_{\text{tab}} \right] \tag{24}$$

$$\text{where } \forall r \in [\![1, R]\!], \qquad \mathbf{Z}_{(l,r)}^J = \sum_{k=1}^K \mathbf{X}_{::k}^l \left(\beta_{(l,r)}^K \right)_k \qquad \left(\mathbf{Z}_{(l,r)}^J \in \mathbb{R}^{n \times d_l} \right) \tag{25}$$

$$\mathbf{Q}^J = \text{Diag}([\|\boldsymbol{\beta}_{(1,1)}^K\|_1^{-1} \mathbb{1}_{d_1}; \; \cdots \; ; \|\boldsymbol{\beta}_{(1,R)}^K\|_1^{-1} \mathbb{1}_{d_1}; \; \cdots \; \cdots \; ; \|\boldsymbol{\beta}_{(L,1)}^K\|_1^{-1} \mathbb{1}_{d_L}; \; \cdots \; ; \|\boldsymbol{\beta}_{(L,R)}^K\|_1^{-1} \mathbb{1}_{d_L}; \; \mathbb{1}_M]) \tag{26}$$

The demonstration of this result is similar to that of the multiway case. Indeed, we note that

$$(\mathbf{x}_{(1)}')_i^T \left[\sum_{r=1}^R \boldsymbol{\beta}_{(1,r)}^K \otimes \boldsymbol{\beta}_{(1,r)}^J; \; \cdots \; ; \sum_{r=1}^R \boldsymbol{\beta}_{(L,r)}^K \otimes \boldsymbol{\beta}_{(L,r)}^J \right] = \sum_{l=1}^L \sum_{r=1}^R \left(\mathbf{x}_{(1)}^l \right)_i^T \left(\boldsymbol{\beta}_{(l,r)}^K \otimes \boldsymbol{\beta}_{(l,r)}^J \right) \tag{27}$$

$$= \sum_{l=1}^L \sum_{r=1}^R \left[\left(\mathbf{x}_{(1)}^l \right)_i^T \left(\boldsymbol{\beta}_{(l,r)}^K \otimes I_{d_l} \right) \right] \boldsymbol{\beta}_{(l,r)}^J \tag{28}$$

$$= \sum_{l=1}^L \sum_{r=1}^R \left(\mathbf{z}_{(l,r)}^J \right)_i^T \boldsymbol{\beta}_{(l,r)}^J \tag{29}$$

An that

$$\sum_{l=1}^L \sum_{r=1}^R \|\boldsymbol{\beta}_{(l,r)}^K \otimes \boldsymbol{\beta}_{(l,r)}^J\|_1 = \|\mathbf{R}_{\text{tens}}^J \boldsymbol{\beta}^J\|_1 \tag{30}$$

$$\text{with} \quad \mathbf{R}_{\text{tens}}^J = \text{Diag}([\|\boldsymbol{\beta}_{(1,1)}^K\|_1 \mathbb{1}_{d_1}; \; \cdots \; ; \|\boldsymbol{\beta}_{(1,R)}^K\|_1 \mathbb{1}_{d_1}; \; \cdots \; \cdots \; ; \|\boldsymbol{\beta}_{(L,1)}^K\|_1 \mathbb{1}_{d_L}; \; \cdots \; ; \|\boldsymbol{\beta}_{(L,R)}^K\|_1 \mathbb{1}_{d_L}; \; \mathbb{1}_M]) \tag{31}$$

We deduce that

$$\left[\mathbf{x}_{(1)_i}'; \; \mathbf{x}_{\text{tab}_i} \right] \boldsymbol{\beta} = (\mathbf{z}_i^J)^T [\boldsymbol{\beta}^J; \boldsymbol{\beta}_{\text{tab}}] \tag{32}$$

$$\text{and} \quad \sum_{l=1}^L \sum_{r=1}^R \|\boldsymbol{\beta}_{(l,r)}^K \otimes \boldsymbol{\beta}_{(l,r)}^J\|_1 + \|\boldsymbol{\beta}_{\text{uni}}\|_1 = \|(\mathbf{Q}^J)^{-1} \boldsymbol{\beta}\|_1 \tag{33}$$

Wich justifies the previous results.

For optimization with respect to $\left[\beta_0; \boldsymbol{\beta}^K; \boldsymbol{\beta}_{\mathrm{tab}}\right]$, the method is analogous. The only difference concerns the form of $\mathbf{Z}^K$. It is written as:

$$\mathbf{Z}^K = [\mathbf{Z}^K_{(1,1)} \cdots \mathbf{Z}^K_{(1,R)} \cdots \cdots \mathbf{Z}^K_{(L,1)} \cdots \mathbf{Z}^K_{(L,R)} \ \mathbf{X}_{\mathrm{tab}}] \tag{34}$$

$$\text{where } \forall r \in [\![1, R]\!], \qquad \mathbf{Z}^K_{(l,r)} = \sum_{j=1}^{d_l} \mathbf{X}^l_{:j:} \left(\beta^J_{(l,r)}\right)_j \qquad \left(\mathbf{Z}^K_{(l,r)} \in \mathbb{R}^{N \times K}\right) \tag{35}$$

The justification of that last result is analogous to the one used in the multiway case.

Notes:

- In the multiway and multibloc logistic regression with lasso, it is possible to allow for the choice of a specific rank for each block. We have not yet studied this model extension, but deriving its equations is straightforward, based on the equations provided in this report.

- We decided to optimize the loss function completely in one direction before turning to the other one instead of alternating one step in each direction because the first procedure was more stable and could be implemented efficiently using the glmnet package in R [5].

Pseudo-code:

In order to clarify the algorithm that we use, we give here the pseudo-code of our implementation. In order to be more readable, we keep the notations that were used during the presentation of the model.

Inputs
- $\epsilon > 0$, $\lambda > 0$, $R \in \mathbb{N}^*$
- $\boldsymbol{\beta}^{K(0)} \in \mathbb{R}^{LRK}$

Treatment
- $q \leftarrow 0$

Repeat
- Construct $\mathbf{Z}^J$ according to eqs. (24) and (25)
- Construct $\mathbf{Q}^J$ according to eq. (26)
- $(\beta_0^{(q)}, \boldsymbol{\beta}^{J(q)}) \longleftarrow \underset{(\beta_0, \boldsymbol{\beta}) \in \mathbb{R} \times \mathbb{R}^{RJ+M}}{\mathrm{argmin}} \left(C(\beta_0, (\mathbf{Q}^J)^{-1}\boldsymbol{\beta}, \mathbf{Z}^J\mathbf{Q}^J, \mathbf{y}, \lambda) \right)$
- Construct $\mathbf{Z}^K$ according to eqs. (34) and (35)
- Construct $\mathbf{Q}^K$ by adapting eq. (26)
- $(\beta_0^{(q)}, \boldsymbol{\beta}^{K(q)}) \longleftarrow \underset{(\beta_0, \boldsymbol{\beta}) \in \mathbb{R} \times \mathbb{R}^{LRK+M}}{\mathrm{argmin}} \left(C(\beta_0, (\mathbf{Q}^K)^{-1}\boldsymbol{\beta}, \mathbf{Z}^K\mathbf{Q}^K, \mathbf{y}, \lambda) \right)$
- $q \leftarrow q + 1$

until $|C^K - C^J| < \epsilon|C^J|$

Return $(\beta_0^{(q)}, \boldsymbol{\beta}^{K(q)}, \boldsymbol{\beta}^{J(q)})$

6   results

Conclusions

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

References

[1] L. Meier, S. Van De Geer, and P. Bühlmann, "The Group Lasso for Logistic Regression," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 70, pp. 53–71, 01 2008.

[2] L. Le Brusquet, G. Lechuga, and A. Tenenhaus, "Régression Logistique Multivoie," in *JdS 2014*, (Rennes, France), p. 6 pages, June 2014.

[3] F. Girka, P. Chevaillier, A. Gloaguen, G. Gennari, G. Dehaene-Lambertz, L. Le Brusquet, and A. Tenenhaus, "Rank-R Multiway Logistic Regression," in *52èmes Journées de Statistique*, (Nice, France), 2021. les 52èmes journées de Statistique 2020 sont reportées ! Elles auront lieu du 7 au 11 Juin 2021.

[4] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Review*, vol. 51, no. 3, pp. 455–500, 2009.

[5] R. Tibshirani, T. Hastie, and J. Friedman, "Regularized paths for generalized linear models via coordinate descent," *Journal of Statistical Software*, vol. 33, 02 2010.