

Machine learning methods to help classify liver tumors from radiomics data

Alexandre SELVESTREL¹

¹Laboratoire des systèmes, Orsay France

²Centrale-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 $\underline{\mathbf{X}} = (x_{ijk})_{i \in \llbracket 1, n \rrbracket, j \in \llbracket 1, J \rrbracket, k \in \llbracket 1, K \rrbracket}$, 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: $\mathbf{X}_{(1)} = [\mathbf{X}_{::1}, \dots, \mathbf{X}_{::K}]$. 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 negligible 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'intérêt de l'étude. Les mixtes notamment.

2 Notations

Nous précisons dans cette section les notations spécifiques utilisées dans ce papier.

- La concaténation de deux matrices \mathbf{A} et \mathbf{B} en juxtaposant leurs colonnes côte à côte est notée $[\mathbf{A} \ \mathbf{B}]$.
- Afin d'éviter de trop utiliser le symbole T , on définit aussi une notation pour désigner la juxtaposition de deux matrices l'une en dessous de l'autre. Ainsi, la matrice définie par bloc avec \mathbf{A} au-dessus de \mathbf{B} est notée $[\mathbf{A}; \mathbf{B}]$. On peut aussi l'écrire $[\mathbf{A}^T \ \mathbf{B}^T]^T$ mais cela multiplie les symboles T , ce qui nuit à la lisibilité.
- Les vecteurs étant des matrices colonnes, avec les mêmes notations, on écrira la concaténation de deux vecteurs \mathbf{u} et \mathbf{v} de la manière suivante: $[\mathbf{u} \downarrow \mathbf{v}]$.
- Le vecteur (colonne) dont les éléments sont $(u_i)_{i \in \llbracket 1, I \rrbracket}$ est aussi noté (u_1, u_2, \dots, u_I) .
- On note $\text{Diag}(\mathbf{u})$ la matrice diagonale dont la diagonale est le vecteur \mathbf{u} .
- Le vecteur de longueur I rempli de 1 est noté $\mathbb{1}_I$.
- Pour les autres notations, nous utilisons les conventions employées par *Kolda and Bader* (2009) [4].

3 State of the art

Lorem ipsum dolor sit amet, consectetur 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, consectetur 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'imputation de données (qui risquerait de biaiser les résultats vu le nombre de données manquantes sur certains individus)

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 $\mathbf{X}_{(1)} = [\mathbf{X}_{:,1}, \dots, \mathbf{X}_{:,K}]$. We then complete this matrix by concatenating (along the columns) the matrix of non tensorial data \mathbf{X}_{tab} (where "tab" stands for "tabular"). By doing so we obtain $\mathbf{X}_{\text{tot}} = [\mathbf{X}_{(1)}, \mathbf{X}_{\text{tab}}]$.

We first train a penalized logistic regression lasso on \mathbf{X}_{tot} . Then, still based on the matrix \mathbf{X}_{tot} , we train a group lasso [1]. The idea of this second model is to group together all the explanatory variables corresponding to the same observation time, in order to give the model a vague description of the tensorial structure of the data. We'll also compare the results with the case where we group the explanatory variables according to their type: first order, shape and texture, and with the case where we group the variables according to the quantity of interest they describe. In all cases, we also group the clinical variables together.

5.2 Multiway logistic regression with lasso

We now turn to tensor approaches. We start by studying a multiway logistic regression penalized by the 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 $\beta_{\text{tens}} \in \mathbb{R}^{JK}$ associated with the tensor explanatory variables of the logistic regression as:

$$\beta_{\text{tens}} = \sum_{r=1}^R \beta_r^K \otimes \beta_r^J \quad (1)$$

with for all $r \in \llbracket 1, R \rrbracket$, $\beta_r^J \in \mathbb{R}^J$ and $\beta_r^K \in \mathbb{R}^K$. To take account of the M clinical variables, which are not tensorial, we associate them with a coefficient $\beta_{\text{uni}} \in \mathbb{R}^M$. In this way, the parameter β of the logistic regression is written: $[(\beta_{\text{tens}})^T (\beta_{\text{uni}})^T]^T$. We then use the logit link function to obtain the following likelihood for the explained Bernoulli variable y_i ($i \in \llbracket 1, n \rrbracket$):... Parler plutôt de proba condit...

$$\mathfrak{L}(y_i | \mathbf{x}_i) = \frac{1}{1 + \exp(-\mathbf{x}_i^T \beta - \beta_0)} \quad (2)$$

where $\beta_0 \in \mathbb{R}$ is the intercept and $\mathbf{x}_i \in \mathbb{R}^{JK}$ is given by $\mathbf{x}_i = (\mathbf{X}_{\text{tot}})_i$:

We set $\beta^J = [(\beta_1^J)^T \dots (\beta_R^J)^T]^T$ and $\beta^K = [(\beta_1^K)^T \dots (\beta_R^K)^T]^T$. 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:

$$\beta_0, \beta^J, \beta^K, \beta_{\text{uni}} = \underset{\beta_0, \beta^J, \beta^K, \beta_{\text{uni}}}{\operatorname{argmin}} \left(-\sum_{i=1}^N \log(\mathfrak{L}(y_i, \beta_0, \beta^J, \beta^K, \beta_{\text{uni}})) + \sum_{r=1}^R \|\beta_r^K \otimes \beta_r^J\|_1 + \|\beta_{\text{uni}}\|_1 \right) \quad (3)$$

Optimization is performed by alternating directions between $\begin{bmatrix} \beta_0 & (\beta^J)^T & (\beta_{\text{uni}})^T \end{bmatrix}^T$ and $\begin{bmatrix} \beta_0 & (\beta^K)^T & (\beta_{\text{uni}})^T \end{bmatrix}^T$. 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:

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

($K_0 \in \mathbb{N}^*$)

optimizing the overall loss function with respect to $\begin{bmatrix} \beta_0 & (\beta^J)^T & (\beta_{\text{uni}})^T \end{bmatrix}^T$ amounts to solve

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

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

$$\mathbf{Z}^J = [\dot{\mathbf{Z}}^J \ \mathbf{X}_{\text{uni}}] \quad (6)$$

$$\text{with } \mathbf{Z}_{\text{tens}}^J = [\mathbf{Z}_1^J \ \dots \ \mathbf{Z}_R^J] \quad (7)$$

$$\text{where } \forall r \in \llbracket 1, R \rrbracket, \quad \mathbf{Z}_r^J = \sum_{k=1}^K \mathbf{X}_{::k} (\beta_r^K)_k \quad (\mathbf{Z}_r^J \in \mathbb{R}^{N \times J}) \quad (8)$$

$$\mathbf{Q}^J = \operatorname{Diag}([\mathbf{u}_{\text{tens}}^T \ \mathbf{1}_M^T]) \quad (9)$$

$$\text{with } \mathbf{u}_{\text{tens}} = (\|\beta_1^K\|_1^{-1} \mathbf{1}_J^T, \dots, \|\beta_R^K\|_1^{-1} \mathbf{1}_J^T)^T \quad (10)$$

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

$$(\mathbf{x}_{\text{tens}})_i^T \left(\sum_{r=1}^R \beta_r^K \otimes \beta_r^J \right) = \sum_{r=1}^R [(\mathbf{x}_{\text{tens}})_i^T (\beta_r^K \otimes \mathbf{I}_J)] \beta_r^J \quad (11)$$

$$= \sum_{r=1}^R (\mathbf{z}_{i,r}^J)^T \beta_r^J \quad (12)$$

$$= (\mathbf{z}_{\text{tens}}^J)_i^T \beta^J \quad \text{On peut potentiellement dégager cette ligne} \quad (13)$$

and that

$$\sum_{r=1}^R \|\beta_r^K \otimes \beta_r^J\|_1 = \|\mathbf{R}_{\text{tens}}^J \beta^J\|_1 \quad (14)$$

$$\text{with } \mathbf{R}_{\text{tens}}^J = \operatorname{Diag}(\|\beta_1^K\|_1 \mathbf{1}_J^T \ \dots \ \|\beta_R^K\|_1 \mathbf{1}_J^T) \quad (15)$$

Thus,

$$(\mathbf{x}_{\text{tot}})_i^T \boldsymbol{\beta} = (\mathbf{z}_i^J)^T \boldsymbol{\beta}^J \quad (16)$$

$$\text{and } \sum_{r=1}^N \|\boldsymbol{\beta}_r^K \otimes \boldsymbol{\beta}_r^J\|_1 + \|\boldsymbol{\beta}_{\text{uni}}\|_1 = \|(\mathbf{Q}^J)^{-1} \boldsymbol{\beta}\|_1 \quad (17)$$

This justifies the previous results

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

$$\mathbf{Z}_{\text{tens}}^K = [\mathbf{Z}_1^K \dots \mathbf{Z}_R^K] \quad (18)$$

$$\text{with } \forall r \in \llbracket 1, R \rrbracket \quad \mathbf{Z}_r^K = \sum_{j=1}^J (\beta_r^J)_j \mathbf{X}_{:j} \quad (19)$$

This is justified by:

$$(\mathbf{x}_{\text{tens}})_i^T \left(\sum_{r=1}^R \boldsymbol{\beta}_r^K \otimes \boldsymbol{\beta}_r^J \right) = \sum_{r=1}^R [(\mathbf{x}_{\text{tens}})_i^T (I_K \otimes \boldsymbol{\beta}_r^J)] \boldsymbol{\beta}_r^K \quad (20)$$

$$= \sum_{r=1}^R (\mathbf{z}_{i,r}^K)^T \boldsymbol{\beta}_r^K \quad (21)$$

$$= (\mathbf{z}_{\text{tens}}^K)_i^T \boldsymbol{\beta}^K \quad (22)$$

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, but adds a separation of the measured quantities into several categories. First-order quantities are placed in a first block, those concerning tumor shape in a second block and those concerning tumor texture in a third block. As clinical quantities are not measured at different times, they are not placed in a 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 $l \in \llbracket 1, L \rrbracket$, let d_l be the number of radiomic quantities of interest in block l . Thus we have :

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

We reorganize the columns of \mathbf{X}_{tot} by grouping together columns associated with quantities from the same block. The resulting matrix is called \mathbf{X}'_{tot} . More precisely, the order of the columns (from left to right) of the matrix \mathbf{X}'_{tot} is lexicographic according to the block number l and then according to the observation time number k . This matrix is then multiplied by $\boldsymbol{\beta}$ in the equation that gives the likelihood (2). Finally, for $l \in \llbracket 1, L \rrbracket$, we call $\underline{\mathbf{X}}^l \in \mathbb{R}^{N \times d_l \times K}$ the tensor obtained by keeping only the radiomic quantities of the l -th block. Consistently with the notations used previously, we call $\mathbf{X}_{(1)}^l$ the matrix obtained by unfolding $\underline{\mathbf{X}}^l$, in which

the individuals are arranged in rows and the order of the columns is lexicographic according to the value of k and then j . The new β structure is defined by blocks. It is:

$$\beta = \left[\left(\sum_{r=1}^R \beta_{(1,r)}^K \otimes \beta_{(1,r)}^J \right)^T \quad \cdots \quad \left(\sum_{r=1}^R \beta_{(L,r)}^K \otimes \beta_{(L,r)}^J \right)^T \quad \beta_{\text{uni}}^T \right]^T \quad (23)$$

With for all $(l, r) \in \llbracket 1, L \rrbracket \times \llbracket 1, R \rrbracket$, $\beta_{(l,r)}^J \in \mathbb{R}^{d_l}$ and $\beta_{(l,r)}^K \in \mathbb{R}^K$

In line with the notations of the multiway logistic model, we note:

$$\beta_{\text{tens}} = \left[\left(\sum_{r=1}^R \beta_{(1,r)}^K \otimes \beta_{(1,r)}^J \right)^T \quad \cdots \quad \left(\sum_{r=1}^R \beta_{(L,r)}^K \otimes \beta_{(L,r)}^J \right)^T \right]^T \quad (24)$$

We now call β^J and β^K the vectors

$$\beta^J = \left[(\beta_{(1,1)}^J)^T \quad \cdots \quad (\beta_{(l=1,r=R)}^J)^T \quad (\beta_{(l=2,r=1)}^J)^T \quad \cdots \quad (\beta_{(l=L,r=R)}^J)^T \right]^T \quad (25)$$

$$\beta^K = \left[(\beta_{(l=1,r=1)}^K)^T \quad \cdots \quad (\beta_{(l=1,r=R)}^K)^T \quad (\beta_{(l=2,r=1)}^K)^T \quad \cdots \quad (\beta_{(l=L,r=R)}^K)^T \right]^T \quad (26)$$

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

$$\beta_0, \beta^J, \beta^K, \beta_{\text{uni}} = \underset{\beta_0, \beta^J, \beta^K, \beta_{\text{uni}}}{\operatorname{argmin}} \left(- \sum_{i=1}^N \log(\mathfrak{L}(y_i, \beta_0, \beta^J, \beta^K, \beta_{\text{uni}})) + \sum_{l=1}^L \sum_{r=1}^R \|\beta_{(l,r)}^K \otimes \beta_{(l,r)}^J\|_1 + \|\beta_{\text{uni}}\|_1 \right) \quad (27)$$

Once again, this problem is solved by alternating optimization directions $\left[\beta_0 \quad (\beta^J)^T \quad (\beta_{\text{uni}})^T \right]^T$ and

$\left[\beta_0 \quad (\beta^K)^T \quad (\beta_{\text{uni}})^T \right]^T$. Each of these two problems can be reduced to a lasso-penalized (matrix) logistic regression.

Indeed, optimizing according to $\left[\beta_0 \quad (\beta^J)^T \quad (\beta_{\text{uni}})^T \right]^T$ is equivalent to searching

$$\underset{(\beta_0, \beta) \in \mathbb{R} \times \mathbb{R}^{R \times J + M}}{\operatorname{argmin}} \left(C(\beta_0, (\mathbf{Q}^J)^{-1} \beta, \mathbf{Z}^J \mathbf{Q}^J, \mathbf{y}, \lambda) \right) \quad (28)$$

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

$$\mathbf{Z}^J = [\mathbf{Z}_{\text{tens}}^J \quad \mathbf{X}_{\text{uni}}] \quad (29)$$

$$\text{with } \mathbf{Z}_{\text{tens}}^J = [\mathbf{Z}_{(l=1,r=1)}^J \quad \cdots \quad \mathbf{Z}_{(l=1,r=R)}^J \quad \cdots \quad \mathbf{Z}_{(l=2,r=1)}^J \quad \cdots \quad \mathbf{Z}_{(l=L,r=R)}^J] \quad (30)$$

$$\text{where } \forall r \in \llbracket 1, R \rrbracket, \quad \mathbf{Z}_{(l,r)}^J = \sum_{k=1}^K (\mathbf{X})_{::k} (\beta_{(l,r)}^K)_k \quad (\mathbf{Z}_{(l,r)}^J \in \mathbb{R}^{n \times d_l}) \quad (31)$$

$$\mathbf{Q}^J = \operatorname{Diag}([\mathbf{u}_{\text{tens}}^T \quad \mathbb{1}_M^T]) \quad (32)$$

$$\text{with } \mathbf{u}_{\text{tens}} = (\|\beta_{(l=1,r=1)}^K\|_1^{-1} \mathbb{1}_{d_1}^T \quad \cdots \quad \|\beta_{(l=1,r=R)}^K\|_1^{-1} \mathbb{1}_{d_1}^T \quad \cdots \quad \|\beta_{(l=2,r=1)}^K\|_1^{-1} \mathbb{1}_{d_2}^T \quad \cdots \quad \|\beta_{(l=L,r=R)}^K\|_1^{-1} \mathbb{1}_{d_L}^T)^T \quad (33)$$

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

$$(\mathbf{x}'_{\text{tens}})_i^T \boldsymbol{\beta}_{\text{tens}} = \sum_{l=1}^L \sum_{r=1}^R (\mathbf{x}_{(1)}^l)_i^T (\boldsymbol{\beta}_{(l,r)}^K \otimes \boldsymbol{\beta}_{(l,r)}^J) \quad (34)$$

$$= \sum_{l=1}^L \sum_{r=1}^R [(\mathbf{x}_{(1)}^l)_i^T (\boldsymbol{\beta}_{(l,r)}^K \otimes I_{d_l})] \boldsymbol{\beta}_{(l,r)}^J \quad (35)$$

$$= \sum_{l=1}^L \sum_{r=1}^R (z_{(l,r)}^J)_i^T \boldsymbol{\beta}_{(l,r)}^J \quad (36)$$

$$= (z_{\text{tens}}^J)_i^T \boldsymbol{\beta}^J \quad (37)$$

An that

$$\sum_{l=1}^L \sum_{r=1}^R \|\boldsymbol{\beta}_{(l,r)}^K \otimes \boldsymbol{\beta}_{(l,r)}^J\|_1 = \|\mathbf{R}_{\text{tens}}^J \boldsymbol{\beta}^J\|_1 \quad (38)$$

$$\text{with } \mathbf{R}_{\text{tens}}^J = \text{Diag}(\|\boldsymbol{\beta}_{(l=1,r=1)}^K\|_1 \mathbb{1}_{d_1}^T \cdots \|\boldsymbol{\beta}_{(l=1,r=R)}^K\|_1 \mathbb{1}_{d_1}^T \cdots \|\boldsymbol{\beta}_{(l=2,r=1)}^K\|_1 \mathbb{1}_{d_2}^T \cdots \cdots \|\boldsymbol{\beta}_{(l=L,r=R)}^K\|_1 \mathbb{1}_{d_L}^T)^T \quad (39)$$

We deduce that

$$(\mathbf{x}'_{(1)})_i^T \boldsymbol{\beta} = (\mathbf{z}_i^J)^T \boldsymbol{\beta}^J \quad (40)$$

$$\text{and } \sum_{l=1}^L \sum_{r=1}^R \|\boldsymbol{\beta}_{(l,r)}^K \otimes \boldsymbol{\beta}_{(l,r)}^J\|_1 + \|\boldsymbol{\beta}_{\text{uni}}\|_1 = \|(\mathbf{Q}^J)^{-1} \boldsymbol{\beta}\|_1 \quad (41)$$

Wich justifies the previous results.

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

$$\mathbf{Z}_{\text{tens}}^K = \left[\mathbf{Z}_{(l=1,r=1)}^K \cdots \mathbf{Z}_{(l=1,r=R)}^K \cdots \mathbf{Z}_{(l=2,r=1)}^K \cdots \cdots \mathbf{Z}_{(l=L,r=R)}^K \right] \quad (42)$$

$$\text{where } \forall r \in \llbracket 1, R \rrbracket, \quad \mathbf{Z}_{(1,r)}^K = \sum_{j=1}^{d_l} (\mathbf{X})_{:j} (\beta_{(l,r)}^J)_j \quad (\mathbf{Z}_{(1,r)}^K \in \mathbb{R}^{N \times K}) \quad (43)$$

$$(44)$$

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

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

Treatment

- $q \leftarrow 0$

Repeat

- Construct \mathbf{Z}^J according to eqs. (29) to (31)
- Construct \mathbf{Q}^J according to eqs. (32) and (33)
- $(\beta_0^{(q)}, \beta^{J(q)}) \leftarrow \underset{(\beta_0, \beta) \in \mathbb{R} \times \mathbb{R}^{RJ+M}}{\operatorname{argmin}} (C(\beta_0, (\mathbf{Q}^J)^{-1}\beta, \mathbf{Z}^J \mathbf{Q}^J, \mathbf{y}, \lambda))$
- Construct \mathbf{Z}^K according to eqs. (42) and (43)
- Construct \mathbf{Q}^K by adapting eqs. (32) and (33)
- $(\beta_0^{(q)}, \beta^{K(q)}) \leftarrow \underset{(\beta_0, \beta) \in \mathbb{R} \times \mathbb{R}^{LRK+M}}{\operatorname{argmin}} (C(\beta_0, (\mathbf{Q}^K)^{-1}\beta, \mathbf{Z}^K \mathbf{Q}^K, \mathbf{y}, \lambda))$
- $q \leftarrow q + 1$

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

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

6 results

Conclusions

Lorem ipsum dolor sit amet, consectetur 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, consectetur 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.