Tensor multiblock logistic regression

SELVESTREL Alexandre

Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des signaux et systèmes

Supervisors : Arthur Tenenhaus, Laurent Lebrusquet

Medical partner : Henri Mondor hospital, radiologist: Sébastien Mulé

January 13, 2025

Liver tumors classification

6^{th} most widespread cancer and 4^{th} mortality cause by cancer

Classification:

- Hepatocellular Carcinoma (HCC): 75% of cases, resection often possible
- CCK = Cholangiocarcinoma (CCK): 6% of cases, resection difficult (possible in 30% of cases)
- Others: benign (18% of cases) or Hepatoblastoma (1% of cases)

Some MRI images

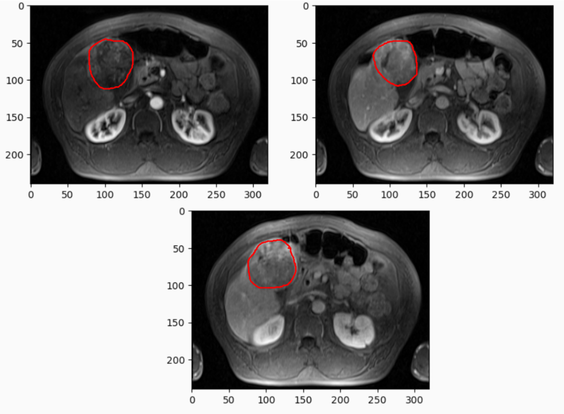


Figure 1: Example of MRI images of a HCC liver tumor (arterial, portal, late) from From Henri Mondor hospital: the 3 images look quite similar

Some MRI images

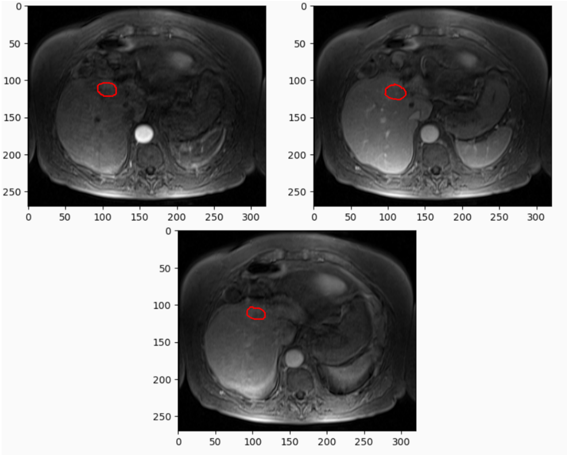


Figure 2: Example of MRI images of a CCK liver tumor (arterial, portal, late) from From Henri Mondor hospital

Available data

- MRI images in 3D of liver tumors (arterial, portal, late)
- gender (63 men, 27 women)
- age at disease (average: 63 years old)

Same variables extracted from each MRI image at 3 times $\rightarrow$ **tensor data**

Features grouped by blocks: grey levels intensity, shape and texture $\rightarrow$ **multiblock data**

Tensor data

A given subject i is represented by a transverse slice in the features tensor

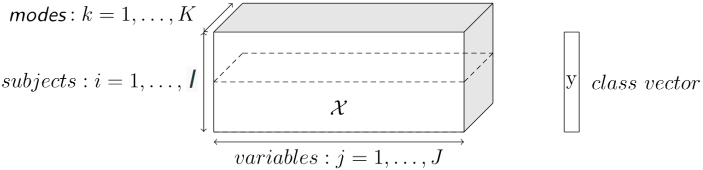


Figure 3: Type of data: tensorial

Warning: Often strong correlation between features from the same variable across different modalities $\rightarrow$ adapt the model to this structure.

Each block is a tensor with its own structure.

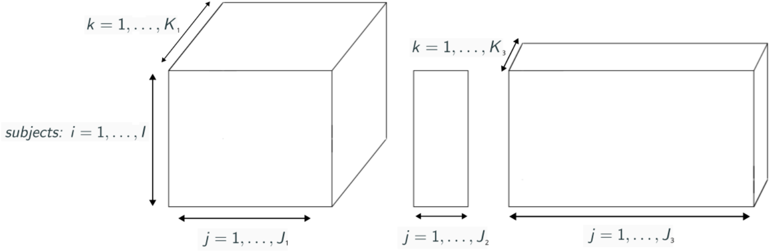


Figure 4: Type of data: multiblock

Table of contents

Logistic model for tensor multiblock data

Logistic regression (recall)

Generalized linear model (GLM) for classification:

x : features vector

Y : binary response (explained variable)

Likelihood for logistic regression

$$P(Y = 1|x) = \frac{\exp(\beta_0 + \boldsymbol{x}^T\boldsymbol{\beta})}{1 + \exp(\beta_0 + \boldsymbol{x}^T\boldsymbol{\beta})}$$

Defines a likelihood function $\mathcal{L}(\boldsymbol{\beta}) = \prod_{i=1}^{I} P(Y_i = y_i|x_i)$.

Naive approach for tensor data: unfolding

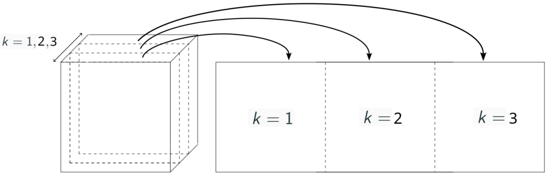


Figure 5: Unfolding a tensor

Naive unfolding

$$\boldsymbol{\beta} = (\beta_{j,k})_{j\in[\![1,J]\!],k\in[\![1,K]\!]} \rightarrow JK \text{ parameters to determine}$$

$$\mathbf{x}^T\boldsymbol{\beta} \rightsquigarrow \sum_{j,k} \beta_{j,k} x_{j,k}$$

Limitation: No consideration of the tensor structure in the likelihood

Lasso penalization

To many features (vs I) $\rightarrow$ penalization to control variance of prediction (overfitting).

Search for easily interpretable model $\rightarrow$ choice of lasso:

Lasso

$$\text{penalization} = \lambda \|\boldsymbol{\beta}\|_1 \qquad (\lambda > 0)$$

Function to maximize :

$$\text{penalized likelihood} = \log(\mathcal{L}(\boldsymbol{\beta})) - \lambda \|\boldsymbol{\beta}\|_1$$

Limitation: No consideration of the tensor structure in the penalization.

Idea: each variable and mode has its own influence on the prediction (i.e. on β) [2].

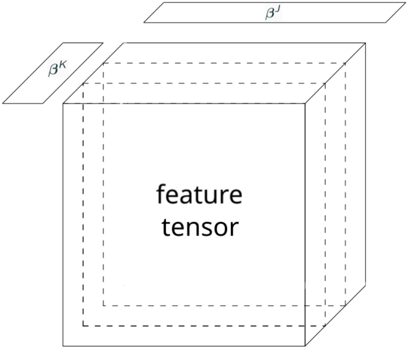


Figure 6: Tensor structure of β

Tensor regression model

Idea: each variable and mode has its own influence on the prediction (i.e. on $\boldsymbol{\beta}$) [2].

Proposed rank 1 model

For J variables observed on K modalities (e.g. times)

$$\boldsymbol{\beta} = \boldsymbol{\beta}^K \circ \boldsymbol{\beta}^J \qquad (\beta_{j,k} = \beta_j^J \beta_k^K)$$

β_j : impact of variable j

β_k : impact of modality k

Only $J + K$ parameters to determine (instead of JK).

Limits of rank 1

$\boldsymbol{\beta} = \boldsymbol{\beta}^K \circ \boldsymbol{\beta}^J$ implies a complete separation between columns and rows:

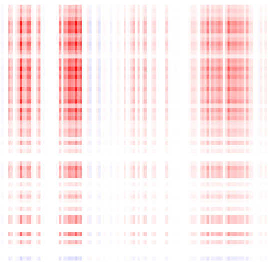


Figure 7: Example of rank 1 matrix

This can be too simplistic.

Extension to rank R [1]

Rank R lasso penalized tensor model

$$\text{Summing rank 1 together : } \boldsymbol{\beta} = \sum_{r=1}^{R} \boldsymbol{\beta}_r^J \circ \boldsymbol{\beta}_r^K$$

$$\text{lasso like penalization} = \lambda \sum_{r=1}^{R} \|\boldsymbol{\beta}_r^J \circ \boldsymbol{\beta}_r^K\|_1 = \lambda \sum_{r=1}^{R} \|\boldsymbol{\beta}_r^J\|_1 \|\boldsymbol{\beta}_r^K\|_1$$

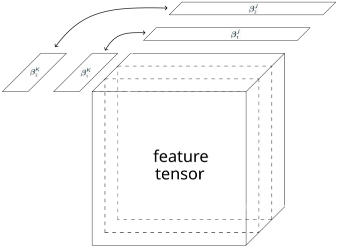


Figure 8: Tensor structure of $\boldsymbol{\beta}$ for rank 2

Blocks of variables

Aim: GLM framework for multiblock tensor data.

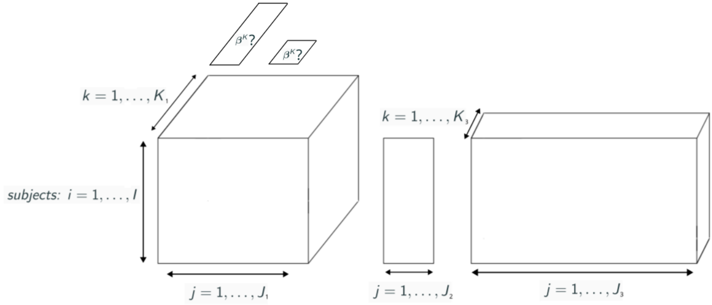


Figure 9: Problem if blocks have different orders or dimensions

Logistic model for tensor multiblock data

Solution: giving each block its own β^J and β^K.

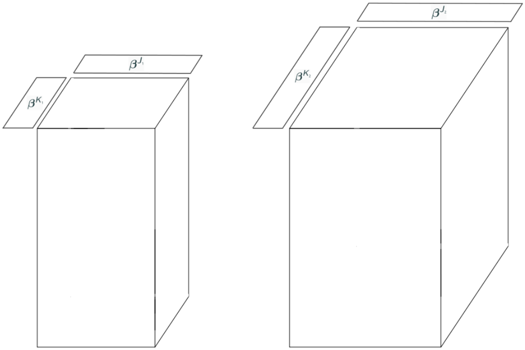


Figure 10: Tensor multiblock model for rank 1

Tensor multiblock logistic regression

Scalar product for L blocks

$$\mathbf{x}^T\boldsymbol{\beta} \rightsquigarrow \sum_{\ell=1}^{L}\sum_{j,k} x^{\ell}_{j,k}(\beta_\ell)_{j,k}$$

Regression coefficient for block ℓ

$\boldsymbol{\beta}_\ell$ can have any rank R_ℓ

$$\boldsymbol{\beta}_\ell = \sum_{r=1}^{R_\ell} \boldsymbol{\beta}^J_{\ell,r} \circ \boldsymbol{\beta}^K_{\ell,r}$$

Penalization

$$\text{Lasso like penalty} = \lambda \sum_{\ell,r} \|\boldsymbol{\beta}^K_{\ell,r} \circ \boldsymbol{\beta}^J_{\ell,r}\|_1 = \lambda \sum_{\ell,r} \|\boldsymbol{\beta}^K_{\ell,r}\|_1 \|\boldsymbol{\beta}^J_{\ell,r}\|_1$$

Maximizing penalized likelihood

Scalar product of x and β

$$\mathbf{x}^T\beta = \sum_{\ell,r,j,k} x^{\ell}_{j,k}(\beta^J_{\ell,r})_j(\beta^K_{\ell,r})_k \qquad (1)$$

$$= \sum_{\ell,r,j}\left(\sum_k x^{\ell}_{j,k}(\beta^K_{\ell,r})_k\right)(\beta^J_{\ell,r})_j \qquad (2)$$

Partial optimization problem

Optimizing along mode J $\Leftrightarrow$ solving a logistic regression on weighted aggregated data $\sum_k x^{\ell}_{j,k}(\beta^K_{\ell,r})_k$

Algorihm

Analgue result for mode K. Possibility to optimize the likelihood by alternating between modes (can be easily adapted for lasso penalization)

Tests on simulated data and application to liver tumor classification

Data generation: example in 2D

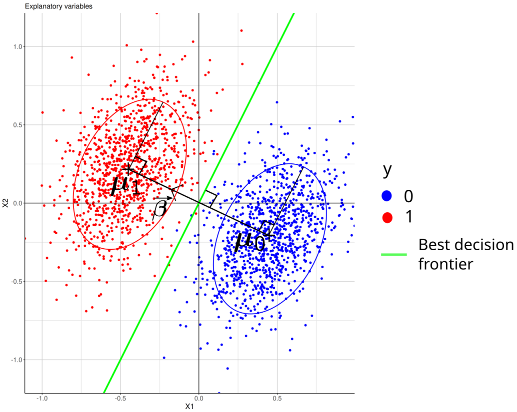


Figure 11: Example of explanatory variables for $\beta = (-2, 1)$

Possibility to choose $(\sigma_\beta, \sigma_{\text{noise}})$ to define the problem difficulty.

AUC on simulated data

Table 1: Cross validated AUC for each model on simulated data for 3000 individuals

$(\sigma_\beta, \sigma_{\text{noise}})$	lasso	g.l. (blocks)	g.l. (mode)	g.l. (var)	tensor	tensor blocks
(0.1,0.5)	0.83	0.86	0.94	0.94	0.99	0.99
(0.1,0.8)	0.63	0.64	0.68	0.68	0.93	0.99

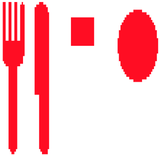
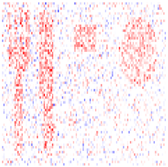
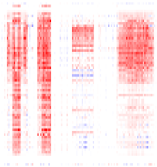


Figure 12: Pictogram of shape 66×117 representing the β to be reconstructed (pixel red is 1, white is 0)

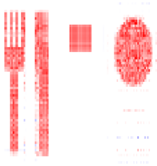
Reconstructed β



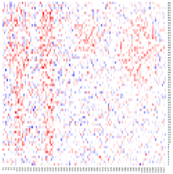
(a) lasso
$(\sigma_\beta, \sigma_{\text{noise}}) = (0.1, 0.5)$



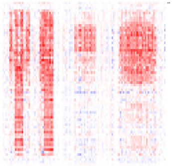
(b) tensor R : 10
$(\sigma_\beta, \sigma_{\text{noise}}) = (0.1, 0.5)$



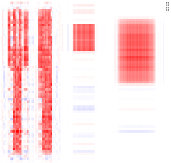
(c) T.M. R : $(12, 1, 10)$
$(\sigma_\beta, \sigma_{\text{noise}}) = (0.1, 0.5)$



(d) lasso
$(\sigma_\beta, \sigma_{\text{noise}}) = (0.1, 0.8)$



(e) tensor R : 10
$(\sigma_\beta, \sigma_{\text{noise}}) = (0.1, 0.8)$



(f) T.M. R : $(6, 1, 1)$
$(\sigma_\beta, \sigma_{\text{noise}}) = (0.1, 0.8)$

Results on liver tumor data

lasso	group lasso (block)	group lasso (time)	group lasso (var)	tensor	tensor blocks
0.74 ± 0.04	0.78 ± 0.03	0.76 ± 0.03	0.73 ± 0.03	0.77 ± 0.03	0.77 ± 0.03

Cross validated AUC on 3D real data

Good results of tensor models on real data, but better explainability (less parameters).

Conclusion

State of the art performances.

Better than state of the art in simulated data when few overlapping between classes.

Scales better than regular logistic regression for high order tensors.

Lowers the complexity of the regression model an therefore reduces overfitting.

Good interpretability (sparse + displays importance of each block, mode and variable in β).

Further work

Testing on other real datasets wether the performances on the simulated dataset can be replicated.

Testing other penalizations (group lasso, elastic net).

Extending the multiblock approach to other classical machine learning algorithms (other GLMs, SVM etc...). Comparing it to CNN.

Improving the optimization, by using coordinate descent (as done in glmnet [3] in R).

Bibliography

- Fabien Girka, Pierrick Chevaillier, Arnaud Gloaguen, Giulia Gennari, Ghislaine Dehaene-Lambertz, Laurent Le Brusquet, and Arthur 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.
- Laurent Le Brusquet, Gisela Lechuga, and Arthur Tenenhaus.
  Régression Logistique Multivoie.
  In *JdS 2014*, page 6 pages, Rennes, France, June 2014.
- Rob Tibshirani, Trevor Hastie, and Jerome Friedman.
  Regularized paths for generalized linear models via coordinate descent.
  Journal of Statistical Software, 33, 02 2010.

Annex

Lasso penalty in fitting algorithm

$$\text{penalty} \propto \sum_{\ell,r} \left(\|\beta^K_{\ell,r}\|_1 \|\beta^J_{\ell,r}\|_1 \right) \tag{3}$$

$$= \sum_{\ell,r} \left(\left\| \|\beta^K_{\ell,r}\|_1 \beta^J_{\ell,r} \right\|_1 \right) = \sum_{\ell,r} \left(\left\| \|\beta^J_{\ell,r}\|_1 \beta^K_{\ell,r} \right\|_1 \right) \tag{4}$$

Lasso penalty in fitting algorithm

$$\text{penalty} \propto \sum_{\ell,r} \left(\|\boldsymbol{\beta}^K_{\ell,r}\|_1 \|\boldsymbol{\beta}^J_{\ell,r}\|_1 \right) \tag{3}$$

$$= \sum_{\ell,r} \left(\left\| \|\boldsymbol{\beta}^K_{\ell,r}\|_1 \boldsymbol{\beta}^J_{\ell,r} \right\|_1 \right) = \sum_{\ell,r} \left(\left\| \|\boldsymbol{\beta}^J_{\ell,r}\|_1 \boldsymbol{\beta}^K_{\ell,r} \right\|_1 \right) \tag{4}$$

Strategy: dilate $\boldsymbol{\beta}^J_{\ell,r}$ by $\|\boldsymbol{\beta}^K_{\ell,r}\|_1$ and $x^\ell_{j,k}$ by $\|\boldsymbol{\beta}^K_{\ell,r}\|_1^{-1}$, so

$$\mathbf{x}^T \boldsymbol{\beta} = \sum_{j,\ell,r} \left(\sum_k x^\ell_{j,k} (\beta^K_{\ell,r})_k \right) (\beta^J_{\ell,r})_j$$

does not change but

$$\|\boldsymbol{\beta}^J\|_1 \rightsquigarrow \sum_{\ell,r} \left(\|\boldsymbol{\beta}^K_{\ell,r}\|_1 \|\boldsymbol{\beta}^J_{\ell,r}\|_1 \right)$$

Lasso penalty in fitting algorithm

$$\text{penalty} \propto \sum_{\ell,r}\left(\|\boldsymbol{\beta}^K_{\ell,r}\|_1\|\boldsymbol{\beta}^J_{\ell,r}\|_1\right) \tag{3}$$

$$= \sum_{\ell,r}\left(\left\|\|\boldsymbol{\beta}^K_{\ell,r}\|_1\boldsymbol{\beta}^J_{\ell,r}\right\|_1\right) = \sum_{\ell,r}\left(\left\|\|\boldsymbol{\beta}^J_{\ell,r}\|_1\boldsymbol{\beta}^K_{\ell,r}\right\|_1\right) \tag{4}$$

Strategy: dilate $\boldsymbol{\beta}^J_{\ell,r}$ by $\|\boldsymbol{\beta}^K_{\ell,r}\|_1$ and $x^\ell_{j,k}$ by $\|\boldsymbol{\beta}^K_{\ell,r}\|_1^{-1}$, so

$$\mathbf{x}^T\boldsymbol{\beta} = \sum_{j,\ell,r}\left(\sum_k x^\ell_{j,k}(\beta^K_{\ell,r})_k\right)(\beta^J_{\ell,r})_j$$

does not change but

$$\|\boldsymbol{\beta}^J\|_1 \rightsquigarrow \sum_{\ell,r}\left(\|\boldsymbol{\beta}^K_{\ell,r}\|_1\|\boldsymbol{\beta}^J_{\ell,r}\|_1\right)$$

Then, do logistic lasso regression with $\mathbf{x} = (x^l_{ijk}\|\boldsymbol{\beta}^K_{\ell,r}\|_1^{-1})$ and λ penalty coefficient to find $\left((\beta^J_{\ell,r})_j\|\boldsymbol{\beta}^K_{\ell,r}\|_1\right)$.

Stopping criterion

Penalized likelihood

$$C = \log(\mathcal{L}(\beta)) - \text{penalty}$$

Before the t-th optimization cycle, its value is C^t and after this cycle it becomes C^{t+1}.

Stopping criterion

$$|C^{t+1} - C^t| < \epsilon |C^t|$$

(typically $\epsilon = 10^{-4}$)

Data generation

Choose the $\boldsymbol{\beta}$ to be reconstructed (pictograms).

Generate the $(\mathbf{x}_i)_{i \in [\![1, I]\!]}$ with 2 multivariate normal laws of means $\boldsymbol{\mu}_0$ and $\boldsymbol{\mu}_1$ and common covariance matrix $\boldsymbol{\Sigma}$ such that:

- $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0$ colinear to $\boldsymbol{\beta}$
- One of the principal axis of $\boldsymbol{\Sigma}$ colinear to $\boldsymbol{\beta}$

Separation of classes linked with eigenvalues of $\boldsymbol{\Sigma}$ (to be compared with $\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0\|$).