Computational Neuro-Modeling of Visual Memory: Multimodal Imaging and Analysis

معهد قطر لبحوث الحوسبة
Qatar Computing Research Institute
Member of Qatar Foundation عضو في مؤسسة قطر

Halima Bensmail, Mohammed El Anbari **Othmane Bouhali** **Aude Oliva, Polina Gollard**

CSAIL

ABSTRACT

The high dimensionality of functional magnetic resonance imaging (fMRI) data presents major challenges to fMRI pattern classification. Directly applying standard classifiers often results in overfitting or singularity, which limits the generalizability of the results. In this paper, we propose a Doubly Regularized LOgistic Regression Algorithm (DR LORA) which penalizes the voxels of the brain that are of no importance for the classification using the Alternating Direction Method of Multipliers (ADMM) and therefore alleviate this overfitting problem. Our algorithm was compared to other classification based algorithms such as Naive Bayes, Random forest and support vector machine. The results show clear performances for our algorithm.

1. Introduction

- Interpreting brain image experiments requires analysis of complex, multivariate data
- We have to develop new machine learning algorithms to train classifiers to decode stimuli, mental states, behaviors and other variables of interest from fMRI data
- We try to answer to the questions:
  - Is there information about a variable of interest? (**pattern discrimination**)
  - Where is the information? (**pattern localization**)
  - How is that information encoded? (**pattern characterization**)

2. Feature Selection

- Want to build a model using a subset of voxels
- Model selection criteria: AIC, BIC, etc.
  - relatively small p (p is the number of predictors)
  - instability (Breiman, 1996)
- Modern data sets: high-dimensional modeling
  - microarrays (the number of genes ($\simeq 10000$)
  - fMRI data ((the number of voxels)($\simeq 250000$)

3. From 4D to 2D arrays

- fMRI data is usually represented as a 4D block of data: 3 spatial dimensions and one of time
- We are most often only interested in working only on the time-series of the voxels in the brain
- We apply a brain mask and go from a 4D array to a 2D array, voxel $\times$ time

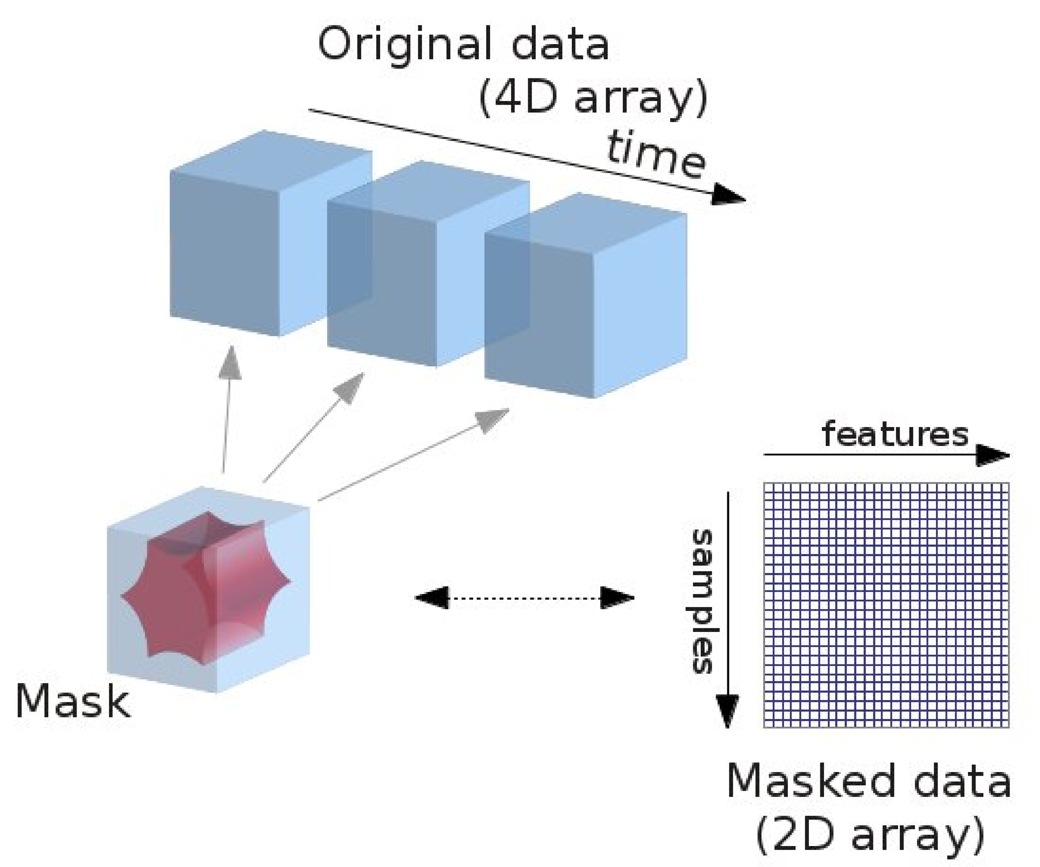

Figure 1: *Applying a mask to go from 4D to 2D arrays.*

4. fMRI Data Pre-Processing & Statistical Analysis using SPM (Matlab)

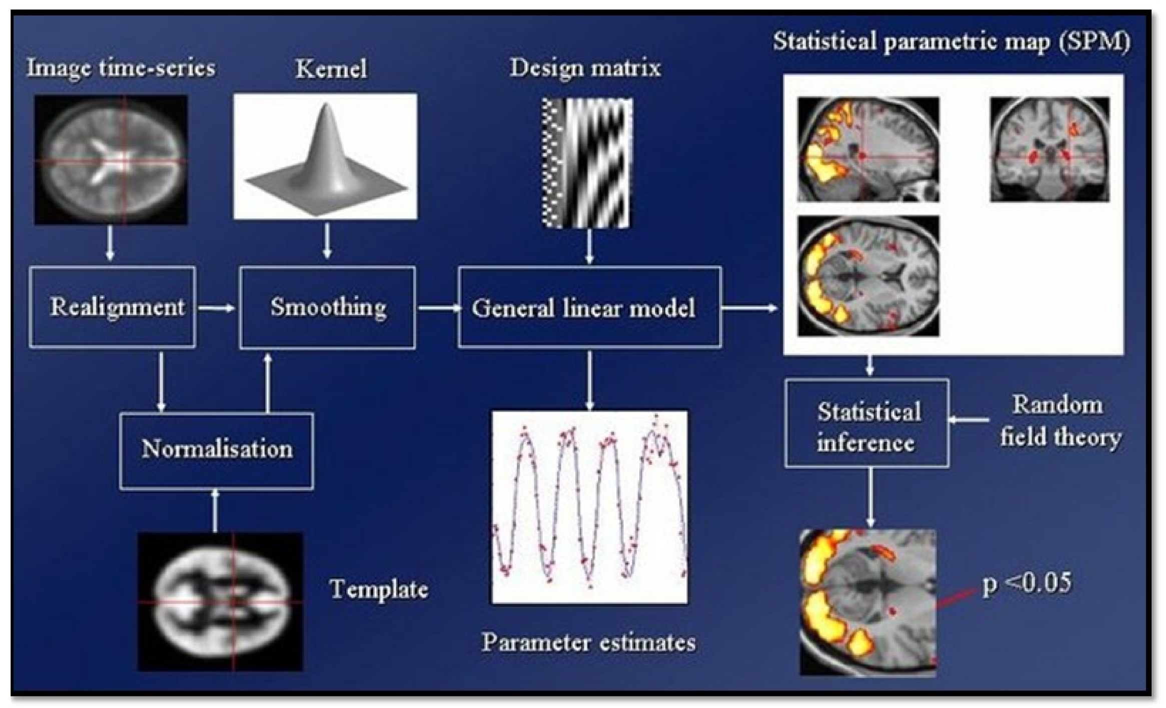

Figure 2: *Steps of Analyzing fMRI data using SPM8 in Matlab.*

5. Doubly regularized logistic regression (DRLORA)

$$(\hat{\alpha},\hat{\beta}) = \arg\min_{\alpha,\beta} \sum_{i=1}^{n}\left[\log(1+\exp(\alpha+\mathbf{x}_i^t\beta)) - y_i(\alpha+\mathbf{x}_i^t\beta)\right] + \lambda_1\sum_{j=1}^{p}|\beta_j| + \lambda_2 P_c(\boldsymbol{\beta}), \quad (1)$$

where $\lambda_1 \geq 0$, $\lambda_2 \geq 0$ are tuning parameters, and

$$P_c(\boldsymbol{\beta}) = \sum_{j=1}^{p-1}\sum_{i>j}\left\{\frac{(\beta_i-\beta_j)^2}{1-\rho_{ij}} + \frac{(\beta_i+\beta_j)^2}{1+\rho_{ij}}\right\},$$

ρ_{ij} denotes the (empirical) correlation between the ith and the jth predictors.

- The l_1 part of the penalty generates a sparse model
- The quadratic part of the penalty
  - encourages grouping effect
  - stabilizes the l_1 regularization path

6. Alternating Direction Method of Multipliers

- ADMM problem form (with f, g convex)

$$\text{minimize} \quad f(\boldsymbol{\beta}) + g(\boldsymbol{\xi})$$
$$\text{subject to} \quad \mathbf{A}\boldsymbol{\beta} + \mathbf{B}\boldsymbol{\xi} = \mathbf{c}$$

- two sets of variables, with separable objective

- The augmented Lagrangian is

$$L_\rho(\boldsymbol{\beta},\boldsymbol{\xi},\boldsymbol{\delta}) = f(\boldsymbol{\beta}) + g(\boldsymbol{\xi}) + \boldsymbol{\delta}^t(\mathbf{A}\boldsymbol{\beta}+\mathbf{B}\boldsymbol{\xi}-\mathbf{c}) + (\rho/2)\|\mathbf{A}\boldsymbol{\beta}+\mathbf{B}\boldsymbol{\xi}-\mathbf{c}\|_2^2$$

- Augmented Lagrangian methods were developed in part to bring robustness to the dual ascent method, and in particular, to yield convergence without assumptions like strict convexity or finiteness of f.

7. Alternating direction method of multipliers (Cont.)

- ADMM:

$$\boldsymbol{\beta}^{k+1} := \arg\min_{\boldsymbol{\beta}} L_\rho(\boldsymbol{\beta},\boldsymbol{\xi}^k,\boldsymbol{\delta}^k) \quad //\boldsymbol{\beta}\text{-minimization}$$
$$\boldsymbol{\xi}^{k+1} := \arg\min_{\boldsymbol{\xi}} L_\rho(\boldsymbol{\beta}^{k+1},\boldsymbol{\xi},\boldsymbol{\delta}^k) \quad //\boldsymbol{\xi}\text{-minimization}$$
$$\boldsymbol{\delta}^{k+1} := \boldsymbol{\delta}^k + \rho(A\boldsymbol{\beta}^{k+1} + B\boldsymbol{\xi}^{k+1} - \mathbf{c}) \quad //\text{dual-update}$$

- In the first step of the ADMM algorithm, we fix ξ and δ and minimize the augmented Lagrangian over β
- In the second step, we fix β and δ and minimize the augmented Lagrangian over ξ
- Finally, we update the dual variable δ

8. Alternating Direction Method of Multipliers for the proposed method

Consider now the generic problem

$$\text{minimize} \quad l(\boldsymbol{\beta}) + \mu\boldsymbol{\beta}^t\mathbf{Q}\boldsymbol{\beta} + \lambda\|\boldsymbol{\beta}\|_1, \quad (2)$$

where l is any convex loss function.
In ADMM form, this problem can be written as

$$\text{minimize}_x \quad \tilde{l}(\boldsymbol{\beta}) + g(\boldsymbol{\xi})$$
$$\text{subject to} \quad \boldsymbol{\beta} - \boldsymbol{\xi} = 0,$$

where $\tilde{l}(\boldsymbol{\beta}) = l(\boldsymbol{\beta}) + \mu\boldsymbol{\beta}^t\mathbf{Q}\boldsymbol{\beta}$ and $g(\boldsymbol{\xi}) = \lambda\|\boldsymbol{\xi}\|_1$.

9. Alternating Direction Method of Multipliers for the proposed method (Cont.)

The ADMM algorithm can be expressed on its scaled dual form as:

$$\boldsymbol{\beta}^{k+1} := \arg\min_{\boldsymbol{\beta}}\left\{\tilde{l}(\boldsymbol{\beta}) + (\tau/2)\|\boldsymbol{\beta}-\boldsymbol{\xi}^k+\boldsymbol{\eta}^k\|_2^2\right\};$$
$$\boldsymbol{\xi}^{k+1} := \arg\min_{\boldsymbol{\xi}}\left\{g(\boldsymbol{\xi}) + (\tau/2)\|\boldsymbol{\beta}^{k+1}-\boldsymbol{\xi}+\boldsymbol{\eta}^k\|_2^2\right\};$$
$$\boldsymbol{\eta}^{k+1} := \boldsymbol{\eta}^k + \boldsymbol{\beta}^{k+1} - \boldsymbol{\xi}^{k+1}.$$

The β-update is proximal operator evaluation. Since $\tilde{l}$ is smooth, this can be done using Newton-Raphson method. The ξ-update has a closed form solation given by

$$\boldsymbol{\xi}^{k+1} := S_{\frac{\lambda}{\tau}}\left(\boldsymbol{\beta}^{k+1} + \boldsymbol{\eta}^k\right),$$

10. Databases

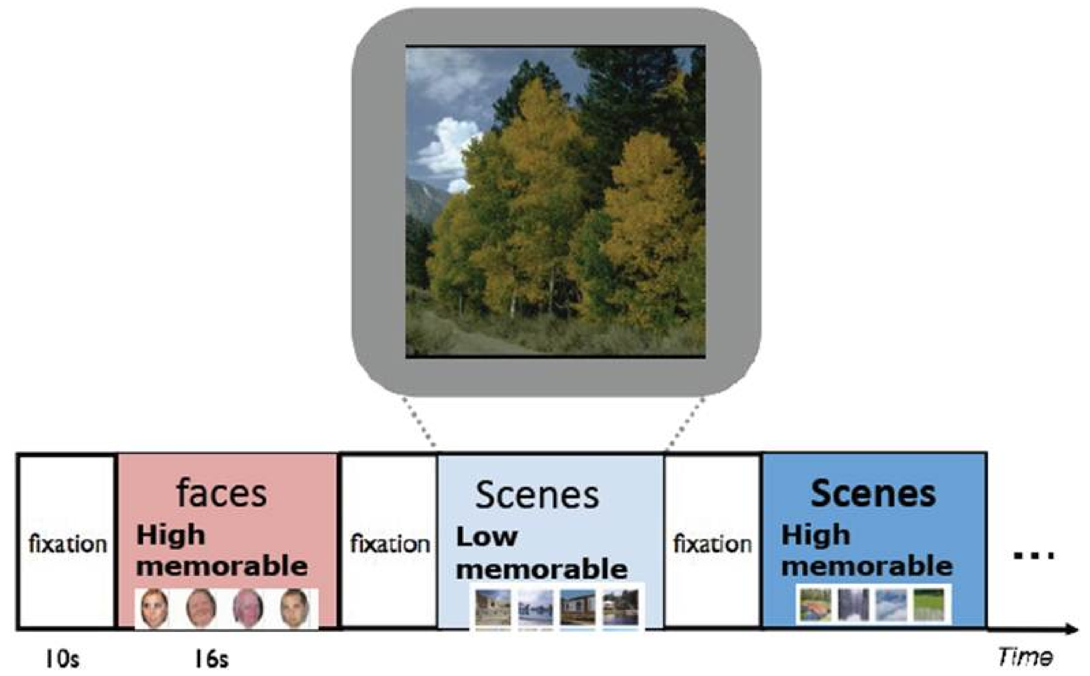

Figure 3: *fMRI Block design.*

- **Block design**:
- **Orthogonal Task**: 1-back repetition detection
- 19 subjects, US citizens
- 3T, 3mm cubic voxels
- 16 functional runs of task (all images in each run)
- Face, scene, visual functional localizers

11. Results

Table 1: *Measures of performance of each method*

Measure of performance	SVM	RFC	K-NN	NBC	DRLORA
Prediction	0.9375	1.00	0.3750	0.4375	1.00
Precision	1.00	1.00	0.67	0.67	1.00
Recall	0.91	1.00	0.31	0.36	1.00

- The best performances are given by the DRLORA, Random forest and support vector machines
- The worst performances are given by k-nearest neighbors and naive Bayes classifier
- In terms of **model complexity**, we note that DRLORA achieves its best performance using a subset of 22961 voxels
- Competitors use the entire set of voxels ($\simeq 105000$)