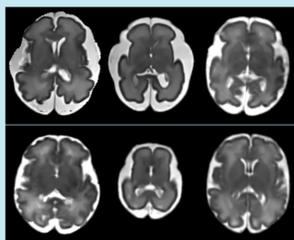
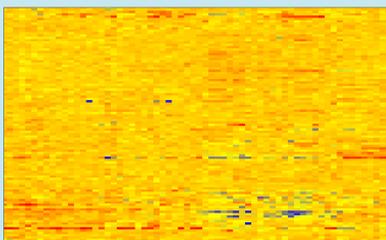


Discriminative and non-parametric sparse feature selection approach for biomedical data analysis

Dr. Julien Pontabry

1. Context

Micro-array of expression from $n = 64$ cells and $p = 6830$ genes
MRI images of $n = 23$ heterogeneous fetal brain with $p \approx 10^6$ voxels



- ▶ what are the *patterns* that may explain variability across data?
- ▶ issues of *high-dimensional* and *low-samples* datasets

2. Seeking for patterns in high-dimensional data

2.1. Feature extraction vs. feature selection

feature extraction creates a new space from original data space

feature selection select a subspace of original data space

- ▶ with extraction, dimensions lose their physical meanings (harder interpretation)

2.2. Feature selection approaches

wrapper scores features with a predictive model

filter ranks features by their ability to capture tendencies (e.g. correlation)

embed scores features with a predictive model built during selection

- ▶ e.g. lasso/elastic-net are embed linear methods based on reconstruction error

3. Methodology

3.1. Discriminative non-parametric model

Data column $Y_j \in \mathbb{R}^p$ and a subset of it $X_j \in \mathbb{R}^{s \ll p}$ are linked by a function f :

$$Y = \begin{pmatrix} v_{1,1} & v_{1,2} & \cdots & v_{1,n} \\ v_{2,1} & v_{2,2} & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ v_{p,1} & v_{p,2} & \cdots & v_{p,n} \end{pmatrix}, \quad Y_j = f(X_j).$$

Function f is approximated by kernel smoothing:

$$\hat{f}_h(x) = \sum_{j=1}^n \omega_j(x, h) Y_j, \quad \omega_j(x, h) = \frac{K_h(x - X_j)}{\sum_{k=1}^n K_h(x - X_k)}.$$

- ▶ *data-driven* approach with *no a-priori* on f
- ▶ *discriminative* because of the *kernel weighting*

3.2. Sparse selection as an optimization problem

Find activation vector γ and smoothing parameter h such that:

$$\{\hat{\gamma}, \hat{h}\} = \arg \min_{\gamma, h} \sum_{j=1}^n \|Y_j - \hat{f}_h(\text{diag}(\gamma) \cdot Y_j)\|_2^2 \quad \text{such that} \quad \|\gamma\|_0 \leq s \ll p$$

- ▶ *sparse* because of the L_0 -norm constraint
- ▶ *discriminative* because *minimal reconstruction error* pushes *maximal separation* between samples
- ▶ particular case of *dictionary learning* methods (here the data is the dictionary)

3.3. Solving the optimization problem

- ▶ *NP-hard* problem
- ▶ solve by L_0 relaxation (similar to Lasso) but sparsity not guaranteed (*shrinkage*)
- ▶ solve by *heuristic*, e.g. a greedy forward algorithm

5. Synthesis

- ▶ *patterns* that may explain dataset variability have been *highlighted* for two kind of biomedical datasets (micro-array of genes expression and MRI fetal brain images)
- ▶ *interpretation* of results is *facilitated* because of the physical meaning conservation

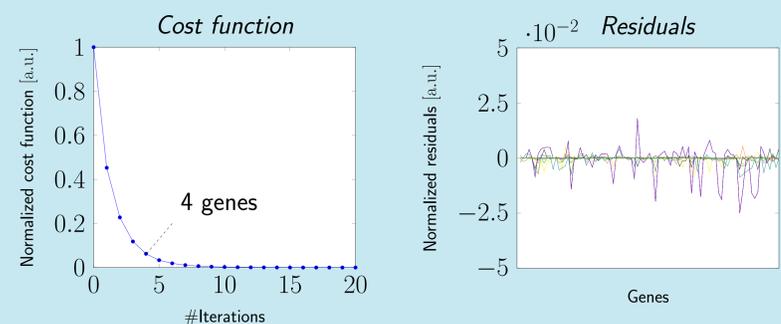
6. Perspectives

- ▶ *relationships between features* could be taken into account (e.g. spatial relationship of cortical points) within a framework similar as the fused Lasso technique
- ▶ *computation speed can be improved* with for instance genetic algorithm or by multi-resolution scheme when features have relationships

4. Results

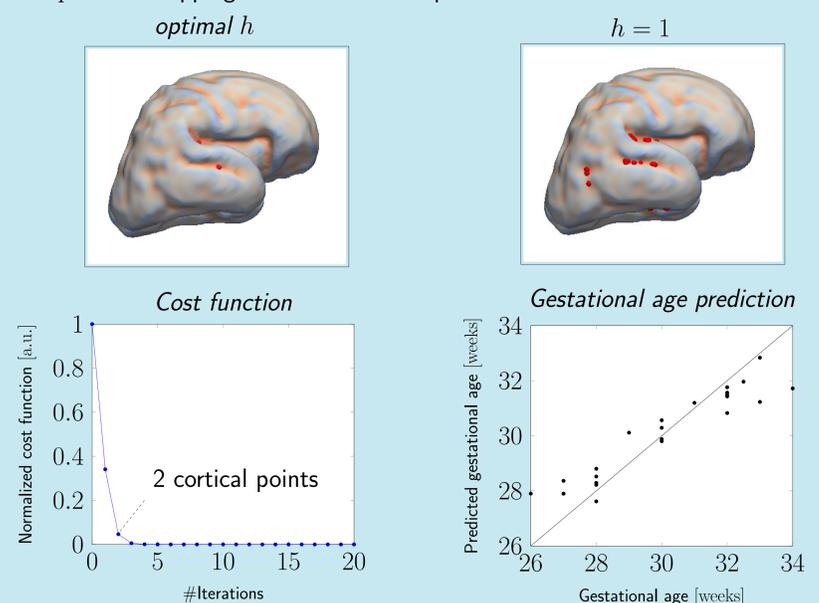
4.1. Micro-array of genes expression

We seek for significant/discriminant genes involved in variability of expression level in a micro-array dataset of $n = 64$ samples and $p = 6830$ genes. The micro-array is directly used as data matrix Y .



4.2. Cortical plate of fetal brain

We seek for significant/discriminant cortical folding patterns characterizing fetal brain development in a population of $n = 23$ subjects. Prior to cortical point selection, the population is normalized into a reference space. The Y data matrix corresponds to the $p \approx 10^6$ mapping vectors of cortical plates.



7. References

- ▶ Pontabry, J., Rousseau, F., Studholme, C., Koob, M. and Dietemann, J.-L. (2016). *A discriminative feature selection approach for shape analysis: application to fetal brain cortical folding*. Medical Image Analysis, accepted.
- ▶ Hastie, T., Tibshirani, R., Friedman, J., & Franklin, J. (2005). *The elements of statistical learning: data mining, inference and prediction*. The Mathematical Intelligencer, 27(2), 83-85.