Multimodal Graphical Models via Group Lasso

Marcelo Fiori
Instituto de Ingeniería Eléctrica, UdelaR, Uruguay
Email: mfiori@fing.edu.uy

Pablo Musé
Instituto de Ingeniería Eléctrica, UdelaR, Uruguay
Email: pmuse@fing.edu.uy

Ahmad Hariri
Institute for Genome Sciences & Policy
Duke University
Email: ahmad.hariri@duke.edu

Guillermo Sapiro
Electrical and Computer Engineering, Duke University
Email: guillermo.sapiro@duke.edu

***Abstract*—Graphical models are a very useful tool to describe and understand natural phenomena, from gene expression and brain networks to climate change and social interactions. In many cases, the data is multimodal. For example, one may want to build one network from several fMRI (functional magnetic resonance imaging) studies from different subjects, or combine different data modalities (as fMRI and questions) for several subjects. To this end, in this work we combine group lasso with graphical lasso, and derive an iterative shrinkage thresholding algorithm for solving the proposed optimization problem.**

The framework is validated with synthetic data and real fMRI data, showing the advantages of combining different modalities in order to infer the underlying network structure.

I. Introduction

The estimation of the inverse of the covariance matrix is a very important problem with applications in a number of fields. The *covariance selection* problem consists in identifying the zero pattern of the precision matrix. Two basic approaches have been developed for estimating the structure of the graphical model when this is sparse, and have been proved successful specially when working with few data points $k < p$, namely, the regression approach [1] and the maximum likelihood, also called Graphical Lasso [2]. The formulation of this latter one is:

$$\max_{\Theta \succ 0} \log\det\Theta - \mathrm{tr}(S\Theta) - \lambda ||\Theta||_{\ell_1},$$

where S is the empirical covariance matrix and $||\Theta||_{\ell_1} = \sum_{i,j} |\Theta_{ij}|$.

Let us suppose now that we want to infer n covariance matrices, but such that they (roughly) share the non-zero pattern. This can be done by means of the group lasso [3], [4]: in this case, we group all the entries (i, j) of the matrices Θ^h, and form an $n-$dimensional vector whose l_2 norm will be penalized in the objective optimization function. This way, the sum of penalty terms for all groups promotes sparsity, in the sense that only a few groups will be active (and so each matrix Θ^h will be sparse), but once a group is active, the corresponding n coefficients (the (i, j) entries for all Θ^h) will be all non-zero in general. The optimization problem to solve is

$$\max_{\Theta \succ 0} \sum_h \log\det\Theta^h - \sum_h \mathrm{tr}(S^h\Theta^h) - \lambda \sum_{i,j} ||\Theta_{ij}||_2 \qquad (1)$$

where (with a slight abuse of notation) Θ_{ij} is now an n-dimensional vector with the (i, j) entries of all matrices Θ^h. This problem is still convex, and we have adapted the ISTA algorithm [5], with the code available in www.fing.edu.uy/~mfiori.

II. Synthetic data examples

In this section the model and algorithm are assessed with two different experiments: in the first one we show how the performance of the grouped methodology improves as the number of groups grows, and we also compare it with concatenating the data instead of the grouping approach. In the second one we show that this methodology is able to mix different kinds of data (e.g. gaussian and discrete).

For the first experiment, we randomly generated six precision matrices with the same support (but different non-zero values), for $p = 60$. For each matrix we simulated Gaussian data $X_h \in k \times p$ for $k = 30$. Figure 1 (left) shows how the performance of the model (1) improves with the number of considered groups, and how the concatenation degrades the performance. In solid black line, estimation using only one dataset (X_1). Below it (dashed black), using the concatenation of different subsets. Above it, using the grouped methodology with: 2, 3, 4, 5 and 6 groups (blue and red).

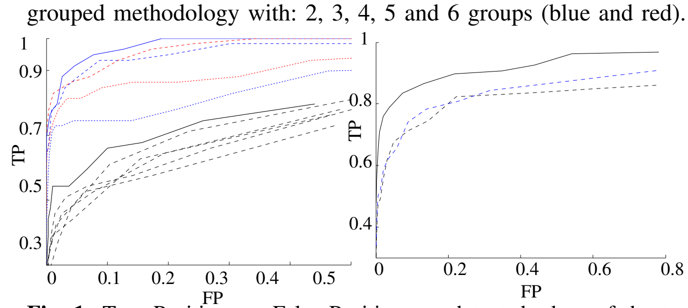

Fig. 1: True Positives vs False Positives on detected edges of the true graph. Left: comparison for several groups. Right: Discrete and Gaussian data. In dashed blue, using only gaussian data X, in dashed black using only discrete data Y, and in solid black using the grouped methodology.

For the second experiment, we generated a Gaussian Graphical Model and a Discrete Graphical model, sharing the same zero-pattern of the inverse covariance matrix, and simulated data from both of them, X and Y respectively. Figure 1 (right) shows the performance when inferring the zero-pattern only from X, only from Y, and with the combination of both via the optimization problem (1).

III. Application to fMRI data

Here we show how this collaborative learning can help to build brain networks for different groups of subjects. For an fMRI study of 155 subjects, we split the dataset into 105 for training and 50 for testing (data from http://www.haririlab.com/brain.php). With the training we built one network for males (A_M) and another one for females (A_F), using the grouped methodology (1). Then, for each subject in the testing set, we built the brain network from the fMRI data, and classified as male of female according to the closest graph adjacency matrix (A_M or A_F). To compare the performance, we also classified each subject with a nearest neighbor criteria with respect to all the subjects in the training set. When classifying with this latter method, a classification performance of 60% is achieved, while a performance of 80% is reached when building one coherent network for each gender.

References

[1] J. Peng, P. Wang, N. Zhou, and J. Zhu, "Partial correlation estimation by joint sparse regression models." *Journal of the American Statistical Association*, vol. 104, no. 486, pp. 735–746, Jun. 2009.

[2] O. Banerjee, L. El Ghaoui, A. D'Aspremont, and G. Natsoulis, "Convex optimization techniques for fitting sparse Gaussian graphical models," *ICML '06*, pp. 89–96, 2006.

[3] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society: Series B*, vol. 68, no. 1, pp. 49–67, 2006.

[4] G. Varoquaux, A. Gramfort, J.-B. Poline, and B. Thirion, "Brain covariance selection: better individual functional connectivity models using population prior," in *NIPS'10*, 2010, pp. 2334–2342.

[5] B. Rolfs, B. Rajaratnam, D. Guillot, I. Wong, and A. Maleki, "Iterative thresholding algorithm for sparse inverse covariance estimation," in *NIPS 2012*, 2012, pp. 1583–1591.