

000
001
002
003
004
005
006
007
008
009
010
011
012
013
014
015
016
017
018
019
020
021
022
023
024
025
026
027
028
029
030
031
032
033
034
035
036
037
038
039
040
041
042
043
044
045
046
047
048
049
050
051
052
053

Inferring Multiple Regulation Networks

Anonymous Author(s)

Affiliation

Address

email

Abstract

Gaussian Graphical Models provide a convenient framework for representing dependencies between variables. Recently, this tool has received a high interest for the discovery of biological networks. The literature focuses on the case where a single network is inferred from a set of measurements, but, as data is typically scarce, several assays, where the experimental conditions affect interactions, are usually merged to infer a single network. In this paper, we describe an approach for estimating several related networks, by rendering the closeness assumption into group penalties. We provide quantitative results demonstrating the benefits of the proposed approach on artificial and real data.

1 Motivations

In systems biology, experimental data is typically scarce, and it is a common practice to merge different experimental conditions to enlarge the number of observations available for inference. However, statistical models assume that the training set is made of identically distributed examples. In the aforementioned paradigm, assuming that all examples are drawn from the same distribution is obviously wrong, and is likely to have detrimental side effects in the estimation process.

We propose to remedy this problem by estimating jointly several interaction networks, each of whom matching a modality of the same variables, that is, a given experimental condition of the same molecule expressions. Our proposal is based on an inductive bias towards commonalities between tasks, aiming at stabilizing the estimation process, hopefully enabling more accurate predictions in small sample size regimes. We couple the inference of several related networks derived from Gaussian Graphical Models (GGM). To our knowledge, this work is the first to cast biological network inference in the multi-task learning framework, even though coupling the estimation of several networks has recently been addressed with penalties related to the Fused-lasso and total variations to model either networks subject to smooth variations or abrupt changes over time [1].

2 Network Inference with Gaussian Graphical Models

We infer the graph of conditional dependencies among variables in the GGM framework. Assuming that X is a p -dimensional Gaussian random variable $X \sim \mathcal{N}(\mathbf{0}, \Sigma)$, we denote by $\mathbf{K} = \Sigma^{-1}$ its concentration matrix. The entry K_{ij} is the partial correlation between the variables X_i and X_j , hence, a null value indicates conditional independence. The support of the concentration matrix is the adjacency matrix of an undirected graph \mathcal{G} depicting conditional dependencies of the p components of X , where each node represents a variable, and edges connect dependent variables.

The GGM approach produces the graph \mathcal{G} from the concentration matrix \mathbf{K} inferred from independent observations (X^1, \dots, X^n) . This matrix cannot be inferred by maximum likelihood that would return a full matrix, yielding a vacuous fully connected graph, but sparsity may be obtained by penalizing the entries of \mathbf{K} by an ℓ_1 -norm [2]. This well-motivated approach produces a sparse,

054 symmetric and positive-definite estimate of \mathbf{K} . However, a cruder procedure has been reported to
 055 be more accurate in terms of edge detection [3]: *neighborhood selection* [4] determines \mathcal{G} via an
 056 iterative estimation of the neighbors of each node. For this purpose, it considers p independent
 057 ℓ_1 -penalized regression problems. Let \mathbf{X} be the $n \times p$ matrix of stacked observations. The edges
 058 adjacent to node i are estimated by the non-zero elements of the solution to

$$059 \min_{\beta \in \mathbb{R}^{p-1}} \frac{1}{n} \|\mathbf{X}_i - \mathbf{X}_{\setminus i} \beta\|_2^2 + \lambda \|\beta\|_1, \quad (1)$$

061 where \mathbf{X}_i is the i th column of \mathbf{X} and $\mathbf{X}_{\setminus i}$ is \mathbf{X} deprived of its i th column: the i th variable is “ex-
 062 plained” by the remaining ones. As the neighborhood of the p variables are selected separately, a
 063 post-symmetrization must be applied to manage inconsistencies between edge selections [4]. Solv-
 064 ing the p regression problems (1) may be interpreted as inferring the concentration matrix in a pe-
 065 nalized, *pseudo* maximum likelihood framework, where the joint distribution of X is approximated
 066 by the product of the p distributions of each variable conditional on the other ones [5].
 067

068 3 Multiple Network Learning

069
 070 In biology, it is a common practice to conduct several assays where the experimental conditions
 071 differ, resulting in T samples measuring the expression of the same molecules. The T samples be-
 072 long to different subpopulations, with different distributions. Assuming that each sample was drawn
 073 independently from a Gaussian distribution $X^{(t)} \sim \mathcal{N}(\mathbf{0}, \Sigma^{(t)})$, each sample may be processed
 074 separately by following the approach described in Section 2. However, when networks have strong
 075 commonalities, sharing information should help estimation, especially for small sample sizes.
 076

077 The Group-lasso encourages sparse solutions with respect to groups, which form a pre-defined par-
 078 tition of variables. It has already inspired some multi-task learning strategies in supervised learn-
 079 ing [6]. In our framework, by grouping the partial correlations across the T graphs for any pairs of
 080 variables, the penalty favors graphs with identical edges. The learning problem is then

$$081 \max_{\{K_{ij}^{(t)}: i \neq j\}_{t=1}^T} \sum_{t=1}^T \mathcal{L}(\mathbf{K}^{(t)} | \mathbf{S}^{(t)}) - \lambda \sum_{i \neq j} \left(\sum_{t=1}^T (K_{ij}^{(t)})^2 \right)^{1/2}. \quad (2)$$

082 where, using the sufficiency of \mathbf{S} for \mathbf{K} , $\mathcal{L}(\mathbf{K} | \mathbf{S})$ denotes the pseudo-log-likelihood.
 083

084
 085 Though the Group-lasso expresses some of our expectations about the commonalities between tasks,
 086 it is not really satisfying regarding the support of $\mathbf{K}^{(t)}$. Indeed, to produce different networks (t, u) ,
 087 there must be at least one pair (i, j) , such that $K_{ij}^{(t)} = 0$ and $K_{ij}^{(u)} \neq 0$, but this event does not occur
 088 with Group-lasso, where variables enter or leave the support group-wise. We remedy this problem
 089 by introducing the Coop-lasso, which better suits our needs, by taking into account that the type of
 090 co-regulation, activation or inhibition, is not likely to be inverted across assays. In terms of partial
 091 correlations, sign swaps are very unlikely. This restraint suggests the learning problem

$$092 \max_{\{K_{ij}^{(t)}: i \neq j\}_{t=1}^T} \sum_{t=1}^T \mathcal{L}(\mathbf{K}^{(t)} | \mathbf{S}^{(t)}) - \lambda \sum_{i \neq j} \left(\left(\sum_{t=1}^T (K_{ij}^{(t)})_+^2 \right)^{1/2} + \left(\sum_{t=1}^T (-K_{ij}^{(t)})_+^2 \right)^{1/2} \right), \quad (3)$$

093 where $(u)_+ = \max(0, u)$. The Group and Coop penalties are identical in the positive and negative
 094 orthant, but the latter is more stringent when signs disagree across tasks: decoupling the positive
 095 and negative parts of the coefficients in the penalty favors solutions with identical signs.
 096
 097
 098

099 4 Experiments

100
 101 In most real-life applications, the major part of the inferred networks are unknown, with little avail-
 102 able information on the existence of edges. We essentially face an unsupervised learning prob-
 103 lem, where no objective criterion allows to compare different solutions. As a result, setting hyper-
 104 parameters is particularly troublesome, alike, say, choosing the number of components in clustering,
 105 and it is a common practice to visualize the networks obtained for several penalties.
 106

107 Regarding the first issue, we chose to present here synthetic and well-known real data that enable
 objective evaluations. Regarding the second issue, penalty parameters can be selected according to

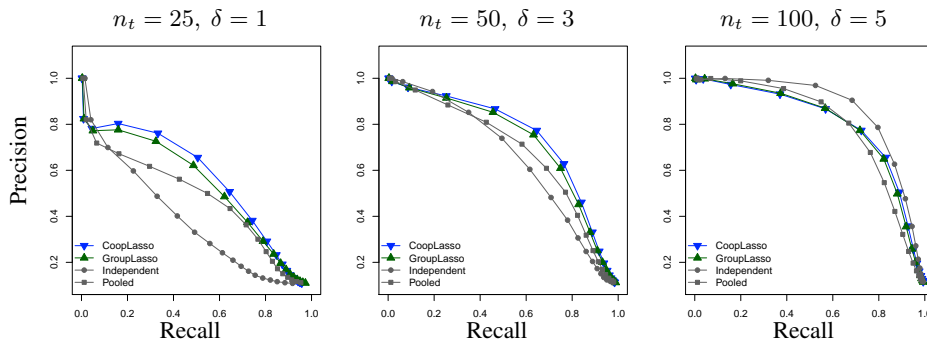


Figure 1: Precision-recall curves for the Coop, Group and the two baseline lasso ($T = 4$ graphs, with $p = 20$ nodes and $k = 20$ edges).

bounds on the rate of false edge discovery [2, 4, 5] or by information criteria concerning the estimation of \mathbf{K} [7]. However, these approaches tend to behave poorly, and the standard practice compares learning algorithms by scanning a series of results, through precision-recall or ROC-curves, letting the choice of penalty parameters as an open question for future research.

Synthetic Data – To generate data from similar graphs, we first draw an “ancestor” graph with p nodes and k edges and a symmetric $p \times p$ matrix of random signs Θ that will eventually define correlation signs. Then, T children graphs are produced by random addition and deletion of δ edges in the ancestor graph. The concentration matrices $\mathbf{K}^{(t)}$ are computed from the normalized graph Laplacians, with the sign of each entry in accordance with Θ . The covariance matrices are then computed by pseudo-inversion of $\mathbf{K}^{(t)}$ and finally used to generate n_t centered Gaussian vectors.

The precision-recall plots display the cumulative number of true and false edge detections among children networks. Precision is the ratio of true selected edges to selected edges; recall is the ratio of true selected edges to true edges. To ensure representativeness, the precision-recall statistics are averaged over 100 random draws of the ancestor graph. Averaging is performed at each trial value of the penalization parameter λ . We compare the multi-task methods, Group and Coop lasso, to two baselines: neighborhood selection applied separately to each data set of size n_t (annotated “independent”), or computed on the overall data merging all data sets (annotated “pooled”).

We provide here a summary of a comprehensive experimental design carried out for $T = 4$ children networks with $p = 20$ nodes and $k = 20$ edges. Figure 1 displays precision-recall plots for three prototypical situations. The left graph represents the small/medium-sample low-perturbation situation, where merging data sets is a good strategy compared to the independent analysis. Multi-task approaches dominate the pooled strategy, with the Coop-lasso being superior to the Group-lasso. When the sample size or the perturbation increases, as in the center of Figure 1, pooling is beneficial for high recalls, but there is a slight advantage for the independent analysis for low recalls (when less degrees of freedom are to be determined). Multi-task approaches dominate the best strategy, with the same ordering as above, but smaller differences. All methods perform similarly for medium-sample low-perturbation and large-sample medium-perturbation (not shown). Finally, the right graph in Figure 1 represents the large-sample high-perturbation situation, where merging data is a bad strategy, as networks differ significantly and the amount of data permits their accurate estimation. Coop and Group-lasso behave well for high recalls, but for low recalls (high penalty parameters λ), they eventually become slightly worse than the pooled estimation. On any edge, the children networks have a probability smaller than $1/3$ to have the same sparsity pattern: the commonality assumption does not conform to reality and the group constraints should thus not be enforced too heavily.

Protein Signaling Network – Very few real data sets come with a reliable and exhaustive ground-truth allowing quantitative assessments. We make use of a data set pertaining to a well-studied signaling pathway [8], involving 11 molecules, with 20 interactions described in the literature. The signaling network is perturbed by activating or inhibiting the production of a given molecule. Fourteen assays have been conducted to reveal different parts of the network. Here, we used four assays: inhibition of PKC, activation of PKC, inhibition of AKT, activation of PKA.

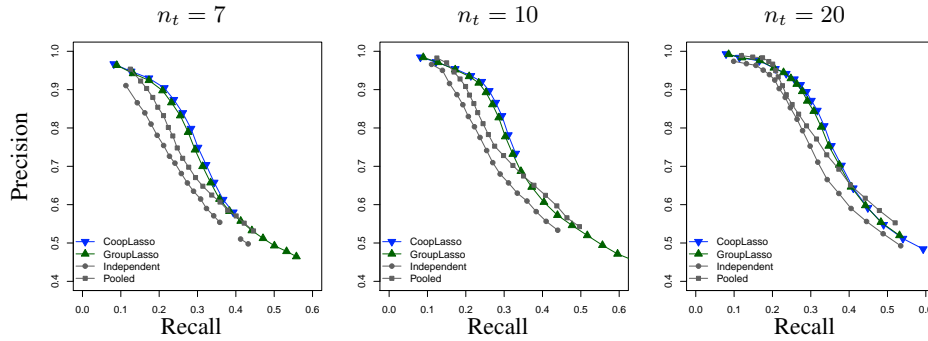


Figure 2: Precision-recall curves for the Coop, Group and the two baseline lasso, for learning the networks on four assays of Sachs’ data from four samples of size $n_t = 7, 10$ and 20 .

All inference methods perform about equally well for large samples (each assay consists here of about 1000 repeated measurements). We report in Figure 2 the results obtained for smaller subsamples, with precision-recall plots averaged over 100 independent random draws of the subsamples. The results accord with the previous experiments: for all subsample sizes, multi-task learning dominates the best baseline strategy, which is here pooled estimation.

5 Conclusion

We presented an approach for learning multiple networks in the Gaussian Graphical Model framework. We target situations routinely occurring with omics data, where measurements are scarce and recorded in different experimental conditions. In this setup, the two baseline approaches consist in either handling the learning problems separately or as a single one by merging all available data sets.

Our strategy, based on neighborhood selection, relaxes the uniqueness constraint that is implicit when the tasks are processed as a single one, merely biasing the results towards a common answer. We introduce the Coop-lasso, a group-penalty that favors similar networks with homogeneous dependencies between any given pair of variables. Homogeneity is quantified by the magnitude and sign of partial correlations. The Coop-lasso contrasts the Group-lasso in its ability to infer different networks for each task. Our experimental results show that, among the baselines, pooling is usually a good strategy, except for large samples with major discrepancies between networks. The Group-lasso multi-task approach performs better, especially in the targeted situation where data are scarce, and the Coop-lasso further improves results.

References

- [1] M. K. Kolar, A. A. Le Song, and E. P. Xing. Estimating time-varying networks. *Ann. Appl. Stat.*, 2009.
- [2] O. Banerjee, L. El Ghaoui, and A. d’Aspremont. Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data. *J. Mach. Learn. Res.*, 9:485–516, 2008.
- [3] F. Villers, B. Schaeffer, C. Bertin, and S. Huet. Assessing the validity domains of graphical Gaussian models in order to infer relationships among components of complex biological systems. *Stat. Appl. Genet. Mol. Biol.*, 7(2), 2008.
- [4] N. Meinshausen and P. Bühlmann. High-dimensional graphs and variable selection with the lasso. *Ann. Statist.*, 34(3):1436–1462, 2006.
- [5] C. Ambroise, J. Chiquet, and C. Matias. Inferring sparse Gaussian graphical models with latent structure. *Electron. J. Stat.*, 3:205–238, 2009.
- [6] A. Argyriou, T. Evgeniou, and M. Pontil. Convex multi-task feature learning. *Mach. Learn.*, 73(3):243–272, 2008.
- [7] M. Yuan and Y. Lin. Model selection and estimation in the Gaussian graphical model. *Biometrika*, 94(1):19–35, 2007.
- [8] K. Sachs, O. Perez, D. Pe’er, D.A. Lauffenburger, and G.P. Nolan. Causal protein-signaling networks derived from multiparameter single-cell data. *Science*, 308:523–529, 2005.