Precision matrix estimation in sparse Gaussian graphical models

gLasso and gSLOPE approach

Michał Makowski

Faculty of Mathematics and Computer Science

Master's thesis

XXX@outlook.com

Uniwersytet Wrocławski

theoretical foundations of machine learning, Kraków

tfml2019

Abstract

In large data sets relationships between variables could be represented as graphs, where nodes represent variables, and edges connect variables which are conditionally dependent, given all other variables. In related Gaussian graphical models edges correspond to nonzero elements of a precision matrix, which is an inverse of a covariance matrix. In a case when the number of observations in a database is comparable to or smaller than the number of variables, classical maximum likelihood estimates of the precision matrix do not exist or have the very large variance. Then the solution are regularization-based methods like *Graphical Lasso*, which stabilize the performance of MLE. We presented novel approach, *Graphical SLOPE*, based on Ordered L-One Norm. We implemented ADMM algorithm for gSLOPE problem and compared performance of gLasso and gSLOPE in synthetic experiments, mainly in terms of *False Discovery Rate* [FDR]. The simulations showed promising results in terms of FDR control, moreover gSLOPE systematically outperforms gLASSO with respect to ROC curves.

Introduction

Probabilistic graphical models provide a useful framework for building parsimonious models for high-dimensional data. They are based on an interplay between probability theory and graph theory, in which the properties of an underlying graph specify the conditional independence properties of a set of random variables. In typical applications, the structure of this graph is not known, and it is of interest to estimate it based on samples, a problem known as graphical model selection. We have focused on Gaussian Graphical Models, where the data follow multivariate random distribution.

Background

Gaussian Graphical Models

Any non-degenerated Gaussian distribution can be reformulated into *so-called* canonical parameters of the form

$$\gamma = \Sigma^{-1}\mu \quad \text{and} \quad \Theta = \Sigma^{-1},$$

If X factorizes according to some graph G, $\theta_{st} = 0$ for any pair $(s,t) \notin E$, which sets up correspondence between the zero pattern of the matrix Θ and pattern of the underlying graph. In particular, if the $\theta_{st} = 0$, then variables s and t are conditionally independent, given the other variables.

Figure 1 – The graph and corresponding precision matrix

The problem of graph selection

Let $X = \{x_1, \dots, x_N\}$ be a collection of samples from a graphical model with unknown structure of the graph, then log-likelihood function of precision matrix is given by

$$\mathbb{L}(\Theta, X) = \log\det\Theta - \operatorname{tr}(S\Theta). \tag{1}$$

For problems frequently arising in practice MLE does not exist or we want to constraint the number of edges. The solution is to penalize MLE

$$\widehat{\Theta}_{ML} \in \underset{\Theta \in S^p_+}{\arg\max} \{\log\det\Theta - \operatorname{tr}(S\Theta) - J_\lambda(\Theta)\}.$$

The natural choice of function J_λ seems to be ℓ_0 norm, but this produces a nonconvex, hard to solve, problem. Its ℓ_1 convex relaxation is known as Graphical Lasso.

Graphical SLOPE

We proposed novel approach motivated by [2] where the function J_λ is OL1 norm. This method was designed by Bogdan et al. to control FDR in linear models.

$$\mathrm{FDR} = \mathbb{E}\left[\frac{\#[\text{False positive}]}{\#[\text{False positive}] + \#[\text{True positive}]}\right]$$

$$\mathrm{localFDR} = \mathbb{E}\left[\frac{\#[\text{False positive outside the component}]}{\#[\text{False positive}] + \#[\text{True positive}]}\right]$$

Choice of parameter

Choice of parameter λ is crucial for performance of both methods

- lambda for gLasso based on the Bonferroniego correction [1],
- lambda series for gSLOPE based on the Holm correction [4],
- lambda series for gSLOPE based on the BH correction [4].

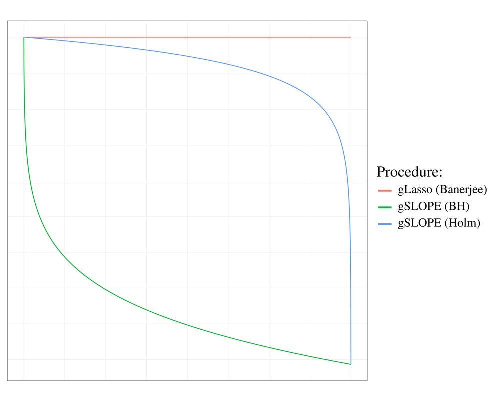

Figure 2 – The change of as the coefficients decrease

ADMM

For solving the Graphical SLOPE problem we used the *Alternating direction method of multipliers* [3], it can solve convex problems of the form

$$\begin{aligned} &\text{minimize} && f(x) + g(y) \\ &\text{subject to} && Ax + By = c. \end{aligned}$$

Results

We conducted a number of synthetic simulations which unveiled desirable properties of gSLOPE. The simulations was done on 3 types of graphs, although the theoretical results were only proved for cluster-type dependency graphs. Number of variables is equal to 100, the desired FDR control is equal to 0.2.

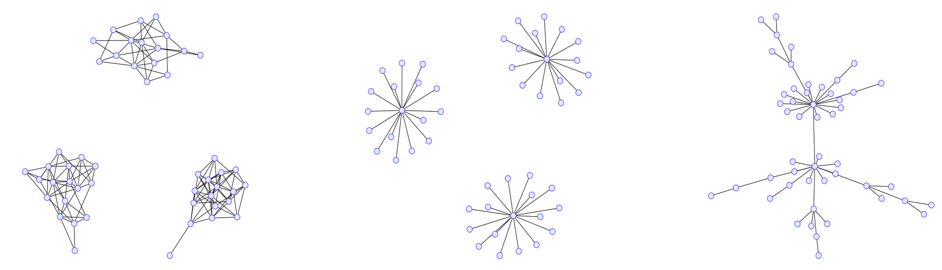

Figure 3 – Left to right: cluster, hub, and scale-free graph.

Cluster-graph

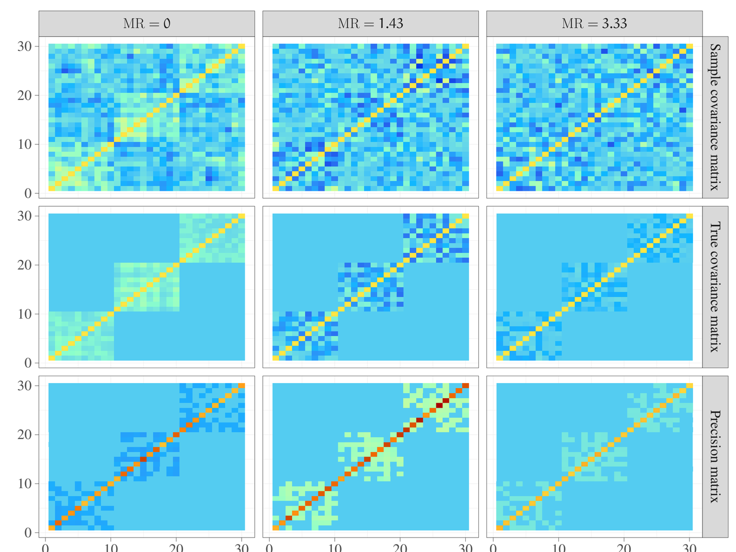

Figure 4 – Impact of MR parameter on generated data in cluster graphs

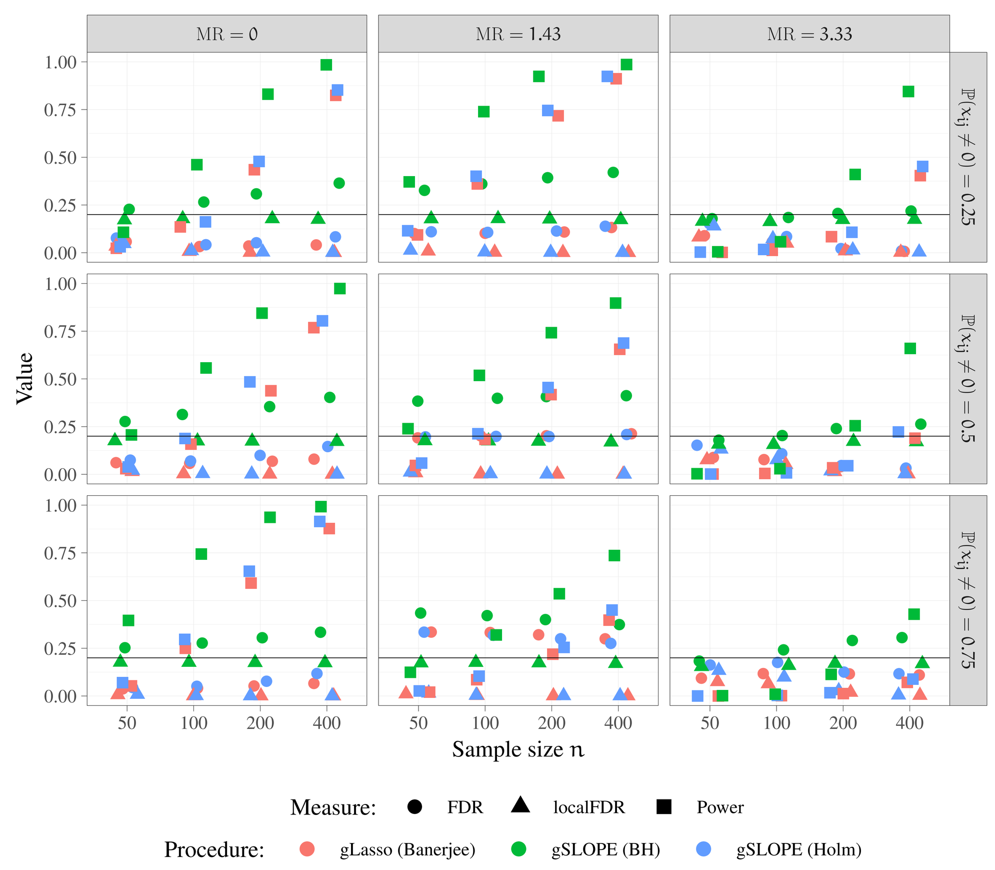

Figure 5 – Results for cluster-type graphs

Hub-graph

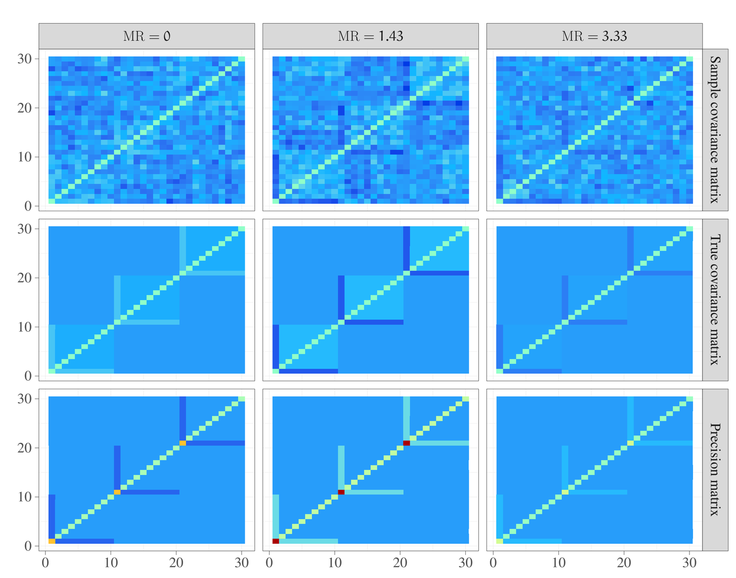

Figure 6 – Impact of MR parameter on generated data in hub graphs

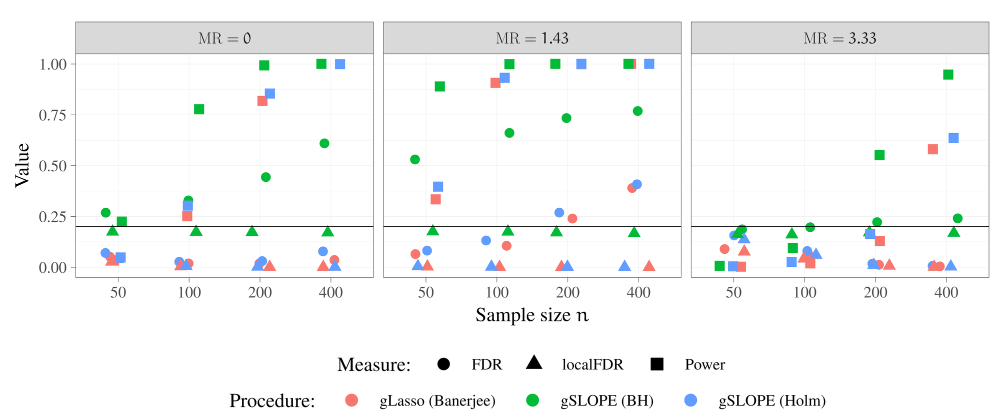

Figure 7 – Results for cluster-type graphs

Scale-free-graph

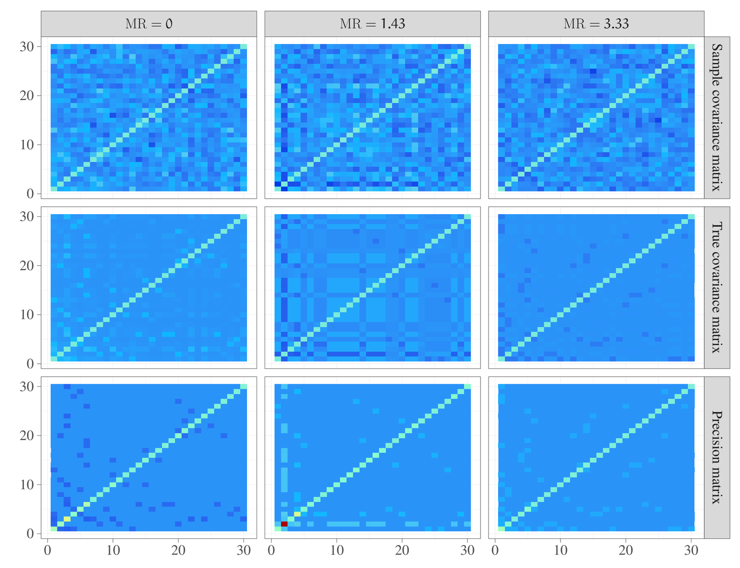

Figure 8 – Impact of MR parameter on generated data in scale-free graphs

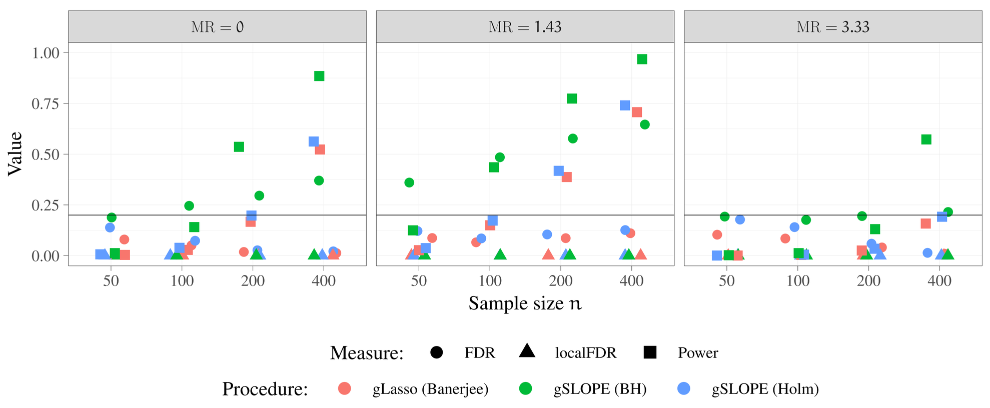

Figure 9 – Results for scale-free-type graphs

ROC curves

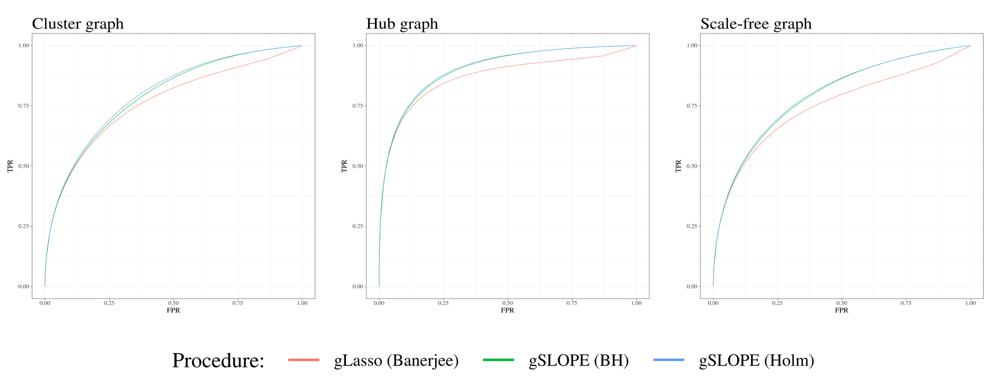

Figure 10 – Left to right: cluster, hub, and scale-free ROC curves; both OL1-based methods slightly outperform gLASSO in terms of ROC curves

Conclusions

- Although theoretical results was proven only for cluster-type graphs, in hub-type graphs and scale-free-type graphs results obtained by both methods are satisfactory.
- Generally, in Gaussian settings gSLOPE with BH correction outperforms gLasso in terms of power having low FDR and local FDR. The Holm correction gives similiar results to gLasso.
- With respect to ROC curves gSLOPE outperforms gLasso, what illustrate compromise between specificity and sensitivity.

References

[1] Onureena Banerjee, Laurent El Ghaoui, and Alexandre d'Aspremont. Model Selection Through Sparse Maximum Likelihood Estimation for Multivariate Gaussian or Binary Data. *The Journal of Machine Learning Research*, 9:485–516, 2008.

[2] Małgorzata Bogdan, Ewout van den Berg, Chiara Sabatti, Weijie Su, and Emmanuel J Candès. SLOPE - Adaptive variable selection via convex optimization. *The annals of applied statistics*, 9(3):1103–1140, 2015.

[3] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. *Machine Learning*, 3(1):1–122, 2010.

[4] Piotr Sobczyk. *Identifying low-dimensional structures through model selection in high-dimensional data*. PhD thesis, Wrocław University of Science and Technology, (not published).

Acknowledgements

I would like to thank Malgorzata Bogdan, PhD, and Piotr Sobczyk for helpful suggestions.

Questions?

Contact me through email or directly at the conference; hope the photo will help you find me. :)