Meta Learning for Support Recovery in High-dimensional Precision Matrix Estimation

Qian Zhang, Yilin Zheng, Jean Honorio

Purdue University

Introduction

- Precision (or inverse covariance) matrix estimation
  - Gaussian graphical models: the set of off-diagonal non-zero entries (support) of the precision matrix $\Longleftrightarrow$ the set of edges of the graph.
  - Sign-consistency: the estimated precision matrix has the same support and sign of entries with respect to the true matrix.
- Challenges:
  - High-dimensionality: n (sample size) $\ll N$ (dimension)
    - assume sparsity; use ℓ_1-regularized log-determinant Bregman divergence minimization.
  - Heterogeneity: multiple tasks with different precision matrices
    - introduce **random precision matrices**;
    - consider **meta learning** and **improper estimation**.

Model

- Multivariate sub-Gaussian distributions with **random precision matrices**. For $k \in \{1, .., K\}, t \in \{1, .., n^{(k)}\}, i \in \{1, .., N\}$:
  - $X_1^{(k)}, \dots, X_{n^{(k)}}^{(k)} \in \mathbb{R}^N$ i.i.d.;
  - $\mathbb{E}\left[X_t^{(k)} \middle| \bar{\Omega}^{(k)}\right] = 0$, Cov $\left(X_t^{(k)} \middle| \bar{\Omega}^{(k)}\right) = \bar{\Sigma}^{(k)} := \left(\bar{\Omega}^{(k)}\right)^{-1}$;
  - $\bar{\Omega}^{(k)} = \bar{\Omega} + \Delta^{(k)}$ with $\bar{\Omega}, \Delta^{(k)} \in \mathbb{R}^{N \times N}$;
  - $\bar{\Omega} \succ 0$ deterministic, $\Delta^{(k)}$ **i.i.d. from distribution** P **satisfying some conditions**;
  - $X_1^{(k)}, \dots, X_{n^{(k)}}^{(k)}$ conditionally independent given $\bar{\Omega}^{(k)}$;
  - $X_{t,i}^{(k)}$ conditioned on $\bar{\Omega}^{(k)}$ is sub-Gaussian.
- $\bar{\Omega}$: true common precision matrix.
  $S :=$ supp$(\bar{\Omega})$: support union.

Method

- Meta learning – a two-step method:
  - estimate the **support union** from K auxiliary tasks;
  - estimate the precision matrix of the novel task (the $(K+1)$-th task) with the knowledge of the support union.
- Improper estimation:
  - pool **all the samples** from auxiliary tasks to estimate a **single** "common precision matrix" to recover the support union.

Method

- Support union recovery:

$$\hat{\Omega} = \underset{\Omega \succ 0}{\arg\min} \left(\sum_{k=1}^{K} T^{(k)} \left(\langle \hat{\Sigma}^{(k)}, \Omega \rangle - \log\det(\Omega) \right) + \lambda \|\Omega\|_1 \right), \tag{1}$$

$T^{(k)} \propto n^{(k)}, \quad \hat{\Sigma}^{(k)} := \frac{1}{n^{(k)}} \sum_{t=1}^{n^{(k)}} X_t^{(k)} \left(X_t^{(k)} \right)^\top.$

- Support recovery for novel task:

$$\hat{\Omega}^{(K+1)} = \underset{\Omega \succ 0}{\arg\min} \left(\langle \hat{\Sigma}^{(K+1)}, \Omega \rangle - \log\det(\Omega) + \lambda \|\Omega\|_1 \right)$$
$$\text{s.t. } \operatorname{supp}(\Omega) \subseteq \operatorname{supp}(\hat{\Omega}), \quad \operatorname{diag}(\Omega) = \operatorname{diag}(\hat{\Omega}). \tag{2}$$

For simplicity, assume $n^{(k)} = n$ for $1 \leq k \leq K$.

Theoretical Results

- Support union recovery
  - Under some conditions, with probability at least
    $$1 - O\left(N^2 \exp\{-nK\}\right), \tag{3}$$
    the estimator $\hat{\Omega}$ is sign-consistent.
  - Sufficient sample complexity for support union recovery: $n \in O((\log N)/K)$.

Theoretical Results

- Support union recovery
  - For some family of N-dimensional random multivariate sub-Gaussian distributions of size K with support union S and any estimate $\hat{S}$ of S, we have

$$\mathbb{P}\{\hat{S} \neq S\} \geq 1 - O\left(\frac{nK}{\log N}\right) \tag{4}$$

  - Necessary sample complexity for support union recovery: $n \in \Omega((\log N)/K)$.

Theoretical Results

- Support recovery for novel task
  - Suppose the support union S is recovered. For a novel task of multivariate sub-Gaussian distribution with precision matrix $\bar{\Omega}^{(K+1)}$ such that $\mathrm{supp}(\bar{\Omega}^{(K+1)}) \subseteq S$, under some conditions, with probability at least,

$$1 - O\left(|S_{\mathsf{off}}| \exp\left\{-n^{(K+1)}\right\}\right), \tag{5}$$

    the estimator $\hat{\Omega}^{(K+1)}$ is sign-consistent.
- Overall sufficient sample complexity:
  - $n \in O(\log(N)/K)$ for each auxiliary task;
  - $n^{(K+1)} \in O(\log(|S_{\mathsf{off}}|))$ for the novel task.

Theoretical Results

- Support recovery for novel task
  - Consider n samples generated from some N-dimensional multivariate sub-Gaussian distribution whose precision matrix has support $S^{(K+1)} \subset S$. For any estimate $\hat{S}^{(K+1)}$ of $S^{(K+1)}$, we have

$$\mathbb{P}\{\hat{S}^{(K+1)} \neq S^{(K+1)}\} \geq 1 - O\left(\frac{n}{\log |S_{\mathsf{off}}|}\right) \tag{6}$$

  - Necessary sample complexity for the novel task: $n^{(K+1)} \in \Omega(\log(|S_{\mathsf{off}}|))$.
- Our method is **minimax optimal**.

Related Work

Table 1: Sufficient sample complexity for support union recovery.

Method	Sample Complexity
Our meta learning method	$n \in O(\log(N)/K)$
Multi-task [Honorio et al., 2012]	$n \in O(\log K + \log N)$
Multi-task [Guo et al., 2011]	$n \in O((N \log N)/K)$
Multi-task [Ma and Michailidis, 2016]	$n \in O(K + \log N)$

Table 2: Sufficient sample complexity for support recovery on the novel task.

Method	Sample Complexity
Our meta learning method	$n^{(K+1)} \in O(\log(\lvert S_{\mathsf{off}} \rvert))$
Single-task [Ravikumar et al., 2011]	$n^{(K+1)} \in O(\log N)$

Thank You!