Network exploration via the adaptive LASSO and SCAD penalties

Yang Feng

Princeton University

A joint work with Prof. Jianqing Fan and Prof. Yichao Wu

Importance of Precision Matrix

- represents graphical models.
- genetic networks.
- gene classification.
- optimal portfolio allocation.
- ...

Sparse Precision Matrix

Zero element of the precision matrix implies conditional independence between these two variables.

Problem and Notations

- n: number of observations.
- p: number of dimension.
- $\mathbf{\Sigma}$: $p \times p$ covariance matrix.
- $\mathbf{\Omega}$: $p \times p$ precision matrix, and $\mathbf{\Omega} = \mathbf{\Sigma}^{-1}$.

Given $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n \sim N(0, \mathbf{\Sigma})$

We want to estimate the precision matrix $\mathbf{\Omega}$.

Penalized log-likelihood:

First, let's denote the generic penalty function by $p(.)$. The problem can be recast as:

$$\max_{\mathbf{\Omega}\in S_p} \log\det\mathbf{\Omega} - < \hat{\mathbf{\Sigma}}, \mathbf{\Omega} > - \sum_{i=1}^{n}\sum_{j=1}^{n} p_{\lambda_{ij}}(\omega_{ij}), \qquad (1)$$

where ω_{ij} is the (i, j)-element of matrix $\mathbf{\Omega}$. It reduces to the cardinality penalty when $p_{\lambda_{ij}}(\omega_{ij}) = \lambda L_0(\omega_{ij})$.

LASSO penalty

Proposed by Tibshirani(1996).
Friedman et al. (2008) applied this penalty and proposed the efficient graphical lasso algorithm for solving the corresponding optimization problem

$$\max_{\boldsymbol{\Omega} \in S_p} \log\det \boldsymbol{\Omega} - < \hat{\boldsymbol{\Sigma}}, \boldsymbol{\Omega} > - \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_{ij} |\omega_{ij}| \tag{2}$$

using a coordinate-descent procedure. We use their algorithm as our starting point to solve the penalized likelihood problem.

- Convex and increases linearly in the magnitude of its argument.
- Produces substantial biases in the estimation for large regression coefficients.

SCAD and adaptive LASSO

- SCAD penalty(Fan (1997), Fan and Li (2001))
- adaptive LASSO penalty (Zou (2006))
- Both have oracle properties.

We resort to the local linear approximation (LLA) as in Zou and Li (2007). At each step k, denote the current solution by $\boldsymbol{\Omega}^{(k)}$.We are optimizing, up to a constant,

$$\max_{\boldsymbol{\Omega} \in S_p} \log \det \boldsymbol{\Omega} - < \hat{\boldsymbol{\Sigma}}, \boldsymbol{\Omega} > - \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} |\omega_{ij}|, \tag{3}$$

where $w_{ij} = p'_\lambda(|\omega_{ij}^{(k)}|)$.

Summary

- Introduce the adaptive LASSO and SCAD penalties to the precision matrix estimation.
- Use the Local Linear Approximation to adapt the non-concave penalty function like SCAD to a weighted L_1 penalty.
- Simulation results and asymptotic properties are provided to justify the proposed methods.

Thank you very much!