Data Sciences – CentraleSupelec
Advance Machine Learning
Course VII - Inference on Graphical Models

Emilie Chouzenoux

Center for Visual Computing
CentraleSupelec
emilie.chouzenoux@centralesupelec.fr

Graphical models

∗ A graph $\mathcal{G}$ consists of a pair $(\mathcal{V}, \mathcal{E})$, with $\mathcal{V}$ the set of vertices and $\mathcal{E}$ the set of edges.

∗ In graphical models, each vertex represents a random variable, and the graph gives a visual way of understanding the joint distribution P of a set of random variables X:

$$X = (X^{(1)}, \ldots, X^{(p)}) \sim P$$

Graphical models

* A graph $\mathcal{G}$ consists of a pair $(\mathcal{V}, \mathcal{E})$, with $\mathcal{V}$ the set of vertices and $\mathcal{E}$ the set of edges.
* In graphical models, each vertex represents a random variable, and the graph gives a visual way of understanding the joint distribution P of a set of random variables X:

$$X = (X^{(1)}, \dots, X^{(p)}) \sim P$$

* In an **undirected** graph, the edges have no directional arrows. We say that the **pairwise Markov property** holds if, for every $(j, k) \in \mathcal{V}^2$, the absence of an edge between $X^{(j)}$ and $X^{(k)}$ **is equivalent to** the conditionally independence of the corresponding random variables, given the other variables:

$$X^{(j)} \perp X^{(k)} | X^{(\mathcal{V} \setminus \{j,k\})}.$$

* Undirected + pairwise Markov = conditional independence graph model.

Gaussian graphical model

∗ A **Gaussian graphical model** (GGM) is a conditional independence graph with a multivariate Gaussian distribution:

$$X = (X^{(1)}, \ldots, X^{(p)}) \sim \mathcal{N}(0, \Sigma)$$

with positive definite covariance matrix $\Sigma \in \mathbb{R}^{p \times p}$.

Gaussian graphical model

* A **Gaussian graphical model** (GGM) is a conditional independence graph with a multivariate Gaussian distribution:

$$X = (X^{(1)}, \dots, X^{(p)}) \sim \mathcal{N}(0, \Sigma)$$

with positive definite covariance matrix $\Sigma \in \mathbb{R}^{p \times p}$.

* The **partial correlation** between $X^{(j)}$ and $X^{(k)}$ given $X^{(\mathcal{V} \setminus \{j,k\})}$ equals:

$$\rho_{jk|\mathcal{V}\setminus\{j,k\}} = -\frac{K_{jk}}{\sqrt{K_{jj}K_{kk}}} \quad \text{with} \quad K = \Sigma^{-1}$$

Gaussian graphical model

∗ A **Gaussian graphical model** (GGM) is a conditional independence graph with a multivariate Gaussian distribution:

$$X = (X^{(1)}, \dots, X^{(p)}) \sim \mathcal{N}(0, \Sigma)$$

with positive definite covariance matrix $\Sigma \in \mathbb{R}^{p \times p}$.

∗ The **partial correlation** between $X^{(j)}$ and $X^{(k)}$ given $X^{(\mathcal{V} \setminus \{j,k\})}$ equals:

$$\rho_{jk|\mathcal{V}\setminus\{j,k\}} = -\frac{K_{jk}}{\sqrt{K_{jj}K_{kk}}} \quad \text{with} \quad K = \Sigma^{-1}$$

∗ Consider the **linear regression**: $X^{(j)} = \beta_k^{(j)} X^{(k)} + \sum_{r \in \mathcal{V}\setminus\{j,k\}} \beta_r^{(j)} X^{(r)} + \epsilon^{(j)}$ with $\epsilon^{(j)}$ zero-mean and independant from $X^{(r)}$, $r \in \mathcal{V} \setminus \{j\}$. Then,

$$\beta_k^{(j)} = -K_{jk}/K_{jj}, \quad \beta_j(k) = -K_{jk}/K_{kk}$$

Gaussian graphical model

∗ A **Gaussian graphical model** (GGM) is a conditional independence graph with a multivariate Gaussian distribution:

$$X = (X^{(1)}, \ldots, X^{(p)}) \sim \mathcal{N}(0, \Sigma)$$

with positive definite covariance matrix $\Sigma \in \mathbb{R}^{p \times p}$.

∗ The **partial correlation** between $X^{(j)}$ and $X^{(k)}$ given $X^{(\mathcal{V} \setminus \{j,k\})}$ equals:

$$\rho_{jk|\mathcal{V}\setminus\{j,k\}} = -\frac{K_{jk}}{\sqrt{K_{jj} K_{kk}}} \quad \text{with} \quad K = \Sigma^{-1}$$

∗ Consider the **linear regression**: $X^{(j)} = \beta_k^{(j)} X^{(k)} + \sum_{r \in \mathcal{V}\setminus\{j,k\}} \beta_r^{(j)} X^{(r)} + \epsilon^{(j)}$ with $\epsilon^{(j)}$ zero-mean and independant from $X^{(r)}$, $r \in \mathcal{V} \setminus \{j\}$. Then,

$$\beta_k^{(j)} = -K_{jk}/K_{jj}, \quad \beta_j(k) = -K_{jk}/K_{kk}$$

∗ The edges in a GGM are then related to Σ, K and β through:

$$(j,k) \text{ and } (k,j) \in \mathcal{E} \Leftrightarrow \Sigma^{-1}_{jk} \neq 0 \Leftrightarrow \rho_{jk|\mathcal{V}\setminus\{j,k\}} \neq 0 \Leftrightarrow \beta_k^{(j)} \neq 0 \text{ and} \beta_j^{(k)} \neq 0$$

Nodewise regression

$*$ We aim at **inferring the presence of edges** in a GGM. **Nodewise regression** consists in performing many regressions [Meinshausen et al., 2006], relying on the fact that:

$$X^{(j)} = \sum_{r \neq j} \bar{\beta}_r^{(j)} X^{(r)} + \epsilon^{(j)}, \quad j = 1, \dots, p$$

1) For $j = 1, \dots, p$, apply a **variable selection method** providing an estimate $\hat{S}^{(j)}$ of

$$\bar{S}^{(j)} = \left\{ r | \bar{\beta}_r^{(j)} \neq 0, r = 1, \dots, p, r \neq j \right\}$$

$\rightsquigarrow$ **Lasso regression** of $X^{(j)}$ versus $\left\{ X^{(r)}, r \neq j \right\}$ yields $\hat{\beta}^{(j)}$, which then yields the support estimate $\hat{S}^{(j)} = \left\{ r | \hat{\beta}^{(j)} \neq 0 \right\}$.

2) Build an estimate of the **graph structure**, using AND/OR rule:
Edge present between nodes j and k $\Leftrightarrow$ $k \in \hat{S}^{(j)}$ AND/OR $j \in \hat{S}^{(k)}$

Graphical LASSO

∗ We aim at **inferring GGM parameters** (μ, Σ) from n i.i.d realizations: $X_1, \ldots, X_n$ of $\mathcal{N}(\mu, \Sigma)$ with $\mu \in \mathbb{R}^p$ and $\Sigma \in \mathbb{R}^{p \times p}$ sdp. We introduce the sample mean and the empirical covariance matrix:

$$\hat{\mu} = n^{-1} \sum_{i=1}^{n} X_i, \quad S = n^{-1} \sum_{i=1}^{n} (X_i - \hat{\mu})(X_i - \hat{\mu})^\top.$$

Then, the negative Gaussian log-likelihood reads

$$-n^{-1} \ell(\Sigma^{-1} | X_1, \ldots, X_n) = -\log \det \Sigma^{-1} + \mathrm{trace}(S \Sigma^{-1}) + \mathrm{constant}.$$

∗ GLASSO is an estimator of Σ^{-1} based on the use of ℓ_1 penalty:

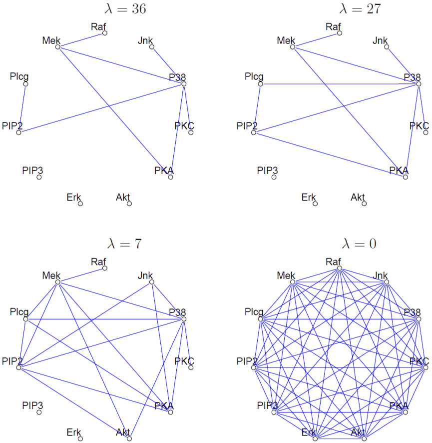
$$\hat{\Sigma}^{-1} = \mathrm{argmin}_{\Sigma^{-1} \succ 0} - \log \det \Sigma^{-1} + \mathrm{trace}(S \Sigma^{-1}) + \lambda \| \Sigma^{-1} \|_1$$

with $\| \Sigma^{-1} \|_1 = \sum_{j<k} |\Sigma^{-1}_{jk}|$, and $\lambda > 0$ regularization parameter.

∗ Convex optimization problem. Several solvers available.

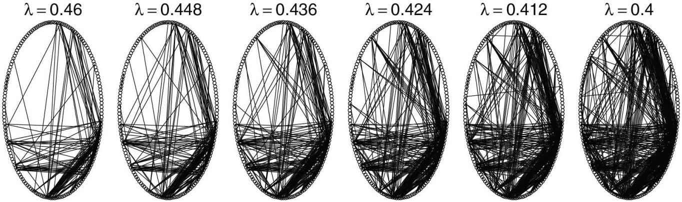
Example: ADMM algorithm.

Example



Four different GLASSO solutions for the flow-cytometry data with $p = 11$ proteins measured on $n = 7466$ cells [Sachs et al., 2003].

Example



Six different GLASSO solutions for the genomic dataset about riboflavin production with *Bacillus subtilis*, $p = 160$ and $n = 115$. [Meinshausen et al., 2010].

Whiteboard

Whiteboard

Whiteboard