Graphical Modeling of High-Throughput Metabolomics Data

Differential Networks pertaining to Alzheimer's Disease

Carel F.W. Peeters
Dept. of Epidemiology & Biostatistics
VU University medical center, Amsterdam
cf.peeters@vumc.nl

Brain Network Seminar
Dept. of Clinical Neurophysiology
VU University Medical Center, Amsterdam
September 12, 2016

Contributors

Wessel N. van Wieringen
Dept. of Epimiology & Biostatistics, VUMC
Dept. of Mathematics, VU University Amsterdam

Anders E. Bilgrau
Novo Nordisk
Dept. of Mathematical Sciences, Aalborg University

Mark A. van de Wiel
Dept. of Epimiology & Biostatistics, VUMC
Dept. of Mathematics, VU University Amsterdam

Contributors

Francien de Leeuw
Alzheimer Center, VUMC

Charlotte Teunissen
Neurology Laboratory, Dept. of Clinical Chemistry, VUMC

Wiesje van der Flier
Alzheimer Center, VUMC
Dept. of Epimiology & Biostatistics, VUMC

Contributors

Thomas Hankemeier
Division for Analytical Biosciences, Leiden University

Herman van Vlijmen
Dept. of Medicinal Chemistry, Leiden University
Molecular Sciences div., Janssen Pharmaceutica

Cornelia van Duijn
Netherlands Institute for Health Sciences
Dept. of Genetic Epimiology, Erasmus MC

Funding

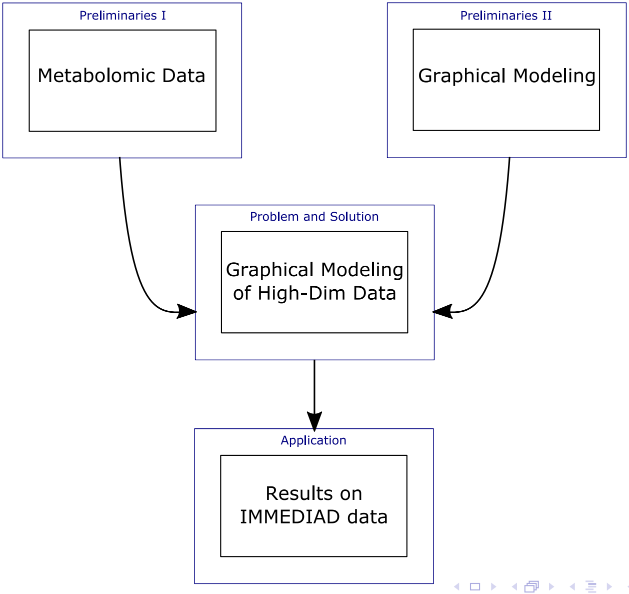
Methodological Developments

European Community Seventh Framework Programme (FP7): grant agreement No. FP7-269553

Data and Analyzes

Janssen Pharmaceutica Stellar funded project (IMMEDIAD): Stellar Neurodegeneration Collaboration Project, Call 2, No. 3

Outline


Preliminaries I
Metabolomic Data
Preliminaries II
Graphical Modeling
Problem and Solution
Graphical Modeling of High-Dim Data
Application
Results on IMMEDIAD data

Omics and omics data

-ome

A totality of some (molecular biological) sort

-omics

Collective quantification of some pool of molecular molecules

Metabolomics

The omics of small-molecule metabolites (of some organism)

Metabolite quantification

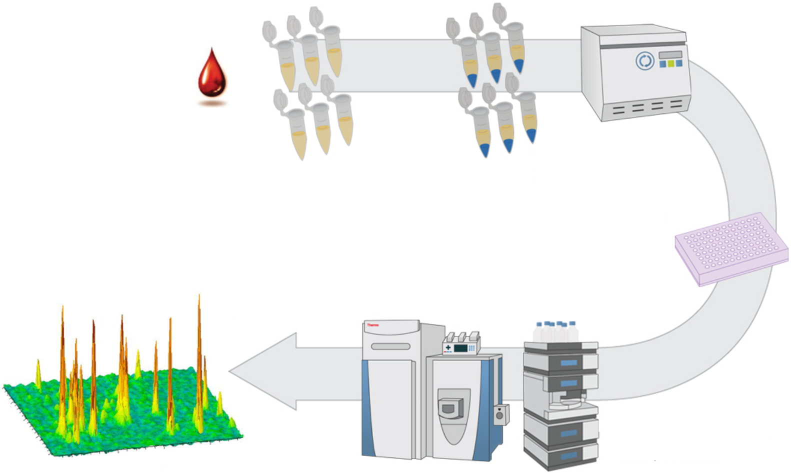


Illustration adapted from: http://planetorbitrap.com/untargeted-metabolomics#.Vzw6yfmLRaQ &

http://metabolomicsplatform.com/metabolomics-overview/

Challenge: Dimensionality metabolomic data

Variables

Observations	1	2	3	⋯	p
1	y_{11}	y_{12}	y_{13}	⋯	y_{1p}
2	y_{21}	y_{22}	y_{23}	⋯	y_{2p}
3	y_{31}	y_{32}	y_{33}	⋯	y_{3p}
4	y_{41}	y_{42}	y_{43}	⋯	y_{4p}
5	y_{51}	y_{52}	y_{53}	⋯	y_{5p}
⋮	⋮	⋮	⋮		⋮
n	y_{n1}	y_{n2}	y_{n3}	⋯	y_{np}

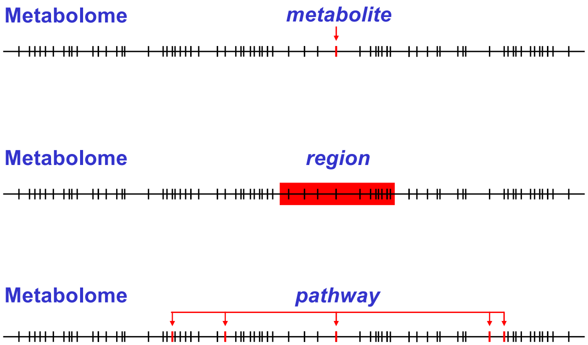
Regular data: $n > p$

Variables (features)

Observations	1	2	3	4	5	⋯	p
1	y_{11}	y_{12}	y_{13}	y_{14}	y_{15}	⋯	y_{1p}
2	y_{21}	y_{22}	y_{23}	y_{24}	y_{25}	⋯	y_{2p}
3	y_{31}	y_{32}	y_{33}	y_{34}	y_{35}	⋯	y_{3p}
⋮	⋮	⋮	⋮	⋮	⋮		⋮
n	y_{n1}	y_{n2}	y_{n3}	y_{n4}	y_{n5}	⋯	y_{np}

Metabolomic data: $p > n$ or $p >> n$

Unit of analysis



Motivation

Desire

- Consider data from multiple metabolomic platforms simultaneously
- Exploratively infer graph (reconstruct metabolic topology)
- Cope with high-dimensional situation
- Take juxtaposition AD vs. SC formally into account

Graphs

Representation

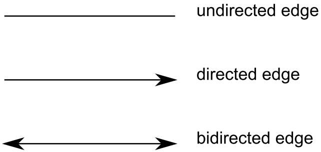
Pathways are represented by a *graph* (or *network*)

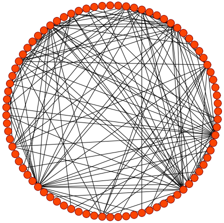
Vertices

◯ *Node* or *vertex* represents molecular feature

Edges

Edge or *arrow* represents some functional relation



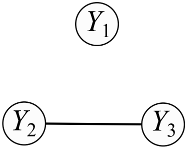


Correlation networks

Example

Three variables: Y_1, Y_2, and Y_3

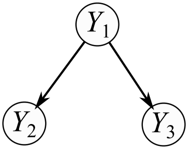
$$\mathrm{cor}(Y_1, Y_2) = 0$$
$$\mathrm{cor}(Y_1, Y_3) = 0$$
$$\mathrm{cor}(Y_2, Y_3) \neq 0$$



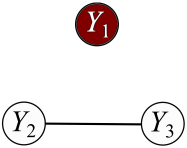
Marginal dependence

Undirected edge represents marginal dependence

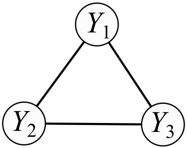
Interpretational danger



True mechanism

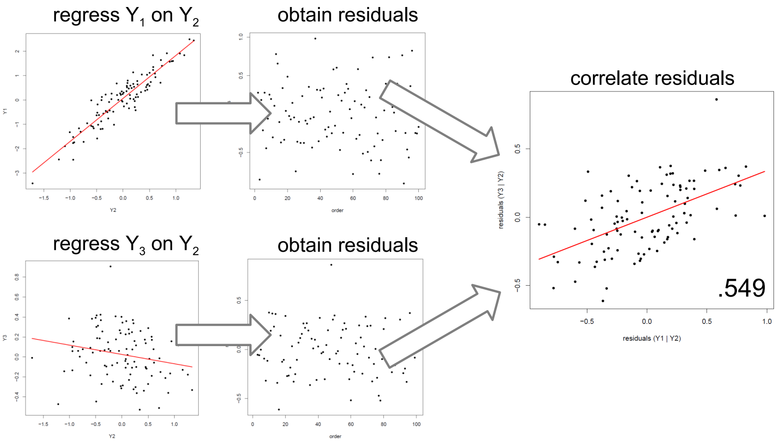


Not observing Y_1:
Spurious association

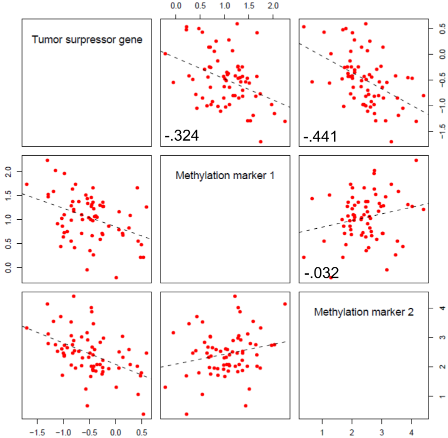


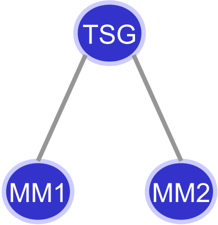
Observing Y_1:
Saturated graph

Solution: Conditioning



Solution: Conditioning





Conditional dependence

Partial correlation

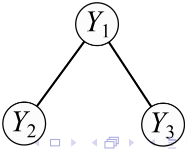
Measures degree of association between two random variables when controlling for third variables

Conditioned correlation

$$\mathrm{cor}(Y_1, Y_2 | Y_3)$$
$$\mathrm{cor}(Y_1, Y_3 | Y_2)$$
$$\mathrm{cor}(Y_2, Y_3 | Y_1)$$

If, e.g., $\mathrm{cor}(Y_2, Y_3|Y_1) = 0$, we say Y_2 and Y_3 are independent *given* Y_1

$$\mathrm{cor}(Y_1, Y_2|Y_3) \neq 0$$
$$\mathrm{cor}(Y_1, Y_3|Y_2) \neq 0$$
$$\mathrm{cor}(Y_2, Y_3|Y_1) = 0$$



Gaussian graphical modeling

Graphical modeling

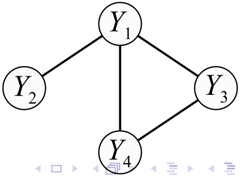
Class of models using graphs to express conditional (in)dependence relations between random variables

Gaussian setting

- Vertices: Correspond to random variables with normal distribution
- Edges: Correspond to the dependence structure
- Say $\mathbf{y} \sim \mathcal{N}_p(\mathbf{0}, \mathbf{\Sigma})$, and define $\mathbf{\Sigma}^{-1} \equiv \mathbf{\Omega}$. Then, for $a, b \in$ vertex set V, $a \neq b$

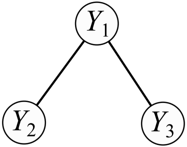
$$-\frac{\omega_{ab}}{\sqrt{\omega_{aa}\omega_{bb}}} = 0 \Longleftrightarrow \omega_{ab} = 0 \Longleftrightarrow a \perp\!\!\!\perp b | V \setminus \{a, b\} \Longleftrightarrow a \not\sim b$$

$$\begin{bmatrix} \omega_{11} & \omega_{12} & \omega_{13} & \omega_{14} \\ \omega_{21} & \omega_{22} & 0 & 0 \\ \omega_{31} & 0 & \omega_{33} & \omega_{34} \\ \omega_{41} & 0 & \omega_{43} & \omega_{44} \end{bmatrix}$$



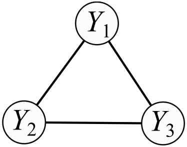
Gaussian graphical modeling

$$\mathbf{\Omega} = \begin{bmatrix} \omega_{11} & \omega_{12} & \omega_{13} \\ \omega_{21} & \omega_{22} & 0 \\ \omega_{31} & 0 & \omega_{33} \end{bmatrix}$$



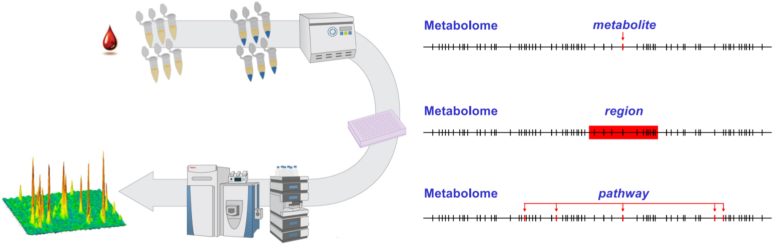
Conditional independence graph

$$\mathbf{\Sigma} = \mathbf{\Omega}^{-1} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix}$$



Correlation graph

Full-Swing back to Challenge: High-dimensional data



Problem

- Let $\mathbf{S} \equiv \hat{\boldsymbol{\Sigma}}$ denote the sample covariance matrix on $\mathbf{y}_i$
- When $p \approx n$ or $p > n$, $\mathbf{S}$ is ill-behaved or singular
- The precision $\mathbf{S}^{-1} \equiv \hat{\boldsymbol{\Omega}}$ is then undefined

Solution: ℓ_2-Penalization

Maximize

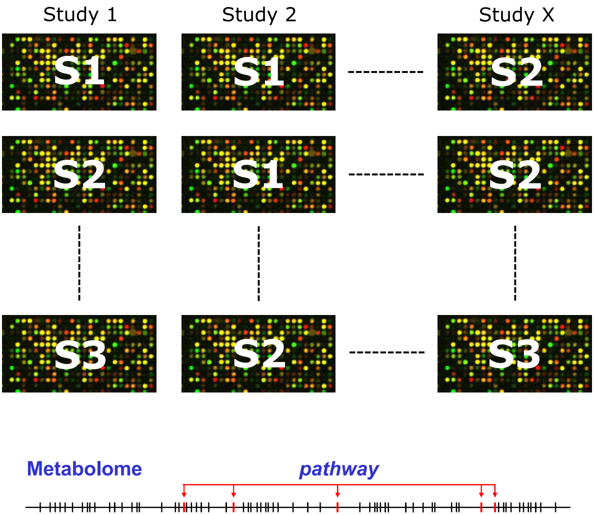
$$\underbrace{\ln|\mathbf{\Omega}| - \mathrm{tr}(\mathbf{S}\mathbf{\Omega})}_{\text{log}-\text{likelihood}} - \underbrace{\frac{\lambda}{2}\|\mathbf{\Omega} - \mathbf{T}\|_2^2}_{\ell_2-\text{penalty}}$$

- $\mathbf{T}$ denotes a p.d. symmetric target matrix
- $\lambda \in (0, \infty)$ denotes a penalty parameter

Analytic penalized ML estimator

$$\hat{\mathbf{\Omega}}(\lambda) = \left\{\left[\lambda \mathbf{I}_p + \frac{1}{4}(\mathbf{S} - \lambda\mathbf{T})^2\right]^{1/2} + \frac{1}{2}(\mathbf{S} - \lambda\mathbf{T})\right\}^{-1}$$

Multiple data classes



Targeted fused ridge estimation

Maximize

$$\mathcal{L}(\{\mathbf{\Omega}_g\};\{\mathbf{S}_g\}) - \frac{\lambda}{2}\sum_{g}\left\|\mathbf{\Omega}_g - \mathbf{T}_g\right\|_F^2 - \frac{\lambda_f}{4}\sum_{g_1,g_2}\left\|(\mathbf{\Omega}_{g_1} - \mathbf{T}_{g_1}) - (\mathbf{\Omega}_{g_2} - \mathbf{T}_{g_2})\right\|_F^2$$

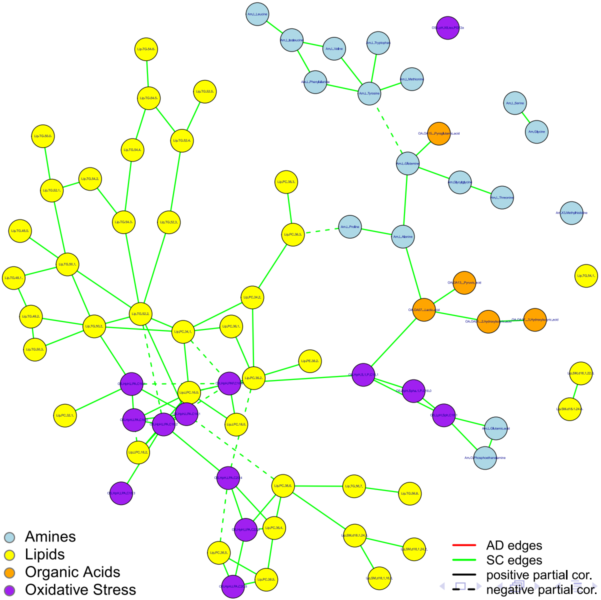
- $\mathbf{T}_g$ indicate class-specific target matrices
- $\lambda \in (0, \infty)$ denotes the ridge penalty parameter
- $\lambda_f \in [0, \infty)$ denotes the fusion penalty parameter

Approach

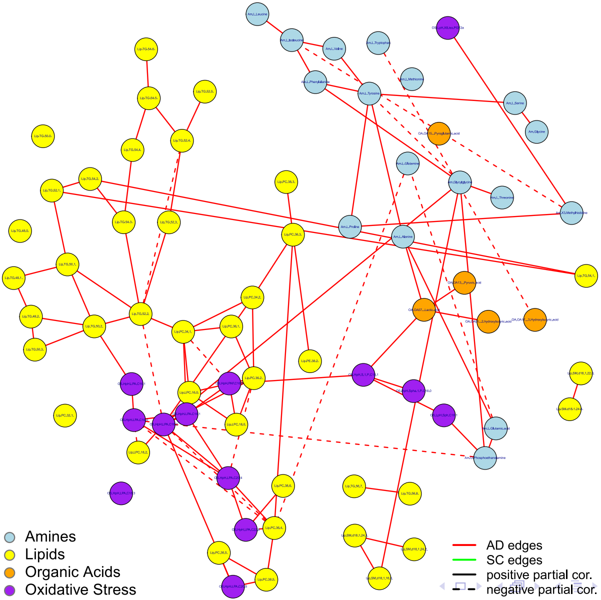
Steps

- Class precision matrices were estimated by ridge fusion
- Optimal penalty parameters were determined by cross-validation
- Class-specific target matrices expressed little prior information
- Support determination on the basis of thresholding
  - retain strongest edges: corresponds to posterior probability thresholding
- Subsequent focus on differential network visualization and analysis

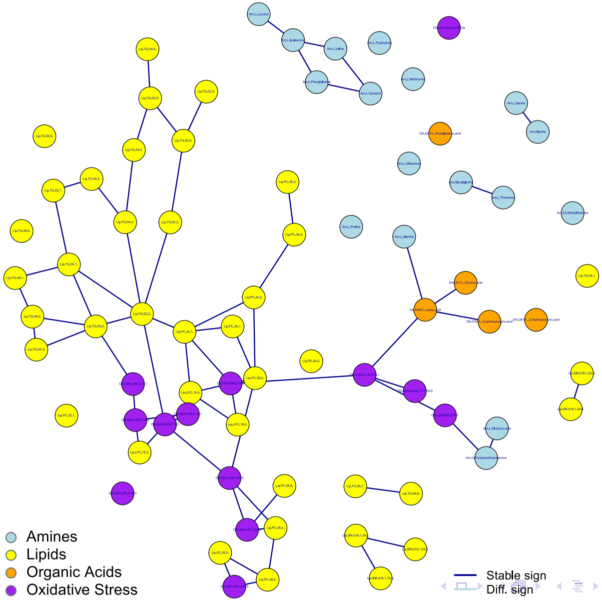
SC Connections



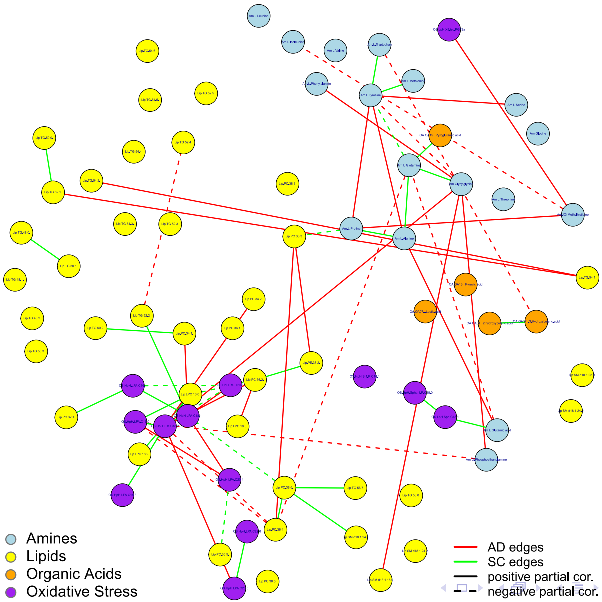
AD Connections



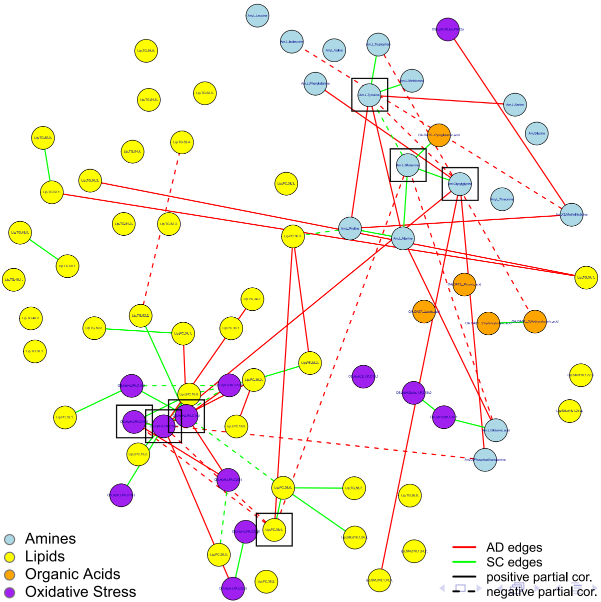
Shared Connections



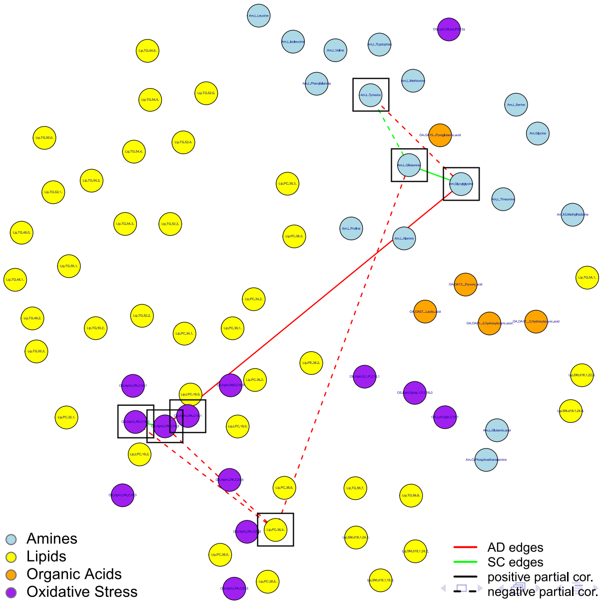
Differential Connections



Differential Connections



Differential Connections



RAGS2RIDGES

The Hoff

Manual/Download

- Peeters, C.F.W., Bilgrau, A.E., & van Wieringen, W.N. (2016). "rags2ridges: Ridge Estimation of Precision Matrices from High-Dimensional Data". R package, version 2.1.1 URL: https://cran.r-project.org/package=rags2ridges.

Theory/Methodology

- Peeters*, C.F.W., Bilgrau*, A.E., Eriksen, P.S., Boegsted, M., & van Wieringen, W.N. (2015). "Targeted Fused Ridge Estimation of Inverse Covariance Matrices from Multiple High-Dimensional Data Classes". arXiv:1509.07982v1 [stat.ME].
- Peeters, C.F.W., van Wieringen, W.N., & van de Wiel, M.A. (in preparation). "Directed Cyclic Mixed Graph Modeling for High-Dimensional Genomic Data Integration".
- van Wieringen, W.N. & Peeters, C.F.W. (2016). "Ridge Estimation of Inverse Covariance Matrices from High-Dimensional Data". Computational Statistics & Data Analysis, 103: 284-303. arXiv:1403.0904v3 [stat.ME].

Software

- Peeters, C.F.W., van de Wiel, M.A., & van Wieringen, W.N. (2016) "The Spectral Condition Number Plot for Regularization Parameter Determination". arXiv:1608.04123v1 [stat.CO].
- van Wieringen, W.N. & Peeters, C.F.W. (2015). "Application of a New Ridge Estimator of the Inverse Covariance Matrix to the Reconstruction of Gene-Gene Interaction Networks". In: di Serio, C., Lio, P., Nonis, A., and Tagliaferri, R. (Eds.) 'Computational Intelligence Methods for Bioinformatics and Biostatistics'. Lecture Notes in Computer Science, vol. 8623. Springer, pp. 170–179.

Explaining the inverse

The scalar inverse

- Let a denote a number (excluding 0)
- The inverse is then the number b such that $a \times b = 1$
- Clearly, $b = \frac{1}{a}$

Matrix

A matrix is a generalization of a number, an array of numbers

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{bmatrix}$$

Explaining the inverse

The Matrix Inverse

Consider the matrix $\mathbf{A}$. Its inverse $\mathbf{B} = \mathbf{A}^{-1}$ is defined such that

$$\mathbf{A}\mathbf{B} = \mathbf{I},$$

where

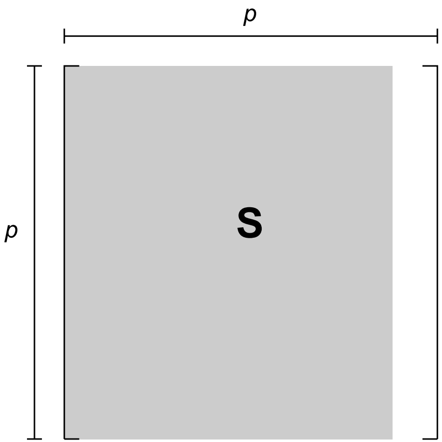
$$\mathbf{I} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

Solution

$$\mathbf{A}^{-1} = \left[\begin{array}{cc} \mathbf{A}_{11}^{-1} + \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{Q}^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{Q}^{-1} \\ -\mathbf{Q}^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{Q}^{-1} \end{array} \right],$$

with $\mathbf{Q} = \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}$ denoting the Schur complement.

Singularity



Recall: ridge estimation

Analytic maximizing argument

$$\hat{\mathbf{\Omega}}(\lambda) = \left\{ \left[\lambda \mathbf{I}_p + \frac{1}{4}(\mathbf{S} - \lambda \mathbf{T})^2 \right]^{1/2} + \frac{1}{2}(\mathbf{S} - \lambda \mathbf{T}) \right\}^{-1}$$

Properties

Behavior

i. $\hat{\mathbf{\Omega}}(\lambda) \succ 0,$ for all $\lambda \in (0, \infty)$;

ii. $\lim_{\lambda \to 0^+} \hat{\mathbf{\Omega}}(\lambda) = \mathbf{S}^{-1}$ if $p < n$;

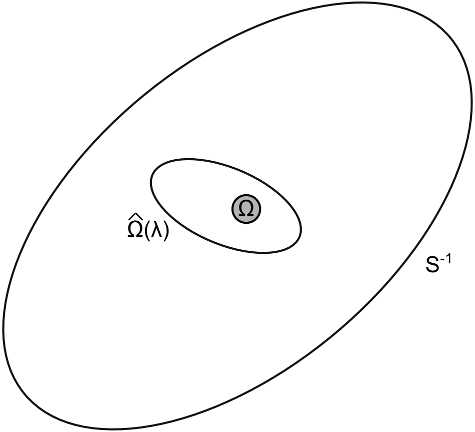
iii. $\lim_{\lambda \to \infty} \hat{\mathbf{\Omega}}(\lambda) = \mathbf{T}$.

Consistency

i. $\lim_{n \to \infty} \mathbb{E}\left[\hat{\mathbf{\Omega}}_n(\lambda_n)\right] \longrightarrow \lim_{n \to \infty} \mathbb{E}\left(\mathbf{S}_n^{-1}\right) = \mathbf{\Omega}$;

ii. $\lim_{n \to \infty} \mathbb{E}\left(\|\hat{\mathbf{\Omega}}_n(\lambda_n) - \mathbf{\Omega}\|_F^2\right) = 0$.

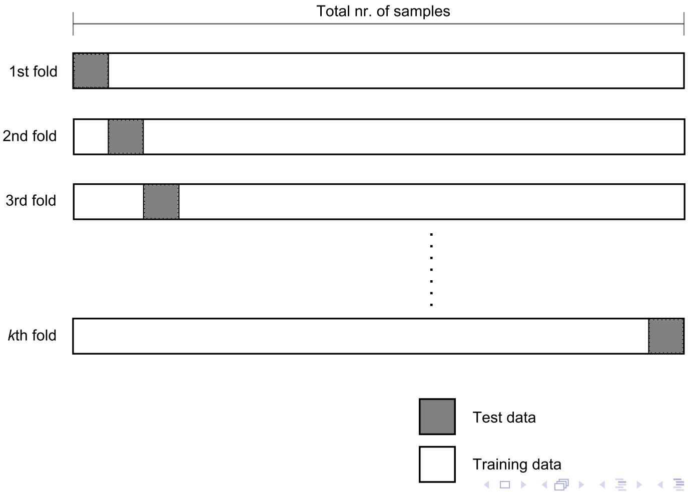
Visual explanation



Choosing the penalty value

K-fold cross-validation (CV)

Single iteration of K-fold CV



Choosing the penalty value

K-fold CV score

$$\varphi^K(\lambda) = \sum_{k=1}^{K} n_k \left\{ - \ln |\hat{\mathbf{\Omega}}(\lambda)_{-k}| + \mathrm{tr}[\hat{\mathbf{\Omega}}(\lambda)_{-k}\mathbf{S}_k] \right\},$$

n_k is the size of subset k, for $k = 1, \ldots, K$ disjoint subsets;
$\mathbf{S}_k$ denotes the sample covariance matrix on *kth* test set;
$\hat{\mathbf{\Omega}}(\lambda)_{-k}$ denotes the estimated regularized precision matrix on kth training set

Highest predictive accuracy

Choose $n_k = 1$, such that $K = n$ (known as leave-one-out CV - LOOCV)

Problem

K-fold CV is computationally demanding for large p and/or large K

Solution

Computationally efficient approximate LOOCV score

Support determination

Scaling

$\hat{\mathbf{P}}(\lambda)$: Regularized precision estimate scaled to partial correlation form

Assume

Nonredundant off-diagonal partial correlation coefficients (indexed by $j < j'$) follow a mixture distribution:

$$f\left\{[\hat{\mathbf{P}}(\lambda^*)]_{jj'}\right\} = \eta_0 f_0\left\{[\hat{\mathbf{P}}(\lambda^*)]_{jj'}; \kappa\right\} + (1-\eta_0) f_{\mathcal{E}}\left\{[\hat{\mathbf{P}}(\lambda^*)]_{jj'}\right\}$$

- $\eta_0 \in [0, 1]$ is the mixture weight
- $f_0\{\cdot\}$ denotes the distribution of a null-edge
- $f_{\mathcal{E}}\{\cdot\}$ denotes the distribution of a present edge

Determine

$$P\left(Y_j \neq Y_{j'} | [\hat{\mathbf{P}}(\lambda^*)]_{jj'}\right)$$

Situation

Data

- G classes of $(n_g \times p)$-dimensional data
- Classes defined by data sets and/or (subtypes of) diseases

Assumption

Precision matrices of constituent classes chiefly share the same structure but potentially differ in a number of locations of interest

Desire

Integrative or meta-analytic Gaussian graphical modeling

Targeted fused ridge estimation: General Formulation

Maximize

$$\underbrace{\mathcal{L}(\{\mathbf{\Omega}_g\};\{\mathbf{S}_g\})}_{\text{log}-\text{likelihood}} - \sum_g \underbrace{\frac{\lambda_{gg}}{2}\left\|\mathbf{\Omega}_g - \mathbf{T}_g\right\|_F^2}_{\text{ridge}-\text{penalty}} - \sum_{g_1,g_2} \underbrace{\frac{\lambda_{g_1 g_2}}{4}\left\|(\mathbf{\Omega}_{g_1} - \mathbf{T}_{g_1}) - (\mathbf{\Omega}_{g_2} - \mathbf{T}_{g_2})\right\|_F^2}_{\text{fusion}-\text{penalty}}$$

- $\mathbf{T}_g$ indicate class-specific target matrices
- $\lambda_{gg} \in (0, \infty)$ denote class-specific ridge penalty parameters
- $\lambda_{g_1 g_2} \in [0, \infty)$ denote pair-specific fusion penalty parameters, $\lambda_{g_1 g_2} = \lambda_{g_2 g_1}$

Penalty matrix

All penalties can be collected into a non-negative symmetric matrix $\mathbf{\Lambda} = [\lambda_{g_1 g_2}]$

Targeted fused ridge estimation

Maximizing argument for class g_0

$$\hat{\mathbf{\Omega}}_{g_0}(\mathbf{\Lambda}, \{\mathbf{\Omega}_g\}_{g \neq g_0}) = \left\{ \left[\bar{\lambda}_{g_0} \mathbf{I}_p + \frac{1}{4} (\bar{\mathbf{S}}_{g_0} - \bar{\lambda}_{g_0} \mathbf{T}_{g_0})^2 \right]^{1/2} + \frac{1}{2} (\bar{\mathbf{S}}_{g_0} - \bar{\lambda}_{g_0} \mathbf{T}_{g_0}) \right\}^{-1},$$

where

$$\bar{\mathbf{S}}_{g_0} = \mathbf{S}_{g_0} - \sum_{g \neq g_0} \frac{\lambda_{gg_0}}{n_{g_0}} (\mathbf{\Omega}_g - \mathbf{T}_g), \quad \text{and} \quad \bar{\lambda}_{g_0} = \frac{\sum_g \lambda_{gg_0}}{n_{g_0}}$$

Properties

Behavior

i. $\hat{\mathbf{\Omega}}_g \succ \mathbf{0}$ for all $\lambda_{gg} \in (0, \infty)$;

ii. $\lim_{\lambda_{gg} \to 0^+} \hat{\mathbf{\Omega}}_g = \mathbf{S}_g^{-1}$ if $\sum_{g' \neq g} \lambda_{gg'} = 0$ and $p \leq n_g$;

iii. $\lim_{\lambda_{gg} \to \infty} \hat{\mathbf{\Omega}}_g = \mathbf{T}_g$ if $\lambda_{gg'} < \infty$ for all $g' \neq g$;

iv. $\lim_{\lambda_{g_1 g_2} \to \infty} (\hat{\mathbf{\Omega}}_{g_1} - \mathbf{T}_{g_1}) = \lim_{\lambda_{g_1 g_2} \to \infty} (\hat{\mathbf{\Omega}}_{g_2} - \mathbf{T}_{g_2})$ if $\lambda_{g_1' g_2'} < \infty$ for all $\{g_1', g_2'\} \neq \{g_1, g_2\}$.

Block coordinate ascent

Input:
Sufficient data: $(\mathbf{S}_1, n_1), \ldots, (\mathbf{S}_G, n_G)$
Penalty matrix: $\mathbf{\Lambda}$
Convergence criterion: $\varepsilon > 0$
Output:
Estimates: $\hat{\mathbf{\Omega}}_1, \ldots, \hat{\mathbf{\Omega}}_G$
procedure RIDGEP.FUSED$(\mathbf{S}_1, \ldots, \mathbf{S}_G, n_1, \ldots, n_G, \mathbf{\Lambda}, \varepsilon)$
    Initialize: $\hat{\mathbf{\Omega}}_g^{(0)}$ for all g.
    for $c = 1, 2, 3, \ldots$ **do**
        for $g = 1, 2, \ldots, G$ **do**
            Update $\hat{\mathbf{\Omega}}_g^{(c)} := \hat{\mathbf{\Omega}}_g\left(\mathbf{\Lambda}, \hat{\mathbf{\Omega}}_1^{(c)}, \ldots, \hat{\mathbf{\Omega}}_{g-1}^{(c)}, \hat{\mathbf{\Omega}}_{g+1}^{(c-1)}, \ldots, \hat{\mathbf{\Omega}}_G^{(c-1)}\right)$
        end for
        if $\max_g \left\{ \frac{\|\hat{\mathbf{\Omega}}_g^{(c)} - \hat{\mathbf{\Omega}}_g^{(c-1)}\|_F^2}{\|\hat{\mathbf{\Omega}}_g^{(c)}\|_F^2} \right\} < \varepsilon$ **then**
            return $\left(\hat{\mathbf{\Omega}}_1^{(c)}, \ldots, \hat{\mathbf{\Omega}}_G^{(c)}\right)$
        end if
    end for
end procedure