Package: latentgraph (via r-universe)

August 30, 2024

Type Package

Title Graphical Models with Latent Variables

Version 1.1

Date 2020-11-29

Author Yanxin Jin, Samantha Yang, Kean Ming Tan

Maintainer Yanxin Jin <yanxinj@umich.edu>

Description Three methods are provided to estimate graphical models with latent variables: (1) Jin, Y., Ning, Y., and Tan, K. M. (2020) (preprint available); (2) Chandrasekaran, V., Parrilo, P. A. & Willsky, A. S. (2012) <doi:10.1214/11-AOS949>; (3) Tan, K. M., Ning, Y., Witten, D. M. & Liu, H. (2016) <doi:10.1093/biomet/asw050>.

License GPL-3

Imports Rcpp, pracma, glmnet, MASS, stats

LinkingTo Rcpp, RcppArmadillo

NeedsCompilation yes

Repository CRAN

Date/Publication 2020-12-10 09:10:09 UTC

Contents

corlatent 2
generate_Gaussian 3
lvglasso 5
semilatent 6

Index **8**

corlatent *Graphical Models with Latent Variables and Correlated Replicates*

Description

Estimate graphical models with latent variables and correlated replicates using the method in Jin et al. (2020).

Usage

```
corlatent(data, accuracy, n, R, p, lambda1, lambda2, lambda3, distribution = "Gaussian",
rule = "AND")
```

Arguments

data	data set. Can be a matrix, list, array, or data frame. If the data set is a matrix, it should have nR rows and p columns. This matrix is formed by stacking n matrices, and each matrix has R rows and p columns. If the data set is a data frame, the dimention and structure are the same as the matrix. If the data set is an array, its dimention is (R, p, n). If the data set is a list, it should have n elements and each element is a matrix with R rows and p columns.
accuracy	the threshhold where algorithm stops. The algorithm stops when the difference between estimaters of the $(k-1)$th iteration and the kth iteration is smaller than the value of accuracy.
n	the number of observations.
R	the number of replicates for each observation.
p	the number of observed variables.
lambda1	tuning parameter that encourages estimated graph to be sparse.
lambda2	tuning parameter that models the effects of correlated replicates. Usually set to be equal to lambda1.
lambda3	tuning parameter that encourages the latent effect to be piecewise constants.
distribution	For a data set with Gaussian distribution, use "Gaussian"; For a data set with Ising distribution, use "Ising". Default is "Gaussian".
rule	rules to combine matrices that encode the conditional dependence relationships between sets of two observed variables. Options are "AND" and "OR". Default is "AND".

Details

The corlatent method has two assumptions. Assumption 1 states that the R replicates are assumed to follow a one-lag vector autoregressive model, conditioned on the latent variables. Assumption 2 states that the latent variables are piecewise constant across replicates. Based on these two assumptions, the method solve the following problem for $1 \leq j \leq p$.

$$\min_{\theta_{j,-j},\alpha_j,\Delta_j} \{-\frac{1}{nR} l(\theta_{j,-j},\alpha_j,\Delta_j) + \lambda\|\theta_{j,-j}\|_1 + \beta\|\alpha_j\|_1 + \gamma\|(I_n \otimes C)\Delta_j\|_1\},$$

where $l(\theta_{j,-j}, \alpha_j, \Delta_j)$ is the log likelihood function, $\theta_{j,-j}$ encodes the conditional dependence relationships between jth observed variable and the other observed variables, α_j models the correlation among replicates, Δ_j encodes the latent effect, λ, β, γ are the tuning parameters, I_n is an n-dimensional identity matrix and C is the discrete first derivative matrix where the ith and $(i+1)$th column of every ith row are -1 and 1, respectively. This method aims at modeling exponential family graphical models with correlated replicates and latent variables.

Value

`omega`	a matrix that encodes the conditional dependence relationships between sets of two observed variables
`theta`	the adjacency matrix with 0 and 1 encoding conditional independence and dependence between sets of two observed variables, respectively
`penalties`	the penalty values

References

Jin, Y., Ning, Y., and Tan, K. M. (2020), 'Exponential Family Graphical Models with Correlated Replicates and Unmeasured Confounders', preprint available.

Examples

```
# Gaussian distribution with "AND" rule
n <- 20
R <- 10
p <- 5
l <- 2
s <- 2
seed <- 1

data <- generate_Gaussian(n, R, p, l, s, sparsityA = 0.95, sparsityobserved = 0.9,
sparsitylatent = 0.2, lwb = 0.3, upb = 0.3, seed)$X

result <- corlatent(data, accuracy = 1e-6, n, R, p,lambda1 = 0.1, lambda2 = 0.1,
lambda3 = 1e+5,distribution = "Gaussian", rule = "AND")
```

`generate_Gaussian` *Generate a Gaussian distributed data set*

Description

This function will generate a Gaussian distributed data set with latent variables and correlated replicates.

Usage

```
generate_Gaussian(n, R, p, l, s, sparsityA, sparsityobserved, sparsitylatent, lwb,
upb, seed)
```

Arguments

n	the number of observations.
R	the number of replicates.
p	the number of observed variables.
l	the number of latent variables.
s	latent effects are generated as s piecewise constant across replicates. The number s should be a factor of R.
sparsityA	proportion of the number of zeros in the transition matrix A. Must be a number from 0 to 1.
sparsityobserved	proportion of the number of zeros in the inverse covariance of the observed variables. Must be a number from 0 to 1.
sparsitylatent	proportion of the number of zeros in the inverse covariances among latent variables and between observed variables and latent variables. Must be a number from 0 to 1.
lwb	lower bound for the elements in the inverse covariance matrix.
upb	upper bound for the elements in the inverse covariance matrix.
seed	the seed for the random number generator.

Details

This function aims to generate a Gaussian distributed data set with latent variables and correlated replicates. For each observation, the latent variables are piecewise constant across replicates, and conditioned on the latent variables, the replicates follow a one-lag vector autoregressive model, where the transition matrix A is sparse with non-zero elements set equal to 0.3. The matrix Σ is the covariance matrix for the observed variables X and the latent variables U, and we partition Σ into matrices that quantify the relationships among the observed variables (Σ_{XX}), between the observed variables and latent variables (Σ_{XU} or Σ_{UX}), and of the latent variables (Σ_{UU}). In general, the data is generated with:

$$X_{i1}|U_{i1} \sim N_p(\Sigma_{XU}\Sigma_{UU}^{-1}U_{i1}, \Sigma_{XX} - \Sigma_{XU}\Sigma_{UU}^{-1}\Sigma_{UX}),$$

$$X_{it}|X_{i(t-1)}, U_{it} \sim N_p(AX_{i(t-1)} + \Sigma_{XU}\Sigma_{UU}^{-1}U_{it}, \Sigma_{XX} - \Sigma_{XU}\Sigma_{UU}^{-1}\Sigma_{UX}),$$

where $1 \leq i \leq n$ and $1 \leq t \leq R$.

Value

X	the generated data, which is a list with n elements and each element is a matrix with R rows and p columns
truegraph	a matrix that encodes the conditional dependence relationships between sets of two observed variables

References

Jin, Y., Ning, Y., and Tan, K. M. (2020), 'Exponential Family Graphical Models with Correlated Replicates and Unmeasured Confounders', preprint available.

Examples

```
data <- generate_Gaussian(n = 50, R = 20, p = 30, l = 2, s = 2, sparsityA = 0.95,
sparsityobserved = 0.9, sparsitylatent = 0.2, lwb = 0.3, upb = 0.3, seed = 1)
```

`lvglasso` *Estimate Gaussian Graphical Models with Latent Variables*

Description

Estimate Gaussian graphical models with latent variables using the method in Chandrasekaran et al. (2012).

Usage

```
lvglasso(data, n, p, lambda1, lambda2, rule = "AND")
```

Arguments

`data`	data set, can be a matrix or data frame with n rows and p columns.
`n`	the number of observations.
`p`	the number of observed variables.
`lambda1`	tuning parameter that encourages estimated graph to be sparse. Lambda1 is proportional to lambda2.
`lambda2`	tuning parameter that encourages the matrix $K_{O,H}(K_H)^{-1}K_{H,O}$ to be low rank, where K_H and $K_{O,H}$ quantify the dependecies among the latent variables and between the observed variables and latent variables, respectively. The matrix $K_{O,H}(K_H)^{-1}K_{H,O}$ summarizes the impact of marginalization over the latent variables.
`rule`	rules to combine the inverse covariance matrices. Options are "AND" and "OR". Default is "AND".

Details

The lvglasso method assumes that all the variables, both observed and latent, are jointly Gaussian, and specifies the conditional distribution of observed variables on the latent variables by a graphical model. Under the high-dimentional setting, this method provides consistent estimators for the conditional graphical model of observed variables conditioned on latent variables.

Value

`omega`	a matrix that encodes the conditional dependence relationships between sets of two observed variables
`theta`	the adjacency matrix with 0 and 1 encoding conditional independence and dependence between sets of two observed variables, respectively
`penalties`	the penalty values

References

Chandrasekaran, V., Parrilo, P. A. & Willsky, A. S. (2012), 'Latent variable graphical model selection via convex optimization', Ann. Statist. 40(4), 1935–1967.

Examples

```
#Gaussian distribution with "AND" rule
n <- 50
R <- 20
p <- 30
l <- 2
s <- 2
data <- generate_Gaussian(n, R, p, l, s, sparsityA = 0.95, sparsityobserved = 0.9,
sparsitylatent = 0.2, lwb = 0.3, upb = 0.3, seed = 1)$X

result <- lvglasso(data, n, p, lambda1 = 0.222, lambda2 = 0.1*0.222, rule = "AND")
```

`semilatent` *Estimate Graphical Models with Latent Variables and Replicates*

Description

Estimate graphical models with latent variables and replicates using the method in Tan et al. (2016).

Usage

```
semilatent(data, n, R, p, lambda, distribution = "Gaussian", rule = "AND")
```

Arguments

`data`	data set. Can be a matrix, list, array, or data frame. If the data set is a matrix, it should have nR rows and p columns. This matrix is formed by stacking n matrices, and each matrix has R rows and p columns. If the data set is a data frame, the dimention and structure are the same as the matrix. If the data set is an array, its dimention is (R, p, n). If the data set is a list, it should have n elements and each element is a matrix with R rows and p columns.
`n`	the number of observations.
`R`	the number of replicates for each observation.
`p`	the number of observed variables.
`lambda`	tuning parameter that encourages estimated graph to be sparse.
`distribution`	For a data set with Gaussian distribution, use "Gaussian"; For a data set with Ising distribution, use "Ising". Default is "Gaussian".
`rule`	rules to combine matrices that encode the conditional dependence relationships between sets of two observed variables. Options are "AND" and "OR". Default is "AND".

Details

The semilatent method has two assumptions. The first one states that the latent variables are constant across replicates. Assumption 2 states that given the latent variables, the replicates are mutually independent. With these two assumptions, the method seeks to solve the following problem for $1 \leq j \leq p$.

$$\min_{\beta_{j,O/j}} \{l_j(\beta_{j,O/j}) + \lambda \|\beta_{j,O/j}\|_1\},$$

where $l_j(\beta_{j,O/j})$ is a nuisance-free loss function, $\beta_{j,O/j}$ is a parameter that represents the conditional dependence relationships between jth observed variable and the other observed variables, and λ is a tuning parameter. This method aims at modeling semiparametric exponential family graphical model with latent variables and replicates.

Value

`omega`	a matrix that encodes the conditional dependence relationships between sets of two observed variables
`theta`	the adjacency matrix with 0 and 1 encoding conditional independence and dependence between sets of two observed variables, respectively
`penalty`	the penalty value

References

Tan, K. M., Ning, Y., Witten, D. M. & Liu, H. (2016), 'Replicates in high dimensions, with applications to latent variable graphical models', Biometrika 103(4), 761–777.

Examples

```
#semilatent Gaussian with "AND" rule
n <- 50
R <- 20
p <- 30
seed <- 1
l <- 2
s <- 2
data <- generate_Gaussian(n, R, p, l, s, sparsityA = 0.95, sparsityobserved = 0.9,
sparsitylatent = 0.2, lwb = 0.3, upb = 0.3, seed)$X

result <- semilatent(data, n, R, p, lambda = 0.1,distribution = "Gaussian",
rule = "AND")
```

Index

corlatent, 2

generate_Gaussian, 3

lvglasso, 5

semilatent, 6