Package 'gSEM'

January 11, 2016

Type Package

Title Semi-Supervised Generalized Structural Equation Modeling

Description Conducts a semi-gSEM statistical analysis (semi-supervised generalized structural equation modeling) on a data frame of coincident observations of multiple predictive or intermediate variables and a final continuous, outcome variable, via two functions sgSEMp1() and sgSEMp2(), representing fittings based on two statistical principles. Principle 1 determines all sensible univariate relationships in the spirit of the Markovian process. The relationship between each pair of variables, including predictors and the final outcome variable, is determined with the Markovian property that the value of the current predictor is sufficient in relating to the next level variable, i.e., the relationship is independent of the specific value of the preceding-level variables to the current predictor, given the current value. Principle 2 resembles the multiple regression principle in the way multiple predictors are considered simultaneously. Specifically, the relationship of the first-level predictors (such as Time and irradiance etc) to the outcome variable (such as, module degradation or yellowing) is fit by a supervised additive model. Then each significant intermediate variable is taken as the new outcome variable and the other variables (except the final outcome variable) as the predictors in investigating the next-level multivariate relationship by a supervised additive model. This fitting process is continued until all sensible models are investigated.

Version 0.4.3.4

Author Junheng Ma <jxm216@case.edu>, Nicholas Wheeler <nrw16@case.edu>, Yifan Xu <ethan.yifanxu@case.edu>, Wenyu Du <wxd97@case.edu>, Abdulkerim Gok <axg515@case.edu>, Jiayang Sun <jiayang.sun@case.edu>

Maintainer Junheng Ma <jxm216@case.edu>

Depends R (>= 2.14.0)

Imports knitr, MASS, htmlwidgets, DiagrammeR

License GPL (>= 2)

LazyData true

NeedsCompilation no

RoxygenNote 5.0.1

Repository CRAN

Date/Publication 2016-01-11 21:18:40

R topics documented:

`acrylic` *A data frame of an acrylic degradation experiment*

Description

This data set is a study of photodegradation of acrylic polymers. In this work, polymeric samples were exposed to different levels of light exposures and resulting optical changes were determined through optical spectroscopy. *IrradTot* (total applied irradiance) is the main predictor (or stressor in this data) and *YI* (yellowness index) is the performance level response. The other columns in the data set (*IAD1*, *IAD2*, *IAD2p*, and *IAD3*) are induced absorbency to dose values extracted from optical absorbency spectra as single metrics and used as intermediate unit level response variables in the gSEM analysis.

Format

A 357 by 6 data frame of continuous variables.

`genInit` *Generate initial values for nls function*

Description

Generate multiple initial vectors for the nls function in sgSEMp1().

Usage

```
genInit(bounds = list(a1 = c(-3, 3), a2 = c(-3, 3), a3 = c(-3, 3)), k = 50)
```

Arguments

bounds	a list of three vectors of length = 2. Each vector gives the upper and lower limits of an interval from which the initial values are randomly generated. The default values list(a1 = c(-3, 3), a2 = c(-3, 3), a3 = c(-3, 3)) sets limits of all three initial values to be (-3, 3).
k	a positive integer (default = 50). The number of initial vectors to generate.

Details

Currently the non-linearizable function included in sgSEMp1() is y = a + b * exp(c * x), where a, b and c are coefficients to be estimated. Thus, an initial vector contains three values. The random initial values are generated by a uniform distribution between the bounds.

Value

A data frame. Each column corresponds to a coefficient. Each row corresponds to a random initial vector.

Examples

```
genInit(list(a1 = c(0,2), a2 = c(4,5), a3 = c(-1, -0.5)), k = 20 )
```

---

gSEM *Semi-supervised Generalized Structure Equation Modeling*

Description

Conducts a semi-gSEM statistical analysis (semi-supervised generalized structural equation modeling) on a data frame of coincident observations of multiple predictive or intermediate variables and a final continuous, outcome variable, via two functions sgSEMp1() and sgSEMp2(), representing fittings based on two statistical principles. Principle 1 determines all sensible univariate relationships in the spirit of the Markovian process. The relationship between each pair of variables, including predictors and the final outcome variable, is determined with the Markovian property that the value of the current predictor is sufficient in relating to the next level variable, i.e., the relationship is independent of the specific value of the preceding-level variables to the current predictor, given the current value. Principle 2 resembles the multiple regression principle in the way multiple predictors are considered simultaneously. Specifically, the relationship of the first-level predictors (such as Time and irradiance etc) to the outcome variable (such as, module degradation or yellowing) is fit by a supervised additive model. Then each significant intermediate variable is taken as the new outcome variable and the other variables (except the final outcome variable) as the predictors in investigating the next-level multivariate relationship by a supervised additive model. This fitting process is continued until all sensible models are investigated.

References

1. Bruckman, Laura S., Nicholas R. Wheeler, Junheng Ma, Ethan Wang, Carl K. Wang, Ivan Chou, Jiayang Sun, and Roger H. French. "Statistical and Domain Analytics Applied to PV Module Lifetime and Degradation Science." IEEE Access 1 (2013): 384-403. doi:10.1109/ACCESS.2013.2267611

2. Bruckman, Laura S., Nicholas R. Wheeler, Ian V. Kidd, Jiayang Sun, and Roger H. French. "Photovoltaic Lifetime and Degradation Science Statistical Pathway Development: Acrylic Degradation." In SPIE Solar Energy+ Technology, 8825:88250D-8. International Society for Optics and Photonics, 2013. doi:10.1117/12.2024717

See Also

sgSEMp1() for implementing principle 1 and sgSEMp2() for implementing principle 2.

`path` *Extract Path Coefficients*

Description

Extract and display an equation of a pairwise path between two variables.

Usage

```
path(x, from, to, round = 3)
```

Arguments

`x`	object of class "sgSEMp1", which is the return value of function sgSEMp1().
`from`	character string. Name of the predictor.
`to`	character string. Name of the response variable.
`round`	a positive integer. The coefficients are rounded to this decimal place.

Details

Extract the "best" model between any two variables. The model name and the model equation are printed on screen. The model coefficients, as well as the model R object are also returned.

Value

A list of the following items: 1) model: the best fitted model, 2) model.print: a character string of the model equation and 3) coefs: Model coefficients vector.

Examples

```
##' ## Load the sample acrylic data set
data(acrylic)

## Run semi-gSEM principle one
ans <- sgSEMp1(acrylic, predictor = "IrradTot", response = "YI")

## Extract relations between IrradTot and IAD2
cf <- path(ans, from = "IrradTot", to = "IAD2")
print(cf)
```

`plot.sgSEMp1` *Plotting of Principle 1 of Semi-gSEM*

Description

Plot semi-gSEM principle 1 result.

Usage

```
## S3 method for class 'sgSEMp1'
plot(x, ..., cutoff = 0.2, width = NULL, height = NULL,
  filename = NULL)
```

Arguments

`x`	The returned list from sgSEMp1. Plotting uses the first element of this list (res.print) in which the first column of it is the response, the second column is variable and the other columns are the corresponding best functional form, r-squared, adj-r-squared, P-value1, P-value2 and P-value3.
`...`	Other parameters. Currently not used.
`cutoff`	A threshold value for the adjusted R-squared. Solid lines represent a relationship with the adjusted R-sqr greater than the cutoff and dotted lines with less than the cutoff. The default is 0.2.
`width`	A numeric describing the width of the plot output in pixels.
`height`	A numeric describing the height of the plot output in pixels.
`filename`	A character string naming a file to save as an html file.

Details

plot.sgSEMp1 plots a structural equation network model diagram based on the best functional form for each selected pairwise variable.

Value

An html style plot of the pairwise relationship pathway diagram between stressors and responses. Arrows show relationships between each variable with given statistical relations along the connection lines.

Examples

```
# Load acrylic data set
data(acrylic)
# Build a semi-gSEM model with principle 1
ans <- sgSEMp1(acrylic)
# Plot the network model with adjusted-R-squred of 0.1
plot(ans, cutoff = 0.1)
```

`plot.sgSEMp2` *Plotting of Principle 2 of Semi-gSEM*

Description

plot.sgSEMp2 plots a structural equation network model diagram based on best functional form for additive pairwise relationships.

Usage

```
## S3 method for class 'sgSEMp2'
plot(x, ..., cutoff = 0.2, width = NULL, height = NULL,
  filename = NULL, detail = F)
```

Arguments

`x`	The returned list from sgSEMp2. Plotting uses the first element of this list (print) in which the first column of it is the response, the second column is variable and other columns are corresponding statistical model, r-squared, adj-r-squared, P-value, P-value rank, p1ff, r2mark, and markrank.
`...`	Other parameters. Currently not used.
`cutoff`	A threshold value for the adjusted R-squared. Solid lines represent a relationship with adjusted R-sqr of greater than the cutoff and dotted lines with less than the cutoff. The default is 0.2.
`width`	A numeric describing the width of the plot output in pixels.
`height`	A numeric describing the height of the plot output in pixels.
`filename`	A character string naming a file to save as an html file.
`detail`	Logic value indicating whether the detailed information about the full model is displayed. Default is False.

Value

An html style plot of the pairwise relationship pathway diagram between stressors and responses. Arrows show relationships between each variable with given statistical relations along the connection lines.

Examples

```
data(acrylic)
ans <- sgSEMp2(acrylic)
plot(ans, cutoff = 0.2)
```

---

sgSEMp1 *Semi-supervised Generalized Structural Equation Modelling (gSEM) - Principle 1*

Description

This function carries out gSEM principle 1. Principle 1 determines the univariate relationships in the spirit of the Markovian process. The relationship between each pair of system elements, including predictors and the system level response, is determined with the Markovian property that assumes the value of the current predictor is sufficient in relating to the next level variable, i.e., the relationship is independent of the specific value of the preceding-level variable to the current predictor, given the current value.

Usage

```
sgSEMp1(x, predictor = NULL, response = NULL, nlsInits = data.frame(a1 =
  1, a2 = 1, a3 = 1))
```

Arguments

x	A dataframe, requiring at least 2 columns. By default, its first column stores the main or primary influencing predictor, or exogenous variable, e.g. time, or a main predictor. The second column stores the response variable, and other columns store intermediate variables.
predictor	A character string of the column name of the main predictor OR a numeric number indexing the column of the main predictor.
response	A character string of the column name of the main response OR a numeric number indexing the column of the main response.
nlsInits	A data frame of initial vectors for the nonlinear least square procedure, nls(). Each column corresponds to a sequence of initial values for one coefficient. The data frame can be generated by the genInit() function. Each row is one initial vector for all coefficients. Currently the only nls function included is y = a + b * exp(c * x).

Details

sgSEMp1 builds a network model of interfacing multiple continuous variables. Each pair of variables is fitted by one of the optimal relationships selected from 6 pre-determined functional forms, representing the sensible models commonly used in (energy) degradation science. They are:

- 1. Simple Linear(SL): y = a + b * x

- 2. Quadratic(Quad): y = a + b * x + c * x^2
- 3. Simple Quadratic(SQuad): y = a + b * x^2
- 4. Exponential(Exp): y = a + b * exp^x
- 5. Logarithm(Log): y = a + b * log(x)
- 6. Nonlinearizable(nls): y = a + b * exp(c * x)

Adjusted R-squared is used for model selection for every pair.

P-values reported in the "res.print" field of the return list are associated with the tests of the coefficients (a,b) and c as appropriate in the chosen model from the 6 candidates. In the case of polynomial model, the p-values are arranged in the order of increasing exponents. For example, in the quadratic functional form y ~ a + bx + cx^2, the three P-values correspond to those of \hat_a, \hat_b and \hat_c, respectively. If there are less than 3 coefficients to estimate, the extra P-value field is filled with NA's.

Value

An object of class sgSEMp1, which is a list of the following items:

- "Graph": A network graph that contains the univariate relationships between response and predictors determined by principle 1.
- "table": A matrix. For each row, first column is the response variable, second column is the predictor, the other columns show corresponding summary information: The optimal functional form, R-squared, adj-R-squared, P-value1, P-value2 and P-value3. See details.
- "bestModels": A matrix. First dimension indicates predictors. The second dimension indicates response variables. The i-jth cell of the matrix stores the name of the best functional form corresponding to the j-th response variable regressed on the i-th predictor.
- "allModels": A three dimensional array, indexed by [I, J, K], for all the models fitted to the n by p data set. The first dimension "I" indexes the predictor included in the model, and accepts integers 1 to p for one of the p variables; thus a value of "I=i" indicates using the ith variable in the data as the predictor. The second dimension "J" indexes the variable used as the response variable. The third dimension "K" specifies the fitting result of one of the 6 functional forms: 1=SL, 2=Quad, 3=SQuad, 4=Exp, 5=Log, 6=nls. The i-j-k-th cell of the list stores a "lm" object, corresponding to the j-th response, i-th predictor and the k-th functional form.

The object has two added attributes:

- "attr(res.best, "Step")": A vector. For each variable, it shows in which step it is chosen to be significantly related to the response variable.
- "attr(res.best, "diag.Step")": A matrix. First dimension is for predictors; second dimension is for response variables. Each cell shows in which step the pairwise relation is being fitted.

See Also

sgSEMp2() and plot.sgSEMp1()

Examples

```
## Load the built-in sample acrylic data set
data(acrylic)

## Run semi-gSEM principle one
ans <- sgSEMp1(acrylic, predictor = "IrradTot", response = "YI")

## Plot the result
plot(ans) #Default cutoff value for a solid path in the resulting graph is 0.2.

## Plot result with different R-sqr cutoff
plot(ans, cutoff = 0.4)

## Summary
summary(ans)

## Extract relations between IrradTot and YI
cf <- path(ans, from = "IrradTot", to = "YI")
print(cf)

## Print three components of the result
ans$table
ans$bestModels
ans$allModels

## Checking fitting result of YI by IrradTot using the exponential model
summary(ans$allModel[[1,2,4]])
```

---

sgSEMp2 *Semi-supervised Generalized Structural Equation Modelling (gSEM) - Principle 2*

Description

This function builds an gSEM model using gSEM principle 2. Principle 2 resembles the multiple regression principle in the way multiple predictors are considered simultaneously. Specifically, the first-level predictors to the system level variable, such as, Time and unit level variables, acted on the system level variable collectively by an additive model. This collective additive model can be found with a generalized stepwise variable selection (using the step() function in R, which performs variable selection on the basis of AIC) and this proceeds iteratively.

Usage

```
sgSEMp2(x, predictor = NULL, response = NULL)
```

Arguments

x	A dataframe, requiring at least 2 columns. By default its first column stores the main or primary influencing predictor, or exogenous variable e.g.., time,

	or a main predictor, the second column stores the response variable, and other columns store intermediate variables.
predictor	A character string of the column name of the system predictor OR a numeric number indexing the column of the main predictor.
response	A character string of the column name of the main response OR a numeric number indexing the column of the system response.

Details

Data is analysed first using Principle 1 to find the best models. If needed, transformations based on the best models are applied to the predictors. Starting from the system response variable, each variable is regressed on all other variables except for the system response in an additive multiple regression model, which is reduced by a stepwise selection using stepAIC(). Then, for each selected variable, fitted regression for 6 selected functional forms and pick the best.

Value

A list of the following items:

- "Graph": A network graph that contains the group and individual relationships between response and predictors determined by principle 2.
- "res.print": A matrix. For each row, first column is the response variable, second column is the predictor, the other columns show corresponding summary information.

See Also

sgSEMp1() and plot.sgSEMp2()

Examples

```
# Using built-in dataset
data(acrylic)
ans <- sgSEMp2(acrylic)
ans$res.print
plot(ans)

## Not run:
# Using simulated data
x4=runif(100,0,2)
x3=1+2.5*x4+rnorm(100,0,0.5)
x1=runif(100,1,4)
x2=-1-x1+x3+rnorm(100,0,0.3)
y=2+2*exp(x1/3)+(x2-1)^2-x3+rnorm(100,0,0.5)
# Check the pairwise plot
sim=cbind(x4,y,x1,x2,x3)
pairs(sim)
ans <- sgSEMp2(as.data.frame(sim))
plot(ans)

## End(Not run)
```

summary.sgSEMp1 *Summary of Semi-gSEM*

Description

Summarizes the gSEM principle 1 result.

Usage

```
## S3 method for class 'sgSEMp1'
summary(object, ...)
```

Arguments

object	An object of class "sgSEMp1", the returned list from sgSEMp1().
...	Other arguments. Currently not used.

Details

summary.sgSEMp1 gives a summary about the gSEm-Principle 1 analysis.

Value

NULL. A summary of data and fitting result is printed on screen.

Examples

```
data(acrylic)
ans <- sgSEMp1(acrylic)
summary(ans)
```

summary.sgSEMp2 *Summary of Semi-gSEM*

Description

Summarizes the gSEM principle 2 result.

Usage

```
## S3 method for class 'sgSEMp2'
summary(object, ...)
```

Arguments

object	An object of class "sgSEMp2", the returned list from sgSEMp2().
...	Other arguments. Currently not used.

Details

summary.sgSEMp1 gives a brief summary about the gSEM Principle 2 analysis.

Value

NULL. A summary of data and fitting result is printed on screen.

Examples

```
data(acrylic)
ans <- sgSEMp2(acrylic)
summary(ans)
```

Index