

Package ‘cvsem’

August 13, 2022

Title SEM Model Comparison with K-Fold Cross-Validation

Version 1.0.0

Description The goal of 'cvsem' is to provide functions that allow for comparing Structural Equation Models (SEM) using cross-validation. Users can specify multiple SEMs using 'lavaan' syntax. 'cvsem' computes the Kullback Leibler (KL) Divergence between 1) the model implied covariance matrix estimated from the training data and 2) the sample covariance matrix estimated from the test data described in Cudeck, Robert & Browne (1983) <[doi:10.18637/jss.v048.i02](https://doi.org/10.1186/10.18637/jss.v048.i02)>. The KL Divergence is computed for each of the specified SEMs allowing for the models to be compared based on their prediction errors.

License GPL (>= 3)

Encoding UTF-8

RoxygenNote 7.2.0

Imports lavaan, stats, Rdpack

RdMacros Rdpack

Suggests testthat (>= 3.0.0)

Config/testthat/edition 3

NeedsCompilation no

Author Anna Wysocki [aut, cre],
Danielle Siegel [aut],
Cameron allen [aut],
Philippe Rast [aut]

Maintainer Anna Wysocki <awysocki@ucdavis.edu>

Repository CRAN

Date/Publication 2022-08-13 12:30:05 UTC

R topics documented:

cvgather	2
cvsem	3
fd	4

gls	5
KL_divergence	6
print.cvsem	6

Index	8
--------------	----------

cvgather	<i>Gather lavaan model objects into a list</i>
----------	--

Description

Gather lavaan model objects to be compared via CV. Function returns a named list.

Usage

```
cvgather(...)
```

Arguments

... Names of lavaan model objects

Value

Named list

Author(s)

philippe

Examples

```
example_data <- lavaan::HolzingerSwineford1939
colnames(example_data) <- c("id", "sex", "ageyr", "agemo", 'school', "grade",
"visualPerception", "cubes", "lozenges", "comprehension",
"sentenceCompletion", "wordMeaning", "speededAddition",
"speededCounting", "speededDiscrimination")

model1 <- 'comprehension ~ sentenceCompletion + wordMeaning'

model2 <- 'comprehension ~ wordMeaning
sentenceCompletion ~ wordMeaning

comprehension ~~ 0.5*wordMeaning'

model_list <- cvgather(model1, model2)
```

Description

Do model comparison on SEM models using cross-validation as described in (Cudeck and Browne 1983) and (Browne and Cudeck 1992). Cross-validation is based on the discrepancy between the sample covariance matrix and the model implied matrix. Currently, `cvsem` supports 'KL-Divergence', Frobenius Distance and Generalized Least Squares 'GLS' as discrepancy metrics.

Usage

```
cvsem(
  data = NULL,
  Models,
  discrepancyMetric = "KL-Divergence",
  k = 5,
  lavaanFunction = "sem",
  echo = TRUE,
  ...
)
```

Arguments

<code>data</code>	Data
<code>Models</code>	A collection of models, specified in lavaan syntax. Provide Models with the <code>cvgather()</code> function.
<code>discrepancyMetric</code>	Specify which discrepancy metric to use (one of 'KL-Divergence', 'FD', 'GLS'). Default is KL Divergence.
<code>k</code>	The number of folds. Default is 5.
<code>lavaanFunction</code>	Specify which lavaan function to use. Default is "sem". Other options are "lavaan" and "cfa"
<code>echo</code>	Provide feedback on progress to user, defaults to TRUE. Set to FALSE to suppress.
<code>...</code>	Not used

Value

A list with the prediction error for each model.

References

Browne MW, Cudeck R (1992). "Alternative Ways of Assessing Model Fit." *Sociological Methods & Research*, **21**, 230–258.

Cudeck R, Browne MW (1983). “Cross-Validation Of Covariance Structures.” *Multivariate Behavioral Research*, **18**, 147–167. doi:10.1207/s15327906mbr1802_2, https://www.tandfonline.com/doi/abs/10.1207/s15327906mbr1802_2.

Examples

```
example_data <- lavaan::HolzingerSwineford1939
colnames(example_data) <- c("id", "sex", "ageyr", "agemo", 'school', "grade",
"visualPerception", "cubes", "lozenges", "comprehension",
"sentenceCompletion", "wordMeaning", "speededAddition",
"speededCounting", "speededDiscrimination")

model1 <- 'comprehension ~ meaning

      ## Add some latent variables:
      meaning =~ wordMeaning + sentenceCompletion
      speed =~ speededAddition + speededDiscrimination + speededCounting
      speed ~~ meaning'

model2 <- 'comprehension ~ wordMeaning + speededAddition'
model3 <- 'comprehension ~ wordMeaning + speededAddition'

models <- cvgather(model1, model2, model3)

fit <- cvsem( data = example_data, Models = models, k = 10, discrepancyMetric = "KL-Divergence")
```

 fd

Frobenius Matrix Discrepancy

Description

Frobenius Distance as described in (Biscay et al. 1997) or (Amendola and Storti 2015).

Usage

```
fd(implied_Sigma, test_S)
```

Arguments

`implied_Sigma` Model implied covariances matrix from training set
`test_S` Sample covariance matrix from test set

Value

FD discrepancy

References

Amendola A, Storti G (2015). “Model Uncertainty and Forecast Combination in High-Dimensional Multivariate Volatility Prediction.” *Journal of Forecasting*, **34**, 83–91. doi:10.1002/for.2322.

Biscay R, Rodríguez LM, Díaz-Frances E (1997). “Cross-validation of covariance structures using the frobenius matrix distance as a discrepancy function.” *Journal of Statistical Computation and Simulation*, **58**, 195–215. doi:10.1080/00949659708811831, <https://www.tandfonline.com/doi/abs/10.1080/00949659708811831>.

gls

Generalized Least Squares Discrepancy Function

Description

Generalized Least Squares (GLS) Discrepancy as defined in (Cudeck and Browne 1983).

Usage

```
gls(implicit_sigma, test_S)
```

Arguments

`implicit_sigma` Model implied covariances matrix from training set
`test_S` Sample covariance matrix from test set

Value

GLS discrepancy

References

Cudeck R, Browne MW (1983). “Cross-Validation Of Covariance Structures.” *Multivariate Behavioral Research*, **18**, 147–167. doi:10.1207/s15327906mbr1802_2, https://www.tandfonline.com/doi/abs/10.1207/s15327906mbr1802_2.

KL_divergence	<i>Compute KL-Divergence on two covariance matrices. KL-Divergence corresponds to the Maximum Wishart Likelihood (MWL) discrepancy described in (Cudeck and Browne 1983).</i>
---------------	---

Description

Compute KL-Divergence on two covariance matrices. KL-Divergence corresponds to the Maximum Wishart Likelihood (MWL) discrepancy described in (Cudeck and Browne 1983).

Usage

```
KL_divergence(implied_sigma, test_S)
```

Arguments

implied_sigma	Model implied covariances matrix from training set
test_S	Sample covariance matrix from test set

Value

KL-Divergence index

References

Cudeck R, Browne MW (1983). "Cross-Validation Of Covariance Structures." *Multivariate Behavioral Research*, **18**, 147–167. doi:10.1207/s15327906mbr1802_2, https://www.tandfonline.com/doi/abs/10.1207/s15327906mbr1802_2.

print.cvsem	<i>Print cvsem object</i>
-------------	---------------------------

Description

Return the ordered list of models where the model with the smallest discrepancy metric is listed first.

Usage

```
## S3 method for class 'cvsem'
print(x, digits = 2, ...)
```

Arguments

x	cvsem object
digits	Round to (default 2) digits.
...	not used

print.cvsem

7

Value

Formatted cvsem object

Index

cvgather, 2
cvsem, 3

fd, 4

gls, 5

KL_divergence, 6

print.cvsem, 6