Package 'SFPL'

January 20, 2025

Type Package

Title Sparse Fused Plackett-Luce

Imports pracma, gtools

Version 1.0.0

Maintainer Sjoerd Hermes <sjoerd.hermes@wur.nl>

Description Implements the methodological developments found in Hermes, van Heerwaarden, and Behrouzi (2024) <doi:10.48550/arXiv.2308.04325>, and allows for the statistical modeling of multi-group rank data in combination with object variables. The package also allows for the simulation of synthetic multi-group rank data.

License GPL-3

Encoding UTF-8

LazyData true

Depends R (>= 3.10)

NeedsCompilation no

Author Sjoerd Hermes [aut, cre]

Repository CRAN

Date/Publication 2024-07-26 19:50:02 UTC

Contents

data_sim *Rank data simulation*

Description

Simulates (partial) rank data for multiple groups together with object variables.

Usage

```
data_sim(m, M, n, p, K, delta, eta)
```

Arguments

m	Length of the partial ranking for each observation.
M	Total number of objects.
n	Number of observations (rankers) per group.
p	Number of object variables.
K	Number of groups.
delta	Approximate fraction of different coefficients across the $\beta^{(k)}$.
eta	Approximate fraction of sparse coefficients in $\beta^{(k)}$ for all k.

Value

y	A list consisting of K matrices with each matrix containing (partial) rankings across n observations for group k.
x	A $M \times p$ matrix containing the values for the p objects variables across the M objects.
beta	A $p \times K$ matrix containing the true value of β, which was used to generate y.

Author(s)

Sjoerd Hermes
Maintainer: Sjoerd Hermes <sjoerd.hermes@wur.nl>

References

1. Hermes, S., van Heerwaarden, J., and Behrouzi, P. (2024). Joint Learning from Heterogeneous Rank Data. arXiv preprint, arXiv:2407.10846

Examples

```
data_sim(3, 10, 50, 5, 2, 0.25, 0.25)
```

ghana *Ranking data*

Description

This is a real dataset containing information on 5 object variables describing the properties of 13 different sweet potato varieties. In addition, the dataset contains partial rankings made by men and women from Ghana.

Usage

```
data("ghana")
```

Format

A list with three dataframes. The first consists of the rankings made by men, the second consists of the rankings made by women and the third contain the object variables.

Details

Contains a subset of the data used in the Hermes et al. (2024) paper.

Source

Data from the Hermes et al. (2024) paper is based on Moyo et al. (2021).

References

1. Hermes, S., van Heerwaarden, J., and Behrouzi, P. (2024). Joint Learning from Heterogeneous Rank Data. arXiv preprint, arXiv:2407.10846
2. Moyo, M., R. Ssali, S. Namanda, M. Nakitto, E. K. Dery, D. Akansake, J. Adjebeng-Danquah, J. van Etten, K. de Sousa, H. Lindqvist-Kreuze, et al. (2021). Consumer preference testing of boiled sweetpotato using crowdsourced citizen science in Ghana and Uganda. Frontiers in Sustainable Food Systems 5, 620363.

Examples

```
data(ghana)
```

`sfpl` *Sparse Fused Plackett-Luce*

Description

Contains the main function of this package that is used to estimate the parameter of interest β. The inner workings of the function are described in Hermes et al., (2024).

Usage

```
sfpl(x, y, ls_vec, lf_vec, epsilon, verbose)
```

Arguments

`x`	A $M \times p$ matrix containing the values for the p objects variables across the M objects.
`y`	A list consisting of K matrices with each matrix containing (partial) rankings across n observations for group k.
`ls_vec`	Vector containing shrinkage parameters.
`lf_vec`	Vector containing fusion penalty parameters.
`epsilon`	Small positive value used to ensure that the penalty function is differentiable. Typically set at 10^{-5}.
`verbose`	Boolean that returns the process of the parameter estimation.

Value

`beta_est`	A list of length ls_vec$\times$lf_vec that contains the parameter estimates $\hat{beta}$ for each combination of ls_vec and lf_vec.

Author(s)

Sjoerd Hermes
Maintainer: Sjoerd Hermes <sjoerd.hermes@wur.nl>

References

1. Hermes, S., van Heerwaarden, J., and Behrouzi, P. (2024). Joint Learning from Heterogeneous Rank Data. arXiv preprint, arXiv:2407.10846

Examples

```
# we first obtain the rankings and object variables
data(ghana)
y <- list(ghana[[1]], ghana[[2]])
x <- ghana[[3]]

# our next step consists of creating two vectors for the penalty parameters
ls_vec <- lf_vec <- c(0, 0.25)

# we choose epsilon to be small: 10^(-5), as we did in Hermes et al., (2024)
# now we can fit our model
epsilon <- 10^(-5)
verbose <- FALSE

result <- sfpl(x, y, ls_vec, lf_vec, epsilon, verbose)
```

`sfpl_approx` *Approximate Sparse Fused Plackett-Luce*

Description

Contains an approximate (typically faster) version of the main function of this package that is used to estimate the parameter of interest β. We recommend this version due to its (relatively) fast convergence.

Usage

```
sfpl_approx(x, y, ls_vec, lf_vec, epsilon, verbose)
```

Arguments

`x`	A $M \times p$ matrix containing the values for the p objects variables across the M objects.
`y`	A list consisting of K matrices with each matrix containing (partial) rankings across n observations for group k.
`ls_vec`	Vector containing shrinkage parameters.
`lf_vec`	Vector containing fusion penalty parameters.
`epsilon`	Small positive value used to ensure that the penalty function is differentiable. Typically set at 10^{-5}.
`verbose`	Boolean that returns the process of the parameter estimation.

Value

`beta_est`	A list of length ls_vec$\times$lf_vec that contains the parameter estimates $\hat{beta}$ for each combination of ls_vec and lf_vec.

Author(s)

Sjoerd Hermes
Maintainer: Sjoerd Hermes <sjoerd.hermes@wur.nl>

References

1. Hermes, S., van Heerwaarden, J., and Behrouzi, P. (2024). Joint Learning from Heterogeneous Rank Data. arXiv preprint, arXiv:2407.10846

Examples

```
# we first obtain the rankings and object variables
data(ghana)
y <- list(ghana[[1]], ghana[[2]])
x <- ghana[[3]]

# our next step consists of creating two vectors for the penalty parameters
ls_vec <- lf_vec <- c(0, 0.25)

# we choose epsilon to be small: 10^(-5), as we did in Hermes et al., (2024)
# now we can fit our model
epsilon <- 10^(-5)
verbose <- FALSE

result <- sfpl_approx(x, y, ls_vec, lf_vec, epsilon, verbose)
```

---

sfpl_select *Model selection for SFPL*

Description

This function selects the "best" fitted SFPL model using either the AIC or the BIC, see Hermes et al., (2024).

Usage

```
sfpl_select(beta_est, x, y, ls_vec, lf_vec)
```

Arguments

beta_est	A list of length ls_vec$\times$lf_vec that contains the parameter estimates $\hat{beta}$, using either sfpl or sfpl_approx, for each combination of ls_vec and lf_vec.
x	A $M \times p$ matrix containing the values for the p objects variables across the M objects.

y	A list consisting of K matrices with each matrix containing (partial) rankings across n observations for group k.
ls_vec	Vector containing shrinkage parameters.
lf_vec	Vector containing fusion penalty parameters.

Value

model_aic	A $p \times K$ matrix containing the parameter estimates using the penalty parameters λ_s, λ_f as chosen by the AIC.
model_bic	A $p \times K$ matrix containing the parameter estimates using the penalty parameters λ_s, λ_f as chosen by the BIC.

Author(s)

Sjoerd Hermes
Maintainer: Sjoerd Hermes <sjoerd.hermes@wur.nl>

References

1. Hermes, S., van Heerwaarden, J., and Behrouzi, P. (2024). Joint Learning from Heterogeneous Rank Data. arXiv preprint, arXiv:2407.10846

Examples

```
# we first obtain the rankings and object variables
data(ghana)
y <- list(ghana[[1]], ghana[[2]])
x <- ghana[[3]]

# our next step consists of creating two vectors for the penalty parameters
ls_vec <- lf_vec <- c(0, 0.25)

# we choose epsilon to be small: 10^(-5), as we did in Hermes et al., (2024)
# now we can fit our model
epsilon <- 10^(-5)
verbose <- FALSE

result <- sfpl_approx(x, y, ls_vec, lf_vec, epsilon, verbose)

# now we select the best models using our model selection function
sfpl_select(result, x, y, ls_vec, lf_vec)
```

Index