Package ‘mmpca’

March 20, 2020

Title  Integrative Analysis of Several Related Data Matrices
Version  2.0.1
Author  Jonatan Kallus [aut, cre]
Maintainer  Jonatan Kallus <kallus@chalmers.se>
Description  A generalization of principal component analysis for integrative
             analysis. The method finds principal components that describe single matrices
             or that are common to several matrices. The solutions are sparse. Rank of
             solutions is automatically selected using cross validation. The method is
             described in Kallus, Johansson, Nelander and Jörnsten (2019)
Depends  R (>= 3.3.0)
Imports  digest (>= 0.6.0), gsl (>= 1.9)
LinkingTo  Rcpp, RcppEigen
License  GPL-3
LazyData  true
NeedsCompilation  yes
Encoding  UTF-8
RoxygenNote  7.0.2
Repository  CRAN
Date/Publication  2020-03-20 12:40:03 UTC

R topics documented:

mmpca ................................................................. 2

Index  4
mmpca

Multiview principal component analysis

Description

Analyzes several related matrices of data.

Usage

mmpca(x, inds, k, lambda = NULL, trace = 0, init_theta = NULL, cachepath = NULL, enable_rank_selection = TRUE, enable_sparsity = TRUE, enable_variable_selection = FALSE, parallel = TRUE)

Arguments

x List of matrices to analyze
inds Matrix containing view indices. The matrix should have two columns and the same number of rows as the length of x. The first (second) column contains the view index of the rows (columns) of the corresponding matrix.
k Integer giving the maximum rank of the analysis, i.e. the maximum number of principal components for each view.
lambda Vector or matrix of lambda values. The length (or width if it is a matrix) depends on the number of active penalties (2, 3 or 4). If it is a matrix, try different lambda values (one try for each row). Default: a matrix where each column is the sequence exp(seq(-6,0))).
trace Integer selecting the amount of log messages. 0 (default): no output, 3: all output.
init_theta NULL, functions or numeric. NULL (default) use initial values based on ordinary SVD. If init_theta is a list of three functions (CMF, matrix_to_triplets and getCMFopts from package CMF) use the supplied functions to find initial values with collaborative matrix factorization (CMF). If init_theta is a numeric vector it is used as initial value.
cachepath Character vector with path to directory to store intermediate results. If NULL (default) intermediate results are not stored. For caching to work it is required that the random number generation seed is constant between calls to mmpca, so set.seed needs to be called before mmpca.
enable_rank_selection Boolean deciding if the second penalty that imposes a low rank model should be enabled.
enable_sparsity Boolean deciding if the third penalty that imposes sparsity in V should be enabled.
enable_variable_selection Boolean deciding if the fourth penalty that increases the tendence for sparsity structure of different V columns to be similar. Defaults to FALSE meaning this penalty is not used.
parallel Boolean deciding if computations should be run on multiple cores simultaneously.

Value

A list with components

initial initial values used in optimization

cmf solution found with CMF (if init_theta == c(CMF, matrix_to_triplets, getCMFopts))

training solutions for different values of lambda

solution solution for optimal lambda value

Author(s)

Jonatan Kallus, <kallus@chalmers.se>

Examples

```r
x <- list(matrix(rnorm(110), 10, 11), matrix(rnorm(120), 10, 12))
inds <- matrix(c(1, 1, 2, 3), 2, 2)
result <- mmpca(x, inds, 3, parallel=FALSE)
```
Index

*Topic models
  mmrca, 2
*Topic multivariate
  mmrca, 2
*Topic pca
  mmrca, 2

mmrca, 2