Package ‘pcaone’

October 29, 2022

Title Randomized Singular Value Decomposition Algorithms with 'RcppEigen'

Version 1.0.0

Date 2022-10-25

Author Zilong Li [aut, cre]

Maintainer Zilong Li <zilong.dk@gmail.com>

Description Randomized Singular Value Decomposition (RSVD) methods proposed in the 'PCAone' paper by Li (2022) <doi:10.1101/2022.05.25.493261>, where we implement and propose two RSVD methods. One is based on Yu (2017) <arXiv:1704.07669> single pass RSVD but with power iteration scheme. The other is our new window based RSVD.

License GPL (>= 3)

Encoding UTF-8

Depends R (>= 3.6.0)

Imports Rcpp

LinkingTo Rcpp, RcppEigen (>= 0.3.3.3.0)

SystemRequirements C++17

RoxygenNote 7.2.1

Suggests testthat (>= 3.0.0)

Config/testthat/edition 3

URL https://github.com/Zilong-Li/PCAoneR

BugReports https://github.com/Zilong-Li/PCAoneR/issues

NeedsCompilation yes

Repository CRAN

Date/Publication 2022-10-29 08:57:35 UTC

R topics documented:

pcaone 2

Index **5**

pcaone — *Randomized Singular Value Decomposition Algorithm with Window-based Power Iterations from PCAone (Li et al 2022).*

Description

The Randomized Singular Value Decomposition (RSVD) computes the near-optimal low-rank approximation of a rectangular matrix using a fast probablistic algorithm.

Usage

```
pcaone(
  A,
  k = NULL,
  p = 7,
  q = 10,
  sdist = "normal",
  method = "alg2",
  windows = 64,
  shuffle = FALSE
)
```

Arguments

`A` array_like;
a real/complex (m, n) input matrix (or data frame) to be decomposed.

`k` integer;
specifies the target rank of the low-rank decomposition. k should satisfy $k << min(m, n)$.

`p` integer, optional;
number of additional power iterations (by default $p = 7$).

`q` integer, optional;
oversampling parameter (by default $q = 10$).

`sdist` string $c('unif',' normal')$, optional;
specifies the sampling distribution of the random test matrix:
$'unif'$: Uniform `[-1,1]`.
$'normal'$ (default) : Normal `~N(0,1)`.

`method` string $c('alg1',' agl2')$, optional;
specifies the different variation of the randomized singular value decomposition :
$'alg1'$: single pass RSVD with power iterations in PCAone refered to algorithm1.
$'alg2'$ (default): window based RSVD in PCAone refered to algorithm2.

`windows` integer, optional;
the number of windows for 'alg2' method. must be a power of 2 (by default $windows = 64$).

`shuffle` logical, optional;
if shuffle the rows of input tall matrix or not. recommended for algorithm 2 (by default $shuffle = FALSE$).

Details

The singular value decomposition (SVD) plays an important role in data analysis, and scientific computing. Given a rectangular (m, n) matrix A, and a target rank $k << min(m, n)$, the SVD factors the input matrix A as

$$A = U_k diag(d_k) V_k^\top$$

The k left singular vectors are the columns of the real or complex unitary matrix U. The k right singular vectors are the columns of the real or complex unitary matrix V. The k dominant singular values are the entries of d, and non-negative and real numbers.

q is an oversampling parameter to improve the approximation. A value of at least 10 is recommended, and $q = 10$ is set by default.

The parameter p specifies the number of power (subspace) iterations to reduce the approximation error. The power scheme is recommended, especially when the singular values decay slowly. However, computing power iterations increases the computational costs. Even though most RSVD implementations recommend $p = 3$ power iterations by default, it's always sufficient to run only few power iterations where our window-based power iterations $('alg2')$ come to play. We recommend using $windows = 64$ and $p >= 7$ for pcaone algorithm2. As it is designed for large dataset, we recommend using $'alg2'$ when $max(n, m) > 5000$.

If $k > (min(n, m)/4)$, a deterministic partial or truncated `svd` algorithm might be faster.

Value

`pcaone` returns a list containing the following three components:

d array_like;
singular values; vector of length (k).

u array_like;
left singular vectors; (m, k) or (m, nu) dimensional array.

v array_like;
right singular vectors; (n, k) or (n, nv) dimensional array.

Note

The singular vectors are not unique and only defined up to sign. If a left singular vector has its sign changed, changing the sign of the corresponding right vector gives an equivalent decomposition.

Author(s)

Zilong Li <zilong.dk@gmail.com>

References

- Z. Li, J Meisner, A Albrechtsen. "PCAone: fast and accurate out-of-core PCA framework for large scale biobank data" (2022) doi: 10.1101/2022.05.25.493261.

Examples

```
library('pcaone')
mat <- matrix(rnorm(100*20000), 100, 20000)
res <- pcaone(mat, k = 10, p = 7, method = "alg2")
str(res)
res <- pcaone(mat, k = 10, p = 7, method = "alg1")
str(res)
```

Index

pcaone, 2

svd, *3*