Package: BigQuic (via r-universe)

<div align="center">October 1, 2024</div>

Type Package

Title Big Quadratic Inverse Covariance Estimation

Version 1.1-13

Date 2022-11-19

Author Khalid B. Kunji [aut, cre], Cho-Jui Hsieh [ctb], Matyas A. Sustik [ctb], Inderjit S. Dhillon [ctb], Pradeep Ravikumar [ctb], Tuo Zhao [ctb], Xingguo Li [ctb], Han Liu [ctb], Kathryn Roeder [ctb], John Lafferty [ctb], Larry Wasserman [ctb], George Karypis [ctb], Melissa O'Neill [ctb], Richard Henderson [ctb]

Maintainer Khalid B. Kunji <kkunji@hbku.edu.qa>

Repository CRAN

Description Use Newton's method, coordinate descent, and METIS clustering to solve the L1 regularized Gaussian MLE inverse covariance matrix estimation problem.

License GPL (>= 3) | file LICENSE

Copyright The included METIS library is Copyright 1997, Regents of the University of Minnesota

URL https://www.r-project.org,

https://bigdata.oden.utexas.edu/software/1035/

http://glaros.dtc.umn.edu/gkhome/views/metis

https://www.pcg-random.org/download.html

https://gcc.gnu.org/projects/gomp/

Imports Rcpp (>= 0.12.1), Matrix, scalreg

LinkingTo Rcpp

Depends R (>= 3.2.2), methods

NeedsCompilation yes

Date/Publication 2022-11-19 22:40:02 UTC

Contents

BigQuic   *Big Quadratic Inverse Covariance Estimation*

Description

Use Newton's method, coordinate descent, and METIS clustering to solve the L1 regularized Gaussian MLE inverse covariance matrix estimation problem.

Usage

```
BigQuic (X = NULL, inputFileName = NULL, outputFileName = NULL, lambda = 0.5,
        numthreads = 4, maxit = 5, epsilon = 1e-3, k = 0, memory_size = 8000,
        verbose = 0, isnormalized = 1, seed = NULL, use_ram = FALSE)
```

Arguments

X	An n rows by p columns matrix of the data without the response vector (Y).
inputFileName	Full path to a file containing X in the format used for input to BigQUIC, p n X An example is installed, you can get its path with: paste(path.package("BigQuic"), "/extdata/testInput", sep = "")
outputFileName	Location and name of output file that will be extrapolated for their naming, e.g. /home/username/test when 3 files are being output will result in /home/username/test.1.output /home/username/test.2.output and /home/username/test.3.output
lambda	The tuning parameter 0 <= lambda <= 1, but small values should not be used for performance reasons, e.g. < .4 or so. A vector of lambdas may also be input, in which case BigQUIC will be run for each lambda. Yes, the examples shows lambda as small as 0.1, but that is only because the testInput matrix is very small so the small lambdas can still finish in a sensible amount of time.
numthreads	Number of threads to use for this computation.
maxit	Maximum number of Newton iterations.
epsilon	Convergence tolerance.
k	Number of memory blocks to use, ideally should be the smallest k such that p/k columns fit in the memory_size.
memory_size	The amount of memory this computation is constrained to.

verbose	Controls how verbose messages should be printed during execution. Valid value range: 0–4. Higher numbers will give more messages for debugging.
isnormalized	Whether or not the input is already normalized.
seed	A seed for the random number generation, useful for replicating results.
use_ram	By default the results are written into files, using this option will load those files back to R and return them instead of their paths (the default behavior). When doing this there is a possibility that R will crash if you don't have enough RAM, use with caution on larger data sets or with many lambdas.

Details

BigQUIC is finally here! The original authors of QUIC and BigQUIC brought QUIC to Matlab (MEX), Standalone (C++), and R, but BigQUIC was delivered for Matlab and Standalone only with no R package. There are also some other features to the package, including sample data generation, inverse selection, and plotting. IMPORTANT: Due to the practicalities of formatting and working with large data sets, files are written to disk at various times when using BigQuic. The locations of the files BigQuic wrote to disk are kept in the object returned by BigQuic. They can be deleted when you're finished with the BigQuic_object manually by using the cleanFiles() function as shown in the examples. There are basiclly 8 cases for file creation, the following will give you an idea of where they are in case R crashes completely and loses the references to the files so you need to delete them manually. Files created in tmp are deleted on reboot, so no worries if you're having trouble finding them.

1. X, output file, use_ram = TRUE length(lambda) output files created in output location 1 file created for X in tmp Note: this is the same as 5, use_ram doesn't matter in this case

2. input file, no output file, use_ram = FALSE length(lambda) output files in location of input file

3. input file, output file, use_ram = FALSE length(lambda)a output files in location of output file Same as 8, use_ram doesn't matter in this case

4. X, no output file, use_ram = FALSE length(lambda) output files in tmp 1 file created for X in tmp Also same as 1 and 5

5. X, output file, use_ram = FALSE length(lambda) output files created in output file location 1 file created for X in tmp

6. X, no output file, use_ram = TRUE 1 file created for X in tmp

7. input file, no output file, use_ram = TRUE no files created

8. input file, output file, use_ram = TRUE length(lambda) output files created in output file location

Value

An object with Reference Class `"BigQuic_object"`

X	The X input for BigQuic, if given
inputFileName	The file name input for BigQuic, if given
isnormalized	Whether or not the input data was previously normalized
k	k used in BigQuic
epsilon	The epsilon that was used in this run of BigQuic

lambda	lambda used in BigQuic
maxit	maxit used in BigQuic
memory_size	memory_size used in BigQuic
numthreads	numthreads used in BigQuic
seed	seed used in BigQuic
use_ram	use_ram used in BigQuic
verbose	level of verbosity used in BigQuic
opt.lambda	The selected optimal lambda value, initially empty, it will be filled in by running BigQuic.select on the object, see the use in the Examples below
precision_matrices	
	The precision matrix for each of the lambdas in a list, so to access the one for the 1st lambda in the example: exampleResult$precision_matrices[[1]]
output_file_names	
	Lists files created by the class
clean	Indicates whether or not cleanFiles() has been called on this object before
inFlag	An internal indicator for the class
outFlag	An internal indicator for the class
getClass	Returns Class method definition
cleanFiles	Deletes files created by the class, except for those intentionally output by specifying an output file name
setX	Used internall to set X
setOptLambda	used internally to set opt.lambda
setSeed	used internally to set the seed
.self	returns the object itself again
.refClassDef	Lists fields and methods of the reference class

Author(s)

Khalid B. Kunji [aut, cre], Cho-Jui Hsieh [ctb], Matyas A. Sustik [ctb], Inderjit S. Dhillon [ctb], Pradeep Ravikumar [ctb], Tuo Zhao [ctb], Xingguo Li [ctb], Han Liu [ctb], Kathryn Roeder [ctb], John Lafferty [ctb], Larry Wasserman [ctb], George Karypis [ctb], Melissa O'Neill [ctb], Richard Henderson [ctb]

Maintainer: Khalid B. Kunji <kkunji@hbku.edu.qa>

References

BIG & QUIC: Sparse Inverse Covariance Estimation for a Million Variables (pdf) C. Hsieh, M. Sustik, I. Dhillon, P. Ravikumar, R. Poldrack. In Neural Information Processing Systems (NIPS), December 2013. (Oral) http://www.cs.utexas.edu/~cjhsieh/hugeQUIC.pdf http://bigdata.ices.utexas.edu/publication/big-quic-sparse-inverse-covariance-estimation-for-a-million-variables-2/

QUIC: Quadratic Approximation for Sparse Inverse Covariance Matrix Estimation (pdf) C. Hsieh, M. Sustik, I. Dhillon, P. Ravikumar. Journal of Machine Learning Research (JMLR), October 2014.

http://jmlr.org/papers/volume15/hsieh14a/hsieh14a.pdf http://bigdata.ices.utexas.edu/publication/quic-quadratic-approximation-for-sparse-inverse-covariance-matrix-estimation-2/

METIS:"A Fast and Highly Quality Multilevel Scheme for Partitioning Irregular Graphs". George Karypis and Vipin Kumar. SIAM Journal on Scientific Computing, Vol. 20, No. 1, pp. 359-392, 1999. http://glaros.dtc.umn.edu/gkhome/fetch/papers/mlSIAMSC99.pdf

PCG: A Family of Simple Fast Space-Efficient Statistically Good Algorithms for Random Number Generation. This paper is currently submitted to ACM Transactions on Mathematical Software, where it is currently under review. http://www.pcg-random.org/pdf/toms-oneill-pcg-family-v1.02.pdf

Examples

```
lambda <- 0.91
exampleResult <- BigQuic(inputFileName = paste(path.package("BigQuic"),
                        "/extdata/testInput", sep = ""),
                        outputFileName = tempfile(pattern =
                        "BigQuic_output_matrix", fileext = ".Bmat"),
                        lambda = lambda, numthreads = 1, memory_size = 512,
                        seed = 1, use_ram = TRUE)
BigQuic.select(exampleResult)
plot(exampleResult)
exampleResult$cleanFiles()
## Not run:
If you have the hdi package installed:
library(hdi)
data(riboflavin)
lambda <- seq(from = 0.9, to = 0.99, by = 0.01)
exampleResult <- BigQuic(as.matrix(riboflavin), lambda = lambda,
                        numthreads = 1, memory_size = 512, seed = 1,
                        use_ram = TRUE)
BigQuic.select(exampleResult)
plot(exampleResult)

## End(Not run)
```

BigQuic.select   *BigQuic Select*

Description

Selects the optimal lambda value from those in the BigQuic_object, i.e. BigQuic Result.

Usage

```
BigQuic.select(BigQuic_result = NULL, stars.thresh = 0.1,
       stars.subsample.ratio = NULL, rep.num = 20, verbose = TRUE,
       verbose2 = 0)
```

Arguments

`BigQuic_result`	A BigQuic_object returned from running BigQuic.
`stars.thresh`	The threshold used in the Stars selection method for choosing a lambda
`stars.subsample.ratio`	
	The ratio giving how large the subsamples will be for Stars, if null there is a heuristic calculation.
`rep.num`	Number of times to do the repetition in Stars.
`verbose`	Controls the level of verbosity in a part of the code.
`verbose2`	Controls the level of verbosity in another section of code.

`BigQuicHelper`   *BigQuic C++ Caller*

Description

Calls the C++ BigQuic algorithm.

`BigQuic_object-class`   *Class* `"BigQuic_object"`

Description

Reference Class that holds all the relevant results of the BigQuic computation.

Extends

All reference classes extend and inherit methods from *"*envRefClass*"*.

Fields

`precision_matrices:` Object of class `list` ~~

`X:` Object of class `matrix` ~~

`inputFileName:` Object of class `character` ~~

`lambda:` Object of class `numeric` ~~

`numthreads:` Object of class `numeric` ~~

`maxit:` Object of class `numeric` ~~

`epsilon:` Object of class `numeric` ~~

`k:` Object of class `numeric` ~~

`memory_size:` Object of class `numeric` ~~

`verbose:` Object of class `numeric` ~~

isnormalized: Object of class `numeric` ~~

seed: Object of class `numeric` ~~

use_ram: Object of class `logical` ~~

clean: Object of class `logical` ~~

output_file_names: Object of class `character` ~~

opt.lambda: Object of class `numeric` ~~

inFlag: Object of class `logical` ~~

outFlag: Object of class `logical` ~~

Methods

cleanFiles(verbose): ~~

setOptLambda(optLambda): ~~

setX(inputX): ~~

setSeed(inputSeed): ~~

Examples

```
showClass("BigQuic_object")
```

`BigQuic_object_builder`

BigQuic Object Builder

Description

Creates reference class objects (... which are really environments) of type BigQuic_object.

`generate_sample`   *Generate Sample*

Description

Generates a sample data set for using with BigQuic, the default seed is 1 for reproducibility. For high dimensional data, choose p much larger than n.

Usage

```
generate_sample(n = 200, p = 150, seed = NULL)
```

Arguments

n	The number of rows in the resulting data set.
p	The number of columns in the resulting data set.
seed	A seed for the random number generator in R.

---

plot.BigQuic_object   *Plot*

Description

Makes plot of the precision matrix showing non-zero values. The diagonal is shown in only black because the agreement with itself is not highly interesting. Negative relations are shown in green and positive in red. The saturation indicates the normalized strength of the relation. The matrix is symmetric and technically only the lower or upper triangle would suffice to provide identical information.

Usage

```
## S3 method for class 'BigQuic_object'
plot(x, ...)
```

Arguments

x	The BigQuic object, which will have its optimal precision matrix plotted.
...	plot can take a variety of arguments depending on the type, that is represented by ...

Index