Package 'TSDFGS'

November 12, 2026

Type Package

Title Training Set Determination for Genomic Selection

Version 3.0.0

Encoding UTF-8

Date 2026-11-05

URL https://www.oumark.net/TSDFGS/, https://github.com/oumarkme/TSDFGS

BugReports https://github.com/oumarkme/TSDFGS/issues

Depends R (>= 4.0)

Description Finds a training set for genomic selection and the size it needs. The r-score of Ou and Liao (2019) <doi:10.1007/s00122-019-03387-0> is derived from Pearson's correlation between the genomic estimated breeding values and the phenotypic values of the test set. optTrain() searches for a good training set, and SSDFGS() searches for its size. Eighteen criteria are available to score or optimize a training set, among them A-, D-, E- and G-optimality, PEVmax, MSPE, CDmean, CDmax, and genomic relationship criteria (Chen et al. (2024) <doi:10.1007/s00122-024-04766-y>; Fernandez-Gonzalez et al. (2023) <doi:10.1007/s00122-023-04265-6>). The computation uses 'Rcpp' and 'RcppEigen', and independent searches can run on several threads with 'OpenMP'.

License GPL (>= 3)

Imports Rcpp, stats

LinkingTo Rcpp, RcppEigen

Suggests ggplot2, testthat (>= 3.0.0)

SystemRequirements C++17

LazyData true

Config/roxygen2/version 8.1.0

NeedsCompilation yes

Author Jen-Hsiang Ou [aut, cre] (ORCID: <https://orcid.org/0000-0001-9305-2931>, affiliation: Department of Medical Genetics, National Taiwan University Hospital), Po-Ya Wu [aut] (ORCID: <https://orcid.org/0000-0002-7342-2867>),

Chen-Tuo Liao [aut, ths] (ORCID:
<https://orcid.org/0000-0001-9777-3701>)

Maintainer Jen-Hsiang Ou <oumark.me@outlook.com>

Repository CRAN

Date/Publication 2026-11-12 16:50:02 UTC

Contents

TSDFGS-package 2
aopt_score 3
avg_grm_minmax_score 4
avg_grm_score 5
avg_grm_self_score 6
cdmax_score 7
cdmean2_score 8
cdmean_score 9
cd_score 10
criteria 10
dopt_score 11
eopt_score 12
FGCM 13
geno 14
gopt_score 14
gv_average_score 15
gv_overall_score 16
mspe_score 17
nt2r 18
optTrain 18
pam_score 20
pevmax_score 21
pev_score 22
r_score 23
SSDFGS 24
subpop 25

Index **26**

`TSDFGS-package` *TSDFGS: Training Set Determination for Genomic Selection*

Description

Finds a training set for genomic selection and the size it needs. The r-score of Ou and Liao (2019) <doi:10.1007/s00122-019-03387-0> is derived from Pearson's correlation between the genomic estimated breeding values and the phenotypic values of the test set. optTrain() searches for a good training set, and SSDFGS() searches for its size. Eighteen criteria are available to score or optimize a training set; see criteria().

Author(s)

Maintainer: Jen-Hsiang Ou <oumark.me@outlook.com> (ORCID) (affiliation: Department of Medical Genetics, National Taiwan University Hospital)

Authors:

- Jen-Hsiang Ou <oumark.me@outlook.com> (ORCID) (affiliation: Department of Medical Genetics, National Taiwan University Hospital)
- Po-Ya Wu <Po-Ya.Wu@hhu.de> (ORCID)
- Chen-Tuo Liao <ctliao@ntu.edu.tw> (ORCID) [thesis advisor]

See Also

Useful links:

- https://www.oumark.net/TSDFGS/
- https://github.com/oumarkme/TSDFGS
- Report bugs at https://github.com/oumarkme/TSDFGS/issues

`aopt_score` *A-optimality score*

Description

$tr[(X'X + \lambda I)^{-1}]$: the trace of the inverse ridge information matrix of the training set. Smaller is better. It uses the training set only.

Usage

```
aopt_score(X, lambda = NULL)
```

Arguments

`X`	A numeric matrix of the training set (row: individual; column: marker or PC).
`lambda`	A positive number, the ridge penalty, or NULL (default) for the default of the criterion.

Details

See `criteria` for the list of criteria and the package website for one page per criterion.

Value

A number.

Author(s)

Jen-Hsiang Ou

References

Akdemir, D., & Isidro-Sanchez, J. (2019). Design of training populations for selective phenotyping in genomic prediction. Scientific Reports, 9, 1446. <doi:10.1038/s41598-018-38081-6>

Examples

```
data(geno)
aopt_score(geno[1:50, ])
```

`avg_grm_minmax_score` *Avg_GRM_MinMax score*

Description

Mean relationship between the training set and the target rows minus the mean relationship within the training set (a = 1, b = 1). Larger is better.

Usage

```
avg_grm_minmax_score(X, X0)
```

Arguments

`X`	A numeric matrix of the training set (row: individual; column: marker or PC).
`X0`	A numeric matrix of the target rows (the test set), with the same columns as X.

Details

See `criteria` for the list of criteria and the package website for one page per criterion.

Value

A number.

Author(s)

Jen-Hsiang Ou

References

Fernandez-Gonzalez, J., Akdemir, D., & Isidro y Sanchez, J. (2023). A comparison of methods for training population optimization in genomic selection. Theoretical and Applied Genetics, 136, 30. <doi:10.1007/s00122-023-04265-6>

Examples

```
data(geno)
avg_grm_minmax_score(geno[1:50, ], geno[51:100, ])
```

`avg_grm_score` *Avg_GRM score*

Description

Mean of the relationship matrix $G = XX'/p$ between the training set and the target rows (a = 1, b = 0 of the generalized Avg_GRM). Larger is better.

Usage

```
avg_grm_score(X, X0)
```

Arguments

`X`	A numeric matrix of the training set (row: individual; column: marker or PC).
`X0`	A numeric matrix of the target rows (the test set), with the same columns as X.

Details

See `criteria` for the list of criteria and the package website for one page per criterion.

Value

A number.

Author(s)

Jen-Hsiang Ou

References

Fernandez-Gonzalez, J., Akdemir, D., & Isidro y Sanchez, J. (2023). A comparison of methods for training population optimization in genomic selection. Theoretical and Applied Genetics, 136, 30. <doi:10.1007/s00122-023-04265-6>

Examples

```
data(geno)
avg_grm_score(geno[1:50, ], geno[51:100, ])
```

`avg_grm_self_score` *Avg_GRM_self score*

Description

Mean of $G = XX'/p$ within the training set. Smaller is better. It uses the training set only.

Usage

```
avg_grm_self_score(X)
```

Arguments

`X`	A numeric matrix of the training set (row: individual; column: marker or PC).

Details

See `criteria` for the list of criteria and the package website for one page per criterion.

Value

A number.

Author(s)

Jen-Hsiang Ou

References

Fernandez-Gonzalez, J., Akdemir, D., & Isidro y Sanchez, J. (2023). A comparison of methods for training population optimization in genomic selection. Theoretical and Applied Genetics, 136, 30. <doi:10.1007/s00122-023-04265-6>

Examples

```
data(geno)
avg_grm_self_score(geno[1:50, ])
```

`cdmax_score` *CDmax score*

Description

Largest, over the target rows, of the coefficient of determination of the ridge model on the columns of `X` (the per-individual value averaged by `cd_score`; default $\lambda = 1$). It protects the best-determined individual, as the CDMAX of STPGA does. Larger is better.

Usage

```
cdmax_score(X, X0, lambda = NULL)
```

Arguments

`X`	A numeric matrix of the training set (row: individual; column: marker or PC).
`X0`	A numeric matrix of the target rows (the test set), with the same columns as X.
`lambda`	A positive number, the ridge penalty, or NULL (default) for the default of the criterion.

Details

See `criteria` for the list of criteria and the package website for one page per criterion.

Value

A number.

Author(s)

Jen-Hsiang Ou

References

Akdemir, D., Sanchez, J. I., & Jannink, J.-L. (2015). Optimization of genomic selection training populations with a genetic algorithm. Genetics Selection Evolution, 47, 38. <doi:10.1186/s12711-015-0116-6>

Examples

```
data(geno)
cdmax_score(geno[1:50, ], geno[51:100, ])
```

`cdmean2_score` *CDmean(v2) score*

Description

Mean of $diag(K_{PT}WK_{TP})/diag(K_{PP})$, the heuristic form of CDmean (GBLUP, $K = XX'/p$, default $\lambda = 1$). Larger is better.

Usage

```
cdmean2_score(X, X0, lambda = NULL)
```

Arguments

`X`	A numeric matrix of the training set (row: individual; column: marker or PC).
`X0`	A numeric matrix of the target rows (the test set), with the same columns as X.
`lambda`	A positive number, the ridge penalty, or NULL (default) for the default of the criterion.

Details

See `criteria` for the list of criteria and the package website for one page per criterion.

Value

A number.

Author(s)

Jen-Hsiang Ou

References

Chen, S.-P., Sung, W.-H., & Liao, C.-T. (2024). Constructing training sets for genomic selection to identify superior genotypes in candidate populations. Theoretical and Applied Genetics, 137, 270. <doi:10.1007/s00122-024-04766-y>

Examples

```
data(geno)
cdmean2_score(geno[1:50, ], geno[51:100, ])
```

cdmean_score *CDmean score (GBLUP form)*

Description

Mean coefficient of determination of the contrasts between each target row and the target mean, for a GBLUP model with $K = XX'/p$, a general mean as the only fixed effect, and penalty $\lambda = \sigma_e^2/\sigma_A^2$ (default 1). Larger is better.

Usage

```
cdmean_score(X, X0, lambda = NULL)
```

Arguments

X	A numeric matrix of the training set (row: individual; column: marker or PC).
X0	A numeric matrix of the target rows (the test set), with the same columns as X.
lambda	A positive number, the ridge penalty, or NULL (default) for the default of the criterion.

Details

See criteria for the list of criteria and the package website for one page per criterion.

Value

A number.

Author(s)

Jen-Hsiang Ou

References

Rincent, R., et al. (2012). Maximizing the reliability of genomic selection by optimizing the calibration set of reference individuals. Genetics, 192, 715-728. <doi:10.1534/genetics.112.141473>
Chen, S.-P., Sung, W.-H., & Liao, C.-T. (2024). Constructing training sets for genomic selection to identify superior genotypes in candidate populations. Theoretical and Applied Genetics, 137, 270. <doi:10.1007/s00122-024-04766-y>

Examples

```
data(geno)
cdmean_score(geno[1:50, ], geno[51:100, ])
```

cd_score *CD-score*

Description

This function calculate CD-score <doi:10.1186/1297-9686-28-4-359> by given training set and test set.

Usage

```
cd_score(X, X0, lambda = NULL)
```

Arguments

X	A numeric matrix. The training set genotypic information matrix can be given as genotype matrix (coded as -1, 0, 1) or principal component matrix (row: sample; column: marker or PC).
X0	A numeric matrix. The test set genotypic information matrix, with the same columns as X.
lambda	A positive number, the ridge penalty, or NULL (default) for the default of the criterion (r-score and PEV: 1 / number of columns; CD: 1).

Value

A number, the CD score (larger is better).

Author(s)

Jen-Hsiang Ou

Examples

```
data(geno)
## Not run: cd_score(geno[1:50, ], geno[51:100, ])
```

criteria *List of training set criteria*

Description

Returns the criteria that `optTrain` can optimize, with the name to give to its `method` argument, the function that scores a given training set, the direction of optimization, and whether the criterion uses a test set.

Usage

```
criteria()
```

Details

Criteria marked `uses_test = FALSE` depend on the training set only. When a test set is given to `optTrain`, such a criterion ignores it (with a warning). One page per criterion on the package website gives the definition, the original reference, and the license review.

Value

A data frame with the columns `method`, `name`, `type`, `score_fun`, `maximize`, and `uses_test`.

Author(s)

Jen-Hsiang Ou

Examples

```
criteria()
```

`dopt_score` *D-optimality score*

Description

$\log\det(X'X + \lambda I)$ of the training set. Larger is better (a larger determinant of the information matrix). It uses the training set only.

Usage

```
dopt_score(X, lambda = NULL)
```

Arguments

`X`	A numeric matrix of the training set (row: individual; column: marker or PC).
`lambda`	A positive number, the ridge penalty, or NULL (default) for the default of the criterion.

Details

See `criteria` for the list of criteria and the package website for one page per criterion.

Value

A number.

Author(s)

Jen-Hsiang Ou

References

Akdemir, D., & Isidro-Sanchez, J. (2019). Design of training populations for selective phenotyping in genomic prediction. Scientific Reports, 9, 1446. <doi:10.1038/s41598-018-38081-6>

Examples

```
data(geno)
dopt_score(geno[1:50, ])
```

---

eopt_score *E-optimality score*

Description

Largest eigenvalue of $(X'X + \lambda I)^{-1}$, that is $1/(\lambda_{min}(X'X) + \lambda)$. Smaller is better. It uses the training set only. When X has more columns than rows, $X'X$ is singular and the score is $1/\lambda$ for every training set.

Usage

```
eopt_score(X, lambda = NULL)
```

Arguments

X	A numeric matrix of the training set (row: individual; column: marker or PC).
lambda	A positive number, the ridge penalty, or NULL (default) for the default of the criterion.

Details

See criteria for the list of criteria and the package website for one page per criterion.

Value

A number.

Author(s)

Jen-Hsiang Ou

References

Akdemir, D., & Isidro-Sanchez, J. (2019). Design of training populations for selective phenotyping in genomic prediction. Scientific Reports, 9, 1446. <doi:10.1038/s41598-018-38081-6>

Examples

```
data(geno)
eopt_score(geno[1:50, ])
```

FGCM *Fit logistic growth curve model*

Description

A function for fitting logisti growth model

Usage

```
FGCM(geno, nt = NULL, n_iter = NULL, multi.threads = FALSE, threads = NULL)
```

Arguments

geno	Genotype information saved as a dataframe. Columns represent variants (SNPs or PCs).
nt	A numerical vector of training set sample size for estimating logistic growth curve parameters
n_iter	Number of simulation of each training set size. Automatically gave a suitable number by default.
multi.threads	Default: FALSE. If TRUE, the simulations run in parallel (OpenMP) on 75% of the available processors (limited by OMP_NUM_THREADS), or half of them when there are 4 or fewer.
threads	Default: NULL. The number of threads; it overrides multi.threads. See nt2r for the random numbers of the parallel run.

Value

Estimation of parameters.

Examples

```
data(geno)
## Not run: FGCM(geno)
```

geno *Genotype information*

Description

A PCA matrix of rice genotype information. This data was published by Zhao et al. (2011) <doi:10.1038/ncomms1467>

Usage

```
geno
```

Format

A numeric matrix (PCA) with 404 rows (sample) and 404 columns (PCs).

Source

http://www.ricediversity.org/data/

Examples

```
data(geno)
```

gopt_score *G-optimality score (GOPTPEV)*

Description

Largest eigenvalue of the prediction error variance matrix of the target rows, $X_0(X'X + \lambda I)^{-1}X_0'$, with an intercept column added. Smaller is better.

Usage

```
gopt_score(X, X0, lambda = NULL)
```

Arguments

X	A numeric matrix of the training set (row: individual; column: marker or PC).
X0	A numeric matrix of the target rows (the test set), with the same columns as X.
lambda	A positive number, the ridge penalty, or NULL (default) for the default of the criterion.

Details

See criteria for the list of criteria and the package website for one page per criterion.

Value

A number.

Author(s)

Jen-Hsiang Ou

References

Akdemir, D., & Isidro-Sanchez, J. (2019). Design of training populations for selective phenotyping in genomic prediction. Scientific Reports, 9, 1446. <doi:10.1038/s41598-018-38081-6>

Examples

```
data(geno)
gopt_score(geno[1:50, ], geno[51:100, ])
```

---

gv_average_score *GVaverage score*

Description

$tr(K_T)$ with $K_T = XX'/p$ (average genomic variation). Larger is better. It uses the training set only.

Usage

```
gv_average_score(X)
```

Arguments

X	A numeric matrix of the training set (row: individual; column: marker or PC).

Details

See `criteria` for the list of criteria and the package website for one page per criterion.

Value

A number.

Author(s)

Jen-Hsiang Ou

References

Chen, S.-P., Sung, W.-H., & Liao, C.-T. (2024). Constructing training sets for genomic selection to identify superior genotypes in candidate populations. Theoretical and Applied Genetics, 137, 270. <doi:10.1007/s00122-024-04766-y>

Examples

```
data(geno)
gv_average_score(geno[1:50, ])
```

`gv_overall_score` *GVoverall score*

Description

$\log\det K_T$ with $K_T = XX'/p$, the relationship matrix of the training set (overall genomic variation). Larger is better. It needs at least as many columns as rows, otherwise K_T is singular and the score is minus infinity.

Usage

```
gv_overall_score(X)
```

Arguments

`X`	A numeric matrix of the training set (row: individual; column: marker or PC).

Details

See `criteria` for the list of criteria and the package website for one page per criterion.

Value

A number.

Author(s)

Jen-Hsiang Ou

References

Chen, S.-P., Sung, W.-H., & Liao, C.-T. (2024). Constructing training sets for genomic selection to identify superior genotypes in candidate populations. Theoretical and Applied Genetics, 137, 270. <doi:10.1007/s00122-024-04766-y>

Examples

```
data(geno)
gv_overall_score(geno[1:50, ])
```

`mspe_score` *MSPE score of ridge regression*

Description

The approximation of the expected mean squared prediction error of ridge regression, $1+tr[X_0AA'X_0']/n_0+tr[(X_0-X_0AX)'(X_0-X_0AX)]$ with $A=X'(XX'+\lambda I)^{-1}$. Smaller is better. The default penalty is $\lambda=1$.

Usage

```
mspe_score(X, X0, lambda = NULL)
```

Arguments

`X`	A numeric matrix of the training set (row: individual; column: marker or PC).
`X0`	A numeric matrix of the target rows (the test set), with the same columns as X.
`lambda`	A positive number, the ridge penalty, or NULL (default) for the default of the criterion.

Details

See `criteria` for the list of criteria and the package website for one page per criterion.

Value

A number.

Author(s)

Jen-Hsiang Ou

References

Chen, S.-P., Sung, W.-H., & Liao, C.-T. (2024). Constructing training sets for genomic selection to identify superior genotypes in candidate populations. Theoretical and Applied Genetics, 137, 270. <doi:10.1007/s00122-024-04766-y>

Examples

```
data(geno)
mspe_score(geno[1:50, ], geno[51:100, ])
```

nt2r *Simulate r-scores of each training set size*

Description

Calculate r-scores (un-target) by in parallel.

Usage

```
nt2r(geno, nt, n_iter = 30, multi.threads = FALSE, threads = NULL)
```

Arguments

geno	A numeric dataframe of genotype, column represent sites (genotype coding as 1, 0, -1)
nt	Numeric. Number of training set size
n_iter	Times of iteration. (default = 30)
multi.threads	Default: FALSE. If TRUE, the iterations run in parallel (OpenMP, any system with OpenMP support) on 75% of the available processors (limited by OMP_NUM_THREADS), or half of them when there are 4 or fewer.
threads	Default: NULL. The number of threads; it overrides multi.threads. The parallel iterations use their own random generators, seeded from R's generator. The values then depend on set.seed() but not on threads, and they are not those of the serial run with the same seed.

Value

A vector of r-scores of each iteration

Examples

```
data(geno)
## Not run: nt2r(geno, 50)
```

optTrain *Optimal training set determination*

Description

This function is designed for determining optimal training set.

Usage

```
optTrain(
  geno,
  cand,
  n.train,
  subpop = NULL,
  test = NULL,
  method = "rScore",
  min.iter = NULL,
  console = TRUE,
  seed = NULL,
  lambda = NULL,
  threads = NULL
)
```

Arguments

geno	A numeric matrix of principal components (rows: individuals; columns: PCs).
cand	An integer vector of which rows of individuals are candidates of the training set in the geno matrix.
n.train	The size of the target training set. This could be determined with the help of the ssdfgp function provided in this package.
subpop	A character vector of sub-population's group name. The algorithm will ignore the population structure if it remains NULL. Without `test`, it has one entry per candidate (same order as cand); with `test`, it has one entry per row of geno.
test	An integer vector of which rows of individuals are in the test set in the geno matrix. The algorithm will use an un-target method if it remains NULL.
method	A criterion name; see `criteria` for the list. "rScore" (default), "PEV" and "CD" are the criteria of v2.4.2.
min.iter	Minimum iteration of all methods can be appointed. One should always check if the algorithm is converged or not. A minimum iteration will set by considering the candidate and test set size if it remains NULL.
console	Default: TRUE. Set it to FALSE if you don't want the function printing out the number count of each iteration.
seed	Default: NULL. An integer passed to `set.seed` before the search, which changes the global random-number state. Leave it NULL to use the current state.
lambda	Default: NULL. A positive ridge penalty; NULL uses the default of each criterion (see `criteria`).
threads	Default: NULL (one thread). The number of threads that Eigen may use inside the matrix products of one search (OpenMP). For the exchange search it only helps when geno has many columns (several hundred; about 1.3 to 1.5 times faster at 404 columns on 2 threads) and the training set found is the same apart from rounding in the last digits. For the targeted r-score and PEV search without sub-populations (a genetic algorithm), the 30 candidate sets of each generation are scored in parallel, and the result does not depend on `threads`. To run many independent searches in parallel use `SSDFGS`.

Details

All twelve search routes of v2.4.2 (with or without sub-populations, targeted or un-targeted, and the three criteria) are kept. The search moves, their random-number order, the default numbers of iterations, and the stopping rule are the same as in v2.4.2, so that a given seed returns the training set of v2.4.2 (apart from the float-to-double change of the scores, see Details of `r_score`). The r-score is the criterion of Ou and Liao (2019) <doi:10.1007/s00122-019-03387-0>.

Only the r-score, PEV, and CD are tested carefully: they reproduce the output recorded from v2.4.2 on all twelve routes. The 15 criteria added in v3 follow their papers and run on all routes, but they are not carefully tested; evaluate them on your own data before relying on them.

Value

This function will return a list with OPTtrain (a vector of chosen optimal training set), TOPscore (highest scores of before iteration), ITERscore (criteria scores of each iteration), n.iter (number of iterations run), and converged (TRUE if the search stopped by the convergence rule, FALSE if it stopped at twice `min.iter`).

Author(s)

Jen-Hsiang Ou

References

Ou, J.-H., & Liao, C.-T. (2019). Training set determination for genomic selection. Theoretical and Applied Genetics, 132, 2781-2792. <doi:10.1007/s00122-019-03387-0>

Examples

```
data(geno)
## Not run: optTrain(geno, cand = 1:404, n.train = 100)
```

---

pam_score *PAM (k-medoids) cost*

Description

Sum, over the target rows, of the Euclidean distance to the nearest training row: the cost that partitioning around medoids minimizes when the training rows are the medoids. Smaller is better.

Usage

```
pam_score(X, X0)
```

Arguments

`X`	A numeric matrix of the training set (row: individual; column: marker or PC).
`X0`	A numeric matrix of the target rows (the test set), with the same columns as X.

Details

See `criteria` for the list of criteria and the package website for one page per criterion.

Value

A number.

Author(s)

Jen-Hsiang Ou

References

Kaufman, L., & Rousseeuw, P. J. (1990). Finding groups in data: an introduction to cluster analysis. Wiley. <doi:10.1002/9780470316801> Guo, T., et al. (2019). Optimal designs for genomic selection in hybrid crops. Molecular Plant, 12, 390-401. <doi:10.1016/j.molp.2018.12.022> Rio, S., et al. (2021). Genomic prediction and training set optimization in a structured Mediterranean oat population. Theoretical and Applied Genetics, 134, 3595-3609. <doi:10.1007/s00122-021-03916-w>

Examples

```
data(geno)
pam_score(geno[1:50, ], geno[51:100, ])
```

---

pevmax_score *PEVmax score*

Description

Largest prediction error variance among the target rows, with an intercept column added: $\max_i x'_{0i}(X'X+\lambda I)^{-1}x_{0i}$. Smaller is better.

Usage

```
pevmax_score(X, X0, lambda = NULL)
```

Arguments

`X`	A numeric matrix of the training set (row: individual; column: marker or PC).
`X0`	A numeric matrix of the target rows (the test set), with the same columns as X.
`lambda`	A positive number, the ridge penalty, or NULL (default) for the default of the criterion.

Details

See `criteria` for the list of criteria and the package website for one page per criterion.

Value

A number.

Author(s)

Jen-Hsiang Ou

References

Akdemir, D., & Isidro-Sanchez, J. (2019). Design of training populations for selective phenotyping in genomic prediction. Scientific Reports, 9, 1446. <doi:10.1038/s41598-018-38081-6>

Examples

```
data(geno)
pevmax_score(geno[1:50, ], geno[51:100, ])
```

---

pev_score *PEV score*

Description

This function calculate prediction error variance (PEV) score <doi:10.1186/s12711-015-0116-6> by given training set and test set.

Usage

```
pev_score(X, X0, lambda = NULL)
```

Arguments

X	A numeric matrix. The training set genotypic information matrix can be given as genotype matrix (coded as -1, 0, 1) or principal component matrix (row: sample; column: marker or PC).
X0	A numeric matrix. The test set genotypic information matrix, with the same columns as X.
lambda	A positive number, the ridge penalty, or NULL (default) for the default of the criterion (r-score and PEV: 1 / number of columns; CD: 1).

Value

A number, the PEV score (smaller is better).

Author(s)

Jen-Hsiang Ou

Examples

```
data(geno)
## Not run: pev_score(geno[1:50, ], geno[51:100, ])
```

`r_score` *r-score*

Description

This function calculate r-score <doi:10.1007/s00122-019-03387-0> by given training set and test set.

Usage

```
r_score(X, X0, lambda = NULL)
```

Arguments

`X`	A numeric matrix. The training set genotypic information matrix can be given as genotype matrix (coded as -1, 0, 1) or principal component matrix (row: sample; column: marker or PC).
`X0`	A numeric matrix. The test set genotypic information matrix, with the same columns as X.
`lambda`	A positive number, the ridge penalty, or NULL (default) for the default of the criterion (r-score and PEV: 1 / number of columns; CD: 1).

Details

The value equals the r-score of v2.4.2 (to its single-precision accuracy of about 1e-7). Since v3 the computation uses p x p matrices when the number of columns p is at most the training set size n, and n x n matrices otherwise, and the result is returned in double precision. The term `1 - 1/n0` (n0: test set size) of v2.4.2 is kept as published in the software; see NEWS.md.

Value

A number, the r-score.

Author(s)

Jen-Hsiang Ou

References

Ou, J.-H., & Liao, C.-T. (2019). Training set determination for genomic selection. Theoretical and Applied Genetics, 132, 2781-2792. <doi:10.1007/s00122-019-03387-0>

Examples

```
data(geno)
## Not run: r_score(geno[1:50, ], geno[51:100, ])
```

SSDFGS *Sample size determination for genomic selection*

Description

This function is designed to generate an operating curve for sample size determination

Usage

```
SSDFGS(geno, nt = NULL, n_iter = NULL, multi.threads = FALSE, threads = NULL)
```

Arguments

geno	A numeric data frame carried genotype information (column: PCs, row: sample)
nt	A numeric vector carried training set sizes for r-score simulation.
n_iter	Number of iterations for estimating parameters.
multi.threads	Default: FALSE. If TRUE, the simulations run in parallel (OpenMP) on 75% of the available processors (limited by OMP_NUM_THREADS), or half of them when there are 4 or fewer.
threads	Default: NULL. The number of threads; it overrides multi.threads. See nt2r for the random numbers of the parallel run.

Value

A list with the operating curve and growth curve figures (OC.fig and GC.fig; ggplot objects, or NULL when the ggplot2 package is not installed), the logistic growth curve parameters, and the operating curve data (OC.fit).

Author(s)

Jen-Hsiang Ou & Po-Ya Wu

Examples

```
data(geno)
## Not run: SSDFGS(geno)
```

subpop *Sub-population information*

Description

Sub-population information of samples. This data was published by Zhao et al. (2011) <doi:10.1038/ncomms1467>

Usage

```
subpop
```

Format

A character vector.

Source

http://www.ricediversity.org/data/

Examples

```
data(subpop)
```

Index

∗ **datasets**
geno, 14
subpop, 25

aopt_score, 3
avg_grm_minmax_score, 4
avg_grm_score, 5
avg_grm_self_score, 6

cd_score, *7*, 10
cdmax_score, 7
cdmean2_score, 8
cdmean_score, 9
criteria, *3–9*, 10, *11*, *12*, *14–17*, *19*, *21*

dopt_score, 11

eopt_score, 12

FGCM, 13

geno, 14
gopt_score, 14
gv_average_score, 15
gv_overall_score, 16

mspe_score, 17

nt2r, *13*, 18, *24*

optTrain, *10*, 18

pam_score, 20
pev_score, 22
pevmax_score, 21

r_score, *20*, 23

SSDFGS, *19*, 24
subpop, 25

TSDFGS *(*TSDFGS-package*)*, 2
TSDFGS-package, 2