Package 'tensorregress'

November 4, 2019

Title Generalized Tensor Regression with Covariates on Multiple Modes

Version 1.0

Date 2019-10-28

Description Implement the generalized tensor regression in Xu, Hu and Wang (2019) <arXiv:1910.09499>. Solve tensor-response regression given covariates on multiple modes with alternating updating algorithm.

Maintainer Zhuoyan Xu <zxu444@stat.wisc.edu>

Imports pracma,speedglm,MASS

Depends rTensor

License GPL (>= 2)

Encoding UTF-8

LazyData true

RoxygenNote 6.1.1

NeedsCompilation no

Author Zhuoyan Xu [aut, cre, cph],
Jiaxin Hu [aut, cph],
Miaoyan Wang [aut, cph]

Repository CRAN

Date/Publication 2019-11-04 16:30:08 UTC

R topics documented:

`sele_rank` *Rank selection*

Description

Estimate the Tucker rank of coefficient tensor based on BIC criterion. The choice of BIC aims to balance between the goodness-of-fit for the data and the degree of freedom in the population model.

Usage

```
sele_rank(tsr, X_covar1 = NULL, X_covar2 = NULL, X_covar3 = NULL,
  rank_range, Nsim = 10, cons = "non", lambda = 0.1, alpha = 1,
  solver = "CG", dist)
```

Arguments

`tsr`	response tensor with 3 modes
`X_covar1`	covariate on first mode
`X_covar2`	covariate on second mode
`X_covar3`	covariate on third mode
`rank_range`	a matrix containing rank candidates on each row
`Nsim`	max number of iterations if update does not convergence
`cons`	the constraint method, "non" for without constraint, "vanilla" for global scale down at each iteration, "penalty" for adding log-barrier penalty to object function.
`lambda`	penalty coefficient for "penalty" constraint
`alpha`	max norm constraint on linear predictor
`solver`	solver for solving object function when using "penalty" constraint, see "details"
`dist`	distribution of response tensor, see "details"

Details

For rank selection, recommend using non-constraint version.

Constraint `penalty` adds log-barrier regularizer to general object function (negative log-likelihood). The main function uses solver in function "optim" to solve the objective function. The "solver" passes to the argument "method" in function "optim".

`dist` specifies three distributions of response tensor: binary, poisson and normal distributions.

Value

a list containing the following:

`rank` a vector with selected rank with minimal BIC

`result` a matrix containing rank candidate and its loglikelihood and BIC on each row

Examples

```
seed=24
dist="binary"
data=sim_data(seed, whole_shape = c(20,20,20),
core_shape=c(3,3,3),p=c(5,5,5),dist=dist, dup=5, signal=4)
rank_range = rbind(c(3,3,3),c(3,3,2),c(3,2,2),c(2,2,2),c(3,2,3))
re = sele_rank(data$tsr[[1]],data$X_covar1,data$X_covar2,data$X_covar3,
 rank_range = rank_range,Nsim=10,cons = 'non',dist = dist)
```

`sim_data` *Simulation of tensor regression models*

Description

Generate response tensors with multiple covariates under different simulation models, specifically for tensors with 3 modes

Usage

```
sim_data(seed, whole_shape = c(20, 20, 20), core_shape = c(3, 3, 3),
  p = c(3, 3, 0), dist, dup, signal, block = rep(FALSE, 3))
```

Arguments

`seed`	seed for generating data
`whole_shape`	a vector containing dimension of the tensor
`core_shape`	a vector containing Tucker rank of the coefficient tensor
`p`	a vector containing numbers of covariates on each mode, see "details"
`dist`	distribution of response tensor, see "details"
`dup`	number of simulated tensors from the same linear predictor
`signal`	a scalar controlling the max norm of the linear predictor
`block`	a vector containing boolean variables, see "details"

Details

By default non-positive entry in `p` indicates no covariate on the corresponding mode of the tensor.

`dist` specifies three distributions of response tensor: binary, poisson and normal distribution.

`block` specifies whether coefficient factor matrix is a membership matrix, set to `TRUE` when utilizing the stochastic block model

Value

a list containing the following:

`tsr` a list of simulated tensors, with the number of replicates specified by `dup`

`X_covar1` a matrix, covariate on first mode

`X_covar2` a matrix, covariate on second mode

`X_covar3` a matrix, covariate on third mode

`W` a list of orthogonal coefficient matrices - one for each mode, with the number of columns given by `core_shape`

`G` an array, core tensor with size specified by `core_shape`

`C_ts` an array, coefficient tensor, Tucker product of `G`,`A`,`B`,`C`

`U` an array, linear predictor,i.e. Tucker product of `C_ts`,`X_covar1`,`X_covar2`,`X_covar3`

Examples

```
seed = 34
dist = 'binary'
data=sim_data(seed, whole_shape = c(20,20,20), core_shape=c(3,3,3),
p=c(5,5,5),dist=dist, dup=5, signal=4)
```

---

tensor_regress *Generalized Tensor Regression*

Description

Tensor-response regression given covariates on multiple modes. Main function in the package. The function takes a response tensor, multiple covariate matrices, and a desired Tucker rank as input. The output is a constrained MLE for the coefficient tensor.

Usage

```
tensor_regress(tsr, X_covar1 = NULL, X_covar2 = NULL,
  X_covar3 = NULL, core_shape, Nsim = 20, cons = c("non", "vanilla",
  "penalty"), lambda = 0.1, alpha = 1, solver = "CG",
  dist = c("binary", "poisson", "normal"))
```

Arguments

`tsr`	response tensor with 3 modes
`X_covar1`	covariate on first mode
`X_covar2`	covariate on second mode
`X_covar3`	covariate on third mode
`core_shape`	the Tucker rank of the regression coefficients
`Nsim`	max number of iterations if update does not convergence

cons	the constraint method, "non" for without constraint, "vanilla" for global scale down at each iteration, "penalty" for adding log-barrier penalty to object function
lambda	penalty coefficient for "penalty" constraint
alpha	max norm constraint on linear predictor
solver	solver for solving object function when using "penalty" constraint, see "details"
dist	distribution of the response tensor, see "details"

Details

Constraint penalty adds log-barrier regularizer to general object function (negative log-likelihood). The main function uses solver in function "optim" to solve the objective function. The "solver" passes to the argument "method" in function "optim".

dist specifies three distributions of response tensor: binary, poisson and normal distribution.

Value

a list containing the following:

W a list of orthogonal coefficient matrices - one for each mode, with the number of columns given by core_shape

G an array, core tensor with the size specified by core_shape

C_ts an array, coefficient tensor, Tucker product of G,A,B,C

U linear predictor,i.e. Tucker product of C_ts,X_covar1,X_covar2,X_covar3

lglk a vector containing loglikelihood at convergence

sigma a scalar, estimated error variance (for Gaussian tensor) or dispersion parameter (for Bernoulli and Poisson tensors)

violate a vector listing whether each iteration violates the max norm constraint on the linear predictor, 1 indicates violation

Examples

```
seed = 34
dist = 'binary'
data=sim_data(seed, whole_shape = c(20,20,20), core_shape=c(3,3,3),
p=c(5,5,5),dist=dist, dup=5, signal=4)
re = tensor_regress(data$tsr[[1]],data$X_covar1,data$X_covar2,data$X_covar3,
core_shape=c(3,3,3),Nsim=10, cons = 'non', dist = dist)
```

Index