Package 'ELPASO'

February 21, 2022

Type Package

Title Ensemble LASSO for Phylogenetic Analysis of Shifts with OU

Version 0.1.0

Author Wensha Zhang [aut], Tobias Kenney [ctb], Lam Si Tung Ho [ctb]

Maintainer Wensha Zhang <wn209685@dal.ca>

Description

Provides tool functions for phylognetic analysis of evolutionary shifts. It uses OU process and ensemble LASSO model to detect the incontinuous changes along during the evolution process.

License GPL (>= 3)

Encoding UTF-8

LazyData true

Depends R (>= 3.6.0),
glmnet,
ncvreg,
phylolm,
PIGShift,
psych

Imports l1ou

Remotes khabbazian/l1ou

RoxygenNote 7.1.2

R topics documented:

---

backward_selection   *backward_selection*

Description

This function provides forward selection given a variable ranking sequence

Usage

```
backward_selection(
  tree,
  Y,
  seq,
  increasing = TRUE,
  criterion = c("BIC", "pBIC", "pBICess", "AICc", "mBIC"),
  stop = TRUE,
  maxShifts = 10,
  opt
)
```

Arguments

tree	ultrametric tree of class phylo with branch lengths, and edges in postorder
Y	Y trait vector without missing entries
seq	A vector gives that selection order of the shifts. (For example, the ensembled rank)
increasing	If a smaller value in param seq means more important variable, increasing = TRUE, vice versa.
criterion	the model selection criterion
stop	TRUE/FALSE. If the selection process will be stopped if score stops decreasing
maxShifts	the maximum number of shifts
opt	options for l1ou functions

Value

s.c	The position of shifts
score	The criterion score

---

combine_ranking_seqs   *combine_ranking_seqs*

Description

ensemble the ranking sequnces with median/arith.mean/geom.mean/quantile

Usage

```
combine_ranking_seqs(rank_seqs, method = "quantile", q = 0.25)
```

Arguments

rank_seqs	A ranking sequence of all the variables
method	The ensemble method "median"/"arith.mean"/"geom.mean"/"quantile"
q	Required only when method is "quantile", specify the quantile value

Value

A vector (length is number of potential shift postions). The ensembled ranking sequence

---

ELPASO	*ELPASO*

Description

This function is the main function of the package, the inputs of the function are the trait vector and phylogenetic tree, the outputs are the estimated positions and change values of the shifts. And the results are based on ensemble variable selection.

Usage

```
ELPASO(
  tree,
  Y,
  criterion = c("BIC", "pBIC"),
  maxShifts = 20,
  nsamples = 200,
  xtype = c("simpX", "orgX"),
  penalty = c("LASSO", "SCAD"),
  ensemble_method = "quantile",
  q = 0.25
)
```

Arguments

tree	ultrametric tree of class phylo with branch lengths, and edges in postorder
Y	Y trait vector without missing entries
criterion	the criterion for model selection, "BIC"(default),"pBIC"
maxShifts	The max number of shifts. Integer.
nsamples	The number subsamples are generated. Integer.
xtype	type of design matrix, 'simpX'(simple design matrix, default) or 'orgX'(original design matrix)
penalty	Penalty function to be applied. Either "LASSO"(default) or "SCAD"
ensemble_method,	
	the way to ensemble the ranking seqs. 'median','quantile'(default),'arith.mean','geom.mean'
q,	if using "quantile" as the ensemble method, specify the quantile value in this parameter. The default value is 0.25, which is a good choice given by the simulation results

Value

Y	input trait vector/matrix.
tree	input tree.
ensemble_rank	The final ensemble rank of shift positions
shifts	estimated shift positions, i.e. vector of indices of edges where the estimated shifts occur.
shift.means	estimates change of the expectation of the shift values
nShifts	estimated number of shifts.
alpha	maximum likelihood estimate of the adaptation rate α
sigma2	maximum likelihood estimate of the variance rate σ^2
fitted.values	fitted values, i.e. estimated trait means.
residuals	residuals.
logLik	log likelihood of given model
criterion	The criterion for model selection
score	information criterion value of the estimated shift configuration.
penalty	Penalty function to be applied

Examples

```
require(l1ou)
data('lizard.tree')
data('lizard.traits')
tree = lizard.tree
Y = as.vector(lizard.traits[,1])
ELMODEL = ELPASO(tree,Y)
```

---

forward_backward_selection

forward_backward_selection

Description

This function provides forward+bacward selection given a variable ranking sequence

Usage

```
forward_backward_selection(
  tree,
  Y,
  seq,
  increasing = TRUE,
  criterions,
  stops,
  maxShifts,
  direction = c("same", "opposite"),
  opt
)
```

Arguments

tree	ultrametric tree of class phylo with branch lengths, and edges in postorder
Y	Y trait vector without missing entries
seq	A vector gives that selection order of the shifts. (For example, the ensembled rank)
increasing	If a smaller value in param seq means more important variable, increasing = TRUE, vice versa.
criterions	A two-element vector. The model selection criterions for forward selection and backward selection respectively
stops	A two-element vector. Each element is TRUE/FALSE. If the selection process will be stopped if score stops decreasing
maxShifts	the maximum number of shifts
direction	'same'(default)/'opposite'. If the forward selection and the backward selection are produced in the same direction.
opt	options for l1ou functions

Value

s.c	The position of shifts
score	The criterion score

---

forward_selection   *forward_selection*

Description

This function provides forward selection given a variable ranking sequence

Usage

```
forward_selection(
  tree,
  Y,
  seq,
  increasing = TRUE,
  criterion = c("BIC", "pBIC", "pBICess", "AICc", "mBIC"),
  stop = TRUE,
  maxShifts = 20,
  opt
)
```

Arguments

tree	ultrametric tree of class phylo with branch lengths, and edges in postorder
Y	Y trait vector without missing entries
seq	A vector gives that selection order of the shifts. (For example, the ensembled rank)

increasing	If a smaller value in param seq means more important variable, increasing = TRUE, vice versa.
criterion	the model selection criterion
stop	TRUE/FALSE. If the selection process will be stopped if score stops decreasing
maxShifts	the maximum number of shifts
opt	options for l1ou functions

Value

s.c	The position of shifts
score	The criterion score

---

get_prediction_likelihood

get_prediction_likelihood

Description

This function calculates the prediction log likelihood to evaluate the accuracy of estimates

Usage

```
get_prediction_likelihood(
  tree,
  Y,
  selected_variable,
  alpha,
  test_data,
  xtype = c("simpX", "orgX")
)
```

Arguments

tree	ultrametric tree of class phylo with branch lengths, and edges in postorder
Y	Y trait vector without missing entries
selected_variable	
	The variables selected
alpha	The selective rate
test_data	A bunch of testing data given by the true model
xtype	the type of design matrix

Value

The prediction log likelihood value

get_ranking_seqs *get_ranking_seqs*

Description

This function generates subsamples of a given combination of tree and trait values, applys LASSO/SCAD on each subsample, and produces ranking sequence of each subsample

Usage

```
get_ranking_seqs(
  tree,
  Y,
  nsamples,
  size,
  replace = FALSE,
  alpha,
  sigma2,
  sigma2_error = 0,
  xtype = c("simpX", "orgX"),
  penalty = c("LASSO", "SCAD")
)
```

Arguments

tree	ultrametric tree of class phylo with branch lengths, and edges in postorder
Y	Y trait vector without missing entries
nsamples	The number subsamples are generated. Integer.
size	The size of each subsample. Integer.
replace	if replace while subsampling, TRUE/FALSE(default)
alpha	the adaption rate (known or estimated)
sigma2	the variance of OU model (known or estimated)
sigma2_error	the variance of gaussian measurement error
xtype	type of design matrix, 'simpX'(simple design matrix, default) or 'orgX'(original design matrix)
penalty	Penalty function to be applied. Either "LASSO"(default) or "SCAD"

Value

A matrix, each row is a ranking sequence of variables, each column represents the rank sets of each variable

get_test_data   *get_test_data*

Description

This function generates a bunch of testing data given the tree and true shifts

Usage

```
get_test_data(
  tree,
  true_variable,
  beta,
  alpha,
  n_test,
  xtype = c("simpX", "orgX")
)
```

Arguments

tree	ultrametric tree of class phylo with branch lengths, and edges in postorder
true_variable	The true shifts
beta	The coefficients of shifts
alpha	The selective rate
n_test	The number of testing datasets
xtype	the type of design matrix

Value

A matrix, each row is a testing dataset for the true model

Index