

Package ‘longke’

July 7, 2023

Type Package

Title Nonparametric Predictive Model for Sparse and Irregular Longitudinal Data

Version 0.1.0

Maintainer Shixuan Wang <wangs43@miamioh.edu>

Description The proposed method aims at predicting the longitudinal mean response trajectory by a kernel-based estimator. The kernel estimator is constructed by imposing weights based on subject-wise similarity on L2 metric space between predictor trajectories as well as time proximity. Users could also perform variable selections to derive functional predictors with predictive significance by the proposed multiplicative model with multivariate Gaussian kernels.

Imports tidy, bvl, fdapace, mvtnorm, dplyr, purrr

License GPL-3

Encoding UTF-8

RoxygenNote 7.2.3

NeedsCompilation no

Author Shixuan Wang [aut, cre],
Seonjin Kim [aut],
Hyunkeun Cho [aut],
Won Chang [aut]

Repository CRAN

Date/Publication 2023-07-07 15:50:02 UTC

R topics documented:

longke-package	2
datagen	2
FPCA_trajectory	3
KE_bwselection	4
KE_fit	5
KE_trajSCB	6

Index**8**

longke-package	<i>Nonparametric predictive model for sparse and irregular longitudinal data</i>
----------------	--

Description

This package contains the function for fitting nonparametric predictive model for sparse and irregular longitudinal data

References

Wang S, Kim S, Cho H, Chang W. Nonparametric predictive model for sparse and irregular longitudinal data. (2023+)

datagen	<i>Simulate longitudinal data</i>
---------	-----------------------------------

Description

Function used to simulate sample sparse and irregular longitudinal data

Usage

```
datagen(ntotal, ntest, t_all, t_split, seed)
```

Arguments

ntotal	Number of total longitudinal subjects
ntest	Number of total longitudinal subjects in the testing set
t_all	Vector of discrete measurement time (i.e 1,2,3,4,...)
t_split	A measurement time where the longitudinal response is of interest to predict after this t_split
seed	Seed to derive replicable data

Value

A list containing two elements

train A long format data matrix containing one functional response (yy) and two functional predictors (xx,zz) with (ntotal-ntest) subjects

test A long format data matrix containing one functional response (yy) and two functional predictors (xx,zz) with (ntest) subjects

Examples

```
data = datagen(ntotal=350,ntest=50,t_all=1:50,t_split=25,seed=1)
data$test
data$train
```

FPCA_trajectory	<i>FPCA_trajectory</i>
-----------------	------------------------

Description

Function used to perform functional principal component analysis (FPCA) for a single functional variable

Usage

```
FPCA_trajectory(data,...)
```

Arguments

data	A long format data matrix containing 3 columns ordered by time, subject ID, variable where the measurement time of the longitudinal data should be discretized
...	Arguments to be passed to <code>fdapace::FPCA</code>

Value

A list containing two elements

fzca_target	A FPCA object
target_fit	A <code>num.t x num.sub</code> matrix containing the imputed longitudinal trajectories where <code>num.t</code> is the total number of the discrete measurement time and <code>num.sub</code> is the total number of subjects

References

Carroll, C., Gajardo, A., Chen, Y., Dai, X., Fan, J., Hadjipantelis, P. Z., ... & Wang, J. L. (2020). *fdapace: Functional data analysis and empirical dynamics*. R package version 0.5, 4.

Yao, F., Müller, H. G., & Wang, J. L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American statistical association*, 100(470), 577-590.

Examples

```

t_all = 1:50
data = datagen(ntotal=350,ntest=50,t_all=t_all,t_split=25,seed=1)
data.sample = data$test[,c(1,2,3)]
# In this case, num.t=50 and num.sub=50 since we only used 50 subjects in the testing data
data.FPCA = FPCA_trajectory(data.sample,list(dataType='Sparse',
      error=FALSE, kernel='gauss', verbose=FALSE, nRegGrid=length(t_all)))
data.FPCA$target_fit

```

KE_bwselection	<i>KE_bwselection</i>
----------------	-----------------------

Description

Function used to perform leave-one-subject-out cross validation to select optimal time bandwidth (b_s) and trajectory bandwidth (b_w)

Usage

```
KE_bwselection(data,bw_time,bw_subj,T1,T2)
```

Arguments

data	A long format data matrix containing columns ordered by time, subject ID, response, predictor1, predictor2, ... where the measurement time of the longitudinal data should be discretized
bw_time	A numeric vector that contains the candidate time bandwidths
bw_subj	A numeric vector that contains the candidate trajectory bandwidths
T1	A measurement time domain where the functional predictors are measured within
T2	A measurement time domain where the functional response is of interest to predict

Value

A list containing 3 elements

BWSelecStep	Total SSE for each bandwidth combination
optimalBW	A vector containing the optimal time/trajectory bandwidth
RunningTime	Running time of the bandwidth selection

References

Wang S, Kim S, Cho H, Chang W. Nonparametric predictive model for sparse and irregular longitudinal data. (2023+)

Examples

```

t_all = 1:50
T1=c(1,25);T2=c(26,50)
data = datagen(ntotal=350,ntest=50,t_all=t_all,t_split=25,seed=1)
data.sample = data$train
bwsele.toy = KE_bwselection(data=data.sample,
bw_time=c(1,2),bw_subj=c(0.1,0.5),T1=T1,T2=T2)
bwsele.toy$optimalBW

```

KE_fit

KE_fit

Description

Function used to predict response trajectory by nonparametric kernel estimator

Usage

```
KE_fit(train,test,T1,T2,bw_time,bw_subj,alpha=0.05,seed=1,coefCI=FALSE)
```

Arguments

train	A long format data matrix containing columns ordered by time, subject ID, response, predictor1, predictor2, ... where the measurement time of the longitudinal data should be discretized within T1.
test	A long format data matrix containing columns ordered by time, subject ID, response, predictor1, predictor2, ... where the measurement time of the longitudinal data should be discretized within T2.
T1	A measurement time domain where the functional predictors are measured within
T2	A measurement time domain where the functional response is of interest to predict
bw_time	(optimal) time bandwidth
bw_subj	(optimal) trajectory/subject bandwidth
alpha	confidence level for bootstrap CI of alpha_0, alpha_1, ...
seed	A random seed fo producing replicable bootstrap CI of alpha_0, alpha_1, ...
coefCI	Logical statement: TRUE to derive bootstrap CI of alpha0, alpha1, ... default is FALSE

Value

A list containing 6 elements

testTraj	A num.test x num.T2 matrix containing num.test subjects' trajectories where num.T2 is the total number of the discrete measurement time over T2
proxycoeff	Coefficient estimation for the non-negative least square regression. From left to right they are alpha_0, alpha_1, ...
f pca . fit	A list containing FPCA fit for the functional predictors and the functional response
w . hat	A list containing num.test elements where ith element contains the proxy distance/similarity between ith testing subject and other training subjects
bootCI . mean	Bootstrap confidence interval of alpha_0, alpha_1, ...
input . list	A list containing the input arguments

References

Wang S, Kim S, Cho H, Chang W. Nonparametric predictive model for sparse and irregular longitudinal data. (2023+)

Examples

```
t_all = 1:50
T1=c(1,25);T2=c(26,50)
data = datagen(ntotal=350,ntest=50,t_all=t_all,t_split=25,seed=1)
train = data$train
test = data$test
ke.fit = KE_fit(train=train,test=test,T1=T1,T2=T2,bw_time=1,bw_subj = 0.2)
```

KE_trajSCB

KE_trajSCB

Description

Function used to derive simultaneous confidence band (SCB) for the predicted response trajectory

Usage

```
KE_trajSCB(KE.fit.object,nboot=500,alpha=0.05)
```

Arguments

KE.fit.object	An object whose class is KE (you can get it by letting ke = KE.fit())
nboot	Number of bootstrap sample size to construct SCB
alpha	Confidence level for bootstrap SCB of predicted response trajectory

Value

A list containing num.test elements where the num.test represents the number of testing subjects. Within each element, there is a list containing 3 elements:

se	A vector containing standard errors at each discrete measurement time
traj.upper	A vector containing upper bound of the testing subject at each measurement time
traj.lower	A vector containing lower bound of the testing subject at each measurement time

References

Wang S, Kim S, Cho H, Chang W. *Nonparametric predictive model for sparse and irregular longitudinal data*. (2023+)

Kim, S., Ryan Cho, H., & Kim, M. O. (2021). *Predictive generalized varying-coefficient longitudinal model*. *Statistics in Medicine*, 40(28), 6243-6259.

See Also

[KE_fit](#)

Examples

```
t_all = 1:50
T1=c(1,25);T2=c(26,50)
data = datagen(ntotal=350,ntest=50,t_all=t_all,t_split=25,seed=1)
train = data$train
test = data$test
ke.fit = KE_fit(train=train,test=test,T1=T1,T2=T2,bw_time=1,bw_subj = 0.2)
ketraj.toy = KE_trajSCB(KE.fit.object = ke.fit,
                       nboot=10,alpha=0.05)
```

Index

`datagen`, [2](#)

`FPCA_trajectory`, [3](#)

`KE_bwselection`, [4](#)

`KE_fit`, [5](#), [7](#)

`KE_trajSCB`, [6](#)

`longke-package`, [2](#)