Package ‘GPLSVCM’

December 1, 2021

Type Package

Title Generalized Partial Linear Spatially Varying Coefficient Model

Version 0.1.0

Author Liying Jin, Jingru Mu

Maintainer <liyingj@ksu.edu>

Depends R (>= 2.10),
BPST,
grpreg

Description Identify the linear and nonlinear components of the modeland fit the corresponding Generalized Partial Linear Spatially Varying Coefficient Model.

License GPL (>= 2)

Encoding UTF-8

LazyData true

Imports mgcv,
MGLM,
MASS,
Triangulation,
plyr,
graphics,
stats,
plot3D,
boot,
Matrix

RoxygenNote 7.1.1

R topics documented:

`compute_PIs` *Compute the prediction intervals for responses of new test points from a fitted generalized partial linear spatially varying coefficient model.*

Description

`compute_PIs` compute the prediction intervals for responses of new test points from a fitted `gplsvcm` object based on a selected prediction method among Jackknife, Jackknife+ and K-fold cross validation (CV+), and return the prediction intervals for responses of predicted points.

Usage

```
compute_PIs(
  Y_train,
  X_train,
  ind_l,
  ind_nl,
  U_train,
  X_pred,
  U_pred,
  V,
  Tr,
  d = 2,
  r = 1,
  lambda = 10^seq(-6, 6, by = 0.5),
  family,
  off = 0,
  r_theta = c(2, 8),
  eps = 0.01,
  method = "CV+",
  cp = 0.95,
  nfold = 10
)
```

Arguments

`Y_train` The response variable,a `n` by one matrix where `n` is the number of observations in the training data set .

`X_train` The design matrix of `n` by `np` where `np` is the number of covariates (including intercept if intercept exists). Each row is a vector of the covariates for an observation in the training data set.

`ind_l` The vector of the indexes which indicate the columns of linear covariates in `X_train`.

`ind_nl`	The vector of the indexes which indicate the columns of nonlinear covariates in `X_train`.
`U_train`	A `n` by two matrix where each row is the coordinates of an observation in the training data set.
`X_pred`	The design matrix for prediction (including intercept if intercept exists).
`U_pred`	The matrix of coordinates for prediction.
`V`	A `N` by two matrix of vertices of a triangulation, where `N` is the number of vertices and each row is the coordinates for a vertex.
`Tr`	A `n_Tr` by three triangulation matrix, where `n_Tr` is the number of triangles in the triangulation and each row is the indices of vertices in `V`.
`d`	The degree of piecewise polynomials – default is 2.
`r`	The smoothness parameter and `r` < `d` – default is 1.
`lambda`	The vector of the candidates of penalty parameter – default is grid points of 10 to the power of a sequence from -6 to 6 by 0.5.
`family`	The family object which specifies the distribution and link to use (see `glm` and `family`).
`off`	The offset – default is 0.
`r_theta`	The vector of the upper and lower bound of an interval to search for an additional parameter `theta` for negative binomial scenario – default is c(2,8).
`eps`	The error tolerance for the Pearson estimate of the scale parameter, which is as close to 1, when estimating an additional parameter `theta` for negative binomial scenario – default is 0.01.
`method`	The prediction method used in the computation, options are "CV+", "Jackknife" and "Jackknife+" – default is "CV+".
`cp`	The desired coverage level for the prediction intervals – default is 0.95.
`nfold`	The number of folds for CV+ method – default is 10.

Details

The construction of the polynomial spline functions is via `basis`.

Value

A data frame of computed prediction intervals for responses of the predicted points.

References

Barber et al.(2021) Predictive inference with the jackknife+. Ann.Statist.49(1):486-507 `https://projecteuclid.org/journals/annals-of-statistics/volume-49/issue-1/Predictive-inference-with-the` `10.1214/20-AOS1965.full`

Examples

```
# See an example of gplsvcm_fitwMIDF.
```

Crash_Texas — *Crash data in Texas*

Description

A dataset containing the number of car crashes within each census tract in Texas of year 2107 and other variables of 4771 locations.

Usage

```
data(Crash_Texas)
```

Format

A data frame with 4771 rows and 7 variables:

count off-roadway crash frequencies

vmt log of vehicle miles traveled

pop log of total population

old proportion of people age 65 and older

hispanics proportion of Hispanics

lon longitude of a location

lat latitude of a location

Examples

```
data(Crash_Texas)
count <- Crash_Texas$count
hist(count)
summary(count)
```

CV_fit — *Fit the model with K fold cross validation and compute the CV residuals*

Description

This is an internal function of package GPLSVCM.

Usage

```
CV_fit(
  Y_train,
  X_l_train,
  X_nl_train,
  ind_l,
  ind_nl,
  U_train,
  X_l_pred,
```

```
    X_nl_pred,
    U_pred,
    B_pred,
    V,
    Tr,
    d,
    r,
    lambda,
    family,
    off,
    r_theta,
    eps,
    nfold
  )
```

---

cv_gplsvcm — *Calculate the K-fold Cross Validation Mean Square Prediction Error from a fitted generalized partial linear spatially varying coefficient model.*

Description

`cv_gplsvcm` implements K-fold cross-validation from a fitted `gplsvcm` object, and returns the mean squared prediction error (MSPE).

Usage

```
cv_gplsvcm(
  Y,
  X,
  ind_l,
  ind_nl,
  U,
  V,
  Tr,
  d = 2,
  r = 1,
  lambda = 10^seq(-6, 6, by = 0.5),
  family,
  off = 0,
  r_theta = c(2, 8),
  eps = 0.01,
  nfold = 10
)
```

Arguments

`Y`	The response variable,a `n` by one matrix where `n` is the number of observations.
`X`	The design matrix of `n` by `np` where `np` is the number of covariates (including intercept if intercept exists). Each row is a vector of the covariates for an observation.

ind_l	The vector of the indexes which indicate the columns of linear covariates in X.
ind_nl	The vector of the indexes which indicate the columns of nonlinear covariates in X.
U	A n by two matrix where each row is the coordinates of an observation.
V	A N by two matrix of vertices of a triangulation, where N is the number of vertices and each row is the coordinates for a vertex.
Tr	A n_Tr by three triangulation matrix, where n_Tr is the number of triangles in the triangulation and each row is the indices of vertices in V.
d	The degree of piecewise polynomials – default is 2.
r	The smoothness parameter and r < d – default is 1.
lambda	The vector of the candidates of penalty parameter – default is grid points of 10 to the power of a sequence from -6 to 6 by 0.5.
family	The family object which specifies the distribution and link to use (see glm and family).
off	The offset – default is 0.
r_theta	The vector of the upper and lower bound of an interval to search for an additional parameter theta for negative binomial scenario – default is c(2,8).
eps	The error tolerance for the Pearson estimate of the scale parameter, which is as close to 1, when estimating an additional parameter theta for negative binomial scenario – default is 0.01.
nfold	The number of folds for cross validation – default is 10.

Details

The construction of the polynomial spline functions is via basis.

Value

The mean square prediction error (MSPE).

Examples

```
# See an example of gplsvcm_fitwMIDF.
```

Datagenerator *Generating populations for simulation.*

Description

Datagenerator is used to generate samples on horseshoe domain for Scenario 1 (Gaussian), Scenario 2 (Poisson) and Scenario 3 (negative binomial).

Usage

```
Datagenerator(family, ngrid)
```

Arguments

- `family` The family object, specifying the distribution and link to use. Choose "gaussian" for Gaussian distribution, "poisson" for poisson distribution, and "nb_bps" for negative binomial distribution.
- `ngrid` The distance between grid points – default is set to 0.02.

Details

This function used the package `mgcv`, see `fs.boundary` and `fs.test`

Value

A data matrix with a response ('y'), true coefficient functions ('m1' and 'm2'), nonlinear covariates ('x1' and 'x2'), linear covariates ('x3' and 'x4') and locations ('u' and 'v').

Examples

```
family=nb_bps()
ngrid = 0.02
pop = Datagenerator(family, ngrid)
```

---

dev_est *Calculate the generalized deviance of a fitted model*

Description

`dev_est` is an internal function of `gplsvcm_fitwMIDF`.

Usage

```
dev_est(X_l, X_nl, mfit, family)
```

Arguments

- `X_l` The matrix of linear covariates for observations.
- `X_nl` The matrix of nonlinear covariates for observations.
- `mfit` A list of information returned by function `gplsvcm_est`
- `family` The family object which specifies the distribution and link to use.

gplsvcm_aglasso *Fitting the generalized partial linear spatially varying coefficient model with Variable Selection and Model Structure Identification by (Adaptive) group lasso.*

Description

gplsvcm_aglasso perform variable selection and model structure identification at the same time to select the covariates with the linear and nonlinear effects respectively out of a large number of covariates using BIC or 10 fold cross validation and then fit the corresponding generalized partial linear spatially varying coefficient model.

Usage

```
gplsvcm_aglasso(
  Y,
  X,
  uvpop = NULL,
  U,
  index = NULL,
  V,
  Tr,
  d = 2,
  r = 1,
  penalty = "agrLasso",
  lambda1 = exp(seq(log(1e-04), log(1), length.out = 50)),
  lambda2 = exp(seq(log(1e-04), log(1), length.out = 50)),
  family,
  criteria = "BIC",
  lambda = 10^seq(-6, 6, by = 0.5),
  off = 0
)
```

Arguments

Y	The response variable,a n by one matrix where n is the number of observations.
X	The design matrix (without intercept) of n by np where np is the number of covariates. Each row is a vector of the covariates for an observation.
uvpop	The coordinates of population grid points over the domain, default is NULL.
U	A n by two matrix where each row is the coordinates of an observation.
index	The row indexes of the observed data points U in the population grid points uvpop.
V	A N by two matrix of vertices of a triangulation, where N is the number of vertices and each row is the coordinates for a vertex.
Tr	A n_Tr by three triangulation matrix, where n_Tr is the number of triangles in the triangulation and each row is the indices of vertices in V.
d	The degree of piecewise polynomials – default is 2.
r	The smoothness parameter and r $<$ d – default is 1.

penalty	The shrinkage method for variable selection and model structure identification, options are "agrLasso" for adaptive group lasso and "grLasso" for group lasso – default is "agrLasso".
lambda1	The sequence of lambda values for group lasso,also need to be specified for computing the weight if using adaptive group lasso – default is a grid of 50 lambda values that ranges uniformly on the log scale over 0.0001 to 1.
lambda2	The sequence of lambda values used in the adaptive part for adaptive group lasso – default is a grid of 50 lambda values that ranges uniformly on the log scale over 0.0001 to 1.
family	The family object which specifies the distribution and link to use (see glm and family).
criteria	A character string specifying the criteria of selecting lambda for (adaptive) group lasso. "BIC" is to use traditional Bayesian Information Criteria, "CV" is to use 10-fold cross validation – default is "BIC".
lambda	The vector of the candidates of smoothing penalty parameter – default is grid points of 10 to the power of a sequence from -6 to 6 by 0.5.
off	The offset – default is 0.

Details

The gplsvcm_aglasso function is used to fit a generalized partial linear spatially varying coefficient model when there is a large number of covariates and the linear and nonlinear parts of the design matrix X are not known before analysis. The construction of the polynomial spline functions is via basis. It first perform a variable selection and model structure identification through adaptive group lasso via grpreg or cv.grpreg and output the selected model by specifying the parameters ind_l and ind_nl of the function gplsvcm_fit. Then the selected model is fitted by the function gplsvcm_fit.

Value

The function returns a list of fitted object information from S3 class "gplsvcm", see the items of the list from gplsvcm_fit.

References

Wood, S., & Wood, M. S. (2015). Package ‘mgcv’. R package version, 1, 29.

Breheny P (2016).grpreg: Regularization Paths for Regression Models with Grouped Covari-ates.Rpackage version 3.0-2, URLhttps://CRAN.R-project.org/packages=grpreg.

Wang L, Wang G, Li X, Mu J, Yu S, Wang Y, Kim M, Wang J (2019). BPST: Smoothing viaBivariate Spline over Triangulation.Rpackage version 1.0, URLhttps://GitHub.com/funstatpackages/BPST.

Examples

```
# Population:
family=gaussian()
ngrid = 0.02

# Data generation:
pop = Datagenerator(family, ngrid)
N=nrow(pop)

# Triangulations and setup:
```

```
Tr = Tr0; V = V0; n = 1000; d = 2; r = 1;

# set up for smoothing parameters in the penalty term:
lambda_start=0.0001; lambda_end=10; nlambda=10;
lambda=exp(seq(log(lambda_start),log(lambda_end),length.out=nlambda))

# Generate Sample:
ind_s=sample(N,n,replace=FALSE)
data=as.matrix(pop[ind_s,])
Y=data[,1]; X=data[,c(6:9)]; U=data[,c(10:11)];

# True coefficents
alpha=data[,c(2:3)]; beta=data[,c(4:5)];

# Fit the model with model selection based on AIC:
mfit1 = gplsvcm_fitwMIDF(Y, X, U, V, Tr, d , r , lambda,family,k_n=NULL,
method="AIC",off = 0,r_theta = c(2, 8), eps= 0.01)

# Fit the model with model selection based on BIC:
mfit2 = gplsvcm_fitwMIDF(Y, X, U, V, Tr, d , r , lambda,family,k_n=NULL,
method="BIC",off = 0,r_theta = c(2, 8), eps= 0.01)

# prediction intervals:
ind_l=mfit2$ind_l; ind_nl=mfit2$ind_nl;
set.seed(123)
PIs=compute_PIs(Y,X,ind_l,ind_nl,U,X,U,V,Tr,d,r,lambda,family,off = 0,
r_theta = c(2, 8), eps= 0.01,method="CV+", cp=0.95, nfold = 10)

# prediction:
Y_hat = gplsvcm_predict(mfit2, X, U)

# k-fold cross-validation:
set.seed(123)
MSPE = cv_gplsvcm(Y,X,ind_l,ind_nl,U,V,Tr,d,r,lambda,family,off = 0,r_theta =
c(2, 8), eps= 0.01,nfold=10)

# plot the estimated coefficients
gplsvcm_plot(mfit2,gridnumber=100,display=c(1,1),xlab=c("u1","u1"),
ylab=c("u2","u2"),main=c(expression(paste("The Estimated Surface for","
",hat(alpha)[1])),expression(paste("The Estimated Surface for","
",hat(alpha)[2]))))
```

`gplsvcm_est` *Estimation for GPLSCVMs*

Description

This is an internal function of `GPLSVCM` which is used in function `gplsvcm_fit` and `gplsvcm_fitwMIDF`.

Usage

```
gplsvcm_est(Y, X_l, X_nl, U, V, Tr, d, r, lambda, family, off, r_theta, eps)
```

Arguments

`Y`	The response variable,a `n` by one matrix where `n` is the number of observations.
`X_l`	The matrix of linear covariates for observations, of dimension `n` by `np_l` where `n` is number of observations and `np_l` is the number of linear covarites.
`X_nl`	The matrix of nonlinear covariates for observations, of dimension `n` by `np_nl` where `n` is number of observations and `np_nl` is the number of nonlinear covarites.
`U`	A `n` by two matrix where each row is the coordinates of an observation.
`V`	A `N` by two matrix of vertices of a triangulation, where `N` is the number of vertices and each row is the coordinates for a vertex.
`Tr`	A `n_Tr` by three triangulation matrix, where `n_Tr` is the number of triangles in the triangulation and each row is the indices of vertices in `V`.
`d`	The degree of piecewise polynomials.
`r`	The smoothness parameter and `r` < `d`.
`lambda`	The vector of the candidates of penalty parameter.
`family`	The family object which specifies the distribution and link to use (see `glm` and `family`).
`off`	The offset.
`r_theta`	The vector of the upper and lower bound of an interval to search for an additional parameter `theta` for negative binomial scenario.
`eps`	The error tolerance for the Pearson estimate of the scale parameter, which is as close as possible to 1, when estimating an additional parameter `theta` for negative binomial scenario.

Details

The construction of the polynomial spline functions is via `basis`.

`gplsvcm_fit` *Fitting generalized partial linear spatially varying coefficient regression models*

Description

`gplsvcm_fit` fits the generalized partial linear spatially varying coefficient models.

Usage

```
gplsvcm_fit(
  Y,
  X,
  ind_l,
  ind_nl,
  U,
  V,
  Tr,
  d = 2,
```

```
  r = 1,
  lambda = 10^seq(-6, 6, by = 0.5),
  family,
  off = 0,
  r_theta = c(2, 8),
  eps = 0.01
)
```

Arguments

Y	The response variable,a n by one matrix where n is the number of observations.
X	The design matrix of n by np where np is the number of covariates (including intercept if intercept exists). Each row is a vector of the covariates for an observation.
ind_l	The vector of the indexes which indicate the columns of linear covariates in X.
ind_nl	The vector of the indexes which indicate the columns of nonlinear covariates in X.
U	A n by two matrix where each row is the coordinates of an observation.
V	A N by two matrix of vertices of a triangulation, where N is the number of vertices and each row is the coordinates for a vertex.
Tr	A n_Tr by three triangulation matrix, where n_Tr is the number of triangles in the triangulation and each row is the indices of vertices in V.
d	The degree of piecewise polynomials – default is 2.
r	The smoothness parameter and r < d – default is 1.
lambda	The vector of the candidates of penalty parameter – default is grid points of 10 to the power of a sequence from -6 to 6 by 0.5.
family	The family object which specifies the distribution and link to use (see glm and family).
off	The offset – default is 0.
r_theta	The vector of the upper and lower bound of an interval to search for an additional parameter theta for negative binomial scenario – default is c(2,8).
eps	The error tolerance for the Pearson estimate of the scale parameter, which is as close to 1, when estimating an additional parameter theta for negative binomial scenario – default is 0.01.

Details

The gplsvcm_fit function is for fitting the Generalized Partial Linear Spatially Varying Coefficient Models (GPLSVCM) when the model structure is specified before analysis, that is, the parameters ind_l and ind_nl are specified before fitting the model.The construction of the polynomial spline functions is via basis. If the true model structure is not known before model fitting, we recommend using another function gplsvcm_fitwMIDF in this package. Note, if ind_l is specified as a null vector, gplsvcm_fit will fit a glm model, and if ind_nl is specified as a null vector,gplsvcm_fit will fit a gsvcm model.

Value

The function returns a list of fitted object information from S3 class "gplsvcm" with the following items:

alpha_hat	The estimated coefficients for the nonlinear component of the model.
beta_hat	The estimated coefficients for the linear component of the model.
Qtheta	The estimated spline coefficients.
lambda_sel	The selected penalty parameter through generalized cross-validation (GCV) for bivariate penalized spline over trianulation estimation.
gcv	The GCV statistics for lambda_sel.
df	The effective degree of freedom for the model.
theta	The estimated additional parameter theta for negative binomial scenario.
Y	The matrix of responses, of dimension n by one where n is number of observations inside the triangulation.
X_nl	The matrix of nonlinear covariates for observations inside the triangulation, of dimension n by np_l where n is number of observations inside the triangulation and np_l is the number of linear covarites.
X_l	The matrix of linear covariates for observations inside the triangulation, n by np_l where n is number of observations inside the triangulation and np_nl is the number of nonlinear covarites.
U	The matrix of coordinates for observations inside the triangulation, of dimension n by 2 where n is number of observations inside the triangulation and each row is the coordinates of an observation.
ind_l	The vector of the indexes which indicate the columns of linear covariates in X.
ind_nl	The vector of the indexes which indicate the columns of nonlinear covariates in X.
family	The family object.
V	The matrix of vertices of the triangulation, with dimension N by two where N is the number of vertices of the triangulation and each row is the coordinates for a vertex
Tr	The triangulation matrix of of the triangulation, with dimention n_Tr by three, where n_Tr is the number of triangles in the triangulation and each row is the indices of vertices in V.
d	The degree of piecewise polynomials.
r	The smoothness parameter.
B	The spline basis function of dimension n by n_Tr*{(d+1)(d+2)/2}, where n and n_Tr are the number of observations and the number of triangles inside the given triangulation respectively, d is the degree of the spline. If some points do not fall in the triangulation, the generation of the spline basis will not take those points into consideration.
Q2	The Q2 matrix after QR decomposition of the smoothness matrix H.
K	The thin-plate energy function.
ind_inside	A vector contains the indexes of all the points which are inside the triangulation.
tria_all	The area of each triangle within the given triangulation.
lambda	The vector of the candidates of penalty parameter used in fitting the model.
r_theta	The vector of the upper and lower bound of an interval to search for an additional parameter theta used in negative binomial scenario.
off	The offset.
eps	The error tolerance used for the Pearson estimate of the scale parameter for negative binomial scenario.

Examples

```
# Population:
family=poisson()
ngrid = 0.02

# Data generation:
pop = Datagenerator(family, ngrid)
N=nrow(pop)

# Triangulations and setup:
Tr = Tr0; V = V0; n = 1000; d = 2; r = 1;

# set up for smoothing parameters in the penalty term:
lambda_start=0.0001; lambda_end=10; nlambda=10
lambda=exp(seq(log(lambda_start),log(lambda_end),length.out=nlambda))

# Generate Sample:
ind_s=sample(N,n,replace=FALSE)
data=as.matrix(pop[ind_s,])
Y=data[,1]; alpha=data[,c(2:3)]; beta=data[,c(4:5)];
X=cbind(rep(1,length(Y)),data[,c(6:9)]); ind_l=c(1,4,5); ind_nl=c(2,3);
U=data[,c(10:11)];

# Fit the model:
mfit = gplsvcm_fit(Y, X,ind_l,ind_nl,U, V, Tr, d , r , lambda,family,off = 0,
r_theta = c(2, 8), eps= 0.01)
```

`gplsvcm_fitwMIDF` *Fitting the generalized partial linear spatially varying coefficient model with Model Selection*

Description

`gplsvcm_fitwMIDF` perform a model selection procedure to identify the linear and nonlinear components first and then fit the corresponding generalized partial linear spatially varying coefficient model.

Usage

```
gplsvcm_fitwMIDF(
  Y,
  X,
  U,
  V,
  Tr,
  d = 2,
  r = 1,
  lambda = 10^seq(-6, 6, by = 0.5),
  family,
  k_n = NULL,
  method = "BIC",
  off = 0,
```

```
    r_theta = c(2, 8),
    eps = 0.01
  )
```

Arguments

Y	The response variable,a n by one matrix where n is the number of observations.
X	The design matrix of n by np where np is the number of covariates (including intercept if intercept exists). Each row is a vector of the covariates for an observation.
U	A n by two matrix where each row is the coordinates of an observation.
V	A N by two matrix of vertices of a triangulation, where N is the number of vertices and each row is the coordinates for a vertex.
Tr	A n_Tr by three triangulation matrix, where n_Tr is the number of triangles in the triangulation and each row is the indices of vertices in V.
d	The degree of piecewise polynomials – default is 2.
r	The smoothness parameter and r < d – default is 1.
lambda	The vector of the candidates of penalty parameter – default is grid points of 10 to the power of a sequence from -6 to 6 by 0.5.
family	The family object which specifies the distribution and link to use (see glm and family).
k_n	The penalty parameter used in the model selection criteria. It need to be supplied only when the argument method is set to NULL –default is NULL.
method	The type of model selection criteria, options are "AIC","BIC" and NULL which correspond to k_n=2, k_n=log(n) and k_n=k_n respectively – default is "BIC".
off	The offset – default is 0.
r_theta	The vector of the upper and lower bound of an interval to search for an additional parameter theta for negative binomial scenario – default is c(2,8).
eps	The error tolerance for the Pearson estimate of the scale parameter, which is as close as possible to 1, when estimating an additional parameter theta for negative binomial scenario – default is 0.01.

Details

The gplsvcm_fitwMIDF function is used to fit a generalized partial linear spatially varying coefficient model when the linear and nonlinear parts of the design matrix X are not known before analysis. The construction of the polynomial spline functions is via basis. It first perform a model selection based on Generalized Information Criterion (GIC) and output the selected model by specifying the parameters ind_l and ind_nl of the function gplsvcm_fit. Then the selected model is fitted by the function gplsvcm_fit.

Value

The function returns a list of fitted object information from S3 class "gplsvcm", see the items of the list from gplsvcm_fit.

References

Zhang et al.(2010) Regularization Parameter Selections via Generalized Information Criterion.https://www.tandfonline.com/doi/abs/10.1198/jasa.2009.tm08013

Examples

```
# Population:
family=poisson()
ngrid = 0.02

# Data generation:
pop = Datagenerator(family, ngrid)
N=nrow(pop)

# Triangulations and setup:
Tr = Tr0; V = V0; n = 1000; d = 2; r = 1;

# set up for smoothing parameters in the penalty term:
lambda_start=0.0001; lambda_end=10; nlambda=10;
lambda=exp(seq(log(lambda_start),log(lambda_end),length.out=nlambda))

# Generate Sample:
ind_s=sample(N,n,replace=FALSE)
data=as.matrix(pop[ind_s,])
Y=data[,1]; X=data[,c(6:9)]; U=data[,c(10:11)];

# True coefficents
alpha=data[,c(2:3)]; beta=data[,c(4:5)];

# Fit the model with model selection based on AIC:
mfit1 = gplsvcm_fitwMIDF(Y, X, U, V, Tr, d , r , lambda,family,k_n=NULL,
method="AIC",off = 0,r_theta = c(2, 8), eps= 0.01)

# Fit the model with model selection based on BIC:
mfit2 = gplsvcm_fitwMIDF(Y, X, U, V, Tr, d , r , lambda,family,k_n=NULL,
method="BIC",off = 0,r_theta = c(2, 8), eps= 0.01)

# prediction intervals:
ind_l=mfit2$ind_l; ind_nl=mfit2$ind_nl;
set.seed(123)
PIs=compute_PIs(Y,X,ind_l,ind_nl,U,X,U,V,Tr,d,r,lambda,family,off = 0,
r_theta = c(2, 8), eps= 0.01,method="CV+", cp=0.95, nfold = 10)

# prediction:
Y_hat = gplsvcm_predict(mfit2, X, U)

# k-fold cross-validation:
set.seed(123)
MSPE = cv_gplsvcm(Y,X,ind_l,ind_nl,U,V,Tr,d,r,lambda,family,off = 0,r_theta =
c(2, 8), eps= 0.01,nfold=10)

# plot the estimated coefficients
gplsvcm_plot(mfit2,gridnumber=100,display=c(1,1),xlab=c("u1","u1"),
ylab=c("u2","u2"),main=c(expression(paste("The Estimated Surface for","
",hat(alpha)[1])),expression(paste("The Estimated Surface for","
",hat(alpha)[2]))))
```

gplsvcm_plot — *Produces coefficient function plots for a fitted generalized partial linear spatially varying coefficient model.*

Description

gplsvcm_plot produces the plots of the estimated coefficient functions from a fitted gplsvcm object.

Usage

```
gplsvcm_plot(
  mfit,
  gridnumber = 100,
  display = NULL,
  xlab = NULL,
  ylab = NULL,
  main = NULL,
  ...
)
```

Arguments

mfit	A fitted gplsvcm object returned from function gplsvcm_fit or gplsvcm_fitwMIDF.
gridnumber	The number of grid points on one range for plots – default is 100.
display	If supplied then it is the vector for specifying how to display the estimated surfaces for the coefficient functions, used in par(mfrow=).
xlab	If supplied then is the vector of characters where each element is the x label for the estimated surface of one coefficient function.
ylab	If supplied then is the vector of characters where each element is the y label for the estimated surface of one coefficient function.
main	If supplied then is the vector of characters where each element is the title for the estimated surface of one coefficient function.
...	other graphics parameters to pass on to plotting commands. See details in image2D.

Details

This function used package Triangulation and plot3D, see TriPlot and image2D.

Value

None

Examples

```
# See an example of gplsvcm_fitwMIDF.
```

`gplsvcm_predict` *Predictions for responses of new test points from a fitted generalized partial linear spatially varying coefficient model.*

Description

`gplsvcm_predict` is used to make predictions for the responses of predicted points from a fitted `gplsvcm` object.

Usage

```
gplsvcm_predict(mfit, Xpred, Upred)
```

Arguments

`mfit`	A fitted `gplsvcm` object returned from function `gplsvcm_fit` or `gplsvcm_fitwMIDF`.
`Xpred`	The design matrix for prediction (including intercept if intercept exists).
`Upred`	The matrix of cooridinates for prediction.

Details

The construction of the polynomial spline functions is via `basis`

Value

A vector of predicted response.

Examples

```
# See an example of gplsvcm_fitwMIDF.
```

`loo_fit` *Fit the model with the i-th training data point removed and compute the leave one out residuals*

Description

This is an internal function of package `GPLSVCM`.

Usage

```
loo_fit(
  Y_train,
  X_l_train,
  X_nl_train,
  ind_l,
  ind_nl,
  U_train,
  X_l_pred,
```

```
    X_nl_pred,
    U_pred,
    B_pred,
    V,
    Tr,
    d,
    r,
    lambda,
    family,
    off,
    r_theta,
    eps
  )
```

nb_bps *Negative Binomial Family*

Description

Negative Binomial Family

Usage

```
nb_bps(link = "log", theta)
```

Details

This is a built in function in GgAM.

Index