Package 'SpatialVS'

July 21, 2025

Title Spatial Variable Selection

Version 1.1

Author Yili Hong, Li Xu, Yimeng Xie, and Zhongnan Jin

Maintainer Yili Hong <yilihong@vt.edu>

Description Perform variable selection for the spatial Poisson regression model under the adaptive elastic net penalty. Spatial count data with covariates is the input. We use a spatial Poisson regression model to link the spatial counts and covariates. For maximization of the likelihood under adaptive elastic net penalty, we implemented the penalized quasi-likelihood (PQL) and the approximate penalized loglikelihood (APL) methods. The proposed methods can automatically select important covariates, while adjusting for possible spatial correlations among the responses. More details are available in Xie et al. (2018, <doi:10.48550/arXiv.1809.06418>). The package also contains the Lyme disease dataset, which consists of the disease case data from 2006 to 2011, and demographic data and land cover data in Virginia. The Lyme disease case data were collected by the Virginia Department of Health. The demographic data (e.g., population density, median income, and average age) are from the 2010 census. Land cover data were obtained from the Multi-Resolution Land Cover Consortium for 2006.

Depends R (>= 3.3.0)

License GPL-2

Imports MASS,nlme, fields

Encoding UTF-8

LazyData true

RoxygenNote 6.0.1.9000

NeedsCompilation no

Repository CRAN

Date/Publication 2018-11-10 05:10:03 UTC

Contents

---

control.default	*Global variable of spatial variable selection, contains optimization tuning parameters.*

Description

control.default is a list that gives the default setting for the optimization procedures

Usage

```
control.default
```

Format

maxIter maximum number of iterations in the [SpatialVS](#)

iwls convergence criterion for the iterative weighted least squares

tol1 convergence criteria for beta estimation in [SpatialVS](#)

tol2 convergence criteria for theta estimation in [SpatialVS](#)

Examples

```
control.default=list(maxIter=200,iwls=10^(-4),tol1=10^(-3),tol2=10^(-3))
```

---

lyme.svs.eco0.dat	*The Lyme disease dataset with Eco id=0*

Description

The Lyme disease dataset contains case data from 2006 to 2011 demographic data and land cover data in Virginia. Lyme disease case data were collected by Virginia Department of Health. Eco id = 0 represents northern/western subregion, which includes Northern Piedmont, Blue Ridge, Ridge and Valley, and Central Appalachian.

The column names of X are listed as follows. The first column is the intercept column.

x1: Dvlpd_NLCD06; Percentage of developed land in each census tract

x2: Forest_NLCD06; Percentage of forest in each census tract

x3: Herbaceous_NLCD06; Percentage of herbaceous in each census tract

x4: X.Water_NLCD06; Percentage of water in each census tract

x5: Tract_Frag06; Sum of area of forested fragments in each census tract divided by the total area

x6: FragPerim06; Sum of forest fragment perimeters in each census tract divided by the total area

x7: CWED_DF06; CWED of developed-forest edge

x8: TECI_DF06; TECI of developed-forest edge

x9: CWED_FH06; CWED of forest-herbaceous edge

x10: TECI_FH06; TECI of forest-herbaceous edge

x11: CWED_HD06; CWED of herbaceous-developed edge

x12: TECI_HD06; TECI of herbaceous-developed edge

x13: Pop_den; Tract population density in 2010

x14: Median_age; Median age at each census tract in 2010

x15: Mean_income; Mean income (inflation adjusted) at each census tract in 2010

Usage

```
lyme.svs.eco0.dat
```

Format

y integer vector, output count, each element represents the disease count in one census tract.

X Numeric matrix, matrix of covariates, includes percentage of developed land in each census tract, percentage of forest in each census tract, etc.

offset Numeric vector, vector for offset values, each element represents the log of population divided by 20,000 in census tract.

location Numeric matrix, location for each census tract.

geoid Numeric vector, geo id for each census tract.

References

Xie, Y., Xu, L., Li, J., Deng, X., Hong, Y., Kolivras, K., and Gaines, D. N. (2018). Spatial Variable Selection and An Application to Virginia Lyme Disease Emergence. Preprint, arXiv:1809.06418 [stat.AP].

Examples

```
data("lyme.svs.eco0")
lyme.svs.eco0.dat$dist=distmat.compute(location=lyme.svs.eco0.dat$location, dist.min=0.4712249)
```

lyme.svs.eco1.dat	*The Lyme disease dataset with Eco id=1*

Description

The Lyme disease dataset contains case data from 2006 to 2011 demographic data and land cover data in Virginia. Lyme disease case data were collected by Virginia Department of Health. Eco id = 1 represents the southern/eastern subregion, which includes Piedmont, Middle Atlantic Coastal Plain, and Southeastern Plains.

The column names of X are listed as follows. The first column is the intercept column.

x1: Dvlpd_NLCD06; Percentage of developed land in each census tract

x2: Forest_NLCD06; Percentage of forest in each census tract

x3: Herbaceous_NLCD06; Percentage of herbaceous in each census tract

x4: X.Water_NLCD06; Percentage of water in each census tract

x5: Tract_Frag06; Sum of area of forested fragments in each census tract divided by the total area

x6: FragPerim06; Sum of forest fragment perimeters in each census tract divided by the total area

x7: CWED_DF06; CWED of developed-forest edge

x8: TECI_DF06; TECI of developed-forest edge

x9: CWED_FH06; CWED of forest-herbaceous edge

x10: TECI_FH06; TECI of forest-herbaceous edge

x11: CWED_HD06; CWED of herbaceous-developed edge

x12: TECI_HD06; TECI of herbaceous-developed edge

x13: Pop_den; Tract population density in 2010

x14: Median_age; Median age at each census tract in 2010

x15: Mean_income; Mean income (inflation adjusted) at each census tract in 2010

Usage

```
lyme.svs.eco1.dat
```

Format

y integer vector, output count, each element represents the disease count in one census tract.

X Numeric matrix, matrix of covariates, includes percentage of developed land in each census tract, percentage of forest in each census tract, etc.

offset Numeric vector, vector for offset values, each element represents the log of population divided by 20,000 in census tract.

location Numeric matrix, location for each census tract.

geoid Numeric vector, geo id for each census tract.

References

Xie, Y., Xu, L., Li, J., Deng, X., Hong, Y., Kolivras, K., and Gaines, D. N. (2018). Spatial Variable Selection and An Application to Virginia Lyme Disease Emergence. Preprint, arXiv:1809.06418 [stat.AP].

Examples

```
data("lyme.svs.eco1")
lyme.svs.eco1.dat$dist=distmat.compute(location=lyme.svs.eco1.dat$location, dist.min=0.2821849)
```

---

small.test.dat	*A small dataset for fast testing of functions*

Description

Simulated data for fast testing of the functions. A list contains integer responses, model matrix, distance matrix, and offset term.

Usage

```
small.test.dat
```

Format

y Integer vector, the response variable, which is count data.

X Numeric matrix, the model matrix contains covariate information.

offset Numeric vector, vector for offset values.

dist Numeric matrix, a matrix for pairwise distance.

Examples

```
data("small.test")

#Here is a toy example for creating a data object that can be used for
#generating dat.obj for SpatialVS function
n=20
#simulate counts data
y=rpois(n=n, lambda=1)
#simulate covariate matrix
x1=rnorm(n)
x2=rnorm(n)
X=cbind(1, x1, x2)
#compute distance matrix from some simulated locations
loc_x=runif(n)
loc_y=runif(n)
dist=matrix(0,n, n)
for(i in 1:n)
```

```
{
  for(j in 1:n)
  {
    dist[i,j]=sqrt((loc_x[i]-loc_x[j])^2+(loc_y[i]-loc_y[j])^2)
  }
}

#assume offset is all zero
offset=rep(0, n)

#assemble the data object for SpatialVS

dat.obj=list(y=y, X=X, dist=dist, offset=offset)
```

SpatialVS   *Function for spatial variable selection*

Description

Perform variable selection for the spatial Poisson regression model under adaptive elastic net penalty.

Usage

```
SpatialVS(dat.obj, alpha.vec = seq(0.6, 1, by = 0.05),
  lambda.vec = seq(0.15, 1, len = 50), method = "PQL", plots = F,
  intercept = T, verbose = T)
```

Arguments

dat.obj	List, input data. Must contains:
	1. X numeric matrix, the covariates.
	2. y integer vector, the response in counts.
	3. dist numeric matrix, the distance matrix.
	4. offset numeric vector, the offset item.
alpha.vec	numeric vector, a vector of possible values of regularization parameter. The range is [0,1].
lambda.vec	numeric vector, a vector of positive values of regularization parameter.
method	string, the method to be used. Options are:
	1. "PQL" penalized quasi-likelihood method that considers spatial correlation.
	2. "PQL.nocor" penalized quasi-likelihood method that ignores spatial correlation.
	3. "APL" approximate penalized loglikelihood method that considers spatial correlation.

4. "APL.nocor" approximate penalized loglikelihood method that ignores spatial correlation.

plots	bool, if True, contour plot of AIC/BIC values is generated.
intercept	bool, if True, an intercept item will be included in model.
verbose	bool, if True, various updates are printed during each iteration of the algorithm.

Value

A list of 13 items:

1. dat.obj, List, a copy of the dat.obj input.

2. start, Initial values of parameters given by glmmPQL().

3. L.obj, Regression coefficients under each alpha.vec and lambda.vec, under the adaptive elastic net.

4. Lout.obj, AIC and BIC values under each L.obj value, under the adaptive elastic net.

5. contour.out.obj, Object used to generate contour plot as a function of alpha.vec and lambda.vec, with AIC or BIC as the output. Used to choose best penalty parameter, under the adaptive elastic net.

6. L.best.obj, Model fitting results under the best chosen alpha.vec and lambda.vec, under the adaptive elastic net.

7. Lout.best.obj, Best BIC value for L.best.obj.

8. L.EN.obj, Lout.EN.obj, contour.out.EN.obj, L.EN.best.obj, Similar items but under the elastic penalty.

9. lasso.weight, Numeric, specifies the adaptive Lasso weight.

10. method, String, the method used for computing the approximate likelihood function.

References

Xie, Y., Xu, L., Li, J., Deng, X., Hong, Y., Kolivras, K., and Gaines, D. N. (2018). Spatial Variable Selection and An Application to Virginia Lyme Disease Emergence. Preprint, arXiv:1809.06418 [stat.AP].

Examples

```
#use small.test.dat as the input to fit the spatial Poisson regression model.
#a grid of alpha.vec and lambda.vec is typically used.
#Here one point of alpha.vec and lambda.vec is used for fast illustration.

test.fit<-SpatialVS(dat.obj=small.test.dat, alpha.vec=0.5,
lambda.vec=5, method="PQL", intercept=TRUE, verbose=FALSE)
```

SpatialVS.summary   *Function for spatial variable selection's summary*

Description

return the summarized parameter estimates from the SpatialVS procedure.

Usage

```
SpatialVS.summary(obj)
```

Arguments

obj   List, returned by SpatialVS.

Value

A vector contains the final estimates of parameters. The estimates of theta are in log scale.

Examples

```
test.fit<-SpatialVS(dat.obj=small.test.dat, alpha.vec=0.5, lambda.vec=5,
method="PQL", intercept=TRUE, verbose=FALSE)
SpatialVS.summary(test.fit)
```

Index