Package ‘scgwr’

October 14, 2022

Type Package
Title Scalable Geographically Weighted Regression
Version 0.1.2-21
Date 2021-11-11
Author Daisuke Murakami[cre,aut], Narumasa Tsutsumida[ctb], Takahiro Yoshida[ctb], Tomoki Nakaya[ctb], Lu Binbin[ctb]
Maintainer Daisuke Murakami <dmuraka@ism.ac.jp>
License GPL (>= 2)
Encoding UTF-8
Imports FNN, spData, sp, dplyr, parallel, optimParallel
NeedsCompilation no
Repository CRAN
Date/Publication 2021-11-11 07:40:02 UTC

R topics documented:

predict0 .......................................................... 1
scgwr ............................................................. 3
scgwr_p ........................................................... 4

Index

predict0                  Spatial prediction using the scalable GWR model

Description

This function predicts explained variables and spatially varying coefficients at unobserved sites using the scalable GWR model.
predict0

Usage

predict0( mod, coords0, x0 = NULL )

Arguments

mod         Output from the scgwr function
coords0     Matrix of spatial point coordinates at predicted sites (N0 x 2)
x0          Matrix of explanatory variables at predicted sites (N0 x K). If NULL, explained
             variables are not predicted (only spatially varying coefficients are predicted).
             Default is NULL

Value

pred        Vector of predicted values (N0 x 1)
b           Matrix of estimated coefficients (N0 x K)
bse         Matrix of the standard errors for the coefficients (N0 x k)
t           Matrix of the t-values for the coefficients (N0 x K)
p           Matrix of the p-values for the coefficients (N0 x K)

Examples

require(spData)
data(boston)

id_obs <- sample(dim(boston.c)[1], 400)

# data at observed sites
y <- log(boston.c[id_obs, "MEDV"])
x <- boston.c[id_obs, c("CRIM", "INDUS", "ZN", "NOX", "AGE")]
coords <- boston.c[id_obs, c("LON", "LAT")]

# data at predicted sites
x0 <- boston.c[-id_obs, c("CRIM", "INDUS", "ZN", "NOX", "AGE")]
coords0 <- boston.c[-id_obs, c("LON", "LAT")]

mod <- scgwr( coords = coords, y = y, x = x )
pred0 <- predict0( mod = mod, coords0 = coords0, x0 = x0 )
pred <- pred0$pred  # predicted value
b <- pred0$b        # spatially varying coefficients
b[1:5,]

bse <- pred0$bse    # standard error of the coefficients
bt <- pred0$t       # t-values
bp <- pred0$p       # p-values
Scalable Geographically Weighted Regression

Description
This function estimates a scalable geographically weighted regression (GWR) model. See scgwr_p for parallel implementation of the model for very large samples.

Usage
```r
scgwr( coords, y, x = NULL, knn = 100, kernel = "gau", p = 4, approach = "CV", nsamp = NULL)
```

Arguments
- `coords`: Matrix of spatial point coordinates (N x 2)
- `y`: Vector of explained variables (N x 1)
- `x`: Matrix of explanatory variables (N x K). Default is NULL
- `knn`: Number of nearest-neighbors being geographically weighted. Default is 100. Larger `knn` is better for larger samples (see Murakami et al., 2019)
- `kernel`: Kernel to model spatial heterogeneity. Gaussian kernel ("gau") and exponential kernel ("exp") are available
- `p`: Degree of the polynomial to approximate the kernel function. Default is 4
- `approach`: If "CV", leave-one-out cross-validation is used for the model calibration. If "AICc", the corrected Akaike Information Criterion is minimized for the calibration. Default is "CV"
- `nsamp`: Number of samples used to approximate the cross-validation. The samples are randomly selected. If the value is large enough (e.g., 10,000), error due to the random sampling is quite small owing to the central limit theorem. The value must be smaller than the sample size. Default is NULL

Value
- `b`: Matrix of estimated coefficients (N x K)
- `bse`: Matrix of the standard errors for the coefficients (N x k)
- `t`: Matrix of the t-values for the coefficients (N x K)
- `p`: Matrix of the p-values for the coefficients (N x K)
- `par`: Estimated model parameters including a scale parameter and a shrinkage parameter if penalty = TRUE (see Murakami et al., 2018)
- `e`: Error statistics. It includes sum of squared errors (SSE), residual standard error (resid_SE), R-squared (R2), adjusted R2 (adjR2), log-likelihood (logLik), corrected Akaike information criterion (AICc), and the cross-validation (CV) score measured by root mean squared error (RMSE) (CV_score(RMSE))
scgwr_p

**Description**
Parallel implementation of scalable geographically weighted regression for large samples

**Usage**
```
scgwr_p(coords, y, x = NULL, knn = 100, kernel = "gau",
p = 4, approach = "CV", nsamp = NULL, cl = NULL)
```

**Arguments**
- `coords`: Matrix of spatial point coordinates (N x 2)
- `y`: Vector of explained variables (N x 1)
- `x`: Matrix of explanatory variables (N x K). Default is NULL
- `knn`: Number of nearest-neighbors being geographically weighted. Default is 100. Larger knn is better for larger samples (see Murakami et al., 2019)
- `kernel`: Kernel to model spatial heterogeneity. Gaussian kernel ("gau") and exponential kernel ("exp") are available
scgwr_p

**p** Degree of the polynomial to approximate the kernel function. Default is 4

**approach** If "CV", leave-one-out cross-validation is used for the model calibration. If "AICc", the corrected Akaike Information Criterion is minimized for the calibration. Default is "CV"

**nsamp** Number of samples used to approximate the cross-validation. The samples are randomly selected. If the value is large enough (e.g., 10,000), error due to the sampling is quite small owing to the central limit theorem. The value must be smaller than the sample size. Default is NULL

**cl** Number of cores used for the parallel computation. If cl = NULL, which is the default, the number of available cores is detected and used

**Value**

- **b** Matrix of estimated coefficients (N x K)
- **bse** Matrix of the standard errors for the coefficients (N x k)
- **t** Matrix of the t-values for the coefficients (N x K)
- **p** Matrix of the p-values for the coefficients (N x K)
- **par** Estimated model parameters including a scale parameter and a shrinkage parameter if penalty = TRUE (see Murakami et al., 2018)
- **e** Error statistics. It includes sum of squared errors (SSE), residual standard error (resid_SE), R-squared (R2), adjusted R2 (adjR2), log-likelihood (logLik), corrected Akaike information criterion (AICc), and the cross-validation (CV) score measured by root mean squared error (RMSE) (CV_score(RMSE))
- **pred** Vector of predicted values (N x 1)
- **resid** Vector of residuals (N x 1)
- **other** Other objects internally used

**References**


**See Also**

- scgwr, predict0

**Examples**

```r
# require(spData);require(sp)
# data(house)
# dat <- data.frame(coordinates(house), house$data[,c("price","age","rooms","beds","syear")])
# coords <- dat[,c("long","lat")]
# y <- log(dat[,"price"])
# x <- dat[,c("age","rooms","beds","syear")]

# Parallel estimation
```
# res1 <- scgwr_p( coords = coords, y = y, x = x )
# res1

# Parallel estimation + Approximate cross-validation using 10000 samples
# res2 <- scgwr_p( coords = coords, y = y, x = x, nsamp = 10000 )
# res2
Index

predict0, 1, 4, 5
scgwr, 3, 5
scgwr_p, 3, 4, 4