

Package ‘gwzinbr’

June 10, 2024

Type Package

Title Geographically Weighted Zero Inflated Negative Binomial Regression

Version 0.1.0

Maintainer Jéssica Vasconcelos <jehh.vasconcelosabreu@gmail.com>

Description Fits a geographically weighted regression model using zero inflated probability distributions. Has the zero inflated negative binomial distribution (zinb) as default, but also accepts the zero inflated Poisson (zip), negative binomial (negbin) and Poisson distributions. Can also fit the global versions of each regression model.

Da Silva, A. R. & De Sousa, M. D. R. (2023). ``Geographically weighted zero-inflated negative binomial regression: A general case for count data'', Spatial Statistics <doi:10.1016/j.spasta.2023.100790>.

Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1996). ``Geographically weighted regression: a method for exploring spatial nonstationarity'', Geographical Analysis, <doi:10.1111/j.1538-4632.1996.tb00936.x>.

Yau, K. K. W., Wang, K., & Lee, A. H. (2003). ``Zero-inflated negative binomial mixed regression modeling of over-dispersed count data with extra zeros'', Biometrical Journal, <doi:10.1002/bimj.200390024>.

License GPL-3

Encoding UTF-8

Imports sp

RoxygenNote 7.3.1

LazyData true

NeedsCompilation no

Author Jéssica Vasconcelos [aut, cre],
Juliana Rosa [aut],
Alan da Silva [aut]

Depends R (>= 3.5.0)

Repository CRAN

Date/Publication 2024-06-10 17:20:06 UTC

Contents

Golden	2
gwzinbr	4
southkorea_covid19	6
Index	8

Golden	<i>Golden Section Search</i>
--------	------------------------------

Description

Runs a Golden Section Search (GSS) algorithm for determining the optimum bandwidth for the geographically weighted zero inflated negative binomial regression and other spatial regression models.

Usage

```
Golden(
  data,
  formula,
  xvarinf = NULL,
  weight = NULL,
  lat,
  long,
  globalmin = TRUE,
  method,
  model = "zinb",
  bandwidth = "cv",
  offset = NULL,
  force = FALSE,
  maxg = 100,
  distancekm = FALSE
)
```

Arguments

data	name of the dataset.
formula	regression model formula as in <code>lm</code> .
xvarinf	name of the covariates for the zero inflated part of the model, default value is <code>NULL</code> .
weight	name of the variable containing the sample weights, default value is <code>NULL</code> .
lat	name of the variable containing the latitudes in the dataset.
long	name of the variable containing the longitudes in the dataset.
globalmin	logical value indicating whether to find a global minimum in the optimization process, default value is <code>TRUE</code> .

method	indicates the method to be used for the bandwidth calculation (adaptive_bsq or fixed_g).
model	indicates the model to be used for the regression (zinb, zip, negbin, poisson), default value is "zinb".
bandwidth	indicates the criterion to be used for the bandwidth calculation (cv, aic), default value is "cv".
offset	name of the variable containing the offset values, if null then is set to a vector of zeros, default value is NULL.
force	logical value indicating whether to force the indicated model even if it is not the best fit for the data, default value is FALSE.
maxg	integer indicating the maximum number of iterations for the zero inflated part of the model, default value is 100.
distancekm	logical value indicating whether to calculate the distances in km, default value is FALSE.

Value

A list that contains:

- h_values - Initial values tested for the bandwidth.
- iterations - All bandwidth values tested and respective cv/aic results for each Golden Section Search executed.
- gss_results - Optimum bandwidth found and respective cv/aic.
- min_bandwidth - Optimum bandwidth.

Examples

```
## Data

data(southkorea_covid19)

## GSS algorithm

gss <- Golden(data = southkorea_covid19[1:122, ],
  formula = n_covid1~diff_sd,
  xvarinf = NULL, weight = NULL, lat = "y", long = "x",
  offset = NULL, model = "poisson", method = "fixed_g",
  bandwidth = "cv", globalmin = FALSE, distancekm = FALSE,
  force=FALSE)

## Bandwidth
gss$min_bandwidth

## Iterations
gss$iterations
```

 gwzinbr

Geographically Weighted Zero Inflated Negative Binomial Regression

Description

Fits a geographically weighted regression model using zero inflated probability distributions. Has the zero inflated negative binomial distribution (zinb) as default, but also accepts the zero inflated Poisson (zip), negative binomial (negbin) and Poisson distributions. Can also fit the global versions of each regression model.

Usage

```
gwzinbr(
  data,
  formula,
  xvarinf = NULL,
  weight = NULL,
  lat,
  long,
  grid = NULL,
  method,
  model = "zinb",
  offset = NULL,
  distancekm = FALSE,
  force = FALSE,
  int_inf = TRUE,
  maxg = 100,
  h = NULL
)
```

Arguments

data	name of the dataset.
formula	regression model formula as in <code>lm</code> .
xvarinf	name of the covariates for the zero inflated part of the model, default value is <code>NULL</code> .
weight	name of the variable containing the sample weights, default value is <code>NULL</code> .
lat	name of the variable containing the latitudes in the dataset.
long	name of the variable containing the longitudes in the dataset.
grid	name of the dataset containing the coordinates for the model locations, default value is <code>NULL</code> .
method	indicates the method to be used for the bandwidth calculation (<code>adaptive_bsq</code> or <code>fixed_g</code>).
model	indicates the model to be used for the regression (<code>zinb</code> , <code>zip</code> , <code>negbin</code> , <code>poisson</code>), default value is <code>"zinb"</code> .

offset	name of the variable containing the offset values, if null then is set to a vector of zeros, default value is NULL.
distancekm	logical value indicating whether to calculate the distances in km, default value is FALSE.
force	logical value indicating whether to force the indicated model even if it is not the best fit for the data, default value is FALSE.
int_inf	logical value indicating whether to include an intercept in the zero inflated part of the model, default value is TRUE.
maxg	integer indicating the maximum number of iterations for the zero inflated part of the model, default value is 100.
h	integer indicating the bandwidth value (obtained from <code>golden()</code>), default value is NULL.

Value

A list that contains:

- `bandwidth` - Bandwidth value.
- `measures` - Goodness of fit statistics and other measures.
- `qntls_gwr_param_estimates` - Quantiles of GWR parameter estimates.
- `descript_stats_gwr_param_estimates` - Descriptive statistics of GWR parameter estimates.
- `t_test_gwr_param_estimates` - Results for the parameters significance t tests.
- `qntls_gwr_se` - Quantiles of GWR standard errors.
- `descript_stats_gwr_se` - Descriptive statistics of GWR standard errors.
- `qntls_gwr_zero_infl_param_estimates` - Quantiles of GWR zero inflated parameter estimates.
- `descript_stats_gwr_zero_infl_param_estimates` - Descriptive statistics of GWR zero inflated parameter estimates.
- `t_test_gwr_zero_infl_param_estimates` - Results for the zero inflated parameters significance t tests.
- `qntls_gwr_zero_infl_se` - Quantiles of GWR zero inflated standard errors.
- `descript_stats_gwr_zero_infl_se` - Descriptive statistics of GWR zero inflated standard errors.
- `non_stationary_test` - Results for the Non-Stationary Test for GWR parameter estimates.
- `non_stationary_test_zero_infl` - Results for the Non-Stationary Test for GWR zero inflated parameter estimates.
- `global_param_estimates` - Parameter estimates for the global model.
- `analysis_max_like_zero_infl_param_estimated` - Analysis of Maximum Likelihood Zero Inflation Parameter Estimates.
- `analysis_max_like_gof_measures` - Goodness of fit measures for the Analysis of Maximum Likelihood Zero Inflation Parameter Estimates.
- `variance_covariance_matrix` - Variance-covariance matrix.

- residuals - Model residuals.
- param_estimates_grid - GWR parameter estimates using grid dataset.
- alpha_estimates - Estimates for the alpha parameter (for zinb and negbin).
- gwr_param_estimates - GWR parameter estimates.

Examples

```
## Data

data(southkorea_covid19)

## Model

mod <- gwzinbr(data = southkorea_covid19,
  formula = n_covid1~Morbidity+high_sch_p+Healthcare_access+
  diff_sd+Crowding+Migration+Health_behavior,
  lat = "x", long = "y", offset = "ln_total", method = "adaptive_bsq",
  model = "negbin", distancekm = TRUE, h=230, force=TRUE)

## Bandwidth
mod$bandwidth

## Goodness of fit measures
mod$measures
```

southkorea_covid19 *South Korea COVID-19 dataset*

Description

COVID-19 data for South Korea from January 20th 2019 to March 20th 2020.

Usage

```
data(southkorea_covid19)
```

Format

A data frame with with 244 observations on the following 11 variables:

- n_covid1 - number of COVID-19 cases in the early phase of the pandemic (prequarantine)
- Morbidity - area morbidity rate
- high_sch_p - percentage of high school educated people
- Healthcare_access - access to healthcare

- *diff_sd* - difficulty to social distancing
- *Crowding* - area crowding
- *Migration* - population mobility
- *Health_behavior* - an index calculated based on habits as alcohol drinking, current smoking, etc
- *x* - a numeric vector of x coordinates
- *y* - a numeric vector of y coordinates
- *ln_total* - log transformation of the province's total population

Index

* datasets

southkorea_covid19, [6](#)

Golden, [2](#)

gwzinbr, [4](#)

southkorea_covid19, [6](#)