

Package ‘mgwnbr’

March 26, 2024

Type Package

Title Multiscale Geographically Weighted Negative Binomial Regression

Version 0.2.0

Maintainer Juliana Rosa <julianamrosa00@gmail.com>

Description Fits a geographically weighted regression model with different scales for each covariate. Uses the negative binomial distribution as default, but also accepts the normal, Poisson, or logistic distributions. Can fit the global versions of each regression and also the geographically weighted alternatives with only one scale, since they are all particular cases of the multiscale approach.

Hanchen Yu (2024). “Exploring Multiscale Geographically Weighted Negative Binomial Regression”, *Annals of the American Association of Geographers* <doi:10.1080/24694452.2023.2289986>.

Fotheringham AS, Yang W, Kang W (2017). “Multiscale Geographically Weighted Regression (MGWR)”, *Annals of the American Association of Geographers* <doi:10.1080/24694452.2017.1352480>.

Da Silva AR, Rodrigues TCV (2014). “Geographically Weighted Negative Binomial Regression - incorporating overdispersion”, *Statistics and Computing* <doi:10.1007/s11222-013-9401-9>.

License GPL-3

Encoding UTF-8

Imports sp

RoxygenNote 7.3.1

LazyData true

NeedsCompilation no

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

Depends R (>= 3.5.0)

Repository CRAN

Date/Publication 2024-03-26 15:10:02 UTC

R topics documented:

georgia	2
mgwnbr	3
Index	6

georgia	<i>Georgia dataset</i>
---------	------------------------

Description

The Georgia census data set from Fotheringham et al. (2002) in dataframe format.

Usage

```
data(georgia)
```

Format

A data frame with with 159 observations on the following 13 variables:

- `AreaKey` - an identification number for each county
- `Latitude` - the latitude of the county centroid
- `Longitude` - the longitude of the county centroid
- `TotPop90` - population of the county in 1990
- `PctRural` - percentage of the county population defined as rural
- `PctBach` - percentage of the county population with a bachelors degree
- `PctEld` - percentage of the county population aged 65 or over
- `PctFB` - percentage of the county population born outside the US
- `PctPov` - percentage of the county population living below the poverty line
- `PctBlack` - percentage of the county population who are black
- `ID` - a numeric vector of IDs
- `X` - a numeric vector of x coordinates
- `Y` - a numeric vector of y coordinates

Description

Fits a geographically weighted regression model with different scales for each covariate. Uses the negative binomial distribution as default, but also accepts the normal, Poisson, or logistic distributions. Can fit the global versions of each regression and also the geographically weighted alternatives with only one scale, since they are all particular cases of the multiscale approach.

Usage

```
mgwnbr(
  data,
  formula,
  weight = NULL,
  lat,
  long,
  globalmin = TRUE,
  method,
  model = "negbin",
  mgwr = TRUE,
  bandwidth = "cv",
  offset = NULL,
  distancekm = FALSE,
  int = 50,
  h = NULL
)
```

Arguments

<code>data</code>	name of the dataset.
<code>formula</code>	regression model formula as in <code>lm</code> .
<code>weight</code>	name of the variable containing the sample weights, default value is <code>NULL</code> .
<code>lat</code>	name of the variable containing the latitudes in the dataset.
<code>long</code>	name of the variable containing the longitudes in the dataset.
<code>globalmin</code>	logical value indicating whether to find a global minimum in the optimization process, default value is <code>TRUE</code> .
<code>method</code>	indicates the method to be used for the bandwidth calculation (<code>adaptive_bsq</code> , <code>fixed_bsq</code> , <code>fixed_g</code>).
<code>model</code>	indicates the model to be used for the regression (<code>gaussian</code> , <code>poisson</code> , <code>negbin</code> , <code>logistic</code>), default value is <code>"negbin"</code> .
<code>mgwr</code>	logical value indicating if multiscale should be used (<code>TRUE</code> , <code>FALSE</code>), default value is <code>TRUE</code> .

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.
distancekm	logical value indicating whether to calculate the distances in km, default value is FALSE.
int	integer indicating the number of iterations, default value is 50.
h	integer indicating a predetermined bandwidth value, default value is NULL.

Value

A list that contains:

- general_bandwidth - General bandwidth value.
- band - Bandwidth values for each covariate.
- measures - Goodness of fit statistics.
- ENP - Effective number of parameters.
- mgwr_param_estimates - MGWR parameter estimates.
- qntls_mgwr_param_estimates - Quantiles of MGWR parameter estimates.
- descript_stats_mgwr_param_estimates - Descriptive statistics of MGWR parameter estimates.
- p_values - P-values for the t tests on parameter significance.
- t_critical - Critical values for the t tests on parameter significance.
- mgwr_se - MGWR standard errors.
- qntls_mgwr_se - Quantiles of MGWR standard errors.
- descript_stats_se - Descriptive statistics of MGWR standard errors.
- global_param_estimates - Parameter estimates for the global model.
- t_test_dfs - Denominator degrees of freedom for the t tests.
- global_measures - Goodness of fit statistics for the global model.

Examples

```
## Data

data(georgia)

for (var in c("PctFB", "PctBlack")){
  georgia[, var] <- as.data.frame(scale(georgia[, var]))
}

## Model

mod <- mgwnbr(data=georgia, formula=PctBach~PctBlack+PctFB,
```

```
lat="Y", long="X", globalmin=FALSE, method="adaptive_bsq", bandwidth="cv",  
  model="gaussian", mgwr=FALSE, h=136)
```

```
## Bandwidths  
mod$general_bandwidth
```

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

Index

* **datasets**
 georgia, [2](#)
georgia, [2](#)
mgwnbr, [3](#)