Package 'gwrr'

February 20, 2015

Type Package

Title Fits geographically weighted regression models with diagnostic tools

Version 0.2-1

Date 2013-06-11

Author David Wheeler

Maintainer David Wheeler <dcwheels@gmail.com>

Depends fields, lars

Description Fits geographically weighted regression (GWR) models and has tools to diagnose and remediate collinearity in the GWR models. Also fits geographically weighted ridge regression (GWRR) and geographically weighted lasso (GWL) models.

License GPL (>= 2)

LazyLoad yes

NeedsCompilation no

Repository CRAN

Date/Publication 2013-06-19 18:13:24

R topics documented:

gwrr-package *Geographically weighted regression models with penalties and diagnostic tools*

Description

Fits geographically weighted regression (GWR) models and has tools to diagnose collinearity in the GWR models. Also fits geographically weighted ridge regression (GWRR) and geographically weighted lasso (GWL) models.

Details

Package:	gwrr
Type:	Package
Version:	0.2-1
Date:	2013-06-11
License:	GPL (>=2)
LazyLoad:	yes

Author(s)

David Wheeler

Maintainer: David Wheeler <dcwheels@gmail.com>

References

Wheeler DC (2007) Diagnostic tools and a remedial method for collinearity in geographically weighted regression. Environment and Planning A, 39: 2464-2481.

Wheeler DC (2009) Simultaneous coefficient penalization and model selection in geographically weighted regression: The geographically weighted lasso. Environment and Planning A, 41: 722-742.

Examples

```
data(columbus)
locs <- cbind(columbus$x, columbus$y)
col.gwr <- gwr.est(crime ~ income + houseval, locs, columbus, "exp")
plot(col.gwr$beta[2,], col.gwr$beta[3,])
col.vdp <- gwr.vdp(crime ~ income + houseval, locs, columbus, col.gwr$phi, "exp")
hist(col.vdp$condition)
```

columbus *Columbus crime*

Description

Crime rate in planning neighborhoods in Columbus, Ohio in 1980

Usage

```
data(columbus)
```

Format

A data frame with 49 observations on the following 6 variables.

`houseval` a numeric vector

`income` a numeric vector

`crime` a numeric vector

`distcbd` a numeric vector

`x` a numeric vector

`y` a numeric vector

Details

The data consist of variables for mean housing value, mean household income, residential and vehicle thefts combined per thousand people for 1980, distance to the central business district (CBD), and x and y spatial coordinates of neighborhood centroids.

Source

Anselin L (1988) Spatial Econometrics: Methods and Models. Kluwer, Dordrecht

References

Wheeler DC (2007) Diagnostic tools and a remedial method for collinearity in geographically weighted regression. Environment and Planning A, 39: 2464-2481

Examples

```
data(columbus)
plot(columbus$x, columbus$y)
```

gwl.est *Geographically weighted lasso*

Description

This function fits a geographically weighted lasso (GWL) model

Usage

```
gwl.est(form, locs, data, kernel = "exp", cv.tol)
```

Arguments

form	A regression model forumula, as in the functions lm and glm
locs	A matrix of spatial coordinates of data points, where the x coordinate is first, then the y coordinate; coordinates are assumed to not be latitude and longitude, as Euclidean distance is calculated from coordinates
data	A data frame with data to fit model
kernel	A kernel weighting function, either exp or gauss, where exponential function is default
cv.tol	A stopping tolerance in terms of cross-validation error for the bi-section search routine to estimate the kernel bandwidth using cross-validation; if missing an internally calculated value is used

Details

This function estimates penalized spatially varying coefficients using the geographically weighed regression and lasso approaches. Spatial kernel weights are applied to observations using the estimated kernel bandwidth to estimate local models at each data point. The kernel bandwidth and lasso solutions are currently estimated using cross-validation with an exponential or Gaussian kernel function. Some regression coefficients may be penalized to zero. The function estimates regression coefficients, the outcome variable values, and the model fit.

Value

A list with the following items:

phi	Kernel bandwidth
RMSPE	Root mean squared prediction error from bandwidth estimation
beta	Matrix of estimated regression coefficients, where a row contains the coefficients for one regression term for all data points
yhat	Estimated outcome variable values
RMSE	Root mean squared error from estimation
rsquare	Approximate R-square for GWR model

Author(s)

David Wheeler

References

Wheeler DC (2009) Simultaneous coefficient penalization and model selection in geographically weighted regression: The geographically weighted lasso. Environment and Planning A, 41: 722-742

See Also

gwrr.est

Examples

```
data(columbus)
locs <- cbind(columbus$x, columbus$y)
col.gwl <- gwl.est(crime ~ income + houseval, locs, columbus, "exp")
plot(col.gwl$beta[2,], col.gwl$beta[3,])
plot(columbus$x, columbus$y, cex=col.gwl$beta[1,]/10)
```

---

gwr.bw.est *Cross-validation estimation of kernel bandwidth*

Description

Estimate the kernel function bandwidth with cross-validation

Usage

```
gwr.bw.est(form, locs, data, kernel = "exp", cv.tol)
```

Arguments

form	A regression model forumula, as in the functions lm and glm
locs	A matrix of spatial coordinates of data points, where the x coordinate is first, then the y coordinate; coordinates are assumed to not be latitude and longitude, as Euclidean distance is calculated from coordinates
data	A data frame with data to fit model
kernel	A kernel weighting function, either exp or gauss, where exponential function is default
cv.tol	A stopping tolerance in terms of cross-validation error for the bi-section search routine to estimate the kernel bandwidth using cross-validation; if missing an internally calculated value is used

Details

This function estimates the kernel bandwidth in a GWR model with leave-one-out cross-validation. It does not estimate the final regression coefficients or outcome variable.

Value

A list with the following items:

`phi`	Kernel bandwidth
`RMSPE`	Root mean squared prediction error from bandwidth estimation
`cv.score`	Sum of squared prediction errors from bandwidth estimation

Author(s)

David Wheeler

References

Wheeler DC (2007) Diagnostic tools and a remedial method for collinearity in geographically weighted regression. Environment and Planning A, 39: 2464-2481

See Also

`gwr.est`

Examples

```
data(columbus)
locs <- cbind(columbus$x, columbus$y)
col.bw <- gwr.bw.est(crime ~ income + houseval, locs, columbus, "exp")
col.gwr <- gwr.est(crime ~ income + houseval, locs, columbus, "exp", bw=col.bw$phi)
```

---

gwr.est *Geographically weighted regression*

Description

This function fits a geographically weighted regression (GWR) model

Usage

```
gwr.est(form, locs, data, kernel = "exp", bw = TRUE, cv.tol)
```

Arguments

form	A regression model forumula, as in the functions lm and glm
locs	A matrix of spatial coordinates of data points, where the x coordinate is first, then the y coordinate; coordinates are assumed to not be latitude and longitude, as Euclidean distance is calculated from coordinates
data	A data frame with data to fit model
kernel	A kernel weighting function, either exp or gauss, where exponential function is default
bw	Either TRUE to estimate a bandwidth for the kernel function, or the bandwidth to use to fit the model; bandwidth is estimated by default
cv.tol	A stopping tolerance in terms of cross-validation error for the bi-section search routine to estimate the kernel bandwidth using cross-validation; if missing an internally calculated value is used

Details

This function estimates spatially varying coefficients using the GWR approach. Spatial kernel weights are applied to observations using the estimated or supplied kernel bandwidth to estimate local models at each data point. The bandwidth is currently estimated with cross-validation with an exponential or Gaussian kernel function. The function estimates regression coefficients, the outcome variable values, and the model fit.

Value

A list with the following items:

phi	Kernel bandwidth
RMSPE	Root mean squared prediction error from bandwidth estimation
beta	Matrix of estimated regression coefficients, where a row contains the coefficients for one regression term for all data points
yhat	Estimated outcome variable values
RMSE	Root mean squared error from estimation
rsquare	Approximate R-square for GWR model

Author(s)

David Wheeler

References

Wheeler DC (2007) Diagnostic tools and a remedial method for collinearity in geographically weighted regression. Environment and Planning A, 39: 2464-2481

See Also

gwr.bw.est

Examples

```
data(columbus)
locs <- cbind(columbus$x, columbus$y)
col.gwr <- gwr.est(crime ~ income + houseval, locs, columbus, "exp")
plot(col.gwr$beta[2,], col.gwr$beta[3,])
plot(columbus$x, columbus$y, cex=col.gwr$beta[1,]/10)
```

`gwr.vdp` *Collinearity diagnostics for geographically weighted regression*

Description

Uses the collinearity diagnostic tools of variance-decomposition proportions and condition indexes for geographically weighted regression (GWR) models.

Usage

```
gwr.vdp(form, locs, data, phi, kernel = "exp", sel.ci = 30, sel.vdp = 0.5)
```

Arguments

`form`	A regression model forumula, as in the functions lm and glm
`locs`	A matrix of spatial coordinates of data points, where the x coordinate is first, then the y coordinate; coordinates are assumed to not be latitude and longitude, as Euclidean distance is calculated from coordinates
`data`	A data frame with data to fit model
`phi`	The kernel bandwidth used in the GWR model
`kernel`	The kernel weighting function used in the GWR model, either exp or gauss; exp is the default
`sel.ci`	The threshold value to use for the condition index to indicate observations with a collinearity issue; indexes above this value will be flagged; the default is 30
`sel.vdp`	The threshold value to use for the variance-decomposition proportion to indicate observations with a collinearity issue; proportions above this value will be flagged; the default is 0.5

Details

This function calculates the variance-decomposition proportions and the condition indexes for the weighted design matrix used in a GWR model. The kernel function and bandwidth used to estimate the GWR model must be input to this function. Observations with a large condition index and relatively large variance-decomposition proportions for more than one regression term indicate an issue with collinearity.

Value

A list with the following items:

condition	Largest condition index for each observation
vdp	Variance-decomposition proportions for the largest variance component for each observation
flag.cond	True if largest condition index exceeds threshold
flag.vdp	True if variance-decomposition proportions for more than one term exceed threshold
flag.cond.vdp	True if condition index and variance-decompostion proportions exceed thresholds

Author(s)

David Wheeler

References

Wheeler DC (2007) Diagnostic tools and a remedial method for collinearity in geographically weighted regression. Environment and Planning A, 39: 2464-2481

See Also

gwr.bw.est

Examples

```
data(columbus)
locs <- cbind(columbus$x, columbus$y)
col.bw <- gwr.bw.est(crime ~ income + houseval, locs, columbus, "exp")
col.vdp <- gwr.vdp(crime ~ income + houseval, locs, columbus, col.bw$phi, "exp")
hist(col.vdp$condition)
```

---

gwrr.est *Geographically weighted ridge regression*

Description

This function fits a geographically weighted ridge regression (GWRR) model

Usage

```
gwrr.est(form, locs, data, kernel = "exp", bw = TRUE, rd = TRUE, cv.tol)
```

Arguments

form	A regression model forumula, as in the functions lm and glm
locs	A matrix of spatial coordinates of data points, where the x coordinate is first, then the y coordinate; coordinates are assumed to not be latitude and longitude, as Euclidean distance is calculated from coordinates
data	A data frame with data to fit model
kernel	A kernel weighting function, either exp or gauss, where exponential function is default
bw	Either TRUE to estimate a bandwidth for the kernel function, or the bandwidth to use to fit the model; bandwidth is estimated by default
rd	Either TRUE to estimate a ridge shrinkage parameter, or the ridge parameter to use to fit the model; ridge parameter is estimated by default
cv.tol	A stopping tolerance in terms of cross-validation error for the bi-section search routine to estimate the kernel bandwidth using cross-validation; if missing an internally calculated value is used

Details

This function estimates penalized spatially varying coefficients using the GWR and ridge regression approaches. Spatial kernel weights are applied to observations using the estimated or supplied kernel bandwidth to estimate local models at each data point. The bandwidth is estimated with cross-validation with an exponential or Gaussian kernel function. The regression coefficients are penalized with a ridge parameter that is estimated with cross-validation. The function estimates regression coefficients, the outcome variable values, and the model fit.

Value

A list with the following items:

phi	Kernel bandwidth
lambda	Ridge shrinkage parameter
RMSPE	Root mean squared prediction error from bandwidth estimation
beta	Matrix of estimated regression coefficients, where a row contains the coefficients for one regression term for all data points
yhat	Estimated outcome variable values
RMSE	Root mean squared error from estimation
rsquare	Approximate R-square for GWR model

Author(s)

David Wheeler

References

Wheeler DC (2007) Diagnostic tools and a remedial method for collinearity in geographically weighted regression. Environment and Planning A, 39: 2464-2481

See Also

gwr.est

Examples

```
data(columbus)
locs <- cbind(columbus$x, columbus$y)
col.gwrr <- gwrr.est(crime ~ income + houseval, locs, columbus, "exp", bw=2.00, rd=0.03)
plot(col.gwrr$beta[2,], col.gwrr$beta[3,])
plot(columbus$x, columbus$y, cex=col.gwrr$beta[1,]/10)
col.gwr <- gwrr.est(crime ~ income + houseval, locs, columbus, "exp", bw=col.gwrr$phi, rd=0)
```

Index