Binomial_MGWR

May 18, 2020

\#
Multiscale Geographically Weighted Regression - Binomial dependent variable
The model has been explored and tested for multiple parameters on real and simulated datasets. The research includes the following outline with separate notebooks for each part.

Notebook Outline:

Back to the main page

---

1 Introduction

1.1 Introduction to the problem

As prefaced earlier, the Geographically Weighted Regression model in PySAL can currently estimate Gaussian, Poisson and Logistic models though the Multiscale extension of the GWR model is currently limited to only Gaussian models. This part of the project aims to expand the MGWR model to nonlinear local spatial regression modeling techniques where the response outcomes may be binomial (or a Logit model). This will enable a richer and holistic local statistical modeling framework to model multi-scale process heterogeneity for the open source community.

1.2 Statistical Equations

A conventional Logistic regression model with $x_1, x_2, ..., x_k$ as predictors, a binary(Bernoulli) response variable y and l denoting the log-odds of the event that y=1, can be written as:

$$l = log_b(p/(1-p)) = \left(\sum \beta_k x_{k,i}\right) \tag{1}$$

$$\tag{2}$$

where $x_{k,1}$ is the kth explanatory variable in place i, $_{ks}$ are the parameters and p is the probability such that p = P(Y = 1).
By exponentiating the log-odds:

$p / (1-p) = bˆ\{_0+_1 x_1+_2 x_2\} $

It follows from this - the probability that Y = 1 is:

$p = (b^{0+_1x_1+_2x_2})/(b^{0+_1x_1+_2x_2} + 1) = 1/(1 + b^{-0+_1x_1+_2x_2})$

2 Local Scoring Algorithm

Following the technique from (Hastie & Tibshirani, 1986), for logisitic generalized additive models the model was estimated using the local scoring algorithm as follows:

1. Initialize the current estimate of the additive predictor n_i^{old}:$n_i^{old} = \sum \beta_k X_k$ and the probability such P(Y=1): $p_i^{old} = exp(n_i^{old})/(1 + exp(n_i^{old}))$
2. Compute the working response: $z_i = n_i^{old} + (y_i - p_i^{old})/(p_i^{old}(1 - p_i^{old}))$
3. compute weights $w_i = p_i^{old}(1 - p_i^{old})$
4. obtain n_i^{new} by fitting a weighted additive model to z_i. In this the smoothers in the backfitting algorithm incorporate the additional weights and GWR is used for the linear parts.

These steps are repeated until the relative change in the fitted coefficients and the functions is below a tolerance threshold (1e-05 in this case).

Reference for these equations: http://ugrad.stat.ubc.ca/~nancy/526_2003/projects/kazi2.pdf

2.0.1 Further work required:

The parameters for the estimated model using Monte Carlo tests with simulated data are close to expected. Further exploration is required to theoretically justify the model in the context of spatial data models, especially MGWR.

As an exploration, this work includes results from both adding a stochastic error to the model during calibration and without it. Results for both are shown in the notebooks below.

3 Notebooks with Tests

Initial module changes and univariate model check - Setup with libraries - Fundamental equations for Binomial MGWR - Example Dataset - Helper functions - Univariate example - Parameter check - Bandwidths check

Simulated Data example - Setup with libraries - Create Simulated Dataset - Forming independent variables - Creating y variable with Binomial distribution - Univariate example - Bandwidth: Random initialization check - Parameters check - Multivariate example - Bandwidths: Random initialization check - Parameters check - Global model parameter check

Real Data example

- Setup with libraries
- Landslide Dataset
- Univariate example

    - Bandwidth: Random initialization check
    - Parameter check

- Multivariate example

- – Bandwidths: Random initialization check
- MGWR bandwidths
- AIC, AICc, BIC check

3.1 Monte Carlo Tests

3.1.1 Monte Carlo tests for model estimated with error

Monte Carlo Simulation Visualization

- Setup with libraries
- List bandwidths from pickles
- Parameter functions
- GWR bandwidth
- MGWR bandwidths
- AIC, AICc, BIC check

  - – AIC, AICc, BIC Boxplots for comparison

- Parameter comparison from MGWR and GWR

3.1.2 Monte Carlo tests for model estimated without error

Monte Carlo Simulation Visualization

- Setup with libraries
- List bandwidths from pickles
- Parameter functions
- GWR bandwidth
- MGWR bandwidths
- AIC, AICc, BIC check

  - – AIC, AICc, BIC Boxplots for comparison

- Parameter comparison from MGWR and GWR

4 References:

1. Fotheringham, A. S., Yang, W., & Kang, W. (2017). Multiscale Geographically Weighted Regression (MGWR). Annals of the American Association of Geographers, 107(6), 1247–1265. https://doi.org/10.1080/24694452.2017.1352480

2. Yu, H., Fotheringham, A. S., Li, Z., Oshan, T., Kang, W., & Wolf, L. J. (2019). Inference in Multiscale Geographically Weighted Regression. Geographical Analysis, gean.12189. https://doi.org/10.1111/gean.12189

3. Hastie, T., & Tibshirani, R. (1986). Generalized Additive Models. Statistical Science, 1(3), 297–310. https://doi.org/10.1214/ss/1177013604

4. Wood, S. N. (2006). Generalized additive models : an introduction with R. Chapman & Hall/CRC.

Back to the main page