Package ‘minb’

October 1, 2023

Type Package
Title Multiple-Inflated Negative Binomial Model
Version 0.1.0
Description Count data is prevalent and informative, with widespread application in many fields such as social psychology, personality, and public health. Classical statistical methods for the analysis of count outcomes are commonly variants of the log-linear model, including Poisson regression and Negative Binomial regression. However, a typical problem with count data modeling is inflation, in the sense that the counts are evidently accumulated on some integers. Such an inflation problem could distort the distribution of the observed counts, further bias estimation and increase error, making the classic methods infeasible. Traditional inflated value selection methods based on histogram inspection are easy to neglect true points and computationally expensive in addition. Therefore, we propose a multiple-inflated negative binomial model to handle count data modeling with multiple inflated values, achieving data-driven inflated value selection. The proposed approach provides simultaneous identification of important regression predictors on the target count response as well. More details about the proposed method are described in Li, Y., Wu, M., Wu, M., & Ma, S. (2023) <arXiv:2309.15585>.
License GPL-3
Depends R (>= 3.5.0)
Imports MASS, pscl, stats
Encoding UTF-8
LazyData true
NeedsCompilation no
RoxygenNote 7.2.3
Author Yang Li [aut],
Mingcong Wu [aut, cre],
Mengyun Wu [aut],
Shuangge Ma [aut]
Maintainer Mingcong Wu <wumingcong@ruc.edu.cn>
Description

The minb function is the main function to achieve data-driven inflated values selection and identification of important predictors for multiple inflated count data modeling.

Usage

```
minb(X,y,pars_init=NULL,lambda1_set=NULL,lambda2_set=NULL,ntune=10,
maxiter=200,tol=1e-03,vrbs=FALSE)
```

Arguments

- **X**: The design matrix, without an intercept, minb includes an intercept by default.
- **y**: The response count vector.
- **pars_init**: an optional list containing the initial values for the corresponding component. See details.
- **lambda1_set**: A user supplied tuning sequence for inflated values. Typical usage is to have the program compute its own sequence based on ntune. Supplying a value of lambda1_set overrides this.
- **lambda2_set**: A user supplied tuning sequence for regression predictors. Typical usage is to have the program compute its own sequence based on ntune. Supplying a value of lambda2_set overrides this.
- **ntune**: The number of the tuning parameter values, default to 10.
- **maxiter**: defines the maximum number of iterations, default to 200.
- **tol**: specifies the iteration convergency tolerance, default to 1e-03.
- **vrbs**: a boolean variable which defines whether to print the iteration details, default to FALSE.

Details

Initialization values can be supplied or estimated by glm.fit (the default). To supply initial values of the parameters, start should be a list with elements "kappa", "omega","beta", and "phi" containing the starting values for the coefficients of the corresponding component of the model.
minb_SimuData

Value

The minb returns a list containing the following components:

- **omega**: The vector of the estimated mixing proportions of the selected inflated values in the multiple-inflated negative binomial model.
- **kappa**: The vector of selected inflated values.
- **phi**: The value of the dispersion parameter of Negative Binomial distribution.
- **beta**: The vector of estimated non-zero regression coefficients of the Negative Binomial distribution.

Examples

```r
# This is an example of minb with simulated data
data(minb_SimuData)
X = minb_SimuData$X
y = minb_SimuData$y
result = minb(X=X, y=y, ntune=2)
result$beta
result$omega
```

Description

The dataset minb_SimuData contains n = 2000 samples simulated from multiple inflated negative binomial model with p = 15 predictors while the first 10 predictors are informative. The responses are inflated at multiple integers 0, 1, 3, 5, and 10.

Usage

minb_SimuData

Format

A data list containing 2000 samples
Index

minb, 2
minb_SimuData, 3