Package ‘ZIPFA’

March 31, 2020

Type Package
Title Zero Inflated Poisson Factor Analysis
Version 0.8.1
Date 2020-03-30
Author Tianchen Xu [aut, cre] (<https://orcid.org/0000-0002-0102-7630>),
Ryan T. Demmer [aut],
Gen Li [aut]
Maintainer Tianchen Xu <tx2155@columbia.edu>
Description Estimation methods for zero-inflated Poisson factor analysis (ZIPFA) on sparse data.
It provides estimates of coefficients in a new type of zero-inflated regression.
It provides a cross-validation method to determine the potential rank of the data in the ZIPFA
and conducts zero-inflated Poisson factor analysis based on the determined rank.
URL https://zjph602xtc.github.io/ZIPFA/,
BugReports https://github.com/zjph602xtc/ZIPFA/issues
Depends R (>= 3.2.0)
Imports Matrix, doParallel, foreach, optimx, trustOptim
License GPL (>= 2)
Encoding UTF-8
LazyData no
Repository CRAN
RoxygenNote 6.0.1
NeedsCompilation no
Date/Publication 2020-03-31 20:30:02 UTC

R topics documented:

  cv_ZIPFA ......................................................... 2
  EMzeropoisson_mat .......................................... 3
  X .............................................................. 4
  ZIPFA .......................................................... 5
Index

\texttt{cv\_ZIPFA} \hspace{1cm} \textit{Cross validation for Zero Inflated Poisson factor analysis}

\section*{Description}

To conduct a cross validation for Zero Inflated Poisson factor analysis to find the number of factors.

\section*{Usage}

\begin{verbatim}
cv_ZIPFA(X, k, fold, tau = 0.1, cut = 0.8, tolLnlikelihood = 5e-4, iter = 20, tol = 1e-4, maxiter = 100, initialtau = 'iteration', Madj = TRUE, display = TRUE, parallel = FALSE)
\end{verbatim}

\section*{Arguments}

- \texttt{X} \hspace{1cm} The matrix to be decomposed.
- \texttt{k} \hspace{1cm} A vector containing the number of factors to try.
- \texttt{fold} \hspace{1cm} The number of folds used in cross validation.
- \texttt{tau} \hspace{1cm} Initial tau value to fit. Will be overwritten by the first value in initial argument.
- \texttt{cut} \hspace{1cm} To delete columns that has more than 100('Cut')\% zeros. Cut = 1, if no filtering.
- \texttt{tolLnlikelihood} \hspace{1cm} The max percentage of log likelihood differences in two iterations.
- \texttt{iter} \hspace{1cm} Max iterations.
- \texttt{initialtau} \hspace{1cm} A character specifying the way to choose the initial value of tau at the beginning of EM iteration. \texttt{stable}: estimate tau from fitted beta in last round; \texttt{initial}: always use the initially assigned tau in \texttt{tau} or \texttt{initial}. Use the default tau = 0.1 if 'initial' is empty. \texttt{iteration}: use fitted tau in last round.
- \texttt{tol} \hspace{1cm} Percentage of l2 norm change of [tau beta].
- \texttt{maxiter} \hspace{1cm} Max iteration number in the zero inflated poisson regression.
- \texttt{Madj} \hspace{1cm} If TRUE then adjust for relative library size M.
- \texttt{display} \hspace{1cm} If TRUE display the fitting procedure.
- \texttt{parallel} \hspace{1cm} Use \texttt{doParallel} and \texttt{foreach} package to accelerate.

\section*{Details}

The function conducts cross validation on the zero-inflated Poisson factor analysis to determine the rank.

\section*{Value}

The function returns a matrix. Each row the CV likelihood of one fold. Each column is the result of number of factors in \texttt{k}. 
EMzeropoisson_mat

Author(s)
Tianchen Xu

Examples

data(X)
cv_ZIPFA(X, fold = 10, k = c(3,4))

EMzeropoisson_mat | Zero Inflated Possion Regression

Description
The zero inflated possion regression model.

Usage
EMzeropoisson_mat(data, tau = 0.1, initial = NULL, inititaltau = 'iteration',
tol = 1e-4, maxiter = 100, Madj = FALSE, m = NULL,
display = TRUE, intercept = TRUE)

Arguments
- **data**: A matrix with the first columns is y and the rest columns are x.
- **tau**: Initial tau value to fit. Will be overwritten by the first value in initial argument.
- **initial**: A list of initial values for the fitting. c(tau beta).
- **inititaltau**: A character specifying the way to choose the initial value of tau at the beginning of EM iteration. stable: estimate tau from fitted beta in last round; initial: always use the initially assigned tau in tau or initial. Use the default tau = 0.1 if 'initial' is empty. iteration: use fitted tau in last round.
- **tol**: Percentage of l2 norm change of [tau beta].
- **maxiter**: Max iteration number.
- **Madj**: If TRUE then adjust for relative library size M.
- **m**: A vector containing relative library size M.
- **display**: If TRUE display the fitting procedure.
- **intercept**: If TRUE then the model contains an intercept.

Details
The function estimates the coefficients in a new type of zero-inflated Poisson regression where the underlying Poisson rate is negatively associated with true zero probability.
**Value**

The function turns a matrix. Each row is fitted value in each iteration. The last row the final result. The first column is fitted tau. If intercept is ture, then the second column is the intercept, and the rest columns are other coefficients. If intercept is false, the rest columns are other coefficients.

**Author(s)**

Tianchen Xu

**Examples**

```r
n = 5000;
x1 = rnorm(n);
x2 = rnorm(n);
lam = exp(x1 - 2*x2 + 1.5);
y = rpois(n, lam)
tau = .75
p = 1/(1+lam^tau);
Z = rbinom(n, 1, p);
y[as.logical(Z)] = 0;
res = EMzeropoisson_mat(matrix(c(y,x1,x2),ncol=3), Madj = FALSE, intercept = TRUE)
```

---

**X**

*A simulated data X.*

**Description**

For exmaple run.

**Usage**

```r
data("X")
```

**Format**

The format is: int [1:200, 1:100] 1 1 1 0 0 0 0 2 ... - attr(*, "dimnames")=List of 2 ..$ : NULL ..$ : chr [1:100] "V1" "V2" "V3" "V4" ...

**Examples**

```r
data(X)
```
ZIPFA

Zero Inflated Poisson factor analysis

Description
To conduct a Zero Inflated Poisson factor analysis.

Usage
ZIPFA(X, k, tau = 0.1, cut = 0.8, tolLnlikelihood = 5e-4,
iter = 20, tol = 1e-4, maxiter = 100, initialtau = 'iteration',
Madj = TRUE, display = TRUE, missing = NULL)

Arguments
X The matrix to be decomposed.
k The number of factors.
tau Initial tau value to fit. Will be overwritten by the first value in initial argument.
cut To delete columns that has more than 100('Cut')% zeros. Cut = 1, if no filtering.
tolLnlikelihood The max percentage of log likelihood differences in two iterations.
iter Max iterations.
initialtau A character specifying the way to choose the initial value of tau at the beginning
of EM iteration. stable: estimate tau from fitted beta in last round; initial: always use the initially assigned tau in tau or initial. Use the default tau = 0.1 if 'initial' is empty. iteration: use fitted tau in last round.
tol Percentage of l2 norm change of [tau beta].
maxiter Max iteration number in the zero inflated poisson regression.
Madj If TRUE then adjust for relative library size M.
display If TRUE display the fitting procedure.
missing Reserved for cv_ZIPFA.

Details
The function conducts a zero-inflated Poisson factor analysis where the underlying Poisson rate is
negatively associated with true zero probability.

Value
tau Fitted tau value.
Ufit A list containing fitted U matrix in each iteration. The last one is the final fit.
Vfit A list containing fitted V matrix in each iteration. The last one is the final fit.
itr Number of iterations.
Likelihood The likelihood for the training data.
CVLikelihood The likelihood for the testing data (if applicable)
Author(s)
Tianchen Xu

Examples

data(X)
ZIPFA(X, k = 3)
Index

*Topic **cross validation**
  cv_ZIPFA, 2
*Topic **datasets**
  X, 4
*Topic **factor analysis**
  EMzeropoisson_mat, 3
  ZIPFA, 5
*Topic **zero inflated**
  EMzeropoisson_mat, 3
  ZIPFA, 5

cv_ZIPFA, 2

EMzeropoisson_mat, 3

X, 4

ZIPFA, 5