

Package: DBNMFrank (via r-universe)

September 7, 2024

Type Package

Title Rank Selection for Non-Negative Matrix Factorization

Version 0.1.0

Maintainer Yun Cai <Yun.Cai@dal.ca>

Description Given the non-negative data and its distribution, the package estimates the rank parameter for Non-negative Matrix Factorization. The method is based on hypothesis testing, using a deconvolved bootstrap distribution to assess the significance level accurately despite the large amount of optimization error. The distribution of the non-negative data can be either Normal distributed or Poisson distributed.

Imports NMF,pmldecon(>= 0.2.0)

License GPL (>= 3)

Encoding UTF-8

LazyData false

NeedsCompilation no

Author Yun Cai [aut, cre], Hong Gu [aut], Tobias Kenney [aut]

Repository CRAN

Date/Publication 2022-06-03 08:00:10 UTC

Contents

DBrank	2
Index	4

Description

The package estimates the rank parameter for Non-negative Matrix Factorization given the non-negative data and its distribution. The method is based on hypothesis testing, using a deconvolved bootstrap distribution to assess the significance level accurately despite the large amount of optimization error. The distribution of the non-negative data can be either Normal distributed or Poisson distributed.

Usage

```
DBrank(data,k,alpha,distn,sz,inisz)
```

Arguments

data	Matrix. The non-negative data. Its rows are different observations and columns are variables.
k	Optional. The value where the hypothesis test start.
alpha	Optional. The significance level. Default is 0.1.
distn	Character. The distribution of the non-negative data. It should be either "Normal" or "Poisson".
sz	Optional. The bootstrap size.
inisz	Optional. The number of initial values used to obtain the true maximum likelihood for NMF.

Details

Our rank selection for NMF is based on sequentially performing the following hypothesis test:

H_0 : the rank of the feature matrix is k .

H_a : the rank of the feature matrix is at least $k+1$.

After applying the goodness-of-fit test, if H_0 is rejected by significance level 'alpha', let $k=k+1$ and repeat the test until the pvalue is greater than 'alpha'. For our hypothesis test, the test statistic is the likelihood ratio. 'inisz' different initial values are used to get the maximum likelihood for rank 'k' NMF and rank 'k+1' NMF. We use a deconvolved parametric bootstrap to obtain the null distribution of the test statistic. The bootstrap size is 'sz'.

Value

rank	The NMF rank selected by the function.
pvalue	The pvalue for the estimated rank.

Examples

```
library(NMF)
set.seed(45217)
#####generate a rank 2 Poisson NMF data
x=syntheticNMF(50,2,30)
est.rank=DBrank(t(x),k=2,sz=50,inisz=6)
```

Index

DBrank, [2](#)