Package 'SparseFactorAnalysis'

July 23, 2015

Type Package

Title Scaling Count and Binary Data with Sparse Factor Analysis

Version 1.0

Date 2015-07-20

Author Marc Ratkovic, In Song Kim, John Londregan, and Yuki Shiraito

Maintainer Marc Ratkovic <ratkovic@princeton.edu>

Description Multidimensional scaling provides a means of uncovering a latent structure underlying observed data, while estimating the number of latent dimensions. This package presents a means for scaling binary and count data, for example the votes and word counts for legislators. Future work will include an EM implementation and extend this work to ordinal and continuous data.

License GPL (>= 2)

Depends directlabels, proto, ggplot2

Imports Rcpp (>= 0.11.4), MASS, VGAM, truncnorm

LinkingTo Rcpp, RcppArmadillo

NeedsCompilation yes

Repository CRAN

Date/Publication 2015-07-23 07:05:04

R topics documented:

`SparseFactorAnalysis-package` *Sparse factor analysis for mixed binary and count data.*

Description

Multi-dimensional scaling provides a means of uncovering a latent structure underlying observed data. This package presents a means for scaling binary and count data, for example the votes and word counts for legislators.

Details

Package:	SparseFactorAnalysis
Type:	Package
Version:	1.0
Date:	2015-03-21
License:	GPL (>= 2)

Author(s)

Marc Ratkovic, In Song Kim, John Londregan, and Yuki Shiraito Maintainer: Marc Ratkovic (ratkovic@princeton.edu)

References

In Song Kim, John Londregan, and Marc Ratkovic. 2015. "Voting, Speechmaking, and the Dimensions of Conflict in the US Senate." Working paper.

`plot.sfa` *Plotting output from sparse factor analysis.*

Description

Function for plotting coefficients from sparse factor analysis. Returns one of two types of plots: either a summary of estimated dimensionality or a scatterplot of two dimensions.

Usage

```
## S3 method for class 'sfa'
plot(x,...)
```

Arguments

x	Object from output of class sfa.
...	Additional items to pass to plot. Options below.

Details

The function produces either a summary of estimated dimensionality or a scatter plot of points between two dimensions. See options below.

type Either "dim" or "scatter". Whether to return a summary of dimensionality or scatterplot between two dimensions.

main, ylabel, xlabel Main title, y axis label, and x axis label.

dims.scatter Which two dimensions to plot against each other.

scatter.by Either row or col. Whether to produce scatterplot of two rows or of two columns.

topbottom Number of points to label at the extreme of each dimension. Only used for type="scatter".

Author(s)

Marc Ratkovic

References

In Song Kim, John Londregan, and Marc Ratkovic. 2015. "Voting, Speechmaking, and the Dimensions of Conflict in the US Senate." Working paper.

See Also

sfa

Examples

```
##  See sfa() for example.
```

---

sfa *Sparse factor analysis for mixed binary and count data.*

Description

Scaling mixed binary and count data while estimating the underlying latent dimensionality.

Usage

```
sfa(M, missing.mat=NULL, gibbs=100, burnin=100, max.optim=50,
  thin=1, save.curr="UDV_curr", save.each=FALSE, thin.save=25,
  maxdim=NULL)
```

Arguments

M	Matrix to be scaled.
missing.mat	Matrix indicating missing data. Should be the same size as M, with a 1 denoting a missing observation and a 0 otherwise. Defaults to all zeroes.
gibbs	Number of posterior samples to draw
burnin	Number of burnin samples.
max.optim	Number of iterations to fit the cutpoints using optim. This is generally faster than the Hamiltonian Monte Carlo estimates, and is useful for the first part of the burnin phase.
thin	Extent of thinning of the MCMC chain. Only every thin draw is saved to the output.
save.curr	Name of file in which to save object.
save.each	Whether to save with a new name at each thinned draw.
thin.save	How many thinned draws to wait between saving output.
maxdim	Number of latent dimensions to fit. Should be greater than the number of estimated dimensions.

Details

The function sfa is the main function in the package, SparseFactorAnalysis. It takes in a matrix which in *rows* has the same data type–either binary or count. For example, every row may consist of roll call votes or word counts, and the columns may correspond with legislators. The method combines the two data types, scales both, and selects the underlying latent dimensionality.

Value

dim.sparse	Output for sparse estimates of dimensionality.
dim.mean	Non-sparse estimates of posterior mean of dimensionality.
rowdim1	Posterior samples of first dimension of spatial locations for each observation i.
rowdim2	Posterior samples of second dimension of spatial locations for row unit of observation.
coldim1	Posterior samples of first dimension of spatial locations for column unit of observation.
coldim2	Posterior samples of second dimension of spatial locations for column unit of observation.
lambda.lasso	Posterior samples for tuning parameter used for dimension selection.
Z	Posterior mean of fitted values, on a z-scale.
rowdims.all	Posterior mean of all row spatial locations.
coldims.all	Posterior mean of all column spatial locations.

Author(s)

Marc Ratkovic and Yuki Shiraito

References

In Song Kim, John Londregan, and Marc Ratkovic. 2015. "Voting, Speechmaking, and the Dimensions of Conflict in the US Senate." Working paper.

See Also

plot.sfa, summary.sfa

Examples

```
## Not run:
##Sample size and dimensions.
 set.seed(1)
 n.sim<-50
 k.sim<-500

##True vector of dimension weights.
 d.sim<-rep(0,n.sim)
 d.sim[1:3]<-c(2, 1.5, 1)*3

##Formulate true latent dimensions.
 U.sim<-matrix(rnorm(n.sim^2,sd=.5), nr=n.sim, nc=n.sim)
 V.sim<-matrix(rnorm(n.sim*k.sim,sd=.5), nr=k.sim, nc=n.sim)
 Theta.sim<-U.sim%*%diag(d.sim)%*%t(V.sim)

##Generate binary outcome and count data.
 probs.sim<-pnorm((-1+Theta.sim+rep(1,n.sim)%*%t(rnorm(k.sim,sd=.5)) +
   rnorm(n.sim,sd=.5)%*%t(rep(1,k.sim))   ))
 votes.mat<-
    apply(probs.sim[1:25,],c(1,2),FUN=function(x) rbinom(1,1,x))
 count.mat<-
    apply(probs.sim[26:50, ],c(1,2),FUN=function(x) rpois(1,20*x))
 M<-rbind(votes.mat,count.mat)

## Run sfa
 sparse1<-sfa(M, maxdim=10)

##Analyze results.
 summary(sparse1)
 plot(sparse1,type="dim")
 plot(sparse1,type="scatter")

##Compare to true data generating process

plot(sparse1$Z,Theta.sim)
abline(c(0,1))


## End(Not run)
```

summary.sfa *Summaries from sparse factor analysis.*

Description

Returns a summary for the dimensionality estimate, first and second dimensions, and additional dimensions as desired.

Usage

```
## S3 method for class 'sfa'
summary(object,... )
```

Arguments

object	Object of type sfa.
...	Additional items to pass to summary. Options below.

Details

Generates a table for an object of class sfa. Additional arguments to pass summary below.

interval A number between 0 and 1. Length of symmetric posterior credible interval.

topbottom A positive integer. How many observations at the top and bottom of each dimension to print.

print.dims Optional. Additional dimensions besides the first two for which to return the top and bottom topbottom observations.

References

Ratkovic, Marc and Tingley, Dustin. 2015. "Sparse Estimation with Uncertainty: Subgroup Analysis in Large Dimensional Design." Working paper.

See Also

sfa

Examples

```
##  See sfa() for example.
```

Index