

Package ‘sparsereg’

March 10, 2016

Type Package

Title Sparse Bayesian Models for Regression, Subgroup Analysis, and Panel Data

Version 1.2

Date 2016-03-01

Author Marc Ratkovic and Dustin Tingley

Maintainer Marc Ratkovic <ratkovic@princeton.edu>

Description

Sparse modeling provides a mean selecting a small number of non-zero effects from a large possible number of candidate effects. This package includes a suite of methods for sparse modeling: estimation via EM or MCMC, approximate confidence intervals with nominal coverage, and diagnostic and summary plots. The method can implement sparse linear regression and sparse probit regression. Beyond regression analyses, applications include subgroup analysis, particularly for conjoint experiments, and panel data. Future versions will include extensions to models with truncated outcomes, propensity score, and instrumental variable analysis.

License GPL (>= 2)

Depends R (>= 3.0.2), MASS, ggplot2

Imports Rcpp (>= 0.11.0), msm, VGAM, MCMCpack, coda, glmnet, gridExtra, grid, GIGrv

LinkingTo Rcpp, RcppArmadillo

NeedsCompilation yes

Repository CRAN

Date/Publication 2016-03-10 23:32:18

R topics documented:

sparsereg-package	2
difference	2
plot.sparsereg	4
print.sparsereg	5
sparsereg	6
summary.sparsereg	9
violinplot	10

Index**12**

 sparsereg-package *Sparse regression for experimental and observational data.*

Description

Sparse modeling provides a mean selecting a small number of non-zero effects from a large possible number of candidate effects. This package includes a suite of methods for sparse modeling: estimation via EM or MCMC, approximate confidence intervals with nominal coverage, and diagnostic and summary plots. Beyond regression analyses, applications include subgroup analysis, particularly for conjoint experiments, and panel data. Future versions will include extensions to limited dependent variables, models with truncated outcomes, and propensity score and instrumental variable analysis.

Details

Package: sparsereg
 Type: Package
 Version: 1.0
 Date: 2015-03-20
 License: GPL (>= 2)

Author(s)

Marc Ratkovic and Dustin Tingley
 Maintainer: Marc Ratkovic (ratkovic@princeton.edu)

References

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

See Also

FindIt, glmnet

 difference *Plotting difference in posterior estimates from a sparse regression.*

Description

Function for plotting differences in posterior density estimates for separate parameters from sparse regression analysis.

Usage

```
difference(x, type="mode", var1=NULL, var2=NULL, plot.it=TRUE,
  main="Difference", xlabel="Effect", ylabel="Density")
```

Arguments

x	Object of class <code>sparsereg</code> .
type	Whether to difference the posterior mode or posterior mean. Options are "mode" and "mean".
var1, var2	Variables names for the effects to difference.
plot.it	Whether to plot the density of the difference.
main, xlabel, ylabel	Main title, x-axis label, and y-axis label.

Details

Generates a density of the estimated posterior of the difference between the effects of two variables.

References

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

See Also

[sparsereg](#), [plot.sparsereg](#), [summary.sparsereg](#), [violinplot](#), [print.sparsereg](#)

Examples

```
## Not run:
set.seed(1)
n<-500
k<-100
Sigma<-diag(k)
Sigma[Sigma==0]<-.5
X<-mvrnorm(n,mu=rep(0,k),Sigma=Sigma)
y.true<-3+X[,2]*2+X[,3]*(-3)
y<-y.true+rnorm(n)

##Fit a linear model with five covariates.
s1<-sparsereg(y,X[,1:5])
difference(s1,var1=1,var2=2)

## End(Not run)
```

plot.sparsereg *Plotting output from a sparse regression.*

Description

Function for plotting coefficients from sparsereg analysis.

Usage

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

Arguments

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

Details

The function returns up to three plots in one figure. Each plot corresponds with main effects, interaction effects, and two-way interactions. Additional options to pass below.

main1, main2, main3 Main titles for plots of main effects, interactive effects, and two-way interactions.

xlabel Label for x-axis.

plot.one Takes on the value of FALSE or 1, 2, or 3, denoting whether to return a single plot for main effects (1), interactive effects (2), or two-way interactions (3).

References

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

See Also

[sparsereg](#), [plot.sparsereg](#), [summary.sparsereg](#), [violinplot](#), [difference](#)

Examples

```
## Not run:
set.seed(1)
n<-500
k<-100
Sigma<-diag(k)
Sigma[Sigma==0]<- .5
X<-mvrnorm(n,mu=rep(0,k),Sigma=Sigma)
y.true<-3+X[,2]*2+X[,3]*(-3)
y<-y.true+rnorm(n)
```

```
##Fit a linear model with five covariates.
s1<-sparsereg(y,X[,1:5])
plot(s1)

## End(Not run)
```

```
print.sparsereg      A summary of the estimated posterior mode of each parameter.
```

Description

The function prints a summary of the estimated posterior mode of each parameter.

Usage

```
## S3 method for class 'sparsereg'
print(x,... )
```

Arguments

```
x          Object of class sparsereg.
...        Additional arguments to pass to print. None supported in this version.
```

Details

Uses the summary function from the package coda to return a summary of the posterior mode of a sparsereg object.

References

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

See Also

[sparsereg](#), [plot.sparsereg](#), [summary.sparsereg](#), [violinplot](#), [difference](#)

Examples

```
## Not run:
set.seed(1)
n<-500
k<-100
Sigma<-diag(k)
Sigma[Sigma==0]<-.5
```

```

X<-mvrnorm(n,mu=rep(0,k),Sigma=Sigma)
y.true<-3+X[,2]*2+X[,3]*(-3)
y<-y.true+rnorm(n)

##Fit a linear model with five covariates.
s1<-sparsereg(y,X[,1:5])
print(s1)

## End(Not run)

```

sparsereg

Sparse regression for experimental and observational data.

Description

Function for fitting a Bayesian LASSOplus model for sparse models with uncertainty, facilitating the discovery of various types of interactions. Function takes a dependent variable, an optional matrix of (pre-treatment) covariates, and a (optional) matrix of categorical treatment variables. Includes correct calculation of uncertainty estimates, including for data with repeated observations.

Usage

```

sparsereg(y, X, treat=NULL, EM=FALSE, gibbs=200, burnin=200, thin=10,
type="linear", scale.type="none", baseline.vec=NULL,
id=NULL, id2=NULL, id3=NULL, save.temp=FALSE, conservative=TRUE)

```

Arguments

y	Dependent variable.
X	Covariates. Typical vocabulary would refer to these as "pre-treatment" covariates.
treat	Matrix of categorical treatment variables. May be a matrix with one column in the case of there being only one treatment variable.
EM	Whether to fit model via EM or MCMC. EM is much quicker, but only returns point estimates. MCMC is slower, but returns posterior intervals and approximate confidence intervals.
gibbs	Number of posterior samples to save. Between each saved sample, thin samples are drawn.
burnin	Number of burnin samples. Between each burnin sample, thin samples are drawn. These iterations will not be included in the resulting analysis.
thin	Extent of thinning of the MCMC chain. Between each posterior sample, whether burnin or saved, thin draws are made.
type	Type of regression model to fit. Allowed types are linear or probit.

baseline.vec	Optional vector with one entry for each column of the treatment matrix. Each entry gives the baseline condition for that treatment, which then during pre-processing is omitted for estimation so it serves as an excluded category.
id, id2, id3	Vectors the same length of the sample denoting clustering in the data. In a conjoint experiment with repeated observations, these correspond with respondent IDs. Up to three different sets of random effects are allowed.
scale.type	Indicates the types of interactions that will be created and used in estimation. <code>scale.type="none"</code> generates no interactions and corresponds to simply running LASSOplus with no interactions between variables. <code>scale.type="TX"</code> creates interactions between each X variable and each level of the treatment variables. <code>scale.type="TT"</code> creates interactions between each level of separate treatment variables. <code>scale.type="TTX"</code> interacts each X variable with all values generated by <code>scale.type="TT"</code> . Note that users can create their own interactions of interest, select <code>scale.type="none"</code> , to return the sparse version of the user specified model.
save.temp	Whether to save intermediate output in a file named <code>temp_sparsereg</code> . Useful for very long runs.
conservative	Experimental. If set to <code>FALSE</code> , the estimate is less conservative in selecting a variable.

Details

The function `sparsereg` allows for estimation of a broad range of sparse regressions. The method allows for continuous, binary, and censored outcomes. In experimental data, it can be used for subgroup analysis. It pre-processes lower-order terms to generate higher-order interactions terms that are uncorrelated with their lower order component, with pre-processing generated through `scale.type`. In observational data, it can be used in place of a standard regression, especially in the presence of a large number of variables. The method also adjusts uncertainty estimates when there are repeated observations through using random effects. For example, a conjoint design may have the same people make several comparisons, or a panel data regression may have multiple observations on the same unit.

The object contains the estimated posterior for all of the modeled effects, and analyzing the object is facilitated by the functions `plot`, `summary`, `violinplot`, and `difference`.

Value

<code>beta.mode</code>	Matrix of sparse (mode) estimates with rows equal to number of effects and columns for posterior samples.
<code>beta.mean</code>	Matrix of mean estimates with rows equal to number of effects and columns for posterior samples. These estimates are not sparse, but they do predict better than the mode.
<code>beta.ci</code>	Matrix of effects used to calculate approximate confidence intervals.
<code>sigma.sq</code>	Vector of posterior estimate of error variance.
<code>X</code>	Matrix of covariates fit. Includes interaction terms, depending on <code>scale.type</code> .
<code>varmat</code>	Matrix of showing which lower-order terms correspond with which effects. Used in producing figures.

baseline	Vector of baseline categories for treatments.
modeltype	Type of sparsereg model fit. In this case, onestage. Used by summary functions.

References

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

Egami, Naoki and Imai, Kosuke. 2015. "Causal Interaction in High-Dimension." Working paper.

See Also

[plot.sparsereg](#), [summary.sparsereg](#), [violinplot](#), [difference](#), [print.sparsereg](#)

Examples

```
## Not run:
set.seed(1)
n<-500
k<-5
treat<-sample(c("a","b","c"),n,replace=TRUE,pr=c(.5,.25,.25))
treat2<-sample(c("a","b","c","d"),n,replace=TRUE,pr=c(.25,.25,.25,.25))
Sigma<-diag(k)
Sigma[Sigma==0]<-.5
X<-mvrnorm(n,m=rep(0,k),S=Sigma)
y.true<-3+X[,2]*2+(treat=="a")*2 +(treat=="b")*(-2)+X[,2]*(treat=="b")*(-2)+
  X[,2]*(treat=="c")*2
y<-y.true+rnorm(n,sd=2)

##Fit a linear model.
s1<-sparsereg(y, X, cbind(treat,treat2), scale.type="TX")
s1.EM<-sparsereg(y, X, cbind(treat,treat2), EM=TRUE, scale.type="TX")

##Summarize results from MCMC fit
summary(s1)
plot(s1)
violinplot(s1)

##Summarize results from MCMC fit
summary(s1.EM)
plot(s1.EM)

##Extension using a baseline category
s1.base<-sparsereg(y, X, treat, scale.type="TX", baseline.vec="a")

summary(s1.base)
plot(s1.base)
violinplot(s1.base)

## End(Not run)
```

summary.sparsereg *Summaries for a sparse regression.*

Description

The function prints and returns a summary table for a sparsereg object.

Usage

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

Arguments

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

Details

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

interval Length of posterior interval to return. Must be between 0 and 1, default is .9. The symmetric interval is returned.

ci Type of interval to return. Options are "quantile" (default) for quantiles and "HPD" for the highest posterior density interval.

order How to order returned coefficients. Options are "magnitude", sorted by magnitude and omitting zero effects, "sort", sorted by size from highest to lowest and omitting zero effects, and "none" which returns all effects

normal Whether to return the normal approximate confidence interval (default of TRUE) or posterior interval (FALSE).

select Either "mode" or a number between 0 and 1. Whether to select variables for printing off the median of the mode (default) or off the probability of being non-zero.

printit Whether to print a summary table.

stage Currently this argument is ignored.

References

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

See Also

[sparsereg](#), [plot.sparsereg](#), [violinplot](#), [difference](#), [print.sparsereg](#)

Examples

```
## Not run:
set.seed(1)
n<-500
k<-100
Sigma<-diag(k)
Sigma[Sigma==0]<-0.5
X<-mvrnorm(n,mu=rep(0,k),Sigma=Sigma)
y.true<-3+X[,2]*2+X[,3]*(-3)
y<-y.true+rnorm(n)

##Fit a linear model with five covariates.
s1<-sparsereg(y,X[,1:5])
summary(s1)

## End(Not run)
```

violinplot

Function for plotting posterior distribution of effects of interest.

Description

The function produces a violin plot for specified effects. This can be useful for presenting or examining particular marginal effects of interest.

Usage

```
violinplot(x, columns=NULL, newlabels=NULL, type="mode", stage=NULL)
```

Arguments

x	Object of class sparsereg.
columns	A vector of numbers (or strings) corresponding to columns (or column names) to produce plots for.
newlabels	New labels for columns rather than variable names in object. If empty, variable names are used.
type	Options are "mode" and "mean". Whether to plot the posterior mode or mean.
stage	Currently, this argument is ignored.

Details

Generates a violin plot for coefficients from object from class sparsereg. The desired coefficients can be requested using the columns argument and they can be assigned new names through newlabels.

References

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

See Also

[sparsereg](#), [plot.sparsereg](#), [summary.sparsereg](#), [difference](#), [print.sparsereg](#)

Examples

```
## Not run:
set.seed(1)
n<-500
k<-100
Sigma<-diag(k)
Sigma[Sigma==0]<- .5
X<-mvrnorm(n,mu=rep(0,k),Sigma=Sigma)
y.true<-3+X[,2]*2+X[,3]*(-3)
y<-y.true+rnorm(n)

##Fit a linear model with five covariates.
s1<-sparsereg(y,X[,1:5])
violinplot(s1,1:3)

## End(Not run)
```

Index

*Topic **package**

sparsereg-package, 2

difference, 2, 4, 5, 8, 9, 11

plot.sparsereg, 3, 4, 4, 5, 8, 9, 11

print.sparsereg, 3, 5, 8, 9, 11

sparsereg, 3–5, 6, 9, 11

sparsereg-package, 2

summary.sparsereg, 3–5, 8, 9, 11

violinplot, 3–5, 8, 9, 10