

Package ‘grplassocat’

April 1, 2019

Type Package

Title Standardization for Group Lasso Models with Categorical Predictors

Version 1.0

Date 2019-02-28

Author Felicitas Detmer and Martin Slawski

Maintainer Felicitas Detmer <fdetmer@gmu.edu>

Description

Implements the simple and computationally efficient standardization scheme for group lasso models with categorical predictors described in Detmer, Cebal, Slawski (2019) <arXiv:1805.06915>.

Depends grplasso

License GPL-3

NeedsCompilation no

Repository CRAN

Date/Publication 2019-04-01 18:20:06 UTC

R topics documented:

| | |
|---------------------|---|
| fit_grp | 1 |
| internals | 3 |

| | |
|--------------|----------|
| Index | 4 |
|--------------|----------|

| | |
|---------|---|
| fit_grp | <i>Function to fit a group lasso model to a standardized feature matrix</i> |
|---------|---|

Description

Standardizes feature matrix including categorical features and fits a group lasso model

Usage

```
fit_grp(eqn, dat, lambda, model = LinReg(), nonpen = c(), standardize = TRUE, ...)
```

Arguments

| | |
|-------------|---|
| eqn | formula of the penalized variables. The response has to be on the left hand side of ~. If interaction terms are included without main effects, the main effects will automatically be added by the package. |
| dat | data.frame, categorical features need to be of type factor |
| lambda | Penalty parameter (scalar) |
| model | an object of class grpl.model as defined in the package grplasso. |
| nonpen | formula of the nonpenalized features |
| standardize | logical. If true, the design matrix of the continuous features will be centered and standardized to unit norm |
| ... | additional arguments to be passed to the grplasso function in the package of the same name. |

Details

Design matrices of the categorical features and interactions between categorical features are centered and standardized by column-wise scaling. After fitting a group lasso model to the standardized design matrix, coefficients are re-scaled and centered to the original scale of the data. Interactions between categorical and continuous features are standardized by a singular value decomposition.

Value

A dataframe containing the coefficients of the fitted group lasso model that have been re-scaled to the original scale of the data is returned. Coefficients of interaction terms for which no observations are included in dat are returned as NA.

Author(s)

Felicitas Detmer, <fdetmer@gmu.edu>

References

Detmer, Felicitas J., and Martin Slawski. "A Note on Coding and Standardization of Categorical Variables in (Sparse) Group Lasso Regression." arXiv preprint arXiv:1805.06915 (2018).

Examples

```
data(dattest)

#---set datatype of categorical features to factor-----
dattest$X1cut=as.factor(dattest$X1cut)
dattest$X2cut=as.factor(dattest$X2cut)
dattest$X3cut=as.factor(dattest$X3cut)
```

```
table(dattest[,c("X1cut", "X2cut", "X3cut")])

#--fit group lasso models
coefs1=fit_grp(y~X1cut * X2cut +X1cut * X3cut +X2cut * X3cut, dattest, lambda=0.5, model=LinReg())
coefs2=fit_grp(y~X1cut * X2cut +X1cut * X3cut +X2cut * X3cut, dattest, lambda=0.5, model=LinReg(),
               nonpen=~X1cut)
```

internals

Miscellaenous

Description

gen_dum, func_int, and seqWrapper are internal functions not intended to be called directly by the user. dattest is a synthetic data set for testing and demonstration purposes.

Author(s)

Felicitas Detmer, <fdetmer@gmu.edu>

Index

*Topic **models**

fit_grp, 1

*Topic **regression**

fit_grp, 1

datatest (internals), 3

fit_grp, 1

func_int (internals), 3

gen_dum (internals), 3

grplasso-package (fit_grp), 1

internals, 3

seqWrapper (internals), 3