Package ‘lmmlasso’

December 23, 2011

Type Package

Title Linear mixed-effects models with Lasso

Version 0.1-2

Date 2011-04-13

Author Juerg Schelldorfer

Maintainer Lukas Meier <meier@stat.math.ethz.ch>

Description This package fits (gaussian) linear mixed-effects models for high-dimensional data (n<<p) using a Lasso-type approach for the fixed-effects parameter.

Depends methods, emulator, miscTools, penalized

License GPL

LazyLoad yes

R topics documented:

<table>
<thead>
<tr>
<th>R topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>lmmlasso-package</td>
<td>1</td>
</tr>
<tr>
<td>classroomStudy</td>
<td>2</td>
</tr>
<tr>
<td>lmmlasso</td>
<td>3</td>
</tr>
<tr>
<td>lmmlassoControl</td>
<td>6</td>
</tr>
<tr>
<td>plot.lmmlasso</td>
<td>7</td>
</tr>
<tr>
<td>print.lmmlasso</td>
<td>8</td>
</tr>
<tr>
<td>summary.lmmlasso</td>
<td>8</td>
</tr>
</tbody>
</table>

Index

10

Description

Fits (gaussian) linear mixed-effects models with lasso penalty for the fixed effects.

Details
This is an early test version.

Author(s)

Juerg Schelldorfer
Maintainer: Juerg Schelldorfer <schell@stat.math.ethz.ch>

References


classroomStudy

Dataset of students math achievements

Description

This is a subset of the Classroom Study described in West et. al. (2007) with 156 observations and 6 variables.

Usage

data(classroomStudy)

Format

A list with the following four components.

Response variable. The students math achievement gain.

grp

Grouping variable comprising the class of the students.

X Fixed-effect design matrix. The first column is the intercept, then sex, minority, mathkind, ses, yearstea and mathprep.

Z Random-effects design matrix. Only one column in order to fit a random intercept model.

Details

A detailed description of the covariates can be found in West et. al. (2007) and was carried out by Hill et. al. (2005)

Source

http://www-personal.umich.edu/~bwest/classroom.csv
References


Examples

```r
data(classroomStudy)

lmmlasso
```

Function to fit high-dimensional (gaussian) linear mixed-effects models

Description

Fits the solution for a high-dimensional (gaussian) linear mixed-effects models

Usage

```r
lmmlasso(x, ...)
```

## Default S3 method:

```r
lmmlasso(x, y, z = x, grp, weights = NULL, coefInit=NULL, lambda,
startValue = 1, nonpen = 1:dim(z)[[2]], pdMat = c("pdIdent", "pdDiag","pdSym"),
method = "ML", CovOpt = c("nlminb","optimize"), stopSat = TRUE,
standardize = TRUE, control = lmmlassoControl(),ranInd=1:dim(z)[[2]],...)
```

Arguments

- **x**
  - matrix of dimension ntot x p including the fixed-effects covariables. An intercept has to be included in the first column as (1,...,1).
- **y**
  - response variable of length ntot.
- **z**
  - random effects matrix of dimension ntot x q. It has to be a matrix, even if q=1.
- **grp**
  - grouping variable of length ntot
- **weights**
  - weights for the fixed-effects covariates: NA means no penalization, 0 means drop this covariate ; if given, the argument nonpen is ignored. By default each covariate has weight 1
- **coefInit**
  - list with three components used as starting values for the fixed effects, the random effects variance components and the error standard deviation.
- **lambda**
  - positive regularization parameter
- **startValue**
  - Choice of the starting values for the fixed effects using linear regression. 1 means 10-fold cross-validation with L1-penalty, 2 means 10-fold cross-validation Ridge Regression and 0 means that all the covariates are set to zero and the intercept is the mean of the response variable
- **nonpen**
  - index of fixed effects with no penalization, ignored if the argument weights is specified, default is 1, which means that only the intercept (the first column in X) is not penalized.
Immlasso

\texttt{pdMat}\quad \text{Covariance structure for the random effects. pdIdent, pdDiag and pdSym are already implemented. Default to pdIdent. pdIdent: } b_i \sim \mathcal{N}(0, \theta^2 I) \ (1 \text{ parameter}), \ pdDiag: b_i \sim \mathcal{N}(0, \text{diag}(\theta_1^2, \ldots, \theta_q^2)) \ (q \text{ parameters}), \ pdSym: b_i \sim \mathcal{N}(0, \Psi) \quad \text{where } \Psi \text{ is symmetric positive definit} \ (q(q+1)/2 \text{ parameters})

\text{method}\quad \text{Only "ML" is allowed. "REML" is not yet implemented.}

\text{CovOpt}\quad \text{which optimization routine should be used for updating the variance components parameters. optimize or nlminb. nlminb uses the estimate of the last iteration as a starting value. nlminb is faster if there are many Gauss-Seidel iterations.}

\text{stopSat}\quad \text{logical. Should the algorithm stop when ntot > p?}

\text{standardize}\quad \text{Should the x matrix be standardized such that each column has mean 0 and standard deviation 1? Be careful if the x matrix includes dummy variables.}

\text{control}\quad \text{control parameters for the algorithm and the Armijo Rule, see \texttt{lmmlassoControl} for the details}

\text{ranInd}\quad \text{Index of the random effects with respect to the x matrix}

\ldots\quad \text{not used.}

\textbf{Details}

All the details of the algorithm can be found in Schelldorfer et. al. (2010).

\textbf{Value}

A \texttt{lmmlasso} object is returned, for which \texttt{coef, resid, fitted, print, summary, plot methods} exist.

\texttt{coefficients}\quad \text{estimated fixed-effects coefficients } \hat{\beta}

\texttt{pars}\quad \text{free parameters in the covariance matrix } \Psi \text{ of the random effects}

\texttt{sigma}\quad \text{standard deviation } \hat{\sigma} \text{ of the errors}

\texttt{random}\quad \text{vector with random effects, sorted by groups}

\texttt{u}\quad \text{vector with the standardized random effects, sorted by effect}

\texttt{ranef}\quad \text{vector with random effects, sorted by effect}

\texttt{fixef}\quad \text{estimated fixed-effects coefficients } \hat{\beta}

\texttt{fitted.values}\quad \text{The fitted values } \hat{y} = \hat{X} \hat{\beta} + \hat{Z} \hat{b}_i

\texttt{residuals}\quad \text{raw residuals } y - \hat{y}

\texttt{Psi}\quad \text{Covariance matrix } \Psi \text{ of the random effects}

\texttt{corPsi}\quad \text{Correlation matrix of the random effects}

\texttt{logLik}\quad \text{value of the log-likelihood function}

\texttt{deviance}\quad \text{deviance=-2*logLik}

\texttt{npar}\quad \text{Number of parameters. Corresponds to the cardinality of the active set of coefficients plus the number of free parameters in } \Psi

\texttt{aic}\quad \text{AIC}

\texttt{bic}\quad \text{BIC}

\texttt{data}\quad \text{data set, as a list with four components: } x, y, z, \text{ grp (see above)}

\texttt{weights}\quad \text{weights for the fixed-effects covariates}

\texttt{nonpen}\quad \text{nonpenalized covariates. Differ from the input if weights is explicitly given}
Immlasso

coefInit list with three components used as starting values
lambda positive regularization parameter
converged Does the algorithm converge? 0: correct convergence; an odd number means that maxIter was reached; an even number means that the Armijo step was not successful. For each unsuccessful Armijo step, 2 is added to converged. If converged is large compared to the number of iterations counter, you may increase maxArmijo.
counter The number of iterations used.
stopped logical. Does the algorithm stopped due to ntot > p?
pdMat Covariance structure for the random effects
method "ML"
CovOpt optimization routine
control see lmmlassoControl
call call
stopped logical. Does the algorithm stopped due to a too large active set?
ranInd Index of the random effects with respect to the x matrix
objective Value of the objective function at the final estimates

Author(s)

Juerg Schelldorfer, <schell@stat.math.ethz.ch>

References


Examples

# (1) Use the lmmlasso on the classroomStudy data set
data(classroomStudy)
fit1 <- lmmlasso(x=classroomStudy$X, y=classroomStudy$y, z=classroomStudy$Z, grp=classroomStudy$grp, lambda=15, pdMat="pdIdent")
summary(fit1)
plot(fit1)

# (2) Use the lmmlasso on a small simulated data set
set.seed(54)
N <- 20  # number of groups
p <- 6  # number of covariates (including intercept)
q <- 2  # number of random effect covariates
ni <- rep(6, N)  # observations per group
ntot <- sum(ni)  # total number of observations
grp <- factor(rep(1:N, times=ni))  # grouping variable

beta <- c(1, 2, 4, 3, 0, 0, 0)  # fixed-effects coefficients
x <- cbind(1, matrix(rnorm(ntot*p), nrow=ntot), nrow=ntot))  # design matrix
bi1 <- rep(rnorm(N,0,3),times=ni) # Psi=diag(3,2)
bi2 <- rep(rnorm(N,0,2),times=ni)
bi <- rbind(bi1,bi2)

z <- x[,1:2,drop=FALSE]

y <- numeric(ntot)
for (k in 1:ntot) y[k] <- x[k,]%*%beta + t(z[k,])%*%bi[,grp[k]] + rnorm(1)

# correct random effects structure
fit2 <- lmmlasso(x=x,y=y,z=z,grp=grp,lambda=1/zero.noslash,pdMat="pdDiag")
summary(fit2)
plot(fit2)

# wrong random effects structure
fit3 <- lmmlasso(x=x,y=y,z=x[,1:3],grp=grp,lambda=10,pdMat="pdDiag")
summary(fit3)
plot(fit3)

---

**lmmlassoControl**

**Options for the lmmlasso Algorithm**

**Description**

Definition of various kinds of options in the algorithm.

**Usage**

lmmlassoControl(tol = 10^(-4), trace = 1, maxIter = 1000,
maxArmijo = 20, number = 5, a_init = 1, delta = 0.1, rho = 0.001,
gamma = 0, lower = 10^(-6), upper = 10^8, seed = 532,
VarInt = c(0, 10), CovInt = c(-5, 5), thres = 10^(-4))

**Arguments**

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>tol</td>
<td>convergence tolerance</td>
</tr>
<tr>
<td>trace</td>
<td>integer. 1 prints no output, 2 prints warnings, 3 prints the current function values and warnings (not recommended)</td>
</tr>
<tr>
<td>maxIter</td>
<td>maximum number of (outer) iterations</td>
</tr>
<tr>
<td>maxArmijo</td>
<td>maximum number of steps to be chosen in the Armijo Rule. If the maximum is reached, the algorithm continues with optimizing the next coordinate.</td>
</tr>
<tr>
<td>number</td>
<td>integer. Determines the active set algorithm. The zero fixed-effects coefficients are only updated each number iteration. It may be that a smaller number increases the speed of the algorithm. Use 0 ≤ number ≤ 5.</td>
</tr>
<tr>
<td>a_init</td>
<td>α_{init} in the Armijo step. See Schelldorfer et. al. (2010).</td>
</tr>
<tr>
<td>delta</td>
<td>δ in the Armijo step. See Schelldorfer et. al. (2010)</td>
</tr>
<tr>
<td>rho</td>
<td>ρ in the Armijo step. See Schelldorfer et. al. (2010)</td>
</tr>
<tr>
<td>gamma</td>
<td>γ in the Armijo step. See Schelldorfer et. al. (2010)</td>
</tr>
<tr>
<td>lower</td>
<td>lower bound for the Hessian</td>
</tr>
</tbody>
</table>
upper bound for the Hessian
set.seed for calculating the starting value, which performs a 10-fold cross-validation.

VarInt Only for opt="optimize". The interval for the variance parameters used in "optimize". See help("optimize")

CovInt Only for opt="optimize". The interval for the covariance parameters used in "optimize". See help("optimize")

If a variance or covariance parameter has smaller absolute value than thres, the parameter is set to exactly zero.

Details
For the Armijo step parameters, see Bertsekas (2003)

References

---

plot.lmmlasso Diagnostic Plots for a lmmlasso object

Description
Plots six diagnostic plots for checking the model assumptions and supporting model selection for a lmmlasso object

Usage
## S3 method for class 'lmmlasso'
plot(x, ...)

Arguments
x a lmmlasso object
...

not used.

Details
plot.lmmlasso shows six diagnostic plots which support checking the model assumption, model fit and may give hints for another model. 1) The first plot depicts the Tukey-Anscombe plot in order to check the assumptions about the errors. Points with the same color belong to the same group. 2) QQ-Plot of the residuals with equal coloring for each group. 3) QQ-Plot of the predicted random effects. Be careful with the interpretation since the random effects have not been standardized. The color shows which points belong to the same random-effects covariate. 4) Boxplot of the predicted random effects for each random-effects variable. 5) Plot of the fitted y-values against the observed y-values. 6) A histogram of the fixed-effects coefficients.
Examples

```r
data(classroomStudy)
fit <- lmmlasso(x=classroomStudy$X, y=classroomStudy$y, z=classroomStudy$Z, grp=classroomStudy$grp, lambda=15, pdMat = "pdIdent")
plot(fit)

print(fit)
```

Description

Prints a short summary of an lmmlasso object comprising information about the variance components parameters and the number of nonzero fixed-effects coefficients.

Usage

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

Arguments

- `x`: an lmmlasso object
- `...`: not used

See Also

summary.lmmlasso

Examples

```r
data(classroomStudy)
fit <- lmmlasso(x=classroomStudy$X, y=classroomStudy$y, z=classroomStudy$Z, grp=classroomStudy$grp, lambda=15, pdMat = "pdIdent")
print(fit)
```

summary.lmmlasso

Description

Providing an elaborate summary of an lmmlasso object.

Usage

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

Arguments

- `object`: an lmmlasso object
- `...`: not used.
Details

This function shows a detailed summary of a `lmmlasso` object. In the fixed-effects part, (n) right from a fixed-effects coefficient means that this coefficient was not subject to penalization.

Examples

data(classroomStudy)
fit <- lmmlasso(x=classroomStudy$X, y=classroomStudy$y, z=classroomStudy$Z, grp=classroomStudy$grp, lambda=15, pdMat="pdIdent")
summary(fit)
Index

*Topic **datasets**
  classroomStudy, 2

*Topic **hplot**
  plot.lmmlasso, 7

*Topic **methods**
  summary.lmmlasso, 8

*Topic **misc**
  lmmlassoControl, 6

*Topic **models**
  lmmlasso, 3

*Topic **package**
  lmmlasso-package, 1

*Topic **regression**
  lmmlasso, 3

classroomStudy, 2

lmmlasso, 3
lmmlasso-package, 1
lmmlassoControl, 6

plot.lmmlasso, 7
print.lmmlasso, 8

summary.lmmlasso, 8