Package ‘fastglm’

March 8, 2019

Type Package

Title Fast and Stable Fitting of Generalized Linear Models using 'RcppEigen'

Version 0.0.1

Maintainer Jared Huling <jaredhuling@gmail.com>


BugReports https://github.com/jaredhuling/fastglm/issues

License GPL (>= 2)

Encoding UTF-8

Imports Rcpp (>= 0.12.13)

LinkingTo Rcpp, RcppEigen

RoxygenNote 6.1.0

Suggests knitr, rmarkdown, glm2

VignetteBuilder knitr

NeedsCompilation yes

Author Jared Huling [aut, cre],
  Douglas Bates [cph],
  Dirk Eddelbuettel [cph],
  Romain Francois [cph],
  Yixuan Qiu [cph]

Repository CRAN

Date/Publication 2019-03-08 15:42:44 UTC
R topics documented:

deviance.fastglm . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 2
family.fastglm . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 2
fastglm . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 3
fastglmPure . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 5
logLik.fastglm . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 7
predict.fastglm . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 7
print.fastglm . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 8
residuals.fastglm . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 8
summary.fastglm . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 9

Index

<table>
<thead>
<tr>
<th>deviance.fastglm</th>
<th>deviance method for fastglm fitted objects</th>
</tr>
</thead>
</table>

Description

deviance method for fastglm fitted objects

Usage

```r
## S3 method for class 'fastglm'
deviance(object, ...)
```

Arguments

- `object`: fastglm fitted object
- `...`: not used

Value

The value of the deviance extracted from the object

<table>
<thead>
<tr>
<th>family.fastglm</th>
<th>family method for fastglm fitted objects</th>
</tr>
</thead>
</table>

Description

family method for fastglm fitted objects

Usage

```r
## S3 method for class 'fastglm'
family(object, ...)
```
**fastglm**

**Arguments**

- `object`  
  fastglm fitted object
- `...`  
  not used

**Value**

returns the family of the fitted object

---

**fastglm**  
*fast generalized linear model fitting*

**Description**

fast generalized linear model fitting
bigLm default

**Usage**

`fastglm(x, ...)`

## Default S3 method:
`fastglm(x, y, family = gaussian(), weights = NULL, offset = NULL, start = NULL, etastart = NULL, mustart = NULL, method = 0L, tol = 1e-08, maxit = 100L, ...)`

**Arguments**

- `x`  
  input model matrix. Must be a matrix object
- `...`  
  not used
- `y`  
  numeric response vector of length nobs.
- `family`  
  a description of the error distribution and link function to be used in the model. For fastglm this can be a character string naming a family function, a family function or the result of a call to a family function. For fastglmPure only the third option is supported. (See family for details of family functions.)
- `weights`  
  an optional vector of 'prior weights' to be used in the fitting process. Should be a numeric vector.
- `offset`  
  this can be used to specify an a priori known component to be included in the linear predictor during fitting. This should be a numeric vector of length equal to the number of cases
- `start`  
  starting values for the parameters in the linear predictor.
- `etastart`  
  starting values for the linear predictor.
- `mustart`  
  values for the vector of means.
method  an integer scalar with value 0 for the column-pivoted QR decomposition, 1 for
the unpivoted QR decomposition, 2 for the LLT Cholesky, or 3 for the LDLT
Cholesky

tol   threshold tolerance for convergence. Should be a positive real number

maxit maximum number of IRLS iterations. Should be an integer

Value

A list with the elements

coefficients  a vector of coefficients

se           a vector of the standard errors of the coefficient estimates

rank         a scalar denoting the computed rank of the model matrix

df.residual  a scalar denoting the degrees of freedom in the model

residuals    the vector of residuals

s            a numeric scalar - the root mean square for residuals

fitted.values the vector of fitted values

Examples

x <- matrix(rnorm(10000 * 100), ncol = 100)
y <- 1 * (0.25 * x[,1] - 0.25 * x[,3] > rnorm(10000))

system.time(gl1 <- glm.fit(x, y, family = binomial()))

system.time(gf1 <- fastglm(x, y, family = binomial()))

system.time(gf2 <- fastglm(x, y, family = binomial(), method = 1))

system.time(gf3 <- fastglm(x, y, family = binomial(), method = 2))

system.time(gf4 <- fastglm(x, y, family = binomial(), method = 3))

max(abs(coef(gl1) - gf1$coef))
max(abs(coef(gl1) - gf2$coef))
max(abs(coef(gl1) - gf3$coef))
max(abs(coef(gl1) - gf4$coef))
**fastglmPure**  
*fast generalized linear model fitting*

**Description**

fast generalized linear model fitting

**Usage**

```r
fastglmPure(x, y, family = gaussian(), weights = rep(1, NROW(y)),
            offset = rep(0, NROW(y)), start = NULL, etastart = NULL,
            mustart = NULL, method = 0L, tol = 1e-07, maxit = 100L)
```

**Arguments**

- `x` input model matrix. Must be a matrix object
- `y` numeric response vector of length nobs.
- `family` a description of the error distribution and link function to be used in the model. For `fastglmPure` this can only be the result of a call to a family function. (See `family` for details of family functions.)
- `weights` an optional vector of 'prior weights' to be used in the fitting process. Should be a numeric vector.
- `offset` this can be used to specify an a priori known component to be included in the linear predictor during fitting. This should be a numeric vector of length equal to the number of cases
- `start` starting values for the parameters in the linear predictor.
- `etastart` starting values for the linear predictor.
- `mustart` values for the vector of means.
- `method` an integer scalar with value 0 for the column-pivoted QR decomposition, 1 for the unpivoted QR decomposition, 2 for the LLT Cholesky, 3 for the LDLT Cholesky, 4 for the full pivoted QR decomposition, 5 for the Bidiagonal Divide and Conquer SVD
- `tol` threshold tolerance for convergence. Should be a positive real number
- `maxit` maximum number of IRLS iterations. Should be an integer

**Value**

A list with the elements

- `coefficients` a vector of coefficients
- `se` a vector of the standard errors of the coefficient estimates
- `rank` a scalar denoting the computed rank of the model matrix
- `df.residual` a scalar denoting the degrees of freedom in the model
- `residuals` the vector of residuals
- `s` a numeric scalar - the root mean square for residuals
- `fitted.values` the vector of fitted values
Examples

```r
set.seed(1)
x <- matrix(rnorm(1000 * 25), ncol = 25)
eta <- 0.1 + 0.25 * x[,1] - 0.25 * x[,3] + 0.75 * x[,5] - 0.35 * x[,6] #0.25 * x[,1] - 0.25 * x[,3]
y <- 1 * (eta > rnorm(1000))

yp <- rpois(1000, eta ^ 2)
yg <- rgamma(1000, exp(eta) * 1NWU, 1WU)

# binomial
system.time(gl1 <- glm.fit(x, y, family = binomial()))

system.time(gf1 <- fastglmpure(x, y, family = binomial(), tol = 1e-8))

system.time(gf2 <- fastglmpure(x, y, family = binomial(), method = 1, tol = 1e-8))

system.time(gf3 <- fastglmpure(x, y, family = binomial(), method = 2, tol = 1e-8))

system.time(gf4 <- fastglmpure(x, y, family = binomial(), method = 3, tol = 1e-8))

max(abs(coef(gl1) - gf1$coef))

max(abs(coef(gl1) - gf2$coef))

max(abs(coef(gl1) - gf3$coef))

max(abs(coef(gl1) - gf4$coef))

# poisson
system.time(gl1 <- glm.fit(x, yp, family = poisson(link = "log")))

system.time(gf1 <- fastglmpure(x, yp, family = poisson(link = "log"), tol = 1e-8))

system.time(gf2 <- fastglmpure(x, yp, family = poisson(link = "log"), method = 1, tol = 1e-8))

system.time(gf3 <- fastglmpure(x, yp, family = poisson(link = "log"), method = 2, tol = 1e-8))

system.time(gf4 <- fastglmpure(x, yp, family = poisson(link = "log"), method = 3, tol = 1e-8))

max(abs(coef(gl1) - gf1$coef))

max(abs(coef(gl1) - gf2$coef))

max(abs(coef(gl1) - gf3$coef))

max(abs(coef(gl1) - gf4$coef))

# gamma
system.time(gl1 <- glm.fit(x, yg, family = Gamma(link = "log")))

system.time(gf1 <- fastglmpure(x, yg, family = Gamma(link = "log"), tol = 1e-8))

system.time(gf2 <- fastglmpure(x, yg, family = Gamma(link = "log"), method = 1, tol = 1e-8))

system.time(gf3 <- fastglmpure(x, yg, family = Gamma(link = "log"), method = 2, tol = 1e-8))

system.time(gf4 <- fastglmpure(x, yg, family = Gamma(link = "log"), method = 3, tol = 1e-8))
```
logLik.fastglm

### Description

logLik method for fastglm fitted objects

### Usage

```r
## S3 method for class 'fastglm'
logLik(object, ...)  
```

### Arguments

- `object`: fastglm fitted object
- `...`: not used

### Value

Returns an object of class `loglik`

---

predict.fastglm

#### Obtains predictions and optionally estimates standard errors of those predictions from a fitted generalized linear model object.

### Description

Obtains predictions and optionally estimates standard errors of those predictions from a fitted generalized linear model object.

### Usage

```r
## S3 method for class 'fastglm'
predict(object, newdata = NULL, type = c("link", "response"), se.fit = FALSE, dispersion = NULL, ...)  
```
Arguments

object a fitted object of class inheriting from "fastglm".
newdata a matrix to be used for prediction
type the type of prediction required. The default is on the scale of the linear predictors; the alternative "response" is on the scale of the response variable. Thus for a default binomial model the default predictions are of log-odds (probabilities on logit scale) and type = "response" gives the predicted probabilities. The "terms" option returns a matrix giving the fitted values of each term in the model formula on the linear predictor scale.
The value of this argument can be abbreviated.
se.fit logical switch indicating if standard errors are required.
dispersion the dispersion of the GLM fit to be assumed in computing the standard errors. If omitted, that returned by summary applied to the object is used.
... further arguments passed to or from other methods.

print.fastglm print method for fastglm objects

Description

print method for fastglm objects

Usage

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

Arguments

x object to print
... not used

residuals.fastglm residuals method for fastglm fitted objects

Description

residuals method for fastglm fitted objects

Usage

## S3 method for class 'fastglm'
residuals(object, type = c("deviance", "pearson", "working", "response", "partial"), ...)

Arguments

- **object**: fastglm fitted object
- **type**: type of residual to be returned
- **...**: not used

Value

- a vector of residuals

---

**summary.fastglm**  
summary method for fastglm fitted objects

### Description

summary method for fastglm fitted objects

### Usage

```r
## S3 method for class 'fastglm'
summary(object, dispersion = NULL, ...)
```

### Arguments

- **object**: fastglm fitted object
- **dispersion**: the dispersion parameter for the family used. Either a single numerical value or NULL (the default), when it is inferred from object.
- **...**: not used

### Value

- a summary.fastglm object

### Examples

```r
x <- matrix(rnorm(10000 * 10), ncol = 10)
y <- 1 * (0.25 * x[,1] - 0.25 * x[,3] > rnorm(10000))

fit <- fastglm(x, y, family = binomial())

#summary(fit)
```
Index

deviance.fastglm, 2

family, 3, 5
family.fastglm, 2
fastglm, 3
fastglmPure, 5

logLik.fastglm, 7

predict.fastglm, 7
print.fastglm, 8

residuals.fastglm, 8

summary.fastglm, 9