Package ‘Homework1’

November 7, 2013

Type Package

Title HW1 fastlm and dmvnorm

Version 1.0

Date 2013-11-05

Author Detian Deng

Maintainer Detian Deng <ddeng@jhsph.edu>

Description This package is for 140.778 HW1. It has two functions. The fastlm function fits a linear model given the design matrix X and outcome vector Y. The dmvnorm function evaluates the multivariate normal density given the matrix of quantiles, vector of mean, and covariance matrix.

License GPL-2

R topics documented:

`Homework1-package` *HW1 fastlm and dmvnorm*

Description

This package is for 140.778 HW1. It has two functions. The fastlm function fits a linear model given the design matrix X and outcome vector Y. The dmvnorm function evaluates the multivariate normal density given the matrix of quantiles, vector of mean, and covariance matrix.

Details

Package:	Homework1
Type:	Package
Version:	1.0
Date:	2013-11-05
License:	What license is it under?

This package has two functions. The 'fastlm' function fits a linear model given the design matrix X and outcome vector Y. The 'dmvnorm' function evaluates the multivariate normal density given the matrix of quantiles, vector of mean, and covariance matrix.

Author(s)

Detian Deng

Maintainer: Detian Deng <ddeng@jhsph.edu>

References

Nocedal, Jorge, and S. Wright. "Numerical optimization, series in operations research and financial engineering." Springer, New York (2006).

Examples

```
# fastlm
set.seed(2)
## Generate predictor matrix
n <- 100
p <- 15
X <- cbind(1, matrix(rnorm(n * (p - 1)), n, p - 1))

## Coefficents
b <- rnorm(p)

## Response
y <- X %*% b + rnorm(n)

fit <- fastlm(X, y)
str(fit)

# dmvnorm
n <- 5
n2 <- n^2
xg <- seq(0, 1, length = n)
yg <- xg
g <- data.matrix(expand.grid(xg, yg))
D <- as.matrix(dist(g))
phi <- 5

S <- exp(-phi * D)
mu <- rep(0, n2)
set.seed(1)
x <- matrix(rnorm(n2), byrow = TRUE, ncol = n2)

mymvpdf<-dmvnorm(x=x, mu=mu, S=S, log = TRUE)
```

`dmvnorm` *Fast Multivariate Normal Density*

Description

This function evaluates the k-dimensional multivariate Normal density with mean mu and covariance S.

Usage

```
dmvnorm(x, mu, S, log = TRUE)
```

Arguments

x	a n*k matrix of points to be evaluated.
mu	a vector of means of length k for the k-dimensional Normal.
S	a k*k covariance matrix.
log	If log == TRUE, returns the logged density(by default), otherwise, returns the original density.

Details

In this function, Cholesky decomposition was used to compute the Mahalanobis distance term and the determinant of S in the density function. By using the function chol(), the positive definiteness of S is checked. If it is not, the function will return a error message.

Value

The function returns a vector of length n, containing the values of the multivariate Normal density evaluated at the n points.

Author(s)

Detian Deng

References

Nocedal, Jorge, and S. Wright. Numerical optimization, series in operations research and financial engineering. Springer, New York (2006).

Examples

```
n <- 10
n2 <- n^2
xg <- seq(0, 1, length = n)
yg <- xg
g <- data.matrix(expand.grid(xg, yg))
D <- as.matrix(dist(g))
phi <- 5

S <- exp(-phi * D)
mu <- rep(0, n2)
set.seed(1)
x <- matrix(rnorm(n2), byrow = TRUE, ncol = n2)

mymvpdf<-dmvnorm(x=x, mu=mu, S=S, log = TRUE)
```

fastlm *Fast Linear Regression*

Description

fastlm() fits a linear regression model to outcome data y and predictor data in a matrix X. The inputs should be X, a n*p matrix, y, a vector of length n, and na.rm, which indicates whether missing values in X or y should be removed. This function can always run faster than the lm.fit() function in R using the same inputs.

Usage

```
fastlm(X, y, na.rm = FALSE)
```

Arguments

X	a n*p matrix, i.e. the design matrix with n observations and (p-1) features
y	a vector of length n, i.e. the n obervations of outcome variable
na.rm	if na.rm == TRUE, then remove observations with NA, otherwise, keep them.

Details

In this function, Cholesky decomposition was applied to solve the normal equation and to compute the inverse of t(X) X, which was needed for estimating covariance matrix of the estimated regression coefficients.

Value

coefficients	a vector of the regression coefficients estimated using maximum likelihood
vcov	the p*p covariance matrix of the estimated regression coefficients.

Author(s)

Detian Deng

References

Nocedal, Jorge, and S. Wright. Numerical optimization, series in operations research and financial engineering. Springer, New York (2006).

Examples

```
set.seed(2)
## Generate predictor matrix
n <- 100
p <- 15
X <- cbind(1, matrix(rnorm(n * (p - 1)), n, p - 1))

## Coefficents
b <- rnorm(p)

## Response
```

```
y <- X %*% b + rnorm(n)

fit <- fastlm(X, y)
```

Index