

Package ‘Homework1’

November 6, 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	1
dmvnorm	2
fastlm	4

Index	6
--------------	----------

Homework1-package	<i>HW1 fastlm and dmvnorm</i>
-------------------	-------------------------------

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))

## Coefficients
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)
```

Description

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

Usage

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

Arguments

<code>x</code>	a $n \times k$ matrix of points to be evaluated.
<code>mu</code>	a vector of means of length k for the k -dimensional Normal.
<code>S</code>	a $k \times k$ covariance matrix.
<code>log</code>	If <code>log == TRUE</code> , 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 $n \times p$ 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 $n \times p$ matrix, i.e. the design matrix with n observations and $(p-1)$ features
y	a vector of length n , i.e. the n observations of outcome variable
<code>na.rm</code>	if <code>na.rm == TRUE</code> , 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

<code>coefficients</code>	a vector of the regression coefficients estimated using maximum likelihood
<code>vcov</code>	the $p \times p$ 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))

## Coefficients
b <- rnorm(p)

## Response
```

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

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

Index

dmvnorm, [2](#)

fastlm, [4](#)

Homework1 (Homework1-package), [1](#)

Homework1-package, [1](#)