

Package ‘OSMAC’

August 22, 2016

Title Optimal Subsampling for Logistic Regression
Version 0.0.0.0000
Author HaiYing Wang <haiying.wang@unh.edu>
Maintainer HaiYing Wang <haiying.wang@unh.edu>
Description This package implements the Optimal Subsampling procedure Motivated from the A-optimality Criterion.
Depends R (>= 3.3.1)
License GPL
Encoding UTF-8
LazyData true

R topics documented:

adult.test	1
adult.train	2
getMLE	3
twostep	3
Index	5

adult.test	<i>Validation set of the census income data</i>
------------	---

Description

Predict whether income exceeds 50K/yr based on census data.

Format

The validation set contains 16,281 observations.

Details

The variables are the same as the traing set (adult.train).

adult.train

Training set of the census income data

Description

Predict whether income exceeds 50K/yr based on census data.

Format

The training set adult.train contains 32,561 observations.

Details

The variables are as follows:

- >50K, <=50K: whether the income is above 50k per year
- age: continuous
- workclass: Private, Self-emp-not-inc, Self-emp-inc, Federal-gov, Local-gov, State-gov, Without-pay, Never-worked
- fnlwgt: continuous
- education: Bachelors, Some-college, 11th, HS-grad, Prof-school, Assoc-acdm, Assoc-voc, 9th, 7th-8th, 12th, Masters, 1st-4th, 10th, Doctorate, 5th-6th, Preschool
- education-num: continuous
- marital-status: Married-civ-spouse, Divorced, Never-married, Separated, Widowed, Married-spouse-absent, Married-AF-spouse
- occupation: Tech-support, Craft-repair, Other-service, Sales, Exec-managerial, Prof-specialty, Handlers-cleaners, Machine-op-inspct, Adm-clerical, Farming-fishing, Transport-moving, Priv-house-serv, Protective-serv, Armed-Forces
- relationship: Wife, Own-child, Husband, Not-in-family, Other-relative, Unmarried
- race: White, Asian-Pac-Islander, Amer-Indian-Eskimo, Other, Black
- sex: Female, Male
- capital-gain: continuous
- capital-loss: continuous
- hours-per-week: continuous
- native-country: United-States, Cambodia, England, Puerto-Rico, Canada, Germany, Outlying-US(Guam-USVI-etc), India, Japan, Greece, South, China, Cuba, Iran, Honduras, Philippines, Italy, Poland, Jamaica, Vietnam, Mexico, Portugal, Ireland, France, Dominican-Republic, Laos, Ecuador, Taiwan, Haiti, Columbia, Hungary, Guatemala, Nicaragua, Scotland, Thailand, Yugoslavia, El-Salvador, Trinidad&Tobago, Peru, Hong, Holand-Netherlands

getMLE	<i>Calculate the weighted MLE</i>
--------	-----------------------------------

Description

This function calculate the weighted MLE for the input covariate matrix x , response vector y , and weight vector w . It returns a list with three elements: `par`, the weighted MLE; `msg`, the fitting message; `iter`, the number of iterations used.

Usage

```
getMLE(x, y, w)
```

Arguments

<code>x</code>	the input covariate matrix
<code>y</code>	the input response vector
<code>w</code>	the wight vector

Examples

```
library(OSMAC)
dat <- adult.train
X <- as.matrix(dat[,c(1,3,5,12:13)])
X <- t(t(X) / apply(X, 2, sd))
X <- cbind(1, X)
Y <- as.numeric(dat[,15]) - 1
getMLE(X, Y, 1)
```

twostep	<i>The twostep algorithm</i>
---------	------------------------------

Description

This function implement the OSMAC method for the input covariate matrix @param X , response vector Y , first step sample size $r1$, the second step sample size $r2$, and the method to use. It returns a list with three elements: `par`, the weighted MLE; `amse`, the standard errors; `msg`, the fitting message; `iter`, the number of iterations used; `method`, the method used.

Usage

```
twostep(X, Y, r1, r2, method = c("mvc", "mmse", "uni"))
```

Arguments

X	the input covariate matrix
Y	the input response vector
r1	the first step sample size
r2	the second step sample size
method	the method to use

Examples

```
library(OSMAC)
dat <- adult.train
X <- as.matrix(dat[,c(1,3,5,12:13)])
X <- t(t(X) / apply(X, 2, sd))
X <- cbind(1, X)
Y <- as.numeric(dat[,15]) - 1
set.seed(0)
twostep(X, Y, 200, 800, "mmse")
```

Index

*Topic **datasets**

adult.test, [1](#)

adult.train, [2](#)

*Topic **getMLE**

getMLE, [3](#)

twostep, [3](#)

*Topic **twostep**

twostep, [3](#)

adult.test, [1](#)

adult.train, [2](#)

getMLE, [3](#)

twostep, [3](#)