Package 'RMoE'

October 23, 2019

Type Package

Title LASSO Regularized Mixture of Experts Models

Version 0.1.0

Description Regularized Mixture of Experts models using the Lasso penalty.
The main reference is the following one:
Huynh, B. T. (2019) <arXiv:1907.06994>.

URL https://github.com/fchamroukhi/HDME

BugReports https://github.com/fchamroukhi/HDME/issues

License GPL (>= 3)

Depends R (>= 2.10)

Imports methods,
stats,
graphics,
MASS,
base,
plot3D,
doParallel,
foreach

Suggests knitr,
rmarkdown

Collate RMoE-package.R
Pik.R
GEstep.R
Obj.R
SoTh.R
CoorLQk.R
GPMstep.R
GSMstep.R
GLOG.R
GBIC.R
GWrite.R
Fs.R
CoorGateP1.R
CoorGateP.R
GRMoE-class.R
GaussRMoE.R

LPi.R
LEstep.R
LFk.R
LCoorExpP1.R
LCoorExpP.R
LPMstep.R
LLOG.R
LBIC.R
LWrite.R
LRMoE-class.R
LogisticRMoE.R
PInitial.R
PEstep.R
PQk.R
PCoorExpP.R
PPMstep.R
PLOG.R
PBIC.R
PWrite.R
PRMoE-class.R
PoissonRMoE.R
data-cleveland.R
data-gaussian.R
data-housing.R
data-ionosphere.R
data-logistic.R
data-musk1.R
data-poisson.R
data-residential.R

VignetteBuilder knitr

Encoding UTF-8

LazyData true

Roxygen list(markdown = TRUE)

RoxygenNote 6.1.1

R **topics documented:**

RMoE-package   *RMoE: LASSO Regularized Mixture of Experts Models*

Description

RMoE is a package containing regularized Mixture of Experts models using the Lasso penalty.

RMoE contains the following Regularized Mixture-of-Experts models:

- GaussianRMoE: Gaussian Regularixed Mixture of Experts;

- LogisticRMoE: Logistic Regularixed Mixture of Experts;

- PoissonRMoE: Poisson Regularixed Mixture of Experts.

To learn more about RMoE, start with the vignettes: browseVignettes(package = "RMoE")

Author(s)

Maintainer: Bao-Tuyen Huynh <baotuyen.dlu@gmail.com>

Authors:

- Faicel Chamroukhi <faicel.chamroukhi@unicaen.fr> (0000-0002-5894-3103)

References

Huynh B. T., Chamroukhi F. 2019. *Estimation and Feature Selection in Mixtures of Generalized Linear Experts Models.* https://arxiv.org/abs/1907.06994.

See Also

Useful links:

- https://github.com/fchamroukhi/HDME

- Report bugs at https://github.com/fchamroukhi/HDME/issues

---

cleveland   *The Cleveland data set*

Description

The Cleveland data set described on the website UC Irvine Machine Learning Repository.

Usage

```
cleveland
```

Format

A data frame with 297 rows and 15 columns.

Source

https://archive.ics.uci.edu/ml/datasets/heart+Disease

---

gaussian   *A simulated gaussian data set*

Description

A simulated gaussian data set. True parameters for this data set are given by:

- Exp.1 = (0,0,1.5,0,0,0,1);
- Exp.2 = (0,1,-1.5,0,0,2,0);
- Gate = (1,2,0,0,-1,0,0);
- Sigma = 1.

Usage

```
gaussian
```

Format

A data frame with 300 rows and 8 columns.

GaussRMoE   *Penalized MLE for the regularized Mixture of Experts.*

Description

This function provides a penalized MLE for the regularized Mixture of Experts (MoE) model corresponding with the penalty parameters Lambda, Gamma.

Usage

```
GaussRMoE(Xm, Ym, K, Lambda, Gamma, option = FALSE, verbose = FALSE)
```

Arguments

Xm	Matrix of explanatory variables. Each feature should be standardized to have mean 0 and variance 1. One must add the column vector (1,1,...,1) for the intercept variable.
Ym	Vector of the response variable. For the Gaussian case Y should be standardized. For multi-logistic model Y is numbered from 1 to R (R is the number of labels of Y).
K	Number of experts (K > 1).
Lambda	Penalty value for the experts.
Gamma	Penalty value for the gating network.
option	Optional. option = TRUE: using proximal Newton-type method; option = FALSE: using proximal Newton method.
verbose	Optional. A logical value indicating whether or not values of the log-likelihood should be printed during EM iterations.

Value

GaussRMoE returns an object of class GRMoE.

See Also

GRMoE

GRMoE-class   *A Reference Class which contains parameters of a GRMoE model.*

Description

GRMoE contains all the parameters of a Gaussian Regularized Mixture-of-Experts.

Fields

X   The matrix data for the input.

Y   Vector of the response variable.

d   Numeric. Number of explanatory variables (including the intercept variable).

n   Numeric. Length of the response/output vector Y.

K   Number of expert classes.

Lambda   Penalty value for the expert part.

Gamma   Penalty value for the gating network.

wk   Parameters of the gating network. Matrix of dimension $(K-1, d)$, with d the number of explanatory variables (including the intercept).

betak   Regressions coefficients for each expert. Matrix of dimension (d, K).

sigma   Numeric. The standard deviation.

loglik   Numeric. Observed-data log-likelihood of the GRMoE model.

storedloglik   Numeric vector. Stored values of the log-likelihood at each EM iteration.

BIC   Numeric. Value of BIC (Bayesian Information Criterion).

zerocoeff   Matrix. Proportion of zero coefficients obtained during each iteration of the EM. First column gives the number of zero coefficients for wk and the second column for betak.

Cluster   Numeric. Clustering label for each observation.

Methods

plot(what = c("loglik", "zerocoefficients"))   Plot method.

    what   The type of graph requested:
- "loglik" = Value of the log-likelihood for each iteration.
- "zerocoefficients" = Proportion of zero coefficients for each iteration.

    By default, all the above graphs are produced.

---

housing	*The Housing data set*

Description

The Housing data set described on the website UC Irvine Machine Learning Repository. The value MEDV/sd(MEDV) is considered as the predictor variable.

Usage

housing

Format

A data frame with 506 rows and 15 columns.

Source

https://archive.ics.uci.edu/ml/machine-learning-databases/housing

---

`ionosphere`	*The Ionosphere data set*

Description

The Ionosphere data set described on the website UC Irvine Machine Learning Repository.

Usage

```
ionosphere
```

Format

A data frame with 351 rows and 35 columns.

Source

https://archive.ics.uci.edu/ml/datasets/ionosphere

`logistic`	*A simulated logistic data set*

Description

A simulated logistic data set. True parameters for this data set are given by:

- Exp.1 = (0,-1,2,0,0,1.5,0);
- Exp.2 = (0,1,0,0,-2,0,0);
- Gate = (1,0,0,1,0,0,-1.5).

Usage

```
logistic
```

Format

A data frame with 300 rows and 8 columns.

LogisticRMoE   *Penalized MLE for the logistic regularized Mixture of Experts.*

Description

This function provides a penalized MLE for the logistic regularized Mixture of Experts (MoE) model corresponding with the penalty parameters Lambda, Gamma.

Usage

```
LogisticRMoE(Xmat, Ymat, K, Lambda, Gamma, option = FALSE,
  verbose = FALSE)
```

Arguments

Xmat	Matrix of explanatory variables. Each feature should be standardized to have mean 0 and variance 1. One must add the column vector (1,1,...,1) for the intercept variable.
Ymat	Vector of the response variable. For the Gaussian case Y should be standardized. For multi-logistic model Y is numbered from 1 to R (R is the number of labels of Y).
K	Number of experts (K > 1).
Lambda	Penalty value for the experts.
Gamma	Penalty value for the gating network.
option	Optional. `option = TRUE`: using proximal Newton-type method; `option = FALSE`: using proximal Newton method.
verbose	Optional. A logical value indicating whether or not values of the log-likelihood should be printed during EM iterations.

Value

LogisticRMoE returns an object of class [LRMoE](#).

See Also

[LRMoE](#)

LRMoE-class   *A Reference Class which contains parameters of a LRMoE model.*

Description

LRMoE contains all the parameters of a Logistic Regularized Mixture-of-Experts.

Fields

X The matrix data for the input.

Y Vector of the response variable.

d Numeric. Number of explanatory variables (including the intercept variable).

n Numeric. Length of the response/output vector Y.

R Numeric. Maximum value of Y.

K Number of expert classes.

Lambda Penalty value for the expert part.

Gamma Penalty value for the gating network.

wk Parameters of the gating network. Matrix of dimension $(K - 1, d)$, with d the number of explanatory variables (including the intercept).

eta Values of the regression coefficients for each level r = 1,...,R. Array of dimension $(K, R-1, d)$.

loglik Numeric. Observed-data log-likelihood of the LRMoE model.

storedloglik Numeric vector. Stored values of the log-likelihood at each EM iteration.

BIC Numeric. Value of BIC (Bayesian Information Criterion).

zerocoeff Matrix. Proportion of zero coefficients obtained during each iteration of the EM. First column gives the number of zero coefficients for wk and the second column for eta.

Cluster Numeric vector. Clustering label for each observation.

Methods

plot(what = c("loglik", "zerocoefficients")) Plot method.

what The type of graph requested:
- "loglik" = Value of the log-likelihood for each iteration.
- "zerocoefficients" = Proportion of zero coefficients for each iteration.

By default, all the above graphs are produced.

---

musk1 *The Musk-1 data set*

Description

The Musk-1 data set described on the website UC Irvine Machine Learning Repository.

Usage

musk1

Format

A data frame with 476 rows and 168 columns.

Source

https://archive.ics.uci.edu/ml/datasets/Musk+(Version+1)

---

poisson   *A simulated poisson data set*

Description

A simulated poisson data set. True parameters for this data set are given by:

- Exp.1 = (0,1,0,-2,0,1.5,0);
- Exp.2 = (0,0,2,0,-1,0,0);
- Gate = (1,0,0,1,0,-1.5,0).

Usage

```
poisson
```

Format

A data frame with 300 rows and 8 columns.

PoissonRMoE   *Penalized MLE for the Poisson regularized Mixture of Experts.*

Description

This function provides a penalized MLE for the Poisson regularized Mixture of Experts (MoE) model corresponding with the penalty parameters Lambda, Gamma.

Usage

```
PoissonRMoE(Xmat, Ymat, K, Lambda, Gamma, option = FALSE,
  verbose = TRUE)
```

Arguments

Xmat	Matrix of explanatory variables. Each feature should be standardized to have mean 0 and variance 1. One must add the column vector (1,1,...,1) for the intercept variable.
Ymat	Vector of the response variable. For the Gaussian case Y should be standardized. For multi-logistic model Y is numbered from 1 to R (R is the number of labels of Y).
K	Number of experts (K > 1).
Lambda	Penalty value for the experts.
Gamma	Penalty value for the gating network.
option	Optional. option = TRUE: using proximal Newton-type method; option = FALSE: using proximal Newton method.
verbose	Optional. A logical value indicating whether or not values of the log-likelihood should be printed during EM iterations.

Value

PoissonRMoE returns an object of class [PRMoE](#).

See Also

[PRMoE](#)

PRMoE-class   *A Reference Class which contains parameters of a PRMoE model.*

Description

PRMoE contains all the parameters of a Poisson Regularized Mixture-of-Experts.

Fields

X The matrix data for the input.

Y Vector of the response variable.

d Numeric. Number of explanatory variables (including the intercept variable).

n Numeric. Length of the response/output vector Y.

K Number of expert classes.

Lambda Penalty value for the expert part.

Gamma Penalty value for the gating network.

wk Parameters of the gating network. Matrix of dimension $(K - 1, d)$, with d the number of explanatory variables (including the intercept).

betak Regressions coefficients for each expert. Matrix of dimension (d, K).

loglik Numeric. Observed-data log-likelihood of the PRMoE model.

storedloglik Numeric vector. Stored values of the log-likelihood at each EM iteration.

zerocoeff Matrix. Proportion of zero coefficients obtained during each iteration of the EM. First column gives the number of zero coefficients for wk and the second column for betak.

BIC Numeric. Value of BIC (Bayesian Information Criterion).

Cluster Numeric vector. Clustering label for each observation.

Methods

plot(what = c("loglik", "zerocoefficients")) Plot method.

what The type of graph requested:
   • "loglik" = Value of the log-likelihood for each iteration.
   • "zerocoefficients" = Proportion of zero coefficients for each iteration.

By default, all the above graphs are produced.

`residential`   *The Residential Building data set*

Description

The Residential Building data set described on the website UC Irvine Machine Learning Repository. The V-9 variable is used as the predictor variable.

Usage

```
residential
```

Format

A data frame with 372 rows and 109 columns.

Source

http://archive.ics.uci.edu/ml/datasets/Residential+Building+Data+Set

Index