

Package ‘rFTRLProximal’

December 27, 2016

Type Package

Title FTRL-Proximal Algorithm

Version 1.0.0

Author Author: Yachen Yan [aut, cre]

Description An efficient C++ based implementation of "Follow The (Proximally) Regularized Leader" online learning algorithm.

This algorithm was proposed in McMahan et al. (2013) <DOI:10.1145/2487575.2488200>.

URL <http://github.com/yanyachen/rFTRLProximal>

BugReports <http://github.com/yanyachen/rFTRLProximal/issues>

Depends R (>= 2.10)

Imports stats, utils, data.table (>= 1.9.6), Matrix, FeatureHashing, magrittr, foreach, Rcpp, RcppProgress

Suggests rBayesianOptimization, MLmetrics

LinkingTo Rcpp, RcppArmadillo, RcppProgress

License GPL-2

LazyData TRUE

RoxygenNote 5.0.1

NeedsCompilation yes

Maintainer Yachen Yan <yanyachen21@gmail.com>

Repository CRAN

Date/Publication 2016-12-27 15:28:19

R topics documented:

FTRLProx_cv	2
FTRLProx_Hashing	3
FTRLProx_predict	4
FTRLProx_predict_spMatrix	5
FTRLProx_train	5

FTRLProx_train_spMatrix	7
FTRLProx_validate	7
FTRLProx_validate_spMatrix	9
rFTRLProximal	9
slice	10

Index	11
--------------	-----------

FTRLProx_cv	<i>FTRL-Proximal Linear Model Cross Validation</i>
-------------	--

Description

An advanced interface for FTRL-Proximal online learning model cross validation.

Usage

```
FTRLProx_cv(x, y, family = c("gaussian", "binomial", "poisson"),
  params = list(alpha = 0.1, beta = 1, l1 = 1, l2 = 1), epoch = 1, folds,
  eval)
```

Arguments

x	a transposed dgCMatrix.
y	a vector containing labels.
family	link function to be used in the model. "gaussian", "binomial" and "poisson" are available.
params	a list of parameters of FTRL-Proximal Algorithm. <ul style="list-style-type: none"> • alpha alpha in the per-coordinate learning rate • beta beta in the per-coordinate learning rate • l1 L1 regularization parameter • l2 L2 regularization parameter
epoch	The number of iterations over training data to train the model.
folds	list provides a possibility of using a list of pre-defined CV folds (each element must be a vector of fold's indices).
eval	a evaluation metrics computing function, the first argument should be prediction, the second argument should be label.

Value

a list with the following elements is returned:

- dt a data.table with each mean and standard deviation stat for training set and test set
- pred a numerical vector with predictions for each CV-fold for the model having been trained on the data in all other folds.

References

H. B. McMahan, G. Holt, D. Sculley, et al. "Ad click prediction: a view from the trenches". In: *The 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2013, Chicago, IL, USA, August 11-14, 2013*. Ed. by I. S.Dhillon, Y. Koren, R. Ghani, T. E. Senator, P. Bradley, R. Parekh, J. He, R. L. Grossman and R. Uthurusamy. ACM, 2013, pp. 1222-1230. DOI: 10.1145/2487575.2488200. <URL: <http://doi.acm.org/10.1145/2487575.2488200>>.

Examples

```
library(FeatureHashing)
library(data.table)
library(rBayesianOptimization)
library(MLmetrics)
data(ipinyou)
m.train <- FTRLProx_Hashing(~ 0 + ., ipinyou.train[,-"IsClick", with = FALSE],
                           hash.size = 2^13, signed.hash = FALSE, verbose = TRUE)
ftrl_model_cv <- FTRLProx_cv(x = m.train, y = as.numeric(ipinyou.train$IsClick),
                            family = "binomial",
                            params = list(alpha = 0.01, beta = 0.1, l1 = 1.0, l2 = 1.0),
                            epoch = 10,
                            folds = KFold(as.numeric(ipinyou.train$IsClick), nfolds = 5),
                            eval = AUC)
```

FTRLProx_Hashing

Feature Hashing for FTRL-Proximal Model

Description

Feature Hashing for FTRL-Proximal Algorithm. A wrapper of `hashed.model.matrix` function in the `FeatureHashing` package. Please always use this function to generate sparse matrix for training and prediction.

Usage

```
FTRLProx_Hashing(formula = ~., data, hash.size, signed.hash, verbose = TRUE)
```

Arguments

<code>formula</code>	formula or a character vector of column names (will be expanded to a formula)
<code>data</code>	a <code>data.frame</code> or <code>data.table</code> . The original data.
<code>hash.size</code>	positive integer. The hash size of feature hashing.
<code>signed.hash</code>	logical value. Indicating if the hashed value is multiplied by random sign. This will reduce the impact of collision. Disable it will enhance the speed.
<code>verbose</code>	logical value. Indicating if the progress bar is displayed or not.

Value

an object of class "dgCMatrix"

Examples

```
library(data.table)
library(FeatureHashing)
data(ipinyou)
m.train <- FTRLProx_Hashing(~ 0 + ., ipinyou.train[, ~"IsClick", with = FALSE],
                           hash.size = 2^13, signed.hash = FALSE, verbose = TRUE)
m.test <- FTRLProx_Hashing(~ 0 + ., ipinyou.test[, ~"IsClick", with = FALSE],
                          hash.size = 2^13, signed.hash = FALSE, verbose = TRUE)
```

FTRLProx_predict

FTRL-Proximal Linear Model Predicting

Description

An advanced interface for FTRL-Proximal online learning model predicting.

Usage

```
FTRLProx_predict(model, newx)
```

Arguments

model	a FTRL-Proximal linear model object.
newx	a transposed dgCMatrix.

Value

an vector of linear model predicted values

Examples

```
library(data.table)
library(FeatureHashing)
library(MLmetrics)
data(ipinyou)
m.train <- FTRLProx_Hashing(~ 0 + ., ipinyou.train[, ~"IsClick", with = FALSE],
                           hash.size = 2^13, signed.hash = FALSE, verbose = TRUE)
m.test <- FTRLProx_Hashing(~ 0 + ., ipinyou.test[, ~"IsClick", with = FALSE],
                          hash.size = 2^13, signed.hash = FALSE, verbose = TRUE)
ftrl_model <- FTRLProx_train(m.train, y = as.numeric(ipinyou.train$IsClick), family = "binomial",
                           params = list(alpha = 0.01, beta = 0.1, l1 = 1.0, l2 = 1.0),
                           epoch = 10, verbose = TRUE)
pred_ftrl <- FTRLProx_predict(ftrl_model, newx = m.test)
AUC(pred_ftrl, as.numeric(ipinyou.test$IsClick))
```

 FTRLProx_predict_spMatrix

FTRL-Proximal Linear Model Predicting Function

Description

FTRLProx_predict_spMatrix predicts values based on linear model weights. This function is an C++ implementation. This function is used internally and is not intended for end-user direct usage.

Usage

```
FTRLProx_predict_spMatrix(x, w, family)
```

Arguments

x	a transposed dgCMatrix object.
w	an vector of linear model weights.
family	link function to be used in the model. "gaussian", "binomial" and "poisson" are available.

Value

an vector of linear model predicted values

 FTRLProx_train

FTRL-Proximal Linear Model Training

Description

An advanced interface for training FTRL-Proximal online learning model.

Usage

```
FTRLProx_train(x, y, family = c("gaussian", "binomial", "poisson"),
  params = list(alpha = 0.1, beta = 1, l1 = 1, l2 = 1), epoch = 1,
  bagging_seeds = NULL, verbose = TRUE)
```

Arguments

x	a transposed dgCMatrix.
y	a vector containing labels.
family	link function to be used in the model. "gaussian", "binomial" and "poisson" are available.
params	a list of parameters of FTRL-Proximal Algorithm.

	<ul style="list-style-type: none"> • alpha alpha in the per-coordinate learning rate • beta beta in the per-coordinate learning rate • l1 L1 regularization parameter • l2 L2 regularization parameter
epoch	The number of iterations over training data to train the model.
bagging_seeds	a vector containing random seeds for shuffling data. If provided, use parallel foreach to fit each model. Must register parallel before hand, such as doParallel or others.
verbose	logical value. Indicating if the progress bar is displayed or not.

Value

a FTRL-Proximal linear model object

References

H. B. McMahan, G. Holt, D. Sculley, et al. "Ad click prediction: a view from the trenches". In: *The 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2013, Chicago, IL, USA, August 11-14, 2013.* Ed. by I. S.Dhillon, Y. Koren, R. Ghani, T. E. Senator, P. Bradley, R. Parekh, J. He, R. L. Grossman and R. Uthurusamy. ACM, 2013, pp. 1222-1230. DOI: 10.1145/2487575.2488200. <URL: <http://doi.acm.org/10.1145/2487575.2488200>>.

Examples

```
library(data.table)
library(FeatureHashing)
library(MLmetrics)
data(ipinyou)
m.train <- FTRLProx_Hashing(~ 0 + ., ipinyou.train[, -"IsClick", with = FALSE],
                           hash.size = 2^13, signed.hash = FALSE, verbose = TRUE)
m.test <- FTRLProx_Hashing(~ 0 + ., ipinyou.test[, -"IsClick", with = FALSE],
                          hash.size = 2^13, signed.hash = FALSE, verbose = TRUE)
ftrl_model <- FTRLProx_train(m.train, y = as.numeric(ipinyou.train$IsClick),
                           family = "binomial",
                           params = list(alpha = 0.01, beta = 0.1, l1 = 1.0, l2 = 1.0),
                           epoch = 10, verbose = TRUE)
ftrl_model_bagging <- FTRLProx_train(m.train, y = as.numeric(ipinyou.train$IsClick),
                                    family = "binomial",
                                    params = list(alpha = 0.01, beta = 0.1, l1 = 1.0, l2 = 1.0),
                                    epoch = 10, bagging_seeds = 1:10, verbose = FALSE)
pred_ftrl <- FTRLProx_predict(ftrl_model, newx = m.test)
pred_ftrl_bagging <- FTRLProx_predict(ftrl_model_bagging, newx = m.test)
AUC(pred_ftrl, as.numeric(ipinyou.test$IsClick))
AUC(pred_ftrl_bagging, as.numeric(ipinyou.test$IsClick))
```

 FTRLProx_train_spMatrix

FTRL-Proximal Linear Model Fitting Function

Description

FTRLProx_train_spMatrix estimates the weights of linear model using FTRL-Proximal Algorithm. This function is an C++ implementation. This function is used internally and is not intended for end-user direct usage.

Usage

```
FTRLProx_train_spMatrix(x, y, family, params, epoch, verbose)
```

Arguments

x	a transposed dgMatrix.
y	a vector containing labels.
family	link function to be used in the model. "gaussian", "binomial" and "poisson" are available.
params	a list of parameters of FTRL-Proximal Algorithm. <ul style="list-style-type: none"> • alpha alpha in the per-coordinate learning rate • beta beta in the per-coordinate learning rate • l1 L1 regularization parameter • l2 L2 regularization parameter
epoch	The number of iterations over training data to train the model.
verbose	logical value. Indicating if the progress bar is displayed or not.

Value

an vector of linear model weights

 FTRLProx_validate

FTRL-Proximal Linear Model Validation

Description

An advanced interface for FTRL-Proximal online learning model validation.

Usage

```
FTRLProx_validate(x, y, family = c("gaussian", "binomial", "poisson"),
  params = list(alpha = 0.1, beta = 1, l1 = 1, l2 = 1), epoch = 1, val_x,
  val_y, eval, verbose = TRUE)
```

FTRLProx_validate_spMatrix

FTRL-Proximal Linear Model Validation Function

Description

FTRLProx_validate_spMatrix validates the performance of FTRL-Proximal online learning model. This function is an C++ implementation. This function is used internally and is not intended for end-user direct usage.

Usage

```
FTRLProx_validate_spMatrix(x, y, family, params, epoch, val_x, val_y, eval,
    verbose)
```

Arguments

x	a transposed dgCMatrix.
y	a vector containing labels.
family	link function to be used in the model. "gaussian", "binomial" and "poisson" are available.
params	a list of parameters of FTRL-Proximal Algorithm. <ul style="list-style-type: none"> • alpha alpha in the per-coordinate learning rate • beta beta in the per-coordinate learning rate • l1 L1 regularization parameter • l2 L2 regularization parameter
epoch	The number of iterations over training data to train the model.
val_x	a transposed dgCMatrix for validation.
val_y	a vector containing labels for validation.
eval	a evaluation metrics computing function, the first argument should be prediction, the second argument should be label.
verbose	logical value. Indicating if the validation result for each epoch is displayed or not.

Value

a FTRL-Proximal linear model object

rFTRLProximal

rFTRLProximal: FTRL-Proximal Algorithm

Description

An efficient C++ based implementation of "Follow The (Proximally) Regularized Leader" online learning algorithm. This algorithm was proposed in McMahan et al. (2013) <DOI:10.1145/2487575.2488200>.

slice	<i>Subsetting dgCMatrx</i>
-------	----------------------------

Description

Returns subsets of a `dgCMatrx`.

Usage

```
slice(x, i)
```

Arguments

<code>x</code>	a transposed <code>dgCMatrx</code> .
<code>i</code>	logical expression indicating elements or rows to keep.

Value

A transposed `dgCMatrx` containing the subset of rows that are selected.

Index

FTRLProx_cv, [2](#)
FTRLProx_Hashing, [3](#)
FTRLProx_predict, [4](#)
FTRLProx_predict_spMatrix, [5](#)
FTRLProx_train, [5](#)
FTRLProx_train_spMatrix, [7](#)
FTRLProx_validate, [7](#)
FTRLProx_validate_spMatrix, [9](#)

rFTRLProximal, [9](#)
rFTRLProximal-package (rFTRLProximal), [9](#)

slice, [10](#)