

Package ‘bagRboostR’

March 23, 2013

Type Package

Title Ensemble bagging and boosting classifiers

Version 0.0.1

Author Shannon Rush <shannonmrush@gmail.com>

Maintainer Shannon Rush <shannonmrush@gmail.com>

Description bagRboostR is a set of ensemble classifiers for multinomial classification. The bagging function is the implementation of Breiman’s ensemble as described by Opitz & Maclin (1999). The boosting function is the implementation of Stagewise Additive Modeling using a Multi-class Exponential loss function (SAMME) created by Zhu et al (2006). Both bagging and SAMME implementations use randomForest as the weak classifier and expect a character outcome variable. Each ensemble classifier returns a character vector of predictions for the test set.

Imports randomForest

Suggests testthat

License MIT

Collate ‘bagging.R’ ‘helpers.R’ ‘samme.R’

R topics documented:

bagging	2
samme	3
Index	4

bagging*Ensemble bagging classifier for multinomial classification*

Description

Makes multiple predictions on test set using resampled variations of data set. Uses randomForest as the weak classifier and expects a multinomial character outcome variable. Returns a final test set prediction with each forest having an equal vote. Implementation of Breiman's ensemble bagging classifier as described by Obitz & Maclin (1999).

Usage

```
bagging(formula, data, test, m = 5, ntree = 500, mtry = NULL, trace = T)
```

Arguments

formula	A formula expression for the classifier model. Expects the form outcome ~ predictors with a single outcome variable. The outcome variable is expected to be of type character.
data	A data frame consisting of the training data set. Must include all variables described in the formula including the outcome variable.
test	A data frame consisting of the test set to be predicted. Must include all predictor variables. May include outcome variable.
m	Number of classifiers to vote on final prediction. Defaults to 5.
ntree	Number of trees for each randomForest to grow. Defaults to 500.
mtry	Number of variables randomForest randomly samples as candidates at each split. Defaults to the square root of the number of variables in data.
trace	Setting for randomForest do.trace. Defaults to TRUE.

Value

Character vector consisting of a final prediction for each test set sample.

Author(s)

Shannon Rush <shannonmrush@gmail.com>

References

Breiman, L. (1996) "Bagging Predictors", Machine Learning, 26, 123-140.

Opitz, D. and Maclin, R. (1999) "Popular Ensemble Methods: An Empirical Study", Journal of Artificial Intelligence Research, 11, 169-198. <http://www.d.umn.edu/~rmaclin/publications/opitz-jair99.pdf>

See Also

[randomForest](#)

samme*Ensemble boosting classifier for multinomial classification*

Description

Makes multiple predictions on the test set using samples chosen from the data set based on past performance. Uses randomForest as the weak classifier and expects a multinomial character outcome variable. Makes a final prediction on the test set based on votes by the classifiers weighted by performance. Implementation of Zhu et al.'s Stagewise Additive Modeling using a Multi-class Exponential loss function (2006).

Usage

```
samme(formula, data, test, m = 5, trace = T, ntree = 500, mtry = NULL)
```

Arguments

formula	A formula expression for the classifier model. Expects the form outcome ~ predictors with a single outcome variable. The outcome variable is expected to be of type character.
data	A data frame consisting of the training data set. Must include all variables described in the formula including the outcome variable.
test	A data frame consisting of the test set to be predicted. Must include all predictor variables. May include outcome variable.
m	Number of classifiers to vote on final prediction. Defaults to 5.
trace	Setting for randomForest do.trace. Defaults to TRUE.
ntree	Number of trees for each randomForest to grow. Defaults to 500.
mtry	Number of variables randomForest randomly samples as candidates at each split. Defaults to the square root of the number of variables in data.

Value

Character vector consisting of a final prediction for each test set sample.

Author(s)

Shannon Rush <shannonmrush@gmail.com>

References

Zhu, J., Rosset, S., Zou, H. and Hastie, T. (2005). Multiclass AdaBoost. Technical report, Stanford Univ. <http://www.stanford.edu/~hastie/Papers/samme.pdf>

See Also

[randomForest](#)

Index

bagging, [2](#)

randomForest, [2](#), [3](#)

samme, [3](#)