Package 'RFinfer'

November 18, 2015

Type Package

Title Inference for Random Forests

Version 0.1

Date 2015-08-17

Author Cole Brokamp

Maintainer <cole.brokamp@gmail.com>

Description A set of add on tools for the randomForest package

License

LazyData TRUE

R topics documented:

`rfPredVar` *rfPredVar*

Description

Generate predictions and prediction variances from a random forest based on the infinitesimal jackknife.

Usage

```
rfPredVar(random.forest, rf.data, pred.data = rf.data, CI = FALSE,
  tree.type = c("rf", "ci"), prog.bar = FALSE)
```

Arguments

rf.data	The data used to train rf
pred.data	The data used to predict with the forest; defaults to rf.data if not given
CI	Should 95% confidence intervals based on the CLT be returned along with predictions and prediction variances?
tree.type	either 'ci' for conditional inference tree or 'rf' for traditional CART tree
prog.bar	should progress bar be shown? (only applicable when tree.type='ci'
rf	A random forest trained with keep.inbag=TRUE. See details for more information.

Details

The original version of randomForest with the keep.inbag=TRUE only keeps track if each training data point was or was not included in each resample. Install a tweaked version of randomForest that amends this to include the number of times each training data point was included in each resample by running the following code in R: devtools::install_github('cole-brokamp/randomForest')

Note: This function does not use the default predict method for forests produced by cforest. The predictions here are the direct averages of all tree predictions, instead of using the observation weights. Therefore, predictions from this function will likely differ from predict.cforest when using subsampling.

This function currently only works with regression forests – not classification forests.

Value

A data frame with the predictions and prediction variances (and optionally 95% confidence interval)

Examples

```
library(randomForest)
data(airquality)
d <- na.omit(airquality)
rf <- randomForest(Ozone ~ .,data=d,keep.inbag=T,sampsize=30,replace=FALSE,ntree=500)
rfPredVar(rf,rf.data=d,CI=TRUE,tree.type='rf')
```

Index