



R news and tutorials contributed by (573) R bloggers

- [Home](#)
- [About](#)
- [RSS](#)
- [add your blog!](#)
- [R jobs](#) ♦♦♦
- [Contact us](#)

Welcome!

Follow @rbloggers 23.5K

Here you will find daily **news and tutorials about R**, contributed by over 573 bloggers.

There are many ways to **follow us** -

[By e-mail:](#)

19348 readers

BY FEEDBURNER

[On Facebook:](#)

R blogg...
27k likes

Be the first of your friends to like this

If you are an R blogger yourself you are invited to [add your own R content feed to this site](#) (Non-English R bloggers should add themselves- [here](#))

[Jobs for R-users](#)

- [Senior Research Scientist \(University of Texas Health Science Center at Houston\)](#)
- [Data Scientist – Machine Learning \(in booking.com @ Amsterdam\)](#)
- [Data Scientist – Analytics \(in booking.com @ Amsterdam\)](#)
- [Data Scientists \(@ New York \)](#)
- [Principal Analysts x2, Senior Analyst, Analyst](#)

[\(@ Wellington \)](#)

Popular Searches

- [heatmap](#)
- [Web Scraping](#)
- [maps](#)
- [hadoop](#)
- [shiny](#)
- [boxplot](#)
- [twitteR](#)
- [alt=](#)
- [Time series](#)
- [animation](#)
- [trading](#)
- [ggplot2](#)
- [latex](#)
- [ggplot](#)
- [finance](#)
- [Excel](#)
- [RStudio](#)
- [pca](#)
- [how to import image file to R](#)
- [quantmod](#)
- [googlevis](#)
- [market research](#)
- [tutorial](#)
- [eclipse](#)
- [rattle](#)
- [undefined](#)
- [knitr](#)
- [coplot](#)
- [GIS](#)
- [remdr](#)

Recent Posts

- [Replicating NatGeo's "Proper" Earthquake Map in R](#)
- [SparkR quick start that works](#)
- [ReppArmadillo 0.6.100.0.0](#)
- [Imputing missing data with R: MICE package](#)
- [Predicting Titanic deaths on Kaggle VII: More Stan](#)
- [RuleFit: When disassembled trees meet Lasso](#)
- [Recruiting Analysts for dynamic cutting edge public sector team](#)
- [Delta Method Confidence Bands for Gaussian Density](#)
- [A Simpler Explanation of Differential Privacy](#)
- [Hadley](#)

[Wickham's "Ask Me Anything" on Reddit](#)

- [R Open Source and GSoC](#)
- [London Vs Boston](#)
- [Jug: Easily Create R APIs](#)
- [R User Groups Highlight R Creativity](#)
- [ChIPseq data mining with ChIPseeker](#)

Other sites

- [SAS blogs](#)
- [Statistics of Israel](#)
- [Jobs for R-users](#)

caretEnsemble Classification example

March 16, 2013

By [Zachary Mayer](#)

 Like  Share  Tweet  6

(This article was first published on [Modern Toolmaking](#), and kindly contributed to [R-bloggers](#).)

Here's a quick demo of how to fit a binary classification model with caretEnsemble. Please note that I haven't spent as much time debugging caretEnsemble for classification models, so there's probably more bugs than my last post. Also note that multi class models are not yet supported.

```

1 #Setup
2 rm(list = ls(all = TRUE))
3 gc(reset=TRUE)
4 set.seed(1234) #From random.org
5
6 #Libraries
7 library(caret)
8 library(devtools)
9 install_github('caretEnsemble', 'zachmayer') #Install zach's c
10 library(caretEnsemble)
11
12 #Data
13 library(mlbench)
14 dat <- mlbench.xor(500, 2)
15 X <- data.frame(dat$x)
16 Y <- factor(ifelse(dat$classes=='1', 'Yes', 'No'))
17
18 #Split train/test
19 train <- runif(nrow(X)) <= .66
20
21 #Setup CV Folds
22 #returnData=FALSE saves some space
23 folds=5
24 repeats=1
25 myControl <- trainControl(method='cv', number=folds, repeats=r
26                             returnResamp='none', classProbs=TRUE
27                             returnData=FALSE, savePredictions=TR
28                             verboseIter=TRUE, allowParallel=TRUE
29                             summaryFunction=twoClassSummary,
30                             index=createMultiFolds(Y[train], k=f
31 PP <- c('center', 'scale')

```

```

32
33 #Train some models
34 model1 <- train(X[train,], Y[train], method='gbm', trControl=m
35               tuneGrid=expand.grid(.n.trees=500, .interactio
36 model2 <- train(X[train,], Y[train], method='blackboost', trCo
37 model3 <- train(X[train,], Y[train], method='parRF', trControl
38 model4 <- train(X[train,], Y[train], method='mlpWeightDecay',
39 model5 <- train(X[train,], Y[train], method='knn', trControl=m
40 model6 <- train(X[train,], Y[train], method='earth', trControl
41 model7 <- train(X[train,], Y[train], method='glm', trControl=m
42 model8 <- train(X[train,], Y[train], method='svmRadial', trCon
43 model9 <- train(X[train,], Y[train], method='gam', trControl=m
44 model10 <- train(X[train,], Y[train], method='glmnet', trContr
45
46 #Make a list of all the models
47 all.models <- list(model1, model2, model3, model4, model5, mod
48 names(all.models) <- sapply(all.models, function(x) x$method)
49 sort(sapply(all.models, function(x) min(x$results$ROC)))
50
51 #Make a greedy ensemble - currently can only use RMSE
52 greedy <- caretEnsemble(all.models, iter=1000L)
53 sort(greedy$weights, decreasing=TRUE)
54 greedy$error
55
56 #Make a linear regression ensemble
57 linear <- caretStack(all.models, method='glm', trControl=train
58 linear$error
59
60 #Predict for test set:
61 library(caTools)
62 preds <- data.frame(sapply(all.models, function(x){predict(x,
63 preds$ENS_greedy <- predict(greedy, newdata=X[!train,])
64 preds$ENS_linear <- predict(linear, newdata=X[!train,], type='
65 sort(data.frame(colAUC(preds, Y[!train])))

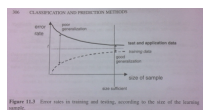
```

Demo2.R hosted with ❤ by GitHub [view raw](#)

Right now, this code fails for me if I try a model like a nnet or an SVM for stacking, so there's clearly bugs to fix.

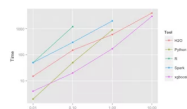
The greedy model relies 100% on the gbm, which makes sense as the gbm has an AUC of 1 on the training set. The linear model uses all of the models, and achieves an AUC of .5. This is a little weird, as the gbm, rf, SVN, and knn all achieve an AUC of close to 1.0 on the training set, and I would have expected the linear model to focus on these predictions. I'm not sure if this is a bug, or a failure of my stacking model.

Related



[Model Exploration using K-sample Plot in Big Data](#)

[Similar post](#)



[Benchmarking Random Forest Implementations In "R bloggers"](#)

[New package for ensembling R models](#)
[Similar post](#)



Like



Tweet

Share

To leave a comment for the author, please follow the link and comment on their blog: [Modern Toolmaking](#).

R-bloggers.com offers [daily e-mail updates](#) about [R](#) news and [tutorials](#) on topics such as: visualization ([ggplot2](#), [Boxplots](#), [maps](#), [animation](#)), programming ([RStudio](#), [Sweave](#), [LaTeX](#), [SQL](#), [Eclipse](#), [git](#), [hadoop](#), [Web Scraping](#)) statistics ([regression](#), [PCA](#), [time series](#), [trading](#)) and more...

If you got this far, why not **subscribe for updates** from the site?
Choose your flavor: [e-mail](#), [twitter](#), [RSS](#), or [facebook](#)...

 Like  Share  Tweet

Comments are closed.

Top 3 Posts from the past 2 days

- [Imputing missing data with R: MICE package](#)
- [In-depth introduction to machine learning in 15 hours of expert videos](#)
- [RuleFit: When disassembled trees meet Lasso](#)

Search & Hit Enter

Top 9 articles of the week

1. [In-depth introduction to machine learning in 15 hours of expert videos](#)
2. [Installing R packages](#)
3. [Scatterplots](#)
4. [Using apply, sapply, lapply in R](#)
5. [Basics of Histograms](#)
6. [Learn R interactively with our new Introduction to R tutorial](#)
7. [How to Make a Histogram with Basic R](#)
8. [Why Big Data? Learning Curves](#)
9. [Adding a legend to a plot](#)

Sponsors





REVOLUTION
ANALYTICS



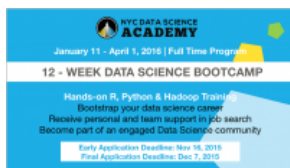
Werden Sie zum Expe[R]ten mit der
R-Akademie von



Beratung | Software
Training | Lösungen



[Plotly: collaborative, publication-quality graphing.](#)



Download a FREE Chapter Today!

The R Series

Parallel Computing for Data Science
With Examples in R, C++ and CUDA

SAVE 25% on All R Books

Promo Code **CZP40**

CRC CRC Press
Taylor & Francis Group
www.crcpress.com

STATWORX

Consulting
Schulung
Data Mining

Mehr erfahren

Try the FASTEST ML for R

Click for a Free Trial

YOTTAMINE ANALYTICS

Become a
Certified Data Scientist

ENROLL NOW

simplilearn

SIGMA

SIGMA

DATA VISUALIZATION CHARTS

[Contact us](#) if you wish to help support R-bloggers, and place your banner here.

Search & Hit Enter

[Full list of contributing R-bloggers](#)

[R-bloggers](#) was founded by [Tal Galili](#), with gratitude to the [R](#) community.

Is powered by [WordPress](#) using a [bavotasan.com](#) design.

Copyright © 2015 **R-bloggers**. All Rights Reserved. [Terms and Conditions](#) for this website