R-bloggers

R news and tutorials contributed by (552) R bloggers

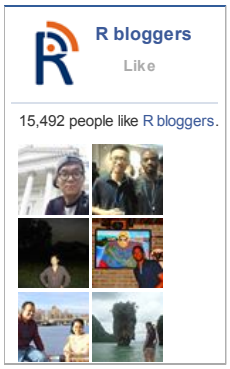
- Home
- About
- add your blog!
- Contact us
- RSS
- R jobs

Welcome!

Here you will find daily **news and tutorials about R**, contributed by over 450 bloggers. You can subscribe for e-mail updates:

Your e-mail here
Subscribe
12560 readers
BY FEEDBURNER

And get updates to your Facebook:



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)

Search & Hit Enter

Jobs for R-users

- SAS Analyst role who has experience with “R” Language
- Data Scientist
- Senior programmer / business analyst for customized solutions
- ATM Data Scientist
- Part-time R Instructor for Data Science Retreat

Popular Searches

- heatmap
- hadoop
- web scraping
- twitter

- maps
- ggplot
- finance
- shiny
- pca
- excel
- ggplot2
- googlevis
- RStudio
- trading
- animation
- boxplot
- eclipse
- LaTeX
- quantmod
- knitr
- time series
- tutorial
- sweave
- alt=
- market research
- how to import image file to R
- rattle
- GIS
- ecdf
- rcmdr

Recent Posts

- Gender gap and visualisation challenge @ useR!2014
- Jun 26-27, 2014 – Introduction to Data Science with R in NYC
- An infographic comparing R, SAS and SPSS
- Shiny 0.10
- Paper: The OpenCPU system – towards a universal interface for scientific computing through separation of concerns
- New R package hazus: Damage functions from FEMA's HAZUS software for use in modeling financial losses from natural disasters
- Rcpp11 conference marathon
- Quickly export multiple R objects to an Excel Workbook
- Multilevel modeling of community composition with imperfect detection
- Interactive documents: An incredibly easy way to use Shiny
- Contest: Prizes for Best R User

Groups Plotting Code
- Keep calm and #include <Rcpp11>
- Conditional Distributions from some Elliptical Vectors
- Free Geometric Mophometrics – a Summary
- Where World Cup players actually come from

Other sites

- SAS blogs
- Statistics of Israel
- Jobs for R-users

caretEnsemble Classification example

March 16, 2013
By Zachary Mayer

(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.

```
#Setup
rm(list = ls(all = TRUE))
gc(reset=TRUE)
set.seed(1234) #From random.org

#Libraries
library(caret)
library(devtools)
install_github('caretEnsemble', 'zachmayer') #
library(caretEnsemble)

#Data
library(mlbench)
dat <- mlbench.xor(500, 2)
X <- data.frame(dat$x)
Y <- factor(ifelse(dat$classes=='1', 'Yes', 'N

#Split train/test
train <- runif(nrow(X)) <= .66

#Setup CV Folds
#returnData=FALSE saves some space
folds=5
repeats=1
myControl <- trainControl(method='cv', number=
                          returnResamp='none',
                          returnData=FALSE, sa
                          verboseIter=TRUE, al
                          summaryFunction=twoC
                          index=createMultiFol
PP <- c('center', 'scale')

#Train some models
model1 <- train(X[train,], Y[train], method='g
```

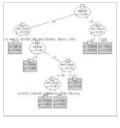
```
                tuneGrid=expand.grid(.n.trees=
model2 <- train(X[train,], Y[train], method='b
model3 <- train(X[train,], Y[train], method='p
model4 <- train(X[train,], Y[train], method='m
model5 <- train(X[train,], Y[train], method='k
model6 <- train(X[train,], Y[train], method='e
model7 <- train(X[train,], Y[train], method='g
model8 <- train(X[train,], Y[train], method='s
model9 <- train(X[train,], Y[train], method='g
model10 <- train(X[train,], Y[train], method='

#Make a list of all the models
all.models <- list(model1, model2, model3, mod
names(all.models) <- sapply(all.models, functi
sort(sapply(all.models, function(x) min(x$resu

#Make a greedy ensemble - currently can only u
greedy <- caretEnsemble(all.models, iter=1000L
sort(greedy$weights, decreasing=TRUE)
greedy$error

#Make a linear regression ensemble
linear <- caretStack(all.models, method='glm',
linear$error

#Predict for test set:
library(caTools)
preds <- data.frame(sapply(all.models, functio
preds$ENS_greedy <- predict(greedy, newdata=X[
preds$ENS_linear <- predict(linear, newdata=X[
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.

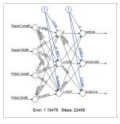
The greedy model relies 100% on the gbm, which makes sense as the gbm has an AUC of 1 on the training set. The linear model uses all of the models, and achieves an AUC of .5. This is a little weird, as the gbm, rf, SVN, and knn all achieve an AUC of close to 1.0 on the training set, and I would have 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 aRticles



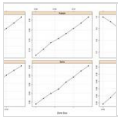
A question of model uncertainty

It has been several months since my last post on classification tree models, because two things have been consuming all of my spare time. ...



Cheat sheet for prediction and classification models in R

Ricky Ho has created a reference a 6-page PDF reference card on Big Data Machine Learning , with examples implemented in the R language. ...



Equivocal Zones

In Chapter 11, equivocal zones were briefly discussed. The idea is that some classification errors are close to the probability boundary (i.e. 50% for ...

Comments are closed.

Top 3 Posts from the past 2 days

- Quickly export multiple R objects to an Excel Workbook
- Prediction model for the FIFA World Cup 2014
- Interactive documents: An incredibly easy way to use Shiny

Search & Hit Enter

Top 9 articles of the week

1. Prediction model for the FIFA World Cup 2014
2. Who will win the World Cup and which prediction model?
3. Using apply, sapply, lapply in R
4. Installing R packages
5. The Most Comprehensive Review of Comic Books Teaching Statistics
6. New Version of RStudio: R Markdown v2 and More
7. A suggestion to Windows-based users of R: It may be time to relocate
8. R Markdown v2
9. Basics of Histograms

Sponsors