We've made a few improvements to the forums. You can read read more on the blog.

Forums / Data Analysis Assignment 2

86/90 Analysis, Feedback/Criticism Welcome

Subscribe for email updates.

Sort replies by: Oldest first Newest first Most popular

No tags yet.+ Add Tag

Marius Mather · 6 days ago

I'm going to throw my hat in the ring to see what people think.

Assignment Writeup

Figure

For the record, while I was happy with some of the ideas I came up with in building the actual prediction model (e.g. looking at the variables the trees were relying on most and creating a new set of trees without those variables), I though the final score I got was possibly overly generous, especially considering I wasn't that careful about specifically addressing the items in the rubric like confounders.

5

Anonymous · 6 days ago

Yes, you basically won the lotto.

4

+ Add New Comment

Anonymous · 6 days ago

Nice work. How did you combine the trees? Any code to share?

0

Marius Mather · 6 days ago

I put all the trees in a list and then used this function I had written to find the majority vote:

```
predict_from_btrees = function(btree_list, data=NA) {
  preds = laply(
    btree_list,
    function(btree) {
        return(as.matrix(predict(btree, newdata=data, type="class")))
    }
  )
  votes = adply(
    preds,
    2,
    function(x) names(table(x))[table(x) == max(table(x))][1]
  )
  return(votes$V1)
}
```

The full code for my bootstrapped tree approach is here.

2

+ Add New Comment

Curtis Lim · 6 days ago

Thank you for posting your paper. Well done. Just curious, could you speak a little as to how you formatted your paper like that? Is it LaTeX? And can you recommend any resources to learn it? Thank you.

0

Marius Mather · 6 days ago

It's sort of LaTex: I wrote my analysis in Markdown and RMarkdown (which are plain text formats), and use a combination of `knitr` and Pandoc to produce the final pdf. Pandoc allows you to get the benefits of LaTex without having to actually learn too much LaTex, you feed it a markdown file and ask it to convert to pdf, and it automatically generates LaTex from the input.

`knitr` is nicely integrated into RStudio, so you can just write up an RMarkdown file and click the "Knit HTML" button, which runs your R code and spits out a markdown file with all the results of your analysis.

The RMarkdown file I used for my results section is here, if you want to see roughly how it works.

2

+ Add New Comment

Thia, Kai Xin · 5 days ago

Nice. I never thought that removing the top5 most common variables to create a new separate bootstrap tree set and recombine with the first will actually increase performance. It will be interesting to test if that will work for randomforest as well

0

+ Add New Comment

Pavan Shinde · 4 days ago

One of the better reports and well deserved marks! Would you be able to share how you made the bar graphs?

0

+ Add New Comment

Daniel Wagner-Schuster · 4 days ago

Well done Marius! The analysis itself and the design look very professional. My question is: did you export the tables directly from R or did you make them new in Excel or any other programme?

0

+ Add New Comment

Eoin P Sharkey · 3 days ago

A very well written paper and a clear method.

Kudos for the use of plyr to build an aggregated model.

Did you consider the large extent of cross-correlation across ,any of the numerical variable ?

The majority voting methods depend on independent models for their boosting effect. Your elimination method may deliver independent models or it may not. I know this because I tried outsome majority voting models myself with little or no boosting effect !

It might have been interesting to try more n-fold splits of your training and validation sets rather than keeping them static.

Your elegant graph looks like a ggplot2 with be theme ?

^ 0 v

+ Add New Comment

Uriel Roque · 3 days ago

it would be nice to see the results of your model in the kaggle competition

^ 0 v

+ Add New Comment

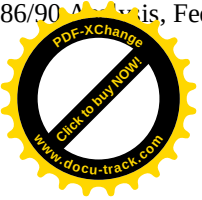
New post

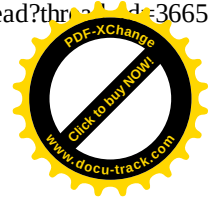
Bold *Italic* Bullets Numbers Link Image Math <HTML>

Make this post anonymous to other students

Subscribe to this thread at the same time

Add post


PDF-XChange
Click to buy NOW!
www.docu-track.com


PDF-XChange
Click to buy NOW!
www.docu-track.com