! This class has been made inactive. No posts will be allowed until an instructor reactivates the class.

question @61 stop following 16 views

Actions

Small tree leads to large error

I have created a few different trees and there are only about 7-10 terminal nodes for each, which leads to poor predictions (the predictions for the tree I submitted had a Kaggle score of 3.31). Is there any way to create a tree with more terminal nodes or one that can otherwise yield a lower score? I have already looked over all the variables and made two log transformations. Should I be making multiple trees for the different factors of a variable?

project

Edit good question | 0 Updated 4 years ago by Charles Hwang

the students' answer, *where students collectively construct a single answer*

Actions

CART usually isn't the best at predicting. Try a random Forest. - which does use multiple trees.

Also, considering that CART is a nonparametric model, and you aren't doing inference, I'm not sure variable transformations are all that effective in getting better predictions. I might be wrong though.

Edit thanks! | 0 Updated 4 years ago by Hannah Butler

followup discussions *for lingering questions and comments*

Resolved Unresolved @61_f1

Actions

Charles Hwang 4 years ago

I did try to fit a random forest but it gave me an error that said `vector memory exhausted (limit reached?)`, even if I tried to use the smaller "merg_train_small.RData" dataset.

helpful! | 0

Reply to this followup discussion

Start a new followup discussion

Compose a new followup discussion