



Do random forest variable importance measures take into account the interactions?

Do random forest measures of variable importance (mean change of accuracy, mean change of Gini index) take the interactions into account? I think I know how we come up with the variable importance plot (by permuting each of the predictors), and it doesn't seem that random forest captures the interaction. Does anybody have another point of view? Thanks.

random-forest importance

edited Jul 10 '14 at 4:39



Nick Stauner

8,550 5 24 53

asked Jul 10 '14 at 3:50



peace7

23 3

2 Answers

The variable importance obtained by permutations is computed only by permuting values for a single variable. Thus, it computes some importance measure of the given variable in the context that all other data is fixed. I think it is reasonable to state that the importance measure includes in the measurement also interactions, if such interactions exists. I mean that I see VI as an impure measure, a measure influenced by the main effect of that variable and also interaction with others.

Gini importance is found often to be in concordance with permutation importance, and I see it as a similar measure.

There is however something called interaction which is measured in random forests, and this measures if a split on a given variable increase or decrease splits on other measure. This can be computed for each pair of measures. It looks like a 2 measure interactions. If one want to measure interactions with more than 2 variables than I suppose it is possible extending the given procedure, but soon becomes too computer intensive.

Last thing called *interactions* is not implemented in R package *randomForests* as far as I know. Take a look on the brief description from the Breiman's page on RF [here](#), and check for Interactions section.

answered Jul 10 '14 at 9:18



rapaio

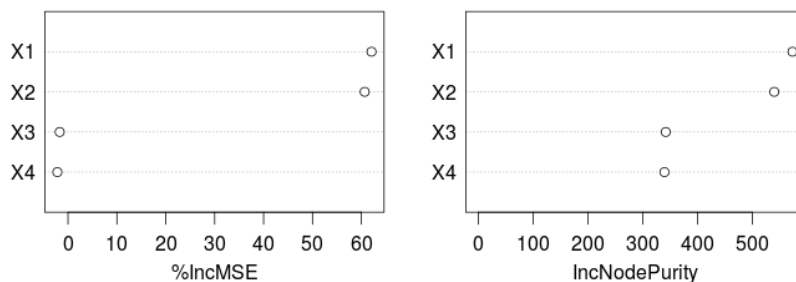
2,926 7 26

Run this code and assert that RF variable importance do incorporate interactions.

```
library(randomForest)
obs=1000
vars =4
X = data.frame(replicate(vars,rnorm(obs)))
ysignal = with(X,sign(X1*X2))
ynoise = 0.1 * rnorm(obs)
y = ysignal + ynoise
RF = randomForest(X,y,importance=T)
varImpPlot(RF)
```

You should see X1 and X2 are found the important and X3 and X4 are not. y is only explained as the interaction between X1 and X2, alone both variables are useless.

RF



answered Apr 1 '15 at 14:38



Soren Havelund Welling

2,871 5 17

7/15/2016

Do random forest variable importance measures take into account the interactions? - Cross Validated

Two thumbs up! Nice work! – [rolando2](#) Apr 1 '15 at 23:45

```
X = data.frame(replicate(vars, rnorm(obs))) ysignal = with(X, sign(X1*X2)) very tight code +1 –  
Qbik Apr 26 at 7:11
```

[Add Another Answer](#)