

Cross Validated is a question and answer site for people interested in statistics, machine learning, data analysis, data mining, and data visualization. It's 100% free, no registration required.

[Take the 2-minute tour](#) ✕

Role of n.minobsinnode parameter of GBM in R [closed]

I wanted to know what the *n.minobsinnode* parameter means in the GBM package. I read the manual, but it is not clear what it does. Should that number be small or large to improve the results?

r gbm

edited Jun 18 '12 at 12:21



mbq ♦

15.8k 7 44 93

asked Jun 18 '12 at 2:11



Benoit_Plante

718 1 8 19

closed as too localized by [cardinal ♦](#), [gung](#), [Andy W](#), [StasK](#), [mpiktas](#) Sep 24 '12 at 1:56

This question is unlikely to help any future visitors; it is only relevant to a small geographic area, a specific moment in time, or an extraordinarily narrow situation that is not generally applicable to the worldwide audience of the internet. For help making this question more broadly applicable, [visit the help center](#).

If this question can be reworded to fit the rules in the [help center](#), please [edit the question](#).

1 Answer

At each step of the GBM algorithm, a new decision tree is constructed. The question when growing a decision tree is 'when to stop?'. The furthest you can go is to split each node until there is only 1 observation in each terminal node. This would correspond to $n.minobsinnode=1$. Alternatively, the splitting of nodes can cease when a certain number of observations are in each node. The default for the R GBM package is 10.

What is the best value to use? It depends on the data set and whether you are doing classification or regression. Since each tree's prediction is taken as the average of the dependant variable of all inputs in the terminal node, a value of 1 probably won't work so well for regression(!) but may be suitable for classification.

Higher values mean smaller trees so make the algorithm run faster and use less memory, which may be a consideration.

Generally, results are not very sensitive to this parameter and given the stochastic nature of GBM performance it might actually be difficult to determine exactly what value is 'the best'. The interaction depth, shrinkage and number of trees will all be much more significant in general.

answered Jun 18 '12 at 2:50



Bogdanovist

1,906 8 21