Bookmarks

- Module 1: The Basics of R and Introduction to the Course
- Entrance Survey
- Module 2: Fundamentals of Probability, Random Variables, Distributions, and Joint Distributions
- Module 3: Gathering and Collecting Data, Ethics, and Kernel Density Estimates
- Module 4: Joint, Marginal, and Conditional Distributions &

Module 11: Intro to Machine Learning and Data Visualization > Machine Learning II > Pruning Decision Trees - Quiz

Pruning Decision Trees - Quiz

Bookmark this page

Question 1

1/1 point (graded)

When fitting a decision tree, which of the following constraints could be made? (Select all that apply)

- [ ] a. Restricting the upper bound of points per leaf
- [x] b. Increasing the lower bound of observations per leaf (node)
- [x] c. Restricting the total number of leafs (nodes) in tree
- [ ] d. Increasing the number of data points for which you are training your algorithm with.

✔

Explanation

B,C. As mentioned by Professor Mullainathan, Increasing the lower bound of points per leaf prevents you from putting one or two points in every leaf and exactly overfitting. Restricting the number of leafs in at tree as well helps in creating too many splits and creating a model that only works in your sample. Restricting the upper bound of points per leaf if anything would encourage overfitting since it

would lead to more splits in a decision tree instead of allowing for generality in other datasets. Increasing the number of data points for which you are training with will not regularize you tree since your tree will then over fit again in the new data.

Submit You have used 1 of 2 attempts

✔ Correct (1/1 point)

Question 2

1/1 point (graded)

Suppose you are doing a predictive model using regressions. Which of the following could you do to avoid overfitting your data?

- a. Limit the number of variables allowed in the regression. ✔
- b. Increase the number of variables allowed in your regression.
- c. Limit the number of data points in your regression.
- d. Increase the number of data points in your regression.

Explanation

Limiting the number of variables allowed in your regression is a way of restricting the complexity of your model. As Prof. Mullainathan explained this will impose a constraint on your maximizer, and therefore help avoid over fitting. Note, however, that restricting the complexity of your model will increases the bias of your estimator (recall the bias-variance tradeoff). Changing the number of data points in your regression will not affect the overfitting problem in your functions because they will just keep overfitting any new data you give them. As a side note, the second method is exactly what LASSO (least absolute shrinkage and selection operators) regressions do in order to increase the predictive ability of your coefficients.

Submit You have used 1 of 2 attempts

✔ Correct (1/1 point)

Question 3

1/1 point (graded)

True or False: The complexity of your model is something that you need to decide on ex ante.

a. True

b. False ✔

Explanation

As Prof. Mullainathan explained, the complexity of your model (C), is decided by your algorithm. In practice, you run your model with different levels of complexity, and you use a separate dataset to choose the level of complexity that maximizes the accuracy of your predictions.

Submit You have used 1 of 1 attempt

✔ Correct (1/1 point)

Discussion
Topic: Module 11 / Pruning Decision Trees - Quiz

Show Discussion


Download on the
App Store
ANDROID APP ON
Google play