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 > Overfitting and the Tuning Problem - Quiz

Overfitting and the Tuning Problem - Quiz

Bookmark this page

Question 1

0/1 point (graded)

True, False, Uncertain: We can use traditional cross validation criterion on our data to find the optimal complexity level for predictive models.

- a. True ✖
- b. False
- c. Uncertain

Explanation

Both False and Uncertain are acceptable answers. Traditional cross validation criterion compare your predicted values against the observed data, in order to assess models' ability to fit your data. So they are precisely a measure of in-sample fit. In prediction problems, you are concerned with out-of-sample fit. If we used the cross-validation criterion on the full dataset to choose the complexity of your model, it will always favor higher complexity, since that will improve your in-sample fit.

Machine learning works by splitting the sample into a “tuning set” and a “training set”. Once you fit your model on the training set, cross validation criteria are used to assess the fit of your model on the set of data which were not used to fit it, i.e the “tuning set”.

Submit You have used 1 of 1 attempt

✖ Incorrect (0/1 point)

Question 2

1/1 point (graded)

Which of the following statements are true? (Select all that apply)

- [ ] a. The higher the complexity of your model, the lower the signal, but the more likely you are to have an overfitting problem.
- [x] b. The higher the complexity of your model, the higher the signal, and the more likely you are to have an overfitting problem.
- [ ] c. The higher the complexity of your model, the lower the signal, and the more likely you are to have an overfitting problem.
- [x] d. The higher the complexity of your model, the higher the signal it will give you, and the worse it will perform out-of-sample.

✔

Explanation
The higher the complexity of your model, the stronger the signal it will provide you. However, the probability of overfitting also increases as your model's complexity increases. So A and C are incorrect. D and B are equivalent: the more your model overfits in-sample, the worse it will be at performing out-of-sample predictions. Recall what happens as you increase the number of nodes in your tree to perfectly fit the data: you can fit the data perfectly, but that will be useless for making out-of-sample predictions.

Submit You have used 1 of 2 attempts

✔ Correct (1/1 point)

Discussion
Topic: Module 11 / Overfitting and the Tuning Problem - Quiz

Show Discussion