

MITx: 14.310x Data Analysis for Social Scientists

Heli



#### 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 &
   Functions of Random
   Variable

Module 11: Intro to Machine Learning and Data Visualization > Machine Learning I > The "Secret Sauce" of Machine Learning - Quiz

# The "Secret Sauce" of Machine Learning - Quiz

☐ Bookmark this page

## Question 1

1.0/1.0 point (graded)

In what sense do we mean estimation "overfits" relative to prediction?

- ullet a. OLS minimizes the in-sample loss  $\checkmark$
- b. OLS minimizes the out-of-sample loss
- c. OLS is low dimensional
- d. OLS is high dimensional

### **Explanation**

By construction, the OLS estimator minimizes the MSE in the sample. That is, it minimizes the distance between the predicted and observed outcomes. So by construction, OLS minimizes the in-sample loss and therefore maximizes the in-sample fit. On the other hand, prediction aims to minimize the out-of-sample loss and maximize the out-of-sample fit, and will thus correct for overfitting.

- Module 5: Moments of a Random Variable,
   Applications to Auctions,
   Intro to Regression
- Module 6: Special
   <u>Distributions, the</u>

   <u>Sample Mean, the</u>
   <u>Central Limit Theorem,</u>
   and Estimation
- Module 7: Assessing and Deriving Estimators -Confidence Intervals, and Hypothesis Testing
- Module 8: Causality,
   Analyzing Randomized
   Experiments, &
   Nonparametric
   Regression
- Module 9: Single and Multivariate Linear Models
- Module 10: Practical Issues in Running

Submit You have used 1 of 2 attempts

### **Question 2**

1/1 point (graded)

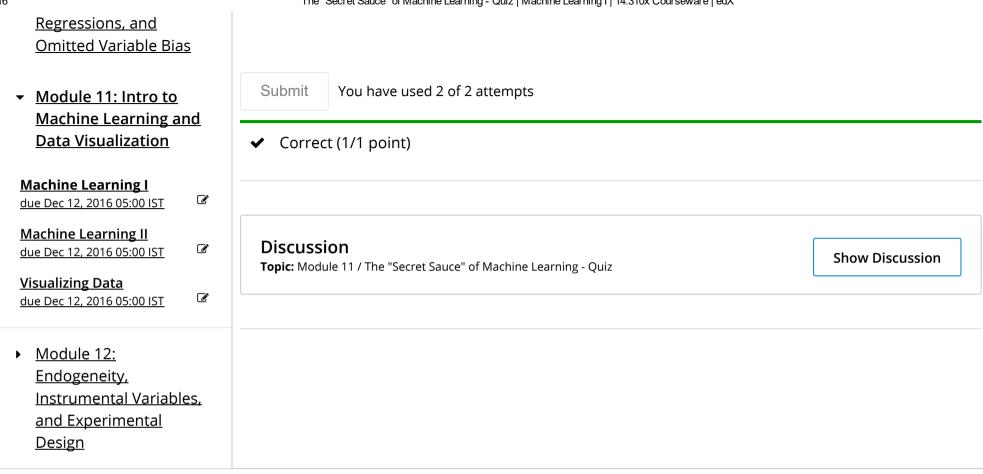
Stata and R report the adjusted R squared. What is the "adjustment" for? (Select all that apply)

- a. It corrects for overfitting.
- b. It corrects for the high variance.
- c. It corrects for the fact R-squared is biased.
- d. It corrects the R-squared, by penalizing the regression for adding more variables.



# **Explanation**

The adjusted R squared corrects for the fact that, by construction, OLS overfits. It is a function of the ratio of the number of observations to the number of variables, so it penalizes you for adding variables. While the R squared will always increase when adding variables, the adjusted R square may increase or decrease depending on whether the gain in the accuracy of prediction is larger or smaller than the penalty for adding one more variable.



© All Rights Reserved



© 2016 edX Inc. All rights reserved except where noted. EdX, Open edX and the edX and Open EdX logos are registered trademarks or trademarks of edX Inc.















