Bookmark this page

3.1.R1

1/1 point (graded)

Why is linear regression important to understand? Select all that apply:

- [ ] The linear model is often correct
- [x] Linear regression is very extensible and can be used to capture nonlinear effects ✔
- [x] Simple methods can outperform more complex ones if the data are noisy ✔
- [x] Understanding simpler methods sheds light on more complex ones ✔

✔

In linear regression, the word "linear" applies to the coefficients: the dependence between Y and the coefficients is linear. This does not mean the dependence between Y and X is linear. Assume X is a one dimensional variable. Basic linear regression is (I omit the noise and intercept for simplicity): Y=βX, But this is still linear regression: Y=β1X+β2X^2+β3log(X), The latter is the same as basic linear regression with feature vector (X,X2,log(X)) instead of X.

https://stats.stackexchange.com/questions/306009/explain-linear-regression-is-very-extensible-and-can-be-used-to-capture-nonline

https://gerardnico.com/data_mining/one_rule

Simple rules often outperformed far more complex methods because some datasets are : (1)really simple (2)so small/noisy/complex that nothing can be learned from them

Explanation

The linear model (and every other model) is hardly ever true, but it is an important piece in many more complex methods.

Submit

Answers are displayed within the problem

3.1.R2

1/1 point (graded)

You may want to reread the paragraph on confidence intervals on page 66 of the textbook before trying this queston (the distinctions are subtle).

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

- [x] A 95% confidence interval is a random interval that contains the true parameter 95% of the time ✔

?????

https://www.reddit.com/r/cheatatmathhomework/comments/2bngst/some_statistics_help_please/
A confidence interval is an interval estimator, and so it is indeed random.

- [ ] The true parameter is a random value that has 95% chance of falling in the 95% confidence interval

- [ ] I perform a linear regression and get a 95% confidence interval from 0.4 to 0.5. There is a 95% probability that the true parameter is between 0.4 and 0.5.

??????????? why false ???

- [x] The true parameter (unknown to me) is 0.5. If I sample data and construct a 95% confidence interval, the interval will contain 0.5 95% of the time. ✔

✔

https://gerardnico.com/data_mining/confidence_interval
A Bayesian would not agree with this statement as it would depend on his/her prior distribution

https://www.reddit.com/r/cheatatmathhomework/comments/2bngst/some_statistics_help_please/
A naive interpretation of the statement would make it true, but formally the statement is false; the probability is unknown to the observer, but it is only one of two values: 0 or 1.

Explanation

Confidence intervals are a "frequentist" concept: the interval, and not the true parameter, is considered random.

Submit

Answers are displayed within the problem