Bookmark this page

5.Q.1

1/1 point (graded)

If we use ten-fold cross-validation as a means of model selection, the cross-validation estimate of test error is:

- biased upward
- biased downward
- unbiased
- potentially any of the above ✔

Explanation

There are competing biases: on one hand, the cross-validated estimate is based on models trained on smaller training sets than the full model, which means we will tend to overestimate test error for the full model.

On the other hand, cross-validation gives a noisy estimate of test error for each candidate model, and we select the model with the best estimate. This means we are more likely to choose a model whose estimate is smaller than its true test error rate, hence, we may underestimate test error. In any given case, either source of bias may dominate the other.

Submit

Answers are displayed within the problem

5.Q.2

1/1 point (graded)

Why can't we use the standard bootstrap for some time series data?

- [x] The data points in most time series aren't i.i.d. ✔
- [ ] Some points will be used twice in the same sample
- [x] The standard bootstrap doesn't accurately mimic the real-world data-generating mechanism ✔

✔

Explanation

The bootstrap always involves using some points more than once in each resample, but that doesn't inherently make it incorrect (unless we are trying to gauge prediction error). The real problem in this case is that the usual bootstrap algorithm samples i.i.d., so there is no serial autocorrelation (unlike what is observed in most time series). This makes the set of resampled time series very very different from the sorts of time series we actually get in the real world.

Submit

Answers are displayed within the problem