

Homework1

Question 1

- ▶ Estimate prediction models of BuildYear using
 - ▶ 8/10 of “Example.csv” (training data)
 - ▶ OLS, RandomForest, Tree with pruning, and SimpleMean algorithm
 - ▶ Cross-validation to select the best algorithm
- ▶ Evaluate its performance using the remainings (2/8) of “Example.csv” (test data)

Question 2

- ▶ In general,
- ▶ Why can not a good prediction model perfectly fit the training data?
- ▶ Why can we not use training data to evaluate the prediction model?
- ▶ Why can we not use test data to select the best prediction algorithm?

Details

- ▶ Submit script-file (Code for Question 1, and comment for Question 2) on ITC-LMS
- ▶ DeadLine: 11/13
- ▶ Tutorial
<https://github.com/tetokawata/BookEmiricalSocialML>