

Lab 4C: Cross-Validation

Directions: Record your responses to the lab questions in the spaces provided.

Predictions

Step 1: train-test split

- Explain these lines of code and describe the train and test data sets.

Aside: `set.seed`

Aside: train-test ratio

Step 2: train the model

Step 3: test the model

- Calculate the MSE in the same way as you did in the previous lab (test MSE is simply MSE of the predictions on the test data).

Recap

Why cross-validate?

Example of overfitting

- Which model does a better job of predicting the 7 training points?

- Which model do you think will do a better job of predicting the rest of the data?

Example of overfitting, continued

- Which model does a better job of generalizing to the rest of the `arm_span` dataset?