

PS9 Hoehne

Jordan Hoehne

April 2018

1 PS9

Question 5

Dimensions of `housing.train`: [1:404]

Thus, the training data is a vector of 404 integers.

Question 6

What is the optimal value of `lambda`?

The optimal value of `lambda` is 0.00288.

What is the in-sample RMSE?

The in-sample RMSE is 0.1881683.

What is the out-of-sample RMSE (i.e. the RMSE in the test data)?

The out-of-sample RMSE is 0.2173507

Question 7

What is the optimal value of `lambda` now?

The optimal value of `lambda` is 0.00314.

What is the in-sample RMSE?

The in-sample RMSE is 0.1899243.

What is the out-of-sample RMSE (i.e. the RMSE in the test data)?

The out-of-sample RMSE is 0.1913142.

Question 8

What are the optimal values of `lambda` and `alpha` after doing 6-fold cross validation?

The optimal value of `lambda` is 0.00606.

The optimal value of `alpha` is 0.582.

What is the in-sample RMSE?

The in-sample RMSE is 0.2008356.

What is the out-of-sample RMSE?

The out-of-sample RMSE is 0.1913142.

Does the optimal value of `alpha` lead you to believe that you should use LASSO or ridge regression for this prediction task?

Given `alpha`, I would recommend using the lasso method for the prediction task as it's RMSE is lower than the RMSE of ridge regression.

Question 9

Why you would not be able to estimate a simple linear regression model on the `housing.train` dataframe.

Because the linear model is so simple, it's not capable of correctly measuring our data. A linear model would lead to a underfit model. Using the RMSE values of each of the tuned models in the previous three questions, comment on where your model stands in terms of the bias-variance tradeoff.

Given our models RMSE's continuously result near .2, we have a highly biased model that exhibits low variance.