PS9_Maddy

Ethan Maddy

April 13, 2021

1 Question no. 7

The dimension of the initial housing_train has 1 dependent variable and 13 independent variables. After prep, the data has 1 dependent variable and 74 independent variables. (61 X variables added)

2 Question no. 8

The optimal value of λ is 0.00222 with the LASSO, 6-fold cross validation. The in-sample RMSE value is 0.0625. The out-of-sample RMSE value is 0.220.

3 Question no. 9

The optimal value of λ is 0.0373 with the Ridge, 6-fold cross validation. The in-sample RMSE value is 0.0694. The out-of-sample RMSE value is 0.219.

4 Question no. 10

No, we cannot estimate a simple linear regression model on a data set that has more columns than rows - there are not enough observations. According to the RMSE values found and low prediction error, it is unlikely that we underfitting or overfitting the data.