

PS9_Dunkleberger

Davis Dunkleberger

April 2023

1 Question 7

Housing_train is a 404 observations of 14 variables (404x14). There are 506 X variables in the original housing data. The housing_train data set has 102 less X variables than the original housing set.

2 LASSO Model Results

The in sample RMSE is 0.172 while the out of sample RMSE was 0.390. The optimal λ is 0.00355648.

3 Ridge Regression Model Results

The out of sample RMSE is 0.173. The optimal λ is 0.001389495.

4 Question 10

You cannot run a simple linear regression on a dataset with more columns than rows. You would use a LASSO or Ridge model to do so. With how close my RMSEs are for both sets of models, the models are different due to variance rather than bias.