

Problem Set 9

Bowei Dong

April 2024

Q7

The dimension of training data is $404 \text{ obs} \times 14 \text{ variables}$.

`housing_test_prepped` has 75 variables, which is 61 more than the original dataset.

Q8

So the in-sample RMSE is 0.413, and the out-of-sample RMSE is 0.390.

The optimal λ , as reported, is 0.00355648.

Q9

The out-of-sample RMSE estimation using ridge regression is still 0.390, which is same with Lasso.

In this case, the optimal λ given is 0.03727594, which is almost 10 times higher than the previous one.

Q10

Estimate a simple linear regression model with more columns than rows has issues related to model **overfitting** and unreliable estimation of coefficients.

Comparing the RMSE values of tuned LASSO and ridge regression models helps us evaluate their bias-variance trade-off. The LASSO model, which performs variable selection by setting some coefficients to zero, typically exhibits lower variance but may have higher bias. In contrast, ridge regression, which shrinks coefficients towards zero without variable selection, tends to have lower bias but potentially higher variance. The choice between LASSO and ridge regression depends on the data's characteristics and the importance of bias versus variance in the prediction task.