PS9

Becky Crouse

April 11, 2023

Question 7:

The dimensions of "housing_train" are 404 rows and 14 columns. The number of X variables in this dataset are the same, but once the training data is prepared (or juiced), 61 new X variables are added bringing the total X variables to 75.

Question 8: LASSO model

- Optimal value of $\lambda = .00139$

- In-sample $RMSE = .0632$

- Out-of-sample $RMSE = .170$

Question 9: Ridge model

- Optimal value of $\lambda = .0373$

- In-sample $RMSE = .0699$

- Out-of-sample $RMSE = .173$

Question 10:

- An OLS model cannot be estimated with a model that has more variables (columns) than observations (rows).

- Both models have a similar out-of-sample RMSE, which suggests that they have a similar bias-variance trade-off. The difference between the in-sample and out-of-sample RMSE is somewhat large in terms of the in-sample error size, but not too large in overall magnitude. Thus I would conclude that these models are appropriately balancing the bias-variance trade-off.