

PS9 Smith

April 9, 2024

1 Question 7 Response

Dimensions of housing-train: 404, 14 Number of variables in original housing data: 14 Number of variables in housing-train: 74

2 Question 8 Response

The Optimal lambda for LASSO is .002196441 The In-sample RMSE for LASSO is .1382056 The out-of-sample RMSE fro LASSO is .1840099

3 Question 9 Response

The Optimal lambda for Ridge is .05833939 The out-of-sample RMSE for Ridge is .1756918

4 Question 10 Response

In a simple linear regression model, the number of columns needs to be less than or equal to the number of rows. If there are more columns than rows, the model becomes overparameterized, and there is not enough information to estimate the coefficients uniquely.

A lower RMSE means better performance of the model. This is done by combining the bias and variance together. Based on this, it is not surprising that the in-sample RMSE from LASSO would perform better than the out of sample RMSE. The LASSO penalty reduces some less important coefficients to zero to help reduce variance, but the bias could be off if the wrong coefficients are selected.

In Ridge regression models, the coefficients are reduced but never set to zero, like done in LASSO. Consequently, the variance is also reduced, but could still be considering coefficients that are good indicators of the model.

Based on the results from Questions 8 and 9, the in-sample RMSE for LASSO is a better fit, but RIDGE is better for out-of-sample fit.