

# PS9\_Wardwell

lwardwell

April 2025

## Question 7

The dimensions of the original training data are  $404 \times 14$ . The dimensions of the preprocessed training data are  $404 \times 75$ . The original number of X variables is 13. The new number of X variables is 74. Therefore, I have 61 more X variables than in the original housing data.

## Question 8

- The optimal value of  $\lambda$  for the LASSO model is .00139.
- The in-sample RMSE is .137.
- The out-of-sample RMSE (on the test data) is .188.

## Question 9

- The optimal value of  $\lambda$  for the Ridge model is .0000000001.
- The in-sample RMSE is .140.
- The out-of-sample RMSE (on the test data) is .181.

## Question 10

Regarding the ability to estimate a simple linear regression model on a dataset with more columns than rows: OLS estimation does not work on data with more columns than rows, and models become subject to overfitting. Based on the RMSE values from the tuned LASSO and Ridge models:

- LASSO out-of-sample RMSE: .188
- Ridge out-of-sample RMSE: .181

My model's position in terms of the bias-variance tradeoff is: The Ridge model chose a very small lambda (0.0000000001), which means the model favors lower bias but possibly higher variance. The LASSO model selected a larger penalty (.00139) which increases bias but reduces variance. The out-of-sample RSE achieved by the Ridge regression is slightly better, suggesting that this model has better bias/variance tradeoff.