

# PS 9

Mengyang Davila

April 11, 2023

## Question 7

The `housing_train` dataset has 404 observations with 14 variables. The recipe creates 61 additional variables in the original housing data, resulting in the `housing_train_prepped` dataset with 404 observations and 75 variables.

## Question 8

Using the ridge regression model, the optimal  $\lambda$  is 0.00139. The in-sample RMSE is 0.137, and the out-of-sample RMSE is 0.188.

## Question 9

Using the LASSO model, the optimal  $\lambda$  is 0.0373. The in-sample RMSE is 0.140, and the out-of-sample RMSE is 0.180.

## Question 10

No, it is not feasible to estimate a simple linear regression model on a data set that has more columns than rows, because it can lead to over-fitting and unreliable results. Regularization techniques such as ridge regression and LASSO can be useful to prevent over-fitting by imposing penalties on the coefficients.

In terms of the bias-variance tradeoff, the ridge regression model may have less bias, as it has a lower in-sample RMSE, indicating that it may fit the data better. However, it may have higher variance, as indicated by the higher out-of-sample RMSE. On the other hand, the LASSO model may have more bias, but it may have lower variance.