PS9$_Z$hang

Haotian Zhang

April 2025

1 Answer for question 7

The dimension of the training data is 404 × 14. There are 60 additional X variables introduced by the recipe steps.

2 Answer for question 8

Optimal value of lambda for LASSO model is 0.002592944; for LASSO model, in-sample RMSE equals to 0.1392827; out-of-sample RMSE equals to 0.1820611.

3 Answer for question 9

Optimal value of lambda for ridge regression is 0.09236709 and out-of-sample RMSE for ridge regression is 0.1726775.

4 Answer for question 10

Q1: A linear regression model on a dataset with more columns than rows will not be reliable and is prone to overfitting.

Q2: Ridge regression with slightly higher RMSE in-sample but better out-of-sample performance, is better balanced in the bias-variance tradeoff. Neither model seems severely overfit, but ridge is slightly more stable across training and test sets, which indicates a better generalization to unseen data.