Problem Set 9

Beth Felkner

April 2024

1 Dimensions of Training Data

The housing train prepped data is 404 rows by 75 columns. It has 61 more variables than the original housing data

2 LASSO Model

- The optimal value of lambda is 0.00139
- The RMSE in-sample is 0.137
- The RMSE out-of-sample is 0.188

3 Ridge Model

- The optimal value of lambda is 0.0373
- The RMSE in-sample is 0.140
- The RMSE out-of-sample is 0.181

4 Conceptual Questions

- I would not be able to estimate a simple linear regression that had more columns than rows because this violates the rank and order condition, which would lead to the model being over-parameterized.

- Considering the magnitude and range of our MEDV that we are predicting, the RMSE (all of them) seems quite small in comparison which would suggest to me that these models are overfit, and thus have too high of reliance on variance according to the bias-variance tradeoff. In both cases we can see that the RMSE within sample is lower than out of sample, which is to be expected. However, the ridge seems to be slightly less overfit, because the out-of-sample error is lower and the difference between in and out of sample error is also smaller.