PS9 Dawkins

Julie Dawkins

April 2024

1 Problem 7

The dimension of the training data is 404x75. This is 61 more X variables than the original housing data set.

2 Problem 8

Using the LASSO model, the optimal λ is 0.00139. In sample, this leads to an RMSE of 0.137 and an R2 of 0.891; out of sample, this leads to an RMSE of 0.188 and an R2 of 0.768.

3 Problem 9

Using the ridge model, the optimal λ is 0.0373. In sample, this leads to an RMSE of 0.140 and an R2 of 0.884; out of sample, this leads to an RMSE of 0.180 and an R2 of 0.787.

4 Problem 10

You cannot estimate a simple linear regression on a data set with more columns than rows due to the rank and order condition. Both models seem to be doing well in the bias/variance trade off. However, the LASSO model has a lower in-sample RMSE and a higher out-of-sample RMSE, implying that it might be higher on the variance (i.e., it is slightly more over-fit) than

the ridge model, which has a higher in-sample RMSE but lower out-of-sample RMSE than the LASSO model.