

Problem Set 9

Yeganeh Karbalaei

April, 2025

- Question 7: What is the dimension of your training data? How many more X variables do you have than in the original housing data?
379, 74 and 506, 14.
- Question 8: What is the optimal value of lambda? What is the in-sample RMSE? What is the out-of-sample RMSE (i.e. the RMSE in the test data)?
Optimal lambda for LASSO: 0.002224789 In-sample RMSE for LASSO: 0.1365926 Out-of-sample RMSE for LASSO: 0.1861817
- Question 9: What is the optimal value of lambda now? What is the out-of-sample RMSE (i.e. the RMSE in the test data)?
Optimal lambda for Ridge: 0.06485379 In-sample RMSE for Ridge: 0.1441057 Out-of-sample RMSE for Ridge: 0.1735581
- Question 10: We cannot be able to estimate the a simple linear regression model on a data set that had more columns than rows because the matrix would become rank deficit.
Ridge regression has a lower in-sample RMSE when I first ran my code, which indicates that it fits the training data better than LASSO. This means that Ridge is less biased on the training data. LASSO fares better according to out-of-sample RMSE which suggests an improvement in generalizing to the unseen data, demonstrating LASSO's coping with variance more effectively. But for the second time, I got different results. In this second run, Ridge seems to have found a better balance in the bias-variance tradeoff.