

Problem Set 9

Haocheng Yang

April 10, 2023

1 Dimension of Training Data and Variables of Original Dataset

The training dataset contains 404 observations with 14 variables.

The original housing dataset contains 14 variables with 506 observations

2 LASSO: optimal value of Lamda and RMSE

The optimal value of Lamda is 0.0676; The out of sample RMSE is 0.170;

3 Ridge: optimal value of Lamda and RMSE

The optimal value of Lamda is 0.0748; The out of sample RMSE is 0.173;

4 Question 10

I don't think that it is a good way to estimate using simple linear regression model on a data set that has more columns than rows. It might lead to model overfitting.

Regarding the bias-variance tradeoff, a model with low bias will fit the training data well, but may overfit and perform poorly on new data.

In the previous questions, we used cross-validation to tune the models and estimate their performance on new data. The optimal RMSE values obtained suggest that the Lasso model with the tuned lambda parameter strikes a good balance between bias and variance. It is able to achieve low out-of-sample RMSE without overfitting the training data.