

PS9

Conrad Polkosnik

April 2, 2018

1 Part 5

The dimension of the training data is 404 observations of 450 variables.

2 Part 6

- The optimal value of lambda is: 0.0355
- The in-sample RMSE is: 0.2183221
- The out-of sample RMSE is: 0.2213377

3 Part 7

- The optimal value of lambda is: 0.0294
- The in-sample RMSE is: 0.2019355
- The out-of sample RMSE is: 0.2202868

4 Part 8

- The optimal value of lambda is: 0.0282
- The optimal value of alpha is: 0.0376
- The in-sample RMSE is: 0.2004339
- The out-of sample RMSE is 0.2181641

Because the optimal value of alpha is so close to 0, it would lead me to believe that I should use ridge regression to make the prediction.

5 Part 9

You would not be able to estimate a simple linear regression model on the `housing.train` data as your prediction would underfit completely. With 450 variables to be considered, the best option would be to regularize the data in order to produce a much more precise prediction with regard to the enormous amount of variables. Regularization is the optimal process for trading off bias and variance. With low RMSE values around .2 for each of the models along with lambda values very close to 0, each of the models seems to have a high measure of complexity which would describe a variance on the high side as well as a low bias.