

PS9

Alex Skipper

April 2018

1 Part 5

The dimensions are 404 observations with 450 variables

2 Part 6

lambda=0.00791
rmse=0.208
out-of-sample=0.1937389

3 Part 7

lambda=0.00636
rmse=0.202
out-of-sample=0.1937635

4 Part 8

lambda=0.006
alpha=0.55
rmse=0.197
out-of-sample=0.1811442
The alpha leads me to believe that we should use LASSO due the the alpha increasing towards 1

5 Part 9

You wouldn't be able to estimate a simple linear regression model because there are so many variables. If you were to run a linear regression the data would be under-fitted and not represent the data accurately. Or it would not capture the outliers effectively or other important information.

The RMSE in the three models are relatively low, which shows that the model accurately fits the data given and could be accurate when applying out of sample data.

The three models fall more on the simplistic side, so the models are not too complex. As a result, the variance is not high, so they are less likely to overfit the data. The trade off is that there might be a lot of bias in the models, so the three models might may be underfitting the data. As a result, important information might not be captured in the model, so data outside the training data may produce lower variance predictions.