

PS9 Mozaffar

Yaseen Mozaffar

April 2020

1 Problem 5

The dimensions of the training data is 404x450

2 Problem 6

Using LASSO regression: the optimal lambda is .0349; The in-sample RMSE is .1970152; The out-of-sample RMSE is .198467

3 Problem 7

Using ridge regression: the optimal lambda is .117; The in-sample RMSE is .1515769; The out-of-sample RMSE is .167843;

4 Problem 8

Using ElasticNet: the optimal lambda is .0721; the optimal alpha is .0646; The in-sample RMSE is .170188; The out-of-sample RMSE is .1171311

5 Problem 9

A simple linear regression would be insufficient because it does not have a mechanism for tuning the model. Because of this, the results are almost entirely dictated by the particular sample used to "train" the model. The GLMNet methods used allow us to average the results of the training data to avoid such bias. Our models, compared to a simple linear regression, are more biased but ultimately provide better estimates. The similarity between the in-sample and out-of-sample RMSE for each of the three models indicates that they are lower in variability than the linear model, but higher in bias.