PS9 McGuire

Joel McGuire

April 2018

1 Dimension

506 rows by 14 columns

2 glmnet Results

Results				
model	λ	α	In RMSE	Out RMSE
Lasso	0.154	0	0.23	0.26
Ridge	0.00723	1	0.01	0.0094
Enet	0.0804	0.307	0.19	0.146

2.1 Comments

The α level for the elastic net model is closer to zero which I take to mean we should prefer the Lasso model, but that seemingly contradicts the results we get for each models RMSE with Ridge having the lowest by a decent margin.

3 Additional Questions

3.1 Why can't we run OLS?

We can not run OLS on the housing.train dataframe because there are more columns than rows, I think it has to do with the assumption of the independent variables being linearly independent, which would obviously be violated if the columns didn't have full rank.

3.2 Bias Variance Trade-off

Linear models generally lean to the side of sacrificing variance for lower bias and I think we see that in the relatively small difference between in and out of sample RMSE. I am not sure how it happened that the out of sample RMSE for the enet model is less than the in sample but I guess there's always room for some noise.