PS9

bradley rann

April 2023

1 Introduction

Housingtrain has the dimensions of 404 x 14. Which is 80 percent of the x variables that we saw from the original data set or roughly 100 less X's. My lasso and ridge models were disappointing and way too overfit. For the lasso my optimal lambda and RMSE (in then out) were: The optimal value of alpha is 311797133923401408.000. In-sample RMSE: 9404088.043212535 Out-of-sample RMSE: 8319409.380164554

For my ridge model: Optimal value of l: 0.1 In-sample RMSE: 33848.130746604016 Out-of-sample RMSE: 2088631.937254573.

This indicates that I have extremely high variance in my models, so I'm high on the variance and low on bias. I believe the problem is that in python there a couple of ways to set up these models and the method I chose ultimately didn't set out the grid of alphas very efficiently. Essentially I need some more practice on setting up these tasks and making sure they're validating properly, because before that everything in my code seems to be working perfectly, I just need to experiment with setting up the model and validating it by troubleshooting the code.

For the last question, no you would not be able to set up a regression with more columns than rows as your estimates would not be reliable.