

Data Science for Economists - Spring 21

Problem Set 9

Syed Waleed Mehmood Wasti

April 15, 2021

1. Done
2. Done
3. Done
4. Done
5. Done
6. Done
7. `housing_train` data has 405 observations for 14 variables while the `housing` data has 506 observations for the same number of variables. The `housing_train_prepped` data though has 75 variables, almost 60 more variables than the original housing data!
8. The optimal λ is 0.00356 and the value for RMSE is 0.197
9. The optimal value of λ is 0.0233 while the RMSE is 0.192.
10. I don't think it would be possible to estimate a simple linear regression model on a data set that has more columns than rows because then it would not provide us with a unique solution.