Problem Set 9

Matt Mullins

April 2018

1 Task 5

Question 1: What is the dimension of your training data?
Answer: The training data dimensions after performing matrix modifications
are 404 X 450 (rows, columns)

2 Task 6

Question 1: What is the optimal value of lambda?
Answer: The optimal value of lambda is 0.0195.
Question 2: What is the in-sample RMSE?
Answer:The in-sample RMSE is 0.178.
Question 3: What is the out-of-sample RM
Answer: The out-of-sample RMSE is 0.184464.

3 Task 7

Question 1: What is the optimal value of lambda now?
Answer: The optimal value of lambda is 0.103
Question 2: What is the in-sample RMSE?
Answer: The in-sample RMSE is 0.163
Question 3: What is the out-of-sample RMSE? 0.1544176

4 Task 8

Question 1: What are the optimal values of lambda and alpha after doing 6-fold
cross validation?
Answer: The optimal value of lambda is 0.143 and the optimal value of alpha
is 0.0684
Question 2: What is the in-sample RMSE?

Answer: The in-sample RMSE is 0.169.

Question 3: Does the optimal value of alpha lead you to believe that you should use LASSO or ridge regression for this prediction?

Answer: Since alpha = 0.0684 ¡ 0.5, this leads me to believe that using the ridge regression is the better choice.

5   Task 9

Question: Explain why you would not be able to estimate a simple linear regression model on the housing train data frame?

Answer: Because of the wide range of variability in housing prices in the sample, a linear model would most likely have a high variance and we would consider it under-fitted. Thus, we would not be properly accounting for the variance-bias problem.

Question: Where do your models stand in terms of the variance-bias trade-off?

Answer: For all 3 of the models, I would think that our models are more biased because of the fine-tuning performed on a relatively low test set. This may cause our predictors to become fit to a highly-specific set of data. In addition, given the relatively small number of input variables (relative to the entire number that predict housing prices), we may have created a model that is too specific. We also seem to support this with seemingly low RMSE.