

Question 1

What is the optimal value of alpha for ridge and lasso regression? What will be the changes in the model if you choose double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

Answer 1

- The most optimal value of alpha for ridge regression is 298 and 499 for lasso regression.
- If we double the value of alpha for both models, the model performance on r squared declines quite a bit for the Lasso model but slightly increases for the Ridge model. This is interesting and indicates that I could have opened up the alpha scores tested with much larger values.
-

Question 2

You have determined the optimal value of lambda for ridge and lasso regression during the assignment. Now, which one will you choose to apply and why?

Answer 2

- Ridge regression with the lambda score listed above performed the best on our r2 metric and is the reason why I would choose to apply this model for learnings. We choose an alpha of 298. While the lasso model performed well on our training data, the Ridge regression model had less variance between test and training evaluation metrics.

Question 3

After building the model, you realized that the five most important predictor variables in the lasso model are not available in the incoming data. You will now have to create another model excluding the five most important predictor variables. Which are the five most important predictor variables now?

Answer 3

- Surprisingly the performance of the Lasso model increased after eliminating those features. This indicates an opportunity for improved feature selection.
 - For the new Lasso model important predictor variables are now:
 - TotRmsAbvGrd
 - GarageCars
 - FullBath
 - Fireplaces
 - OverallCond

Question 4

How can you make sure that a model is robust and generalisable? What are the implications of the same for the accuracy of the model and why?

Answer 4

- In order to make the model more robust and generalisable into the future, I would do several things. I would do additional feature selection pre-processing to identify only the features most important for modeling. It was clear that from our analysis, there is much to be improved upon during our feature selection, eliminating and imputation. I would then spend more time evaluating other parameters within ridge and lasso models that could help influence the models performance. We would want to make sure we don't overfit the data, but I do think a lot more could be done to improve the model's performance and robustness.