

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 optimal value of alpha for ridge and lasso regression

Ridge Alpha - 1 and Lasso Alpha – 10

- Making Ridge Alpha 2 results in slight decrease in R2 score on training data and increase in testing data.
- Making Lasso Alpha 20 results in slight decrease in R2 score on training data and increase in testing data.

- **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

The R2 score of Lasso is slightly higher than Ridge for the test dataset, so we will choose Lasso regression to solve this problem.

Question 3

After building the model, you realised 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

The five most important predictor variables are

- 11stFlrSF-----First Floor square feet
- GrLivArea-----Above grade (ground) living area square feet
- Street_Pave-----Pave road access to property
- RoofMatl_Metal-----Roof material_Metal
- RoofStyle_Shed-----Type of roof(Shed)

- **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

- The model should be generalized so that the test accuracy is not lesser than the training score.
- It should perform well on datasets other than the ones used during training.
- The model should be as simple as possible
- It should have a proper bias variance tradeoff.