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 to double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

Answer:

- The optimal value for ridge regression is 10 and lasso regression is 0.001.
- For the ridge regression, as we double the value of alpha as 20, the coefficients are reduced toward zero not absolute zeros. R2 is reduced & the training error is increased. At the level of alpha 10, the error is minimum, and we have high R2 value.
- For Lasso regression, as we double the value of alpha to 20, the coefficients are shrinked towards zero, reducing the coefficients value. This reduces the number of coefficients making the model simpler. Hence the value of R2 is also reduced
- The most important predictor variables are
  Lasso: Overall Quality, House Style_2story, Overall Condition, 1STFLR SF, Building age
  Ridge: House Style_2story, Overall Quality, Overall Condition, 1STFLR SF, Garage Finish

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:

- We will be applying alpha = 10 for ridge regression and alpha = 0.001 for the lasso regression as the R2 value and error terms are optimal at these levels.
- The r2 value for the Lasso regression is slightly higher than the Ridge Regression in the test dat.
- The error term is also marginally lower in case of Lasso regression.
- **So, will be using the Lasso regression model**

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:

The most important predictor in the model before:

Overall Quality, House Style_2story, Overall Condition, 1STFLR SF, Building age

After removing important predictors, the next five most variables are

Building _Age, Garage Finish, HouseStyle_Others, Sale_Condition_other, BldgType_others

Question 4

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

Answer:

- By having a simpler model, it can work better on test data. There will be high bias and the variance will be low. Selecting a model using variance-bias tradeoff can helps us in having a robust model
- Model can be made generalizable by having the test accuracy same as train accuracy
- Focusing on the number features. Selecting optimal number of relevant features can help in making the model more generalized
- A robust and generalized model works well on both training and test data. So the accuracy will be higher.
- Having complex model, may overfit the data reducing the test accuracy. Regularization will help in handling the issue making the model robust and generalized