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:

- The optimal value of alpha is **80** and **0.0001** for ridge and lasso regression respectively.
- After changing ridge alpha from **9 to 18** and lasso alpha from **0.0004 to 0.0008** accuracy on both test and train data goes down but on the lasso, I observed that test and train are approximately the same which seems odd.
- **OverallQual** for ridge and **BsmtFullBath** for lasso is the most important feature after changing alpha by double

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:

As I observed that R2 for Ridge is **88.7%** and **87.6%** on train and test data
and Lasso is **86.3%** and **85.6%** on the train and test data, so go with Ridge because we have higher R2 and the difference b/w test and train R2 is approx the same for both and also lasso tends to set 0 value param for many input var but in Ridge its at least keeping it which help to achieve more R2 on both test and train.

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:

After creating with removing the top 5 variables, now 1stFlrSF, 2ndFlrSF, LowQualFinSF, BedroomAbvGr, and MiscVal are the top 5 features for the new model.

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:

As we know to get more generalisable we can use Ridge or Lasso regression but for getting a more robust model we need to pick the best alpha value for that, we can Grid Search Cross-Validation method which analysis all provided values and pick the best from them. As we know from the bias-variance tradeoff we get the optimal model near the interaction point and which is achieved from Grid Search CV.