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 for ridge regression is 0.9

The optimal value of alpha for lasso regression is 0.0001

Observations after alpha is doubled on Ridge

- Ridge Model training accuracy decreased from from 0.883 to 0.880
- Ridge Model Top Coefficients predictors are the same , however the co efficient values are decreasing when alpha is doubled
- The number of predictors is same when alpha is doubled.

Observations after alpha is doubled on Lasso

- Lasso Model training accuracy decreased from from 0.885 to 0.883
- Lasso Model Top Coefficients predictors are the same , however the co efficient values are decreasing when alpha is doubled
- Lasso Model Predictors reduced from 62 to 57

Hence we can conclude that the r2 score decreases when the alpha value is doubled.
The most important variables after change are OverllQual_9, MSZoning_RL, MSZoning_RH, MSZoning_FV, MSZoning_RM

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?

Will choose Lambda model as it gives r2score for train as 0.885 and r2score for test as 0.858 with feature selection of 62 predictors with alpha 0.0001.Eventhough ridge is similar, features are more in ridge.

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?

After removing the top five variables from lasso model the new top five model predictor variables are Condition1_PosA, CentralAir_Y, Neighborhood_NridgHt, Neighborhood_Somerst, GarageQual_Gd

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?

The model should be as simple as possible, though its accuracy will decrease but it will be more robust and generalisable. It can be also understood using the Bias-Variance trade-off. The simpler the model the more the bias but less variance and more generalizable. Its implication in terms of accuracy is that a robust and generalisable model will perform equally well on both training and test data i.e. the accuracy does not change much for training and test data.