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 ridge and lasso regression depends on its coefficients calculated. In lasso even for small values of ' λ' the coefficients of features reduce to zero and we need to choose the optimal 'λ' value.

If alpha =0 then there will be no regularisation in the model.

If alpha is high then testing and training score is low.

Similarly if double the alpha it more regularises the model more which means it can more interpret and create under fitted model. The higher the coefficient the more important the variable.

After implementing this change, the most important predictor variables will be those that remain non-zero in the model. These variables are the ones that the model considers most influential in making predictions. The coefficients associated with these variables will indicate their relative importance. The predictors that have been driven to zero are considered less important according to the model.

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 mostly choose lasso regression because it is good for feature selection. If you suspect that only a subset of features are truly important, and you want a model that automatically excludes less important predictors, Lasso regression might be better because of that we are able to select important feature in assignment.

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 looking at coefficient we can state that the higher the coefficient important the feature. The first five important predictor variable will be

1stFlrSF,

TotRmsAbvGrd,

GarageYrBlt,

OverallCond,

GrLivArea.

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: After performing EDA on dataset. What quality of dataset we use, Regularisation using different values of alpha and try to figure out where we get best fitted model.