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?

Ans:

In the case of ridge regression:- if we plot the curve between negative mean absolute error and alpha ,observed that as the value of alpha increase from 0 the error term decrease and the train error is showing increasing trend when value of alpha increases .when the value of alpha is 10 the test error is minimum so optimum value of alpha equal to 10 for ridge regression. For lasso regression the value of alpha is very small is 0.001, when we increase the value of alpha the model try to penalize more and try to make most of the coefficient value zero. Initially it came in negative mean absolute error and alpha.

When we double the value of alpha for our ridge regression model will apply more penalty on the curve and try to make the model more generalized that is making model more simpler and no thinking to fit every data of the data set . we get more error for both test and train. Similarly when we increase the value of alpha for lasso it try to penalize both model and coefficient of the variable will reduced to zero, when increase the value of our r2 square also decreases.

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?

Ans:

- Optimal Value of lamda for ridge : 10
- Optimal Value of lamda for Lasso : 0.001
- Because of Feature selection as well we can choose Lasso regression in this case. it would help in feature elimination and the model will be more robust.
- Top 5 most significant variables in Lasso are having better values than ridge regression

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?

Ans:

After building a model the five most important predictor variables in lasso model are

('GarageFinish_RFn', 0.062),
('GarageFinish_Unf', 0.066),
('SaleCondition_Normal', 0.114),
('SaleCondition_Others', 0.115),
('SaleCondition_Partial', 0.124)]

If we drop these predictor variables, if we create an another model the new important predictor variables are

('GarageType_BuiltIn', 0.084),

('GarageType_Detchd', 0.086),

('GarageType_No Garage', 0.093),

('GarageType_Others', 0.118),

('GarageFinish_No Garage', 0.193)

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?

Ans:

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.

Bias: Bias is error in model, when the model is weak to learn from the data. High bias means model is unable to learn details in the data. Model performs poor on training and testing data. Variance: Variance is error in model, when model tries to over learn from the data. High variance means model performs exceptionally well on training data as it has very well trained on this of data but performs very poor on testing data as it was unseen data for the model. It is important to have balance in Bias and Variance to avoid overfitting and under-fitting of data.