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 = 3.0

Lasso Regression = 100

If we double the value of alpha coefficients modified and

The top 5 variables with high coefficient before and after change are as below in descending order

Ridge

       Before - > OverallQual , GrLiveArea, 2ndFlrSF, 1stFlrSF, GarageCars

       After - > OverallQual, 2ndFlrSF, GrLiveArea, GarageCars, TotRmsAbvGrd

Lasso

       Before - > GrLiveArea, OverallQual , GarageCars, Neighborhood_NridgHt, TotRmsAbvGrd

       After - > GrLiveArea, OverallQual , GarageCars, TotRmsAbvGrd , Neighborhood_NridgHt

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-

Though metrics for both the regressions performed similarly, I would prefer Lasso because it reduces the coefficients to zero in turn reducing features in the model and complexity of the model

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 -

Important variables were

      Before - > GrLiveArea, OverallQual , GarageCars, Neighborhood_NridgHt, TotRmsAbvGrd

      After elimination - > 1stFlrSF, 2ndFlrSF, GarageArea, SaleType_CWD, OverallCon

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 -

 Generalisable means model is neither underfit nor overfit. This can be acheived by making the model as simple as possible. Building model with best accuracy over train followed by Regularization which reduces the complexity of model by eliminating the unwanted features. will result in robust model.