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 :

a) The optimal alpha value in case of Ridge and Lasso is as below:
   Ridge – 2, Lasso - 0.001
b) If we double the values of alpha for both , model may get decrease in regularization and might cause underfitting.
c) Ridge
   R2 After change : R2_TRAIN : 0.92725, R2_Test : 0.884648
   R2 Before change : R2 train : 0.93046, R2 test : 0.886421
   MSE After change : MSE_Train : 0.002011, MSE_Test : 0.001209
   MSE Before change :MSE train : 0.00198 ,MSE test : 0.001156
   Top 5 Variables after change : OverallQual , GrLivArea,OverallCond, TotalBsmtSF,GarageArea
   Lasso
   After change : R2_TRAIN : 0.86512, R2_Test : 0.8846
   Before Change : R2 train : 0.89179 ,R2_test : 0.886421
   After change : MSE_Train : 0.00285, MSE_Test : 0.002242
   Before Change : MSE train : 0.00235 MSE test : 0.001799
   Variables selected : 25
   Top 5 Variables after change : OverallQual , GrLivArea, TotalBsmtSF,GarageArea,
   YearBuilt_Updated

   Conclusion : For both Ridge and Lasso , R2 score decreased and MSE got increased after doubling alpha values

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)
R2 value for Ridge and Lasso for train and test are
R2 Rideg train : 0.93046 , R2 Rideg test : 0.886421
R2 Lasso train : 0.89179 , R2 Lasso test : 0.886421

MSE value for Ridge and Lasso for train and test are
MSE Ridge train : 0.00198 , MSE Ridge test : 0.001156
MSE Lasso train : 0.00235 , MSE Lasso test : 0.001799

The optimal lambda value in case of Ridge and Lasso is as below:
Ridge - 2,
Lasso - 0.001

We can say that Lasso has almost similar R2 values for test and train , almost same MSE as Ridge , Feature elimination to 36 from 174.
Will prefer Lasso over 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?

Ans : Top 5 features after Building Lasso model are :
OverallQual , OverallCond , GrLivArea, TotalBsmtSF, YearBuilt_Updated
After excluding above predictors and rebuilding model we get below mentioned variables
1stFlrSF, 2ndFlrSF, GarageArea, LotArea, Neighborhood_Crawfor

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 :
As per Occams Razor : If two models are given with finite data , we should pick based on below assumptions :

a) Simpler models are less biased and less variance and more generic
b) Simpler models are more robust than complex models
c) Simpler models require less training data to build the model and test.
d) Simpler models are tend towards more errors but complex errors are more prone to Overfitting.