Q1: 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: The optimal value of alpha for ridge regression is 0.2.
The most important predictor variable is “GrLivArea“

Ridge Regression:After doubling alpha
Loss function = OLS + 2*alpha*summation(squared coefficients)

The most important predictor variable after alpha is 0.4 remains the same “GrLivArea“

Lasso Regression:After doubling alpha
Loss function = OLS + 2*alpha*summation(absolute values of magnitude of coefficient)

The optimal value of alpha in case of lasso regression is 0.0002.The most important predictor variable is “GrLivArea“.

The alpha doubles to 0.0004 and the predictor variable remains the same “GrLivArea“

It is observed that the predictor variable coefficient has reduced after doubling the alpha value in both ridge and lasso regression.

Q2: 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: The optimal value for ridge is 0.01 and lasso is 0.0002. Ridge has higher r2_score than laaso. That means ridge describes the model better than lasso. Hence we will choose ridge over lasso.

Q3: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:The top five predictor variable in lasso are:
GrlivArea,OverallQual,Overallcond,TotalBsmtSF,LotArea.
Now that all the above variables are deleted the next top five variables are:

1stFlrSF,2ndFlrSF,Neighborhood_StoneBr,BsmtFinSF1, Neighborhood_Crawfor.

Q4: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:Model accuracy determines the efficiency of the model.

If the diference between R square value of both the training and test data is less than 5% we can say that the model is robust and generalisable with respect to accuracy.