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 :

```
Alpha for Ridge – 20
Alpha for Lasso – 100

When the alpha values were doubled ; the following was observed;
Ridge – The values of r2 for train and test was almost same
Lasso – The values  of r2 reduced for train and test

#Top 5 parameters for Lasso Regression with alpha -200 ;

#OverallQual_10 ( Very Excellent )
#OverallQual_9  ( Excellent )
#Neighborhood_NoRidge  ( indicating location )
#Neighborhood_Crawfor ( indicating location )
#OverallQual_8
```

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?

```
Will chose ridge regression with lambda – 40 and r2 score for train and
test as below ;

0.8731941795475238
0.856196300239418
Ridge regression seems to suite the model considering the multicollinea
rity of input variables , given limited data set , given the r2 scores
```

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?

```
RoofMatl_WdShngl       - roof material is wood Shingles
OverallCond_9          Excellent
Exterior1st_BrkFace    - Exterior covering on house is brick face
```

```
BsmtExposure_Gd         - Basement Exposure is good
Functional_Typ          - Home is of typical functionality
```

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?

1) A model is robust when data variation doesn’t impact the performance
2) A generalisable model would be able to provide consistent performance being adaptable to unforeseen data
3) Model shouldn’t be overfitting for above two to be true
4) A complex model would have high accuracy ; but a robust and generalisable model will have less variance with some bias , having bias would reduce accuracy
5) Using Ridge & Lasso regression /regularisation techniques , we try to bring in a balance on the above mentioned model characteristics