1.Question 1

What is the optimal value of alpha for ridge and lasso regression? What will be the changes in the model i
f 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.optimal value of alpha for ridge is 5.0 and lasso 0.001
   if i change change alpha for lasso is to .002 r2 score is reduced from 0.87 to 0.85
   if i change change alpha for ridge is to 10 r2 score is reduced from 0.87 to 0.85
   most predictor variablesa are
   1.OverallQual :Excellent,good quality materials used for house is good predictable variable.

   2.RoofMatl :roof material such as Standard (Composite) Shingle ,Wood Shakes are good predictable v
ariable.

   3.Neighborhood:Neighborhood with Northridge ,Stone Brook are good predictable variable.

   4.GrLivArea

   5.BsmtExposure.

2.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 alpha for ridge is 5.0 and lasso 0.001.i will choose lasso regression model because i
t includes feature selection along with regularization.And also it performs well with both test and train dat
aset.

Question 3

After building the model, you realised that the five most important predictor variables in the lasso model ar
e not available in the incoming data. You will now have to create another model excluding the five most i
mportant predictor variables. Which are the five most important predictor variables now?

ans:Exterior1st_BrkFace ,Exterior2nd_ImStucc,   LandSlope_Sev, RoofStyle_Mansor, LandContour_HL
S , MSZoning_FV , Condition1_Norm MSSubClass_70 ,HouseStyle_1Story , FireplaceQu_TA. Fireplac
eQu_Gd   i will use these predictor variables

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:A model needs to be made robust and generalizable so that they are not impacted by outliers in the tr
aining data. The model should also be generalisable so that the test accuracy is not lesser than the trainin
g score. The model should be accurate for datasets other than the ones which were used during training.
To ensure that this is not the case, the outlier analysis needs to be done and only those which are relevan
t to the dataset need to be retained. Those outliers which it does not make sense to keep must be remove
d from the dataset. This would help increase the accuracy of the predictions made by the model. If the mo
del is not robust , it cannot be trusted for predictive analysis.