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 and lasso is:

Ridge: 0.4

Lasso: 100

Following are the observations when alpha value doubled.

In Ridge regression after increasing the Alpha value from 0.4 to 0.8 r2 score has decreased for train set and slightly increased for test data.

In Lasso regression after increasing the Alpha values R2score of training & test data has decreased slightly.

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] The R2_score , RSS and MSE are better for lasso regression than Ridge , so Lasso Regression gives the better 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] After dropping following top 5 variables,

1.LotArea

2.OverallQual

3.OverallCond

4.YearBuilt

5.BsmtUnfSF

Next top 5 Predictive variables are,

1. TotalBsmtSF
2. BedroomAbvGr
3. TotRmsAbvGrd
4. GarageCars
5. Total_sqr_footage

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] If models built simpler then it will be more robust. If the model is simpler(robust) then it will have more bias but it variance will be less. So if the model is simpler then the implications are, model can be less accurate and it will have more bias but less variance. Because simpler model will perform well in train data but will have less predictive capability for test data.

So ideally model should be balanced with both bias and variance to get the good predictions from the model. This can be achieved using regularisation.

Regularization helps in managing the model complexity by essentially shrinking the coefficients towards zero. This avoids the model becoming too complex, thus reducing the risk of overfitting and keeps the model optimum simpler, regularization method penalizes the model if it becomes more complex. Regularization also helps to achieve the Bias-Variance trade off. It compromises by increasing bias to a optimum position where Total Error is minimum.