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?

- Optimal Values of alpha for Ridge : 2.0

- Optimal Values of alpha for Lasso : 0.0006

For doubling the Alpha values I have build Ridge and Lasso mpdel with 4.0 and 0.0012 values for alpha respectively. Rsquare values got slighly increased after doubling the alpha as below:

Ridge R-square :

Train : 0.9101260549564558 Test : 0.8746177612301544

Ridge R-square (After doubling) :

Train : 0.9086015016829656 Test : 0.8753942548141012

Lassso R-square :

Train : 0.909323877088995 Test : 0.8756707276852946

Lasso R-square (After doubling) :

Train : 0.9062348911728326 Test : 0.8757045443255819

Most important predictors after change for Lasso :

1 Neighborhood_Crawfor

2 GrLivArea

3 MSZoning_RL

4 Neighborhood_MeadowV

5 MSZoning_FV

6 Neighborhood_NridgHt

7 Neighborhood_StoneBr

8 Neighborhood_BrDale

9 GarageType_Attchd

10 TotalBsmtSF

Most important predictors after change for Ridge :

1 Neighborhood_Crawfor

2 MSZoning_RL

3 Neighborhood_MeadowV

4 GrLivArea

5 MSZoning_FV

6 Neighborhood_NridgHt

7 Neighborhood_StoneBr

8 GarageType_Attchd

9 Neighborhood_BrDale

10 Neighborhood_Somerst

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?

- The R-square value for Lasso is slighty better than that of the Ridge model. Thus it implies that Lasso performance better on test hich is unknown data than Ridge.

- Also Lasso punishes the high value coefficient by reducing them to zero thus featutes get reduces and model gets simpler.

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?

- Five important predictors before :

'Neighborhood_Crawfor',

'GrLivArea',

'MSZoning_RL',

'Neighborhood_MeadowV',

'MSZoning_FV'

- Five important predictors after :

Exterior1st_BrkComm

Exterior2nd_CmentBd

Exterior1st_Stone

Exterior1st_CemntBd

Neighborhood_BrDale

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?

- Robustness and generalisable model decides how broade range of inputes a model can work with. More complex the model iss less the model is generalisable.

- Because a very complex model is very sensative to any change in input data, thus is will not be alble to work well on unseen data. We can make model more robust and genralisable by trading off complexity of model with bias.

- Regularization of model penalizes the model to avoid becoming too complex thus helps to get more robust.

- As we know that to maintain the accuracy total error should be very less. Thus it depend upon the Bias and the variance of the model.

- By keeping a well balance between Bias and Variance we can make total error less and thus accuracy can be impoved keeping robustness of model intact.

-