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 is 0.1 and R2-score associated with it is 0.88,The optimal value of alpha for Lasso is 0.0001 and R2-score associated with it is 0.83

After Doubling,

Alpha for ridge is 0.2 and r2score is 0.78

Alpha for lasso is 0.002 and r2score is 0.87

Ridge Alpha Co-Efficient after Doubling

MSZoning_RL   0.368594

MSZoning_FV   0.363097

MSZoning_RH   0.361355

MSZoning_RM   0.326677

RoofMatl_Membran   0.319116

Lasso Co-Efficient after Doubling

OverallQual   0.102841

GrLivArea 0.101970

Neighborhood_Crawfor   0.068315

Neighborhood_NridgHt   0.057673

Neighborhood_Somerst   0.055731

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: After comparing r2 scores and Mean Sqaure error of Lasso and ridgr model i would choose ridge as it is giving better r2 score compared with lasso.

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: The model should be generalized so that the test accuracy is not lesser than the training score.The model should be accurate for datasets other than the ones which were used during training.

Too much importance should not given to the outliers so that the accuracy predicted by the model is high. To ensure that this is not the case,the outliers analysis needs to be done and only those which

are relavant to the dataset need to be retained. Those outliers which it does not make sense to keep must be removed from the dataset.

If the model is not robust, It cannot be trusted for predictive analysis.