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 value for the ridge is 10 and optimal values of alpha for lasso is 0.001.

When we used double the value of alpha for both ridge and lasso, there is a small change in r2 score and looks like it worked well.

The most important predictor variables after the change is implemented are

1 RoofMatl
2. Neighborhood
3. OverallQual
4. OverallCond
5 Exterior1st

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?

 I will choose Lasso since it helped in feature elimination by making intrecepts 0 and made model simple and reduce the complexity. Comparing to the Lasso regrssion, Ridge regression gave a complex 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?

five most important predictor variables are:

1. RoofMatl

2.Neighborhood

3.OverallCond

4.Exterior2nd

5.Exterior1st

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?

We can say that a model is robust and generalisable, when the mdoel is simple and able to perform well on both train and test data. The simplicity of a model can be determined but Bias Varaince tradeoff, Higher the Bias more the complexity and lower varaince. Lower the Bias less complexity and higher varaince. To increase the accuracy of the model, we need to use more data and treat the missing values and outliers properly.