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: we have found optimal value for Ridge is 10.0 and optimal value of lasso is 0.0001.

Once we double the alpha i.e., Ridge alpha now 20.0 and lasso alpha now 0.0002 we found the top 5 important predictors are: 'GrLivArea', 'MSZoning_RL', 'OverallQual', 'OverallCond', 'MSZoning_RM'

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: We will choose lasso as the R2 score is little better and lasso is a feature elimination model. All together lasso needs 60 fits to come to the same result where ridge takes 135 fits. Using lasso model will be more robust.

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:

```
'TotalBsmtSF','FullBath','SaleCondition_Partial','HalfBath','Neighborhood_Crawfor'.
```

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: A model will robust when it does not get impacted by outliers and it will generalizable when train and test data performance score of the model lies between 5%.

Generally using outliers' treatment, we can make sure a model is robust. In our model we have removed outliers and worked with 92% records.

To make model generalised, we have used Ridge and lasso with multiple alphas 0.000001, 0.00001,0.0001, 0.001, 0.01, 0.1, 1, 10, 100, 500, 1000, 10000 and the R2 score we get using training and test data are 0.92 and 0.88(within 5%). So, we can conclude that model will be not overfitted by unseen data.