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?

Ans: The optimal value for alpha and lasso regression

      Ridge Alpha 0.5

      Lasso Alpha 5

If we choose double the value of alpha for both ridge and lasso regression, model complexity will have a greater contribution to the cost.

After changes, the important predictor variables are – GrLivArea, MSZoning_RM, MSZoning_FV, MSZoning_RH, MSZoning_RL,LotConfig_CulDSac

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?

Ans: Even though Ridge has given good performance (R-Square value), I would choose Lasso model for following reasons.

- Model is giving decent performance.
- Efficiently solved high dimensionality problem by shrinking insignificant coefficients to zero.
- Simpler model and easy for maintenance.

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?
Ans: The 5 most important predictors are MSZoning_RH',
'Neighborhood_NridgHt', 'BsmtExposure_Gd', 'GrLivArea','SaleCondition_Partial'

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?

Ans: By making sure the model is not over fitting and is as simple as possible, we are ensuring that it is robust and generalizable. The accuracy of the model will go up if we try to over fit the model but that no longer makes it generalizable. When the model is generalized the accuracy should be pretty good on both the training and the testing dataset making the model robust.