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 ridge regression : 100.0

Optimal value for lasso regression : 100.0

Doubling the value of alpha in ridge regression results in considering most important variables and gives less weightage to the variables that are contributing less even though they are considered, on the other hand if the alpha in lasso regression is doubled it cuts down on variables heavily and keeps only the critical ones which have direct impact.

In Ridge regression the most important predictors would be the ones who have more linear relation with the target, and in Lasso it would be variables that have more direct impact.

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 would prefer lasso regression as it focuses on simplicity by using key predictors in decision making rather than involving all the variables.

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?

Firstly, I must remove the five important predictors which i have chosen. Then I would be building a new lasso regression model. After training the model I would rank the remaining features by their importance. The coefficients of the current model would show which features are more significant.

The five important predictors after processing are:

'ExterQual'

'KitchenQual'

'BsmtQual'

'Fireplaces'

'LotArea'

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?

Making a model robust and generalizable includes techniques like cross-validation, regularization, feature selection, and data splitting, which can make sure the the. Model is performing well.