Subjective Questions:

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?

Sol:

The optimal value could be 10 for each method. If I choose the double of optimal value the R2 value has little difference towards decreasing. It is observed there decrement in GrLivArea feature.

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?

Sol:

Ridge regression has the better R2 value than Lasso. Multi linear regression fails to recognize the less important features and uses all of them, leading to overfitting. Ridge regression provides the hyperparameter which will infuse bias, for the better fit of the data.

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?

Sol: Lotarea, 1stfl, ExterCond, BsmtCond,, GarageArea

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?

Sol: Distribution of errors is normal - which satisfies the regression assumptions. Model is reasonably fit. Graphs of target and predicted variables are linear.