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?

The optimal value of alpha for Ridge and Lasso regression are `0.05` and **20** respectively.

The most important predictor variables for Ridge are

`'MSZoning_RL', 'Fireplaces', 'GarageArea', 'WoodDeckSF', 'MSZoning_FV`

The most important predictor variables for Lasso are

`'MSZoning_RL', 'Fireplaces', 'GarageArea', 'WoodDeckSF', 'MSZoning_FV'`

after the change

If we choose double value for alpha then

Ridge alpha set to 0.1

The most important predictor variables are

For Ridge

`'MSZoning_RL', 'Fireplaces', 'GarageArea', 'WoodDeckSF', 'MSZoning_FV'`

For Lasso alpha set to 40

The most important predictor variables are

`'HalfBath', 'Fireplaces', 'GarageArea', 'WoodDeckSF', 'MasVnrType_None'`

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 apply ridge regression for this assignment.

A different set of predictors are obtained when we rerun Lasso.This indicates that collinearity is very high.Ridge regression is useful to reduce multicollinearity .

Disadvantage of lasso regression is that it creates a sparse model. All predictors are potentially relevant in this dataset and hence Ridge is better than Lasso.

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?

After excluding the above, the 5 most important predictor variables are

`'2ndFlrSF', 'GrLivArea', 'BsmtFullBath', 'FullBath', 'MSZoning_FV'`

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?

The following are the steps to ensure that a model is robust and generalisable.

1. The data should be cleaned and validated by removing outliers and missing values.
2. Extract relevant features using techniques such as transformation, scaling, encoding.
3. Select and evaluate appropriate models based on accuracy, recall, mse and other metrics
4. Avoid overfitting and underfitting by using regularization techniques such as Ridge Regression and Lasso Regression

The accuracy of the model depends on all the above factors.