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: The optimal value of alpha for ridge and lasso regression depends on the value of alpha.

The alpha for ridge is 8

The alpha for lasso is 0.0001

Ridge regression: If the alpha value for the ridge is doubled , it will increase the regularization and as a result the coefficient of predictor will tend to become smaller. The increase in regularization will lead to a simple model less likely to overfit the training data. It can prevent multicollinearity by encouraging coefficient to be small bit not zero.

Lasso Regression: If the alpha value for lasso is doubled. It will also increase regularization. Lasso uses L1 regularization, which encourages some coefficients to be zero. Increasing alpha inthis model will set the coefficient to be zero. The impact on the model will be sparsity in the coefficient vector. Predictor variables may become irrelevant as they can become zero.

Effectively performing feature selection, only on a subset of the most important predictor variables will have non collinearity.

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: Ridge adds L2 regularization to the linear regression which penalizws the sum of squared coefficients.

It is effective when we believe that most of the features are relevant but we want to prevent multicollinearity and control the magnitude of the coefficients.

Ridge can also be a good choice when we have large no.of features and we want to avoid feature selection.

Lasso: It adds L1 regularization which can lead to spare coefficient vector by setting some coefficient to be exactly zero. It is useful when you suspect the many features are irrelevant and to do automatic feature selection.

It can be suited well , when we have high-dimensional dataset and want to simplify the model by eliminating unimportant predictors. In the assignment model lasso has a slightly higher R2 score on

the test data, it indicates that lasso's feature selection capability might be more suitable for dataset. Reducing the impact of irrelevant predictors.

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: '1stFlrSF', '2ndFlrSF', 'OverallQual', 'OverallCond',
       '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?

Answer: The model should be as simple as possible, though its accuracy will decreased but it will be more robust and generalization. It can also be understand the use the bias – variance trade-off, The simpler model the move the bias but less variance and more generalizable.