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 1:

The optimal value of alpha for ridge and lasso regression is 10 and 100 respectively.

If we choose to double the value of alpha for both ridge and lasso, the penalty term of the model increases.

The most important predictor variables after the change is implemented is
Mean squared error

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 2:

I would choose lasso regression because even though there isn't much difference in the r2 score, the lasso regression makes the coefficients zero which helps in the feature selection whereas ridge regression doesn't make the coefficients zero.

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 3:

The next most important predictor variables will be

MSSubClass

OverallQual

KitchenQual

BsmtQual

LotShape

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 4:

We can make the model robust and generalisable by handling outliers, by regularization and not overfitting the data.

The implications of the same for the accuracy of the model are that generalising would have almost same accuracy for both test and training data whereas more robustness causes reduced accuracy in training data.