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?

Ans: The optimal value of alpha for Ridge is 2 and for Lasso it is 0.001. With these alphas the R2 of the model was approximately 0.86 for ridge and 86 for lasso. After doubling the alpha values in the Ridge and Lasso, the prediction accuracy for ridge remains around 0.86 but it dropped a bit and for the lasso it dropped to 0.84 but there is a small change in the co-efficient values.

Note: You can see the demonstration for this in the notebook

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?

The optimum lambda value in case of Ridge and Lasso is as follows:-

- Ridge – 2
- Lasso – 0.0001

The Mean Squared Error in case of Ridge and Lasso are:

- Ridge - 0.001497474525940253
- Lasso - 0.001672578205248682

The Mean Squared Error of both the models are almost same.

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?

Top 5 variables based on Lasso coefficients:
['Total_sqr_footage', 'MSZoning_C (all)', 'GarageArea', 'age', 'TotRmsAbvGrd']

Top 5 variables based on updated Lasso coefficients:
['LotArea', 'Neighborhood_MeadowV', 'BsmtFinType2_None', 'MSSubClass_30', 'LotFrontage']

MSE for original Lasso model: 0.0015439468939609576
R2 score for original Lasso model: 0.8557437952139043

MSE for updated Lasso model: 0.002188526931863625
R2 score for updated Lasso model: 0.7955184919263238

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?

To make a model more robust and generalisable, make the model simple but not over simplified. We can use the regularisation to make the model more simple and robust.

1. Applying cross check validation will help to generalise the model
2. We can use the feature selection to prevent the model begin over fitted and thus we can make the model more generalised not to rely on one particular variable
3. Complex model have more chance to make error

To increase the model accuracy we need to be more careful in making the model complex and generalisable. If the model is too complex can improve the accuracy but there is a risk of overfitting so we need to be careful while trading off with the accuracy and the generalisation Also there is a chance of being the model underfitting give more simple model could cause this