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:

Optimal value of alpha:

- Ridge – 0.3
- Lasso – 0.0001

R2_score for ridge regression

```
r2_train_ridge: 0.9329814513808519
r2_test_ridge: 0.8987681391856739
```

R2_score for Lasso regression

```
Lasso_train_r2: 0.9329814513808519
Lasso_test_r2: 0.8987681391856739
```

When we double the value of alpha which is:

Ridge: 0.6

Lasso: 0.0002

```
r2_train_ridge after doubling: 0.9321440399754538
r2_test_ridge after doubling: 0.8986651714848642
```

```
Lasso train r2 value after doubling: 0.9089919948933693
Lasso train r2 value after doubling: 0.8860313285075364
```

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:

I have got similar optimal value for both Ridge and Lasso which is 0.3 and 0.0001

But when I double the value of alpha r2_score of lasso got reduced for test data so I would choose lasso as we can still get a chance eliminate some of the featured which will not be helpful.

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:

After the top 5 predictor variables are dropped the next most important predictors will be

- MSSubClass
- OverallQual
- KirchenQual

- 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:

When the test r2_score is not much lesser than train score

When the model is not impacted by the outliers

It is very important to remove outliers and standardize the data

The predicted variables should be signification

This is when we say the model is robust and generalized

Implications:

1. Get more data as much as we can
2. Fixing null values, missing values and outliers
3. Newly derived columns wrt catefgorical variables
4. Cross validation