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:

R2 score with optimal value of alpha:

The optimal value of alpha for ridge is 5

```
Training set r2 score: 0.9536964080483874
Test set r2 score: 0.8835914704189932
```

The optimal value of alpha for Lasso is 100

```
Training set r2 score: 0.9438499881437824
Test set r2 score: 0.8901862148225264
```

R2 score with doubled value of alpha:

The optimal value of alpha for ridge is 10

```
Training set r2 score: 0.9467159463880559
Test set r2 score: 0.8818476728969055
```

The optimal value of alpha for Lasso is 200

```
Training set r2 score: 0.9319109102705608
Test set r2 score: 0.8865098998460733
```

```
By the r2 score we can conclude that after doubling the alpha level for
ridge and Lasso r2 score has slightly decreased.
```

The most important predictor variables after the change is implemented are-:

Positively correlated variables are-:

```
Total SF
GrLivArea
OverallQual_10
OverallQual_9
SaleType_New
Neighborhood_StoneBr
```

```
BsmtFinSF1
GarageCars
OverallQual_8
GarageArea
```

Negatively correlated variables are -:

```
OverallQual_4
KitchenQual_Fa
KitchenQual_Gd
OverallQual_5
BsmtQual_Gd
OverallCond_4
KitchenQual_TA
BsmtQual_TA
KitchenAbvGr
OverallCond_3
```

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:

As we can see from the above value that r2 score of Lasso for test set is greater than r2 score of test set. So I will choose Lasso regression

---

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 "

The five most important predictor variables excluding the five most important predictor variables are-:

```
Neighborhood_StoneBr ,BsmtFinSF1, GarageCars, OverallQual_8, GarageArea
```

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 simple model is robust in nature and when there is not much difference in test and training set accuracy and when the predicted variable is not insignificant then the model can be robust and generalisable.

The more the data, the better the accuracy and less chances of overfitting.When model is underfitted we use cross-validation for a good fit. Splitting the dataset in training and test set can be helpful to check the accuracy of model and see the model accuracy on unseen data, these were the few implications of accuracy of a model.
