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.Optimal Values for Ridge

a. Without RFE applied: 3

```
In [85]: # Check for the best hyperparameter alpha
         print(ridge_model_cv.best_params_)
         print(ridge_model_cv.best_score_)

         {'alpha': 3.0}
         -15025.801538827067
```

b. With RFE applied: 1

```
In [111]: # Printing the best hyperparameter alpha
          print(ridge_model_cv.best_params_)
          print(ridge_model_cv.best_score_)

          {'alpha': 1.0}
          -14901.606391050553
```

2.Optimal Values for Lasso

a. Without RFE applied: 100

```
In [89]: # Check for best hyperparameter alpha value.
         print(lasso_model_cv.best_params_)
         print(lasso_model_cv.best_score_)

         {'alpha': 100}
         -15037.475564276194
```

b. With RFE applied: 8

```
In [115]: print(lasso_model_cv.best_params_)
          print(lasso_model_cv.best_score_)

          {'alpha': 8.0}
          -15050.606664689452
```

In all generated models except Linear,

- R2 score on Training has decreased slightly whereas testing score has slight increase.
- After applying RFE, R2 score differences seems to be minimal in Lasso than compared to Ridge.

After change is implemented with RFE since there are more features to analyse we are considering RFE in suggesting most important features for model generation and most important predictor variables are:

GrLivArea
TotalBsmtSF
OverallQual
RoofStyle_Shed
Street_Pave
OverallCond
BsmtFinSF1
GarageArea
LotArea
YearBuilt

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 per generated R2 scores for Ridge and Lasso, would prefer to choose “Lasso” since difference between train and test related scores are slightly lower and it’s implicit way of handling feature(s) selection makes to create a simplified final model which is important for creating robust and generalized model.

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 regenerating model by dropping five most important predictor variables, below mentioned features can be considered as per “Lasso” score consideration on RFE related:

GrLivArea

RoofStyle_Shed

1stFlrSF

MSZoning_FV

SaleType_Con

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: In case of any model generation, the most important fact to be considered is not having test accuracy lesser than training score and more over generated test accuracy should not be yielding negative value which is termed as “Overfitting” meaning model has memorized data and not working well with unseen test data.
In order to overcome situation of generating less test accuracy we need to have model generation framework which will help in developing “Generalized Model” so that it’s behaviour can be accurate for datasets other than the ones which were used during training.
Robustness of a model is most important criteria for carrying out reliable predictive analysis and it’s not solely based on high test scores but also depends on train scores. Both scores need to be optimal which depends on business case and expectations of model. We need to consider both train and test scores so that model can perform well on unseen data which means that data we may need to retain some outliers which may help with predictions. As demonstrated in assignment accuracy of model will vary depending on way data is processed and how features are selected .There may be no perfect model but different steps are available to ensure developed model is fit for purpose of uniqueness of business case.