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?

Answers:

1a) Optimal Value of Alpha:

The below are the optimal values for Ridge and Lasso Regressions
Optimal alpha (lambda) value for Ridge Regression model is: 100
Optimal alpha (lambda) value for Lasso Regression model is: 0.01

1b) Changes in the model, if you choose double the value of alpha for both ridge and lasso regression:

Below are the comparisons of normal and doubled values alpha for Ridge and Lasso regression.

Ridge Regression: Original Model (alpha=100), Doubled Alpha Model (alpha=200)

```
For Ridge Regression (Original Model, alpha=100) Model Metrics
Tarining Data Metrics:
R2 Score:  0.8876409164998031
RSS Value:  18.032761253651422
MSE Value:  0.017661862148532246
RMSE Value:  0.132897938842302

Test Data Metrics:
R2 Score:  0.8584002871954429
RSS Value:  10.204888961395227
MSE Value:  0.023298833245194 58
RMSE Value:  0.15263955334445453
```

```
For Ridge Regression (Doubled Alpha Model, alpha=2*100=200) Model Metrics
Tarining Data Metrics:
R2 Score:  0.8831921178110138
RSS Value:  18.746758930753824
MSE Value:  0.018361174271061532
RMSE Value:  0.135503410551401

Test Data Metrics:
R2 Score:  0.8564896847906498
RSS Value:  10.342583346532855
MSE Value:  0.023613203987517935
RMSE Value:  0.15366588426686625
```

From the above snap shots below are the observations for Ridge Regression:
• The test accuracy of the ridge regression model (alpha=100) is slightly higher in comparison to the test accuracy of the doubled alpha model (doubled alpha=200).
• MSE test scores comparing similar data of the original dataset and doubled alpha model gives us an idea that it is slightly smaller for the single alpha model than the doubled alpha model.
• Ridge Regression model (single alpha model) seems to perform better on the train and test data in comparison to the doubled alpha Ridge Regression model.
• Increase in the value of alpha in the model lead to a decrease in R2 score but an increase in the MSE (causing more shrinkage of coefficient values). Thus, making the original (single) alpha model a better choice.

(ii) Lasso Regression: Original Model (alpha=0.01), Doubled Alpha Model (alpha=0.02)

```
For Lasso Regression (Original Model, alpha=0.01) Model Metrics
Tarining Data Metrics:
R2 Score:  0.8656033582678043
RSS Value:  21.569618389108392
MSE Value:  0.021125972957011158
RMSE Value:  0.14534776557281903

Test Data Metrics:
R2 Score:  0.852957805882874
RSS Value:  10.597120812499995
MSE Value:  0.024194339754566196
RMSE Value:  0.15554529807926112
```

```
For Lasso Regression (Doubled alpha model, alpha=2*0.01=0.02) Model Metrics
Tarining Data Metrics:
R2 Score:  0.8401048376851421
RSS Value:  25.661933132737083
MSE Value:  0.025134116682406546
RMSE Value:  0.15853742990980568

Test Data Metrics:
R2 Score:  0.8325081319326475
RSS Value:  12.070899592311772
MSE Value:  0.02755913148929628
RMSE Value:  0.1660094319287199
```

From the above snap shots below are the observations for Lasso Regression:
• The test accuracy of the lasso regression model (alpha=0.01) is higher in comparison to the test accuracy of the doubled alpha model (doubled alpha=0.02).
• MSE test scores comparing similar data of the original dataset and doubled alpha model gives us an idea that it is smaller for the single alpha model than the doubled alpha model.
• Lasso Regression model (single alpha model) seems to perform better on the train and test data in comparison to the doubled alpha Lasso Regression model.
• Increase in the value of alpha in the model lead to a decrease in R2 score but an increase in the MSE (causing more shrinkage of coefficient values). In Lasso, the insignificant coefficients that have their values near to 0 correspond to 0 values; performing feature selection in the model. Thus, making the original (single) alpha model a better choice.

(1c) The most important predictor variables after the change is implemented.

Below are the top 10 features which are contributing to the model are as follows:

- Ridge Regression Model (doubled alpha=200) important features are
  [**'OverallQual', 'GrLivArea', '1stFlrSF', 'GarageArea', 'YearRemodAdd', '2ndFlrSF', 'Neighborhood_Crawfor', 'Neighborhood_NridgHt', 'Neighborhood_MeadowV', 'Neighborhood_IDOTRR'**]

- Lasso Regression Model (doubled alpha=0.02) important features are
  [**'OverallQual', 'GrLivArea', 'GarageArea', 'YearBuilt', 'YearRemodAdd', '1stFlrSF', 'GarageType_Attchd', 'FireplaceQu_Not Available', 'Fireplaces','BsmtExposure_Gd'**]

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:

The Optimal Value of Alpha for Ridge and Lasso are,

- The computed optimal value of alpha for Ridge Regression (Original Model): 100
- The computed optimal value of alpha for Lasso Regression (Original Model): 0.01

Ridge and Lasso both the models have almost same test and train accuracy. So it can be said that there is no over fitting.

```
For Ridge Regression (Original Model, alpha=100) Model Metrics
Tarining Data Metrics:
R2 Score:  0.8876409164998031
RSS Value:  18.032761253651422
MSE Value:  0.017661862148532246
RMSE Value:  0.132897938842302

Test Data Metrics:
R2 Score:  0.8584002871954429
RSS Value:  10.204888961395227
MSE Value:  0.02329883324519458
RMSE Value:  0.15263955334445453
```

```
For Lasso Regression (Original Model, alpha=0.01) Model Metrics
Tarining Data Metrics:
R2 Score:  0.8656033582678043
RSS Value:  21.569618389108392
MSE Value:  0.021125972957011158
RMSE Value:  0.14534776557281903

Test Data Metrics:
R2 Score:  0.852957805882874
RSS Value:  10.597120812499995
MSE Value:  0.024194339754566196
RMSE Value:  0.15554529807926112
```

```python
# Checking no. of features in Ridge and Lasso models
lasso_coef = pd.Series(lasso.coef_, index= X_train.columns)
selected_features= len(lasso_coef[lasso_coef != 0])
print('Features selected by Lasso:', selected_features)
print('Features present in Ridge:', X_train.shape[1])
```

```
Features selected by Lasso: 45
Features present in Ridge: 123
```

Lasso and Ridge both have similar r2 score, RSS, MSE and RMSE on test dataset.
But Lasso has eliminated 78 features and final no. of features in Lasso Regression model is 45. Where Ridge has all 123 features. So, our Lasso model is simpler than Ridge with having similar r2 score, MAE and RMSE.

Considering above points we can choose our **Lasso Regression model** as our final 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:

Top five features in original Lasso Model (before removing) were as follows:

```
['OverallQual', 'GrLivArea', 'GarageArea', 'YearRemodAdd', 'YearBuilt']
```

Top five predictor variables in the new model are: (After removing the before mentioned top 5 predictors from the original lasso model):

```
['1stFlrSF', '2ndFlrSF', 'Neighborhood_NridgHt', 'KitchenQual_TA', 'FireplaceQu_No
t Available']
```

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:

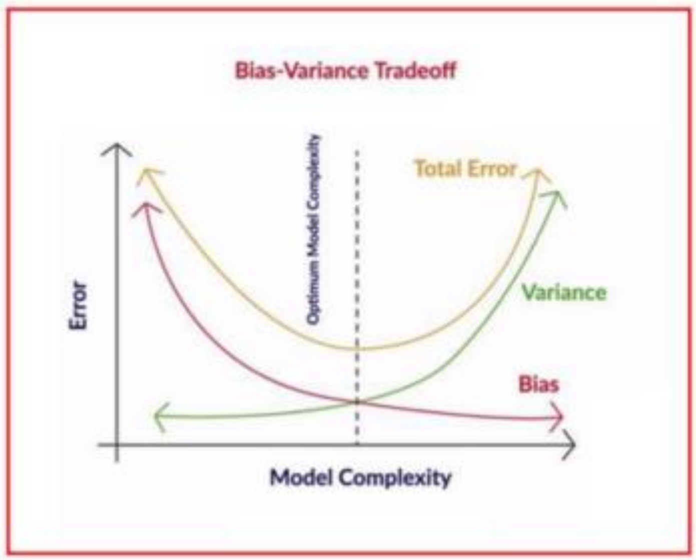
Robustness of a model implies, either the testing error of the model is consistent with the training error, the model performs well with enough stability even after adding some noise to the dataset. Thus, the robustness (or generalizability) of a model is a measure of its successful application to data sets other than the one used for training and testing.

By the implementing regularization techniques, we can control the trade-off between model complexity and bias which is directly connected the robustness of the model. Regularization helps in penalizing the coefficients for making the model too complex; thereby allowing only the optimal amount of complexity to the model. It helps in controlling the robustness of the model by making the model optimal simpler. Therefore, in order to make the model more robust and generalizable, one need to make sure that there is a delicate balance between keeping the model simple and not making it too naive to be of any use. Also, making a model simple leads to BiasVariance Trade-off:

- A complex model will need to change for every little change in the dataset and hence is very unstable and extremely sensitive to any changes in the training data.
- A simpler model that abstracts out some pattern followed by the data points given is unlikely to change wildly even if more points are added or removed.

Bias helps you quantify, how accurate is the model likely to be on test data. A complex model can do an accurate job prediction provided there has to be enough training data. Models that are too naïve, for e.g., one that gives same results for all test inputs and makes no discrimination whatsoever has a very large bias as its expected error across all test inputs are very high. Variance is the degree of changes in the model itself with respect to changes in the training data.

Thus, accuracy of the model can be maintained by keeping the balance between Bias and Variance as it minimizes the total error as shown in the below graph.



Thus, accuracy and robustness may be at the odds to each other as too much accurate model can be prey to over fitting hence it can be too much accurate on train data but fails when it faces the actual data or vice versa.