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:

a) **Optimal values of alpha for**

Ridge regression: 10.21
Lasso regression: 0.001

b) & c) If alpha values are doubled following are the changes

- **Ridge regression:**

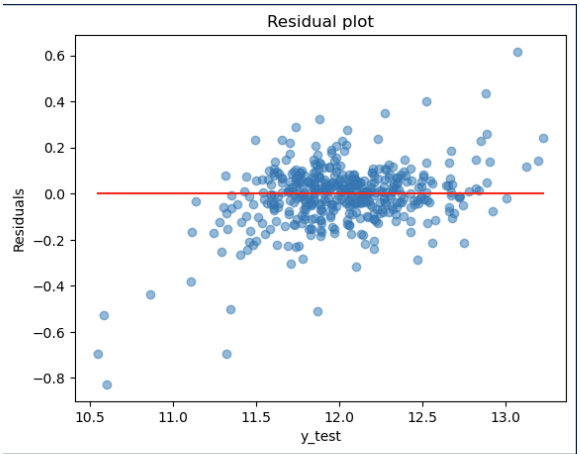
i. **R2-RMSE-MSE:**

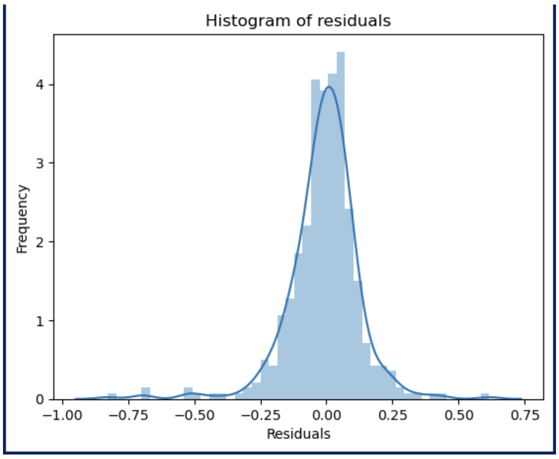
R2 score of train data: 0.9303370373667955
RMSE of train data: 0.10668110703901111
MSE of train data: 0.011380858599068946

R2 score of test data: 0.877137681990487
RMSE of test data: 0.13598510033250516
MSE of test data: 0.018491947512441494

ii. **Residual plot & error variance:**

No Changes observed in the graphs





iii. **Most important Predictor variables:**

Features	Coefficient
OverallCond_3	-0.0782
Functional_Typ	0.0706
Neighborhood_Crawfor	0.0686
OverallQual_9	0.0667
CentralAir_Y	0.0658
Neighborhood_Edwards	-0.0611
OverallCond_4	-0.0574
OverallCond_7	0.0564
Exterior1st_BrkFace	0.0564
OverallCond_8	0.0511

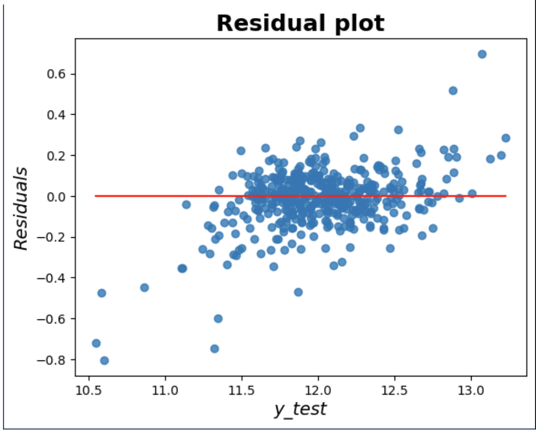
- **Lasso Regression**

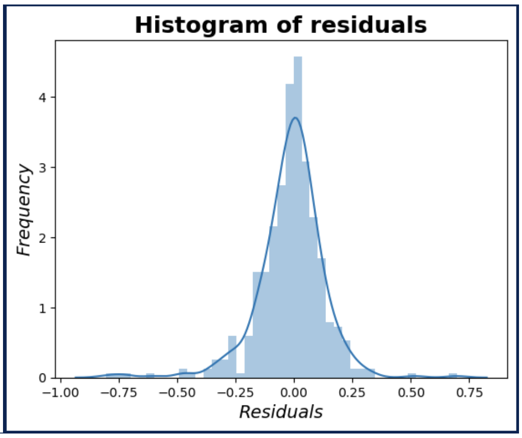
    i. **R2-RMSE-MSE**

    R2 score of train data: 0.9050246565684178
    RMSE of train data: 0.12456383769073949
    MSE of train data: 0.015516149660244893

    R2 score of test data: 0.8601026161101185
    RMSE of test data: 0.14510646768146723
    MSE of test data: 0.021055886962992695

    ii. **Residual Plot & Error variance**





    iii. **Most important Predictor variables:**

Features	Coefficient
OverallCond_3	-0.0911
OverallQual_9	0.0856
CentralAir_Y	0.0788
Neighborhood_Crawfor	0.0745
Functional_Typ	0.0731
OverallCond_4	-0.0721
Neighborhood_Edwards	-0.0447
Condition1_Norm	0.0443
OverallQual_8	0.0408
Exterior1st_BrkFace	0.0397

Question2
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?
Ans:
Optimal Value of Alpha:

- The computed optimal value of alpha for Ridge Regression (Original Model): 10.21
- The computed optimal value of alpha for Lasso Regression (Original Model): 0.001

Ridge Regression

```
R2 score of train data: 0.9303370373667955
RMSE of train data: 0.10668110703901111
MSE of train data: 0.011380858599068946
R2 score of test data: 0.877137681990487
RMSE of test data: 0.13598510033250516
MSE of test data: 0.018491947512441494
```

Lasso Regression

```
R2 score of train data: 0.9050246565684178
RMSE of train data: 0.12456383769073949
MSE of train data: 0.015516149660244893
R2 score of test data: 0.8601026161101185
RMSE of test data: 0.14510646768146723
MSE of test data: 0.021055886962992695
```

- The R2 test score on the Ridge Regression Model is slightly better than that of Lasso Regression Model.
- The MSE for Test set (Ridge Regression) is slightly lower than that of the Ridge Regression Model; implies Ridge Regression performs better on the unseen test data. However, Lasso helps in feature selection (the coefficient values of some of the insignificant predictor variables became 0), implies Lasso Regression has a better edge over Ridge Regression.
- Going by the results of both lasso and ridge for this data model, ridge performed better than lasso, however, while choosing a type of regression in the real world, an analyst has to deal with the lurking and confounding dangers of outliers, non-normality of errors and overfitting especially in sparse datasets among others. Using L2 norm (Ridge) results in exposing the analyst to such risks. Hence, use of L1 norm (Lasso) could be quite beneficial as it is quite robust to fend off such risks to a large extent, thereby resulting in better and robust regression models.

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?

Ans: The five most important predictor variables in the current lasso model is:-

1. OverallCond_3
2. OverallQual_9
3. Neighborhood_Crawfor
4. OverallCond_4
5. CentralAir_Y

We build a Lasso model in the Jupiter notebook after removing these attributes from the dataset.
The R2 of the new model without the top 5 predictors drops to .867
The Mean Squared Error increases to 0.0202

The new Top 5 predictors are:-

Features	Coefficient
OverallCond_9	0.1160
OverallCond_7	0.1123
OverallCond_8	0.1100
Functional_Typ	0.0924
Exterior1st_BrkFace	0.0842

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?

Ans:

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.