Surprise Housing_ Advanced Regression _Assignment

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:

Alpha for ridge: 500

Alpha for Lasso: 0.01

R2 value is decreased

GrLivArea is the important predictor variable

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:

Will choose the lasso regression for alpha 0.01, since the test R2 value is better than ridge

lasso regression Test R2: 0.88

ridge regression Test R2: 0.86

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:

Below are the 5 important variables:

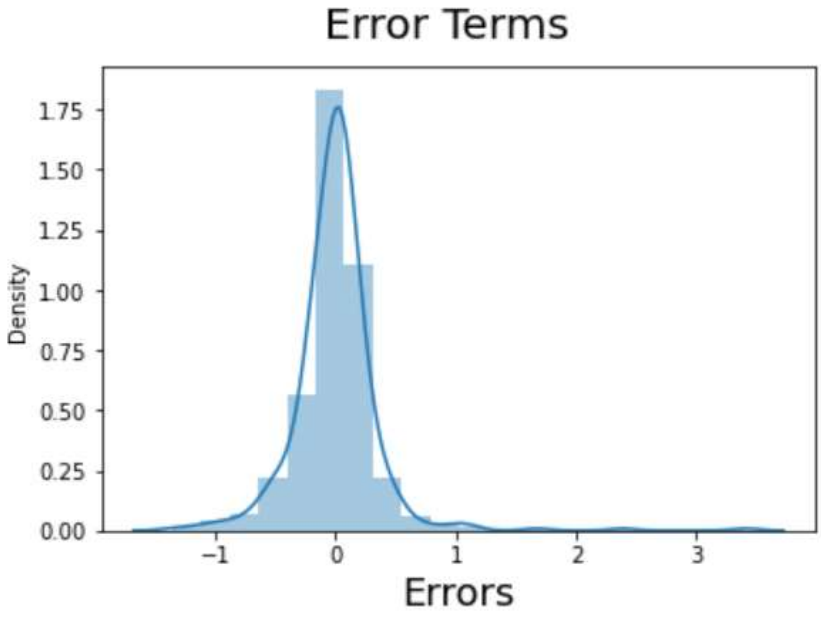
```
In [764]: model['Lasso'].sort_values(ascending = False)[:5]

Out[764]: 2ndFlrSF               1.303245
          1stFlrSF               1.218628
          TotalBsmtSF            1.205612
          Neighborhood_NridgHt   1.100284
          Neighborhood_NoRidge   1.092562
          Name: Lasso, dtype: float64
```

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?

The model is robust and generalisable since the residuals are equally distributed across the 0 axis and error terms are normally distributed

```
Out[735]: Text(0.5, 0, 'Errors')
```



Out[736]: Text(0.5, 0, 'Errors')