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

The optimal value of alpha for ridge and lasso regression

Ridge Alpha 0.5

0.8248066133645755

0.8363338738464898

lasso Alpha 20

0.8249524660059517

0.8371286945504668

After doubling the alpha values

Ridge: Alpha 1

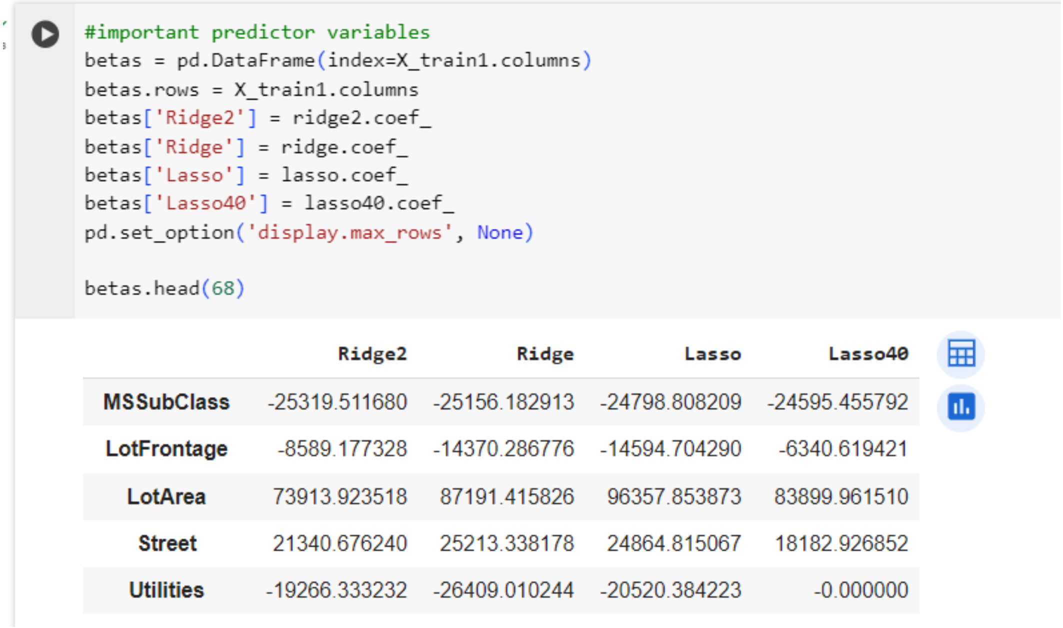
R2-train:0.8239978593545518
R2-Test:0.8353184712938473

Lasso Alpha 40

R2-train: 0.8240959318030928
R2-Test: 0.8367039606712412

R2score on training, testing data has decreased

And coefficients of predictors changed.



```
#important predictor variables
betas = pd.DataFrame(index=X_train1.columns)
betas.rows = X_train1.columns
betas['Ridge2'] = ridge2.coef_
betas['Ridge'] = ridge.coef_
betas['Lasso'] = lasso.coef_
betas['Lasso40'] = lasso40.coef_
pd.set_option('display.max_rows', None)

betas.head(68)
```

	Ridge2	Ridge	Lasso	Lasso40
MSSubClass	-25319.511680	-25156.182913	-24798.808209	-24595.455792
LotFrontage	-8589.177328	-14370.286776	-14594.704290	-6340.619421
LotArea	73913.923518	87191.415826	96357.853873	83899.961510
Street	21340.676240	25213.338178	24864.815067	18182.926852
Utilities	-19266.333232	-26409.010244	-20520.384223	-0.000000

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:

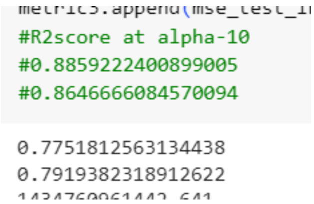
The r2_score of lasso is slightly higher than ridge for the test dataset so we will choose lasso regression to solve this problem

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?

Earlier 5 important predictors are ‘LotArea','OverallQual', ‘1stFlrSF’,’ GrLivArea’,’ MasVnrArea’

Now after dropping these from train and test, R2score of training and testing data has decreased



```
metrics.append(mse_test_lr
#R2score at alpha-10
#0.8859222400899005
#0.8646666084570094
```

```
0.7751812563134438
0.7919382318912622
1434760961442.641
```

New five most important predictor variables

1. TotalBsmtSF
2. 2ndFlrSF
3. RoofMatl
4. GarageCars
5. FullBath

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 model should generalize well to maintain similar accuracy on both training and test datasets, while minimizing the impact of outliers. Outliers should be analyzed and only relevant ones retained, while those not meaningful to the dataset should be removed. A robust model is essential for reliable predictive analysis.