

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- The optimal value of alpha for ridge regression is 20.0 and lasso regression is 0.01.

In our case in ridge regression there is minor change in r2\_score. It got decrease.

```
#finalising ridge regression with optimal alpha value
ridge = Ridge(alpha=40)
ridge.fit(X_train,y_train)

y_train_pred = ridge.predict(X_train)
y_test_pred = ridge.predict(X_test)

print(r2_score(y_true=y_train,y_pred=y_train_pred))
print(r2_score(y_true=y_test,y_pred=y_test_pred))
```

```
0.8798061568805683
```

```
0.8730429012759318
```

Same case in Lasso regression

```
#finalising lasso regression with optimal alpha value
lasso = Lasso(alpha=0.02)
lasso.fit(X_train,y_train)

y_train_pred = lasso.predict(X_train)
y_test_pred = lasso.predict(X_test)

print(r2_score(y_true=y_train,y_pred=y_train_pred))
print(r2_score(y_true=y_test,y_pred=y_test_pred))
```

```
0.866289733619241
```

```
0.8685628998012738
```

After this change in both ridge and lasso regression the most important predictor is same as earlier MSSubClass.

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- In the assignment we have used Lasso Regression as our final model because it out performs Ridge regression in test data set. Although lasso computation is more as compare to ridge but it also performs feature reduction which would be one of the best point.

The main difference in Ridge and Lasso regression is Ridge add sum of squares in the coefficient and Lasso add absolute values in the coefficient.

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- After excluding those variable we top 5 variables are as follows:

39	GarageQual	0.069431
3	LandContour	0.059678
36	GarageType	0.042503
24	Heating	0.041765
35	Fireplaces	0.032029

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- Model can be made robust and generalizable in many ways:

- Check for Overfitting and Underfitting:

Overfitting is the scenario when the model perform well on training dataset but does not from on test dataset. Due to this issue, the model fails to generalize the other data points. This situation occurs when model learn noise from the dataset.

We can remove Overfitting issue with cross validation techniques.

Underfitting is the scenario when the model does not perform well on training dataset.

We can remove the Underfitting by adding more data points.

- Check for outlier: If there are outliers present in the data points, then we need to perform some operation to reduce or remove those data points. Outliers can be deducted with the help of box plot.
- Check for bias and variance for the model  
Variance: How sensitive the model is to input dataset.  
Bias: How much error the model likely to make in the test data.  
We can not remove both the error, therefore we have to choose optimal point where both are compromised. This is called as Bias and Variance Trade Off  
The implication of the same for the model accuracy will increase because our model would be more robust and generalizable.

