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 for ridge and lasso regression are 50 and 0.001.
If we double or increase alpha value, there will be a decrease in Model's output. That is the reason we first look for optimal value to alpha before making predictions.
We should look for predictors post change because overall significance would already be ready reduced. Yes, we can do it to find out variable importance if change of alpha is applied.

[800]:
```
Ridge
Ridge(alpha=100)
```

[801]:
```
## Make predictions
y_train_pred = ridge.predict(X_train)
y_pred = ridge.predict(X_test)
```

[802]:
```
## Check metrics
ridge_metrics = compute_and_display_metrics(y_train, y_train_pred, y_test, y_pred,ridge_metrics)
```

```
R2 Train : 0.9123091311492253
R2 Test : 0.8798749650038801
RSS Train : 11.174919988931812
RSS Test : 6.470821797374479
MSE Train : 0.010945073446554175
MSE. Test : 0.014773565747430318
```

[803]:
```
Lasso
Lasso(alpha=0.002)
```

[804]:
```
## Make predictions
y_train_pred = lasso.predict(X_train)
y_pred = lasso.predict(X_test)
```

[805]:
```
## Check metrics
lasso_metrics = compute_and_display_metrics(y_train, y_train_pred, y_test, y_pred,lasso_metrics)
```

```
R2 Train : 0.91262594840633
R2 Test : 0.8801764776682574
RSS Train : 11.134546258512009
RSS Test : 6.454580097874455
MSE Train : 0.010905530125868765
MSE. Test : 0.0147364842417225
```

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

My preference would be the optimal value of Alpha in Ridge because when we compared the both model output, statistics of Ridge been on the higher side.
Also, with Lasso because of its behavior, it zeroes more features compared to Ridge which reduces the opportunity areas for the business.

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

After dropping our top 5 lasso predictors, we get the following new top 5 predictors:-

GrLivArea 0.11
OverallQual 0.05
TotalBsmtSF 0.05
YearBuilt 0.04
OverallCond 0.03

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

A model is **robust** when any variation in the data does not affect its performance much.
A **generalizable** model is able to adapt properly to new, previously unseen data, drawn from the same distribution as the one used to create the model.
To make sure a model is robust and generalizable, we have to **take care it doesn't overfit**. This is because an overfitting model has very high variance and a smallest change in data affects the model prediction heavily. Such a model will identify all the patterns of a training data, but fail to pick up the patterns in unseen test data.
In other words, the model should not be too complex in order to be robust and generalizable.

- If we look at it from the prespective of **Accuracy**, a too complex model will have a very high accuracy. So, to make our model more robust and generalizable, we will have to decrease variance which will lead to some bias. Addition of bias means that accuracy will decrease.
- In general, we have to find strike some balance between model accuracy and complexity. This can be achieved by Regularization techniques like Ridge Regression and Lasso.