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?

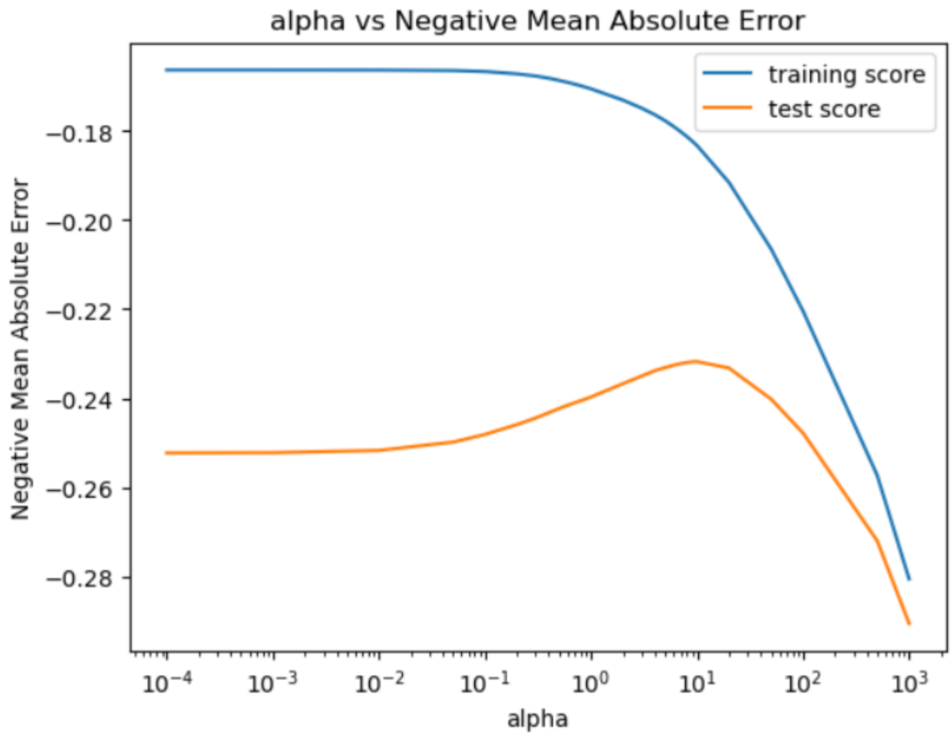
Optimal Alpha Value:

When we have the alpha as zero differentiating the cost function would result into the minimal RSS. As we keep alpha between zero to infinity the error on training dataset they will steadily increase.

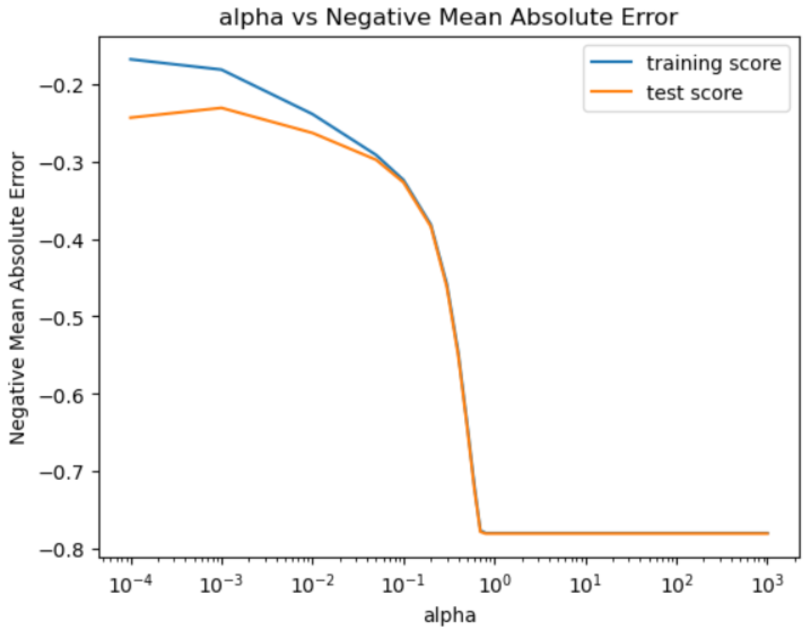
When keep alpha as Zero it doesn't generalize the model so performance on test data will bad. When keep alpha as infinity it will underfit the model which is too simple which would again reduce performance on test data will bad

So, as we move the alpha from 0 to infinity the error on test data will decrease due generalization and regularization. Once it crosses the optimal it will start to increase again

```
#ridge alpha plot
plot_cv_results(model_cv_ridge)
```



```
In [114]: #lasso alpha calculation
          #inorder to find the optimal alpha we need to plot alpha againt the negative mean absolute error
          plot_cv_results(model_cv)
```



Based on the above two charts. Optimal alpha values are

1. Ridge Regression – 10
2. Lasso Regression – 0.001

Changes When we double Alpha

Metric	Ridge Regression	2alpha - Ridge Regression	Ridge Change	Lasso Regression	2alpha- Lasso Regression	Lasso Change
R2 Score (Train)	0.919752	0.912025	Decrease	0.921294	0.910552	Decrease
R2 Score (Test)	0.862683	0.856244	Decrease	0.866495	0.857844	Decrease
RSS (Train)	81.93346	89.82203	Increase	80.35911	91.326012	Increase
RSS (Test)	55.41009	58.00843	Increase	53.87202	57.362585	Increase
MSE (Train)	55.41009	58.00843	Increase	53.87202	57.362585	Increase
MSE (Test)	0.126507	0.132439	Increase	0.122995	0.130965	Increase

When we double the alpha r2 tends to decrease and RSS and MSE increase on both ridge and lasso

We can observe it on both test and training data set.

This is because model starts to underfit once will cross the optimal point

Top Predictors on doubling the Alpha

ridge regression			2alpha Ridge	
MSSubClass	-0.600946		MSSubClass	-0.54657
OverallCond_Good	-0.329487		OverallCond_Good	-0.243018
OverallQual_Fair	0.316185		OverallQual_Fair	0.242213
Neighborhood_Edwards	0.270468		Neighborhood_Edwards	0.215544
Neighborhood_OldTown	0.225742		Neighborhood_OldTown	0.187536

In ridge regression only the co-efficient penalized and co-efficient value reduced we can see the same top 5 predictors

lasso			2alpha lasso	
MSSubClass	-0.739228		MSSubClass	-0.570741
OverallCond_Good	-0.516687		OverallCond_Good	-0.502002
OverallQual_Fair	0.495732		OverallQual_Fair	0.448749
OverallQual_Very Exceller	-0.397536		Neighborhood_Edwards	0.31714
Neighborhood_Edwards	0.350938		BsmtFullBath	0.286159

In lasso we can observe change in co-efficient and one new top predictor. Overall Quality very excellent not in top predictors any more

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?

When we are choosing the right model, we can use model selection principle

1. Keep Model Simple but not too naïve
2. Strike the balance between Bias vs Variance
3. The model with Lesser RMSE

First two points can be achieved by regularization done by both ridge and lasso. Lasso having one additional advantage of feature selection. As the business objective is to identify the variables influence the price. we can choose lasso for this assignment. But if there is a business need to keep all features then we can go with ridge.

On third point we can see the results below based on that lasso performs better on all category

Out[119]:

	Metric	Linear Regression	Ridge Regression	Lasso Regression
0	R2 Score (Train)	9.353249e-01	0.919752	0.921294
1	R2 Score (Test)	-3.926914e+17	0.862683	0.866495
2	RSS (Train)	6.603327e+01	81.933463	80.359109
3	RSS (Test)	1.584587e+20	55.410092	53.872024
4	MSE (Train)	1.584587e+20	55.410092	53.872024
5	MSE (Test)	3.617779e+17	0.126507	0.122995

1. Lasso seems to have better r2 and MSE
2. with lasso we also have addition feature selection

We can choose and apply lasso model

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?

The Top predictors of the lasso model are

MSSubClass	-0.739228
OverallCond_Good	-0.516687
OverallQual_Fair	0.495732
OverallQual_Very Excellent	-0.397536
Neighborhood_Edwards	0.350938

On dropping the top five predictors and rebuilding the model we get following five as the top predictors

LotArea	-0.767325
OverallCond_Poor	-0.59314
OverallQual_Good	0.433092
Neighborhood_Gilbert	0.401882
MSZoning_RL	0.328118

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?

Bias is the difference between actual value and the predicted value. When the bias is very low means model has memorized all the datapoints in the training data set and failed to generalize. When the bias is varying high mean model hasn't learnt anything from training dataset or simplified the model too much

Variance is how well it can explain the spread of the data. When variance is high it will create a complex curve to fit across all training data point causing overfitting which result less bias on training data. But the moment we give unseen data the model will fail to fit

So when we want to create robust and generalisable model, it should be simple model at the same time explain the most of the variance in the data

In order to create robust and generalisable we need to strike a balance between bias and variance

We can achieve it by using regularization involving cross fold validation and hyper parameter tuning

In terms of implication on the accuracy it will less accurate on the training dataset but it will perform better on the unseen data with good accuracy