

# 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?

## Ridge Regression

⇒ Optimal Alpha Value:

```
1 # Printing the best hyperparameter alpha
2 print(model_cv.best_params_)
3 print(model_cv.best_score_)
```

```
{'alpha': 10.0}
-0.07770378415038273
```

⇒ Important Variables for Optimal value

	Attribute	Importance
50	Neighborhood_Crawfor	0.095178
47	Neighborhood_BrkSide	0.064943
136	Functional_Typ	0.060846
34	MSZoning_RL	0.060625
126	CentralAir_Y	0.056995

After double the value:

⇒ Important Variables after doubling alpha value

	Attribute	Importance
50	Neighborhood_Crawfor	0.076134
2	OverallQual	0.057103
47	Neighborhood_BrkSide	0.052765
136	Functional_Typ	0.051520
126	CentralAir_Y	0.046809

## Lasso Regression

⇒ Optimal Alpha Value:

```
: 1 # Printing the best hyperparameter alpha
  2 print(model_cv.best_params_)
  3 print(model_cv.best_score_)

{'alpha': 0.001}
-0.07919508847776727
```

⇒ Important Variables for Optimal value

	Attribute	Importance
50	Neighborhood_Crawfor	0.111993
150	SaleType_New	0.099590
47	Neighborhood_BrkSide	0.079684
2	OverallQual	0.060892
65	Neighborhood_Somerst	0.056038

After double the value:

⇒ Important Variables after doubling alpha value

	Attribute	Importance
50	Neighborhood_Crawfor	0.087334
150	SaleType_New	0.073501
2	OverallQual	0.066496
47	Neighborhood_BrkSide	0.051863
3	OverallCond	0.045531

For Optimal Value				Doubling the Optimal Value			
	Metric	Ridge Regression	Lasso Regression		Metric	Ridge Regression	Lasso Regression
0	R2 Score (Train)	0.935020	0.924053	0	R2 Score (Train)	0.931093	0.914076
1	R2 Score (Test)	0.916304	0.913655	1	R2 Score (Test)	0.917179	0.911204
2	RSS (Train)	8.645909	10.105078	2	RSS (Train)	9.168460	11.432647
3	RSS (Test)	5.684647	5.864580	3	RSS (Test)	5.625191	6.031025
4	MSE (Train)	0.095199	0.102919	4	MSE (Train)	0.098033	0.109471
5	MSE (Test)	0.117750	0.119599	5	MSE (Test)	0.117132	0.121284

## 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?

- ⇒ Ridge Regression can be chosen since it has better predictive (R2 Score) comparing with lasso regression. And, Ridge regression works well with large set of variables.

## 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?

- ⇒ Below are the five important predictors after dropping the five important variables,

	Attribute	Importance
132	Functional_Typ	0.085495
151	SaleCondition_Partial	0.082544
31	MSZoning_FV	0.060104
122	CentralAir_Y	0.055857
2	OverallCond	0.051190

## 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?

- ⇒ Model can be made robust by doing a proper outlier treatment and removing the outliers. If the outliers are not treated the predictive power of the model reduces which affects the accuracy