

## 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?

**Lasso:** From the chart which we built in Lasso, we can see that the Negative Mean Absolute Error is low at  $\alpha = 0.4$  and stabilises after that but we have to choose a low value of alpha to balance the trade-off between Bias-Variance and to get the small coefficients of features. And when we kept our  $\alpha=0.01$ , then we observe that when the alpha was increased, it will try to penalize the model and it sets the irrelevant i.e. most of the coefficients to zero.

**Ridge:** From the chart we can see that test negative mean absolute error is increasing with increase in alpha. Negative Mean Absolute Error stabilizes at  $\alpha = 2$

When we double the alpha value in both ridge and lasso techniques, absolutely alpha will increase the Negative Mean Absolute Error which will eventually penalize our both train and test models.

Most Important Features in Lasso:

1	OverallQual
2	GrLivArea
3	OverallCond
4	GarageArea
5	BsmtFullBath
6	Fireplaces
7	FullBath
8	TotalBsmtSF
9	LotArea
10	MSZoning_RL
11	WoodDeckSF
12	ScreenPorch
13	1stFlrSF
14	BsmtFinSF1
15	HalfBath
16	KitchenAbvGr
17	MSSubClass
18	PoolArea
19	PropertyAgeinYears

Ridge features:

1	MSZoning_RH
2	MSZoning_FV

3	MSZoning_RL
4	SaleType_Oth
5	Neighborhood_Crawfor
6	Neighborhood_StoneBr
7	Exterior1st_BrkFace
8	Foundation_Stone
9	Exterior2nd_CmentBd
10	MSZoning_RM
11	SaleType_New
12	Neighborhood_NridgHt
13	OverallQual
14	Neighborhood_NoRidge
15	RoofStyle_Mansard
16	SaleCondition_Alloca
17	CentralAir_Y
18	GrLivArea
19	SaleCondition_Normal

## 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?

If we have too many variables and we want to immediately take out the important features then we will choose Lasso regression and if we don't want to get too many coefficients and we want reduce the magnitude of those coefficients then we can use Ridge. Here in our case study our focus was to get the features so we have used Lasso for the same.

## 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 Features will be:

1	OverallQual
2	GrLivArea
3	OverallCond
4	GarageArea
5	BsmtFullBath

#### **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?

To make our model robust, we have maintain balance between the bias and variance. Bias is the error in mode. It may underfit the data of the model. High bias will impact the model performance.

High variance is also error in model and the model will try to over learn from the data. It may over fit the data.