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 of alpha for ridge =100 and lasso regression is 0.01

On doubling the alpha value ridge =200 and lasso regression is 0.02

Best Predictor Variable is "Neighborhood" and changing the Alpha value has minimal effect, as the initial Alpha value is 100 and doubling does not change much and is incremental change.

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: As learnt during the session Lasso is better than ridge, as Lasso allows to select the features. In the assignment, it is evident that Lasso has better.

R2 Train for Ridge is .80 and Lasso is .83

R2 Test for Ridge is .72 and Lasso is .75

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: Most important predictor variables are –

Neighborhood_NoRidge,

Neighborhood_NridgHt,

Neighborhood_StoneBr,

RoofMatl_WdShngl

,Exterior1st_BrkFace

Dropping these Predictor Variables and again calculating following are the Top5 P
redictor Variables –

1. 'SaleCondition_AdjLand',
2. 'SaleCondition_Alloca',
3. 'SaleCondition_Family',
4. 'SaleCondition_Normal',
5. 'SaleCondition_Partial'

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 generalisable by choosing the appropriate Hyper parameter.
By doing this, we strike a balance between variance and bias and choosing the
optimum accuracy.

Answer: In the assignment case, with CV and Lasso, we were able to identify the
optimum parameters using appropriate value of alpha. As per the Scatter and
Distplot –

Scatter plot: The residuals are distributed across the X axis and no trend is seen
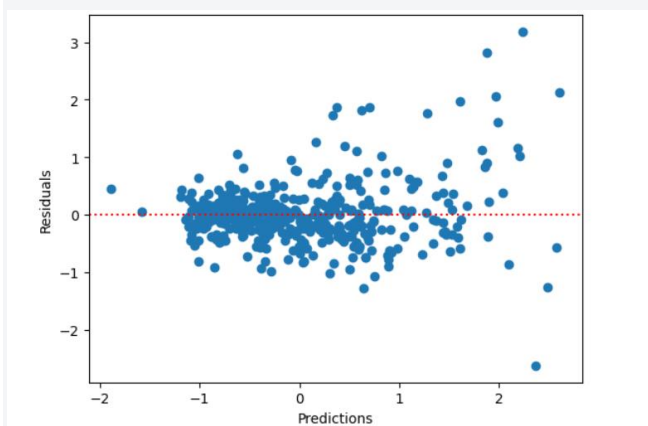
Distplot :Show normal distribution across 0

Normality of error terms/ residuals