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 value for ridge regression obtained through hyperparameter tuning equals to .99 and for lasso regression , the value is 0.0001.

As tested, There will not be significant changes in the current model even if we double the alpha values.

However theoretically, these values when increases can make the model simpler and increase training error.

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:

In this scenario, we will go ahead with lasso regression because it automatically takes care of feature reduction and helps in deciding the better parameters for the prediction and eventual the overall decision making. It tends to work on multi collinear data.

Lasso regression is an effective regularisation method. It helps to reduce the overfitting problem and is highly useful when there are high number of independent variables.

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 next top 5 variables are :

0   BsmtExposureGd

1   MasVnrTypeBrkFace

2 NeighborhoodNridgHt

3   3SsnPorch140

4   OverallQual9

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 ensure that the model is robust and generalisable, we need to check on the following:

☐   The gap between train and test errors are minimal

☐   We need to regularise/add penalty term to avoid concerns of overfitting

To ensure that the assumptions of linear regression are held true , we need to plot residual plots. The errors should be scattered around 0 and should follow no pattern.

     Conduct data transformations if needed.

     The effect on accuracy can be seen as in terms to reduce the variance, we might have to sacrifice on accuracy.