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 values for alpha that I have got are as below,

For Ridge Regression – The value is 0.3

For Lasso Regression – The value is 0.0001

If the value of alpha is doubled the values I got for both regressions is as below,

For Ridge Regression – The value is 0.6

For Lasso Regression – The value is 0.0002

Comparing the Ridge and Lasso after double the vlaues of alpha

```
#Comparing results of Ridge and Lasso

resultTable = {'Metric':["R2 Score Train", "R2Score Test", "RSS Train", "RSS Test", "MSE Train", "MSE Test"],
               'Ridge regression':metric4}
rg_metric = pd.DataFrame(resultTable, columns=["Metric", 'Ridge regression'])
ls_metric = pd.Series(metric5, name ='Lasso regression')
final = pd.concat([rg_metric,ls_metric],axis=1)
final
```

33]

	Metric	Ridge regression	Lasso regression
0	R2 Score Train	0.922869	0.903720
1	R2Score Test	0.891004	0.881357
2	RSS Train	10.720316	13.381783
3	RSS Test	7.205239	7.842954
4	MSE Train	0.011052	0.013796
5	MSE Test	0.017320	0.018853

For lasso and ridge, the values that I got are very similar.

There is some noticeable change in the values of R2 Score test and train after the change is implemented.

The important variables after the change is implemented are as below,

- MSZoning_FV

- MSZoning_RL

- GrLivArea

- OverallQual

- TotalBsmtSF

- Neighborhood_Crawfor

- Foundation_PConc

- Neighborhood_NridgHt

- SaleCondition_Normal

- GarageCars

(The coding part is mentioned in the python file attached in the link)

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:

For the alpha values if I Compare, I have got mostly similar values in case of ridge and lasso regression.,

But the ridge regression does not zero any of the coefficients for the lambda values whereas lasso has one or two zero coefficient values .

The zero values could help me in elimination of some features for further analysis and build models and train the dataframe.

Hence I would chose lasso.

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 :

The five most important predictor variables that I have eliminated are

"MSZoning_FV",

"GrLivArea"

"MSZoning_RL",

"OverallQual"

"Foundation_PConc"

(The coding part is mentioned in the python file attached in the link)

The five most important predictor variables now are,

Overall condition

Lot area

Shape

Condition1

IsRemodeled.

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?

Answer:

The model is robust and generalisable when ,

- Test accuracy and training score are on similar lines. Test accuracy cannot be much lesser than training score.
- The model should not be changeable or impacted by outliers
- The variables that are used for prediction must be significant.
- Overall a simple model can be more robust.

Implications of the accuracy of the model

- Huge Data : The more amount of data, the more easy and better to train the dataset.
- Missing Values and Outliers : Fix the missing values and outliers as this can lead to inaccurate model.

- Value Standardizing : We can use the existing values and get inferred references to get more accurate model.
- Feature Selection : Having knowledge on domain should help in deciding the important and relevant features
- Choosing Algorithm : Applying the right algorithm based on scenario helps in accuracy
- Validation