

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:

Best alpha value for Lasso: {'alpha': 0.01}

R2 Score:

0.8372319420231107

0.8368795587133783

Below Features are best explaining the DataSet:

BsmtFullBath 0.346320

OverallCond 0.242032

Neighborhood_OldTown 0.209550

Neighborhood_NridgHt 0.205890

TotRmsAbvGrd 0.103046

Best alpha value for Ridge: {'alpha':20.0}

R2 Score:

0.9342468536658213

0.814596382108107

Below Features are best explaining the DataSet:

Exterior1st_AsphShn 9.744117

RoofMatl_Metal 9.654680

RoofMatl_Roll 8.972255

RoofMatl_WdShake 8.893163

RoofMatl_Membran 8.748567

Double alpha value for Lasso: {'alpha':0.02 }

R2 score for train and test:

0.8812564101198541

0.8591211119605175

Below Features are best explaining the DataSet:

Exterior1st_AsphShn 0.752243

Neighborhood_NridgHt 0.489745

Neighborhood_OldTown 0.461670

BsmtFullBath 0.330302

Neighborhood_Edwards 0.232571

Double alpha value for Ridge: {'alpha':40.0}

0.8640761878601945

0.8593581699411202

Features for this ridge value:

OverallCond 0.200755

Neighborhood_NridgHt 0.199647

Neighborhood_OldTown 0.197090

BsmtFullBath 0.182501

LowQualFinSF 0.143748

Answer:

Now, most important predictor variables after the change is implemented is:

Exterior1st_AsphShn 0.752243

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:

After creating model in both Ridge and Lasso we can see that the r2_scores are almost same.

Lasso regression is final model, as it penalize more on the dataset and does feature elimination

Best alpha value for Lasso: {'alpha':0.01 }

R2 Score:

0.8372319420231107

0.8368795587133783

Best alpha value for Ridge: {'alpha':20.0}

R2 Score:

0.9342468536658213

0.814596382108107

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:

Next 5 features, if above 5 not available in Input data

Exterior1st_AsphShn 0.752243

Neighborhood_NridgHt 0.489745

Neighborhood_OldTown 0.461670

BsmtFullBath 0.330302

Neighborhood_Edwards 0.232571

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:

If we manage outliers and dependent variable even if input variable varied insignificantly we can make it robust.

Robust model impacts the accuracy of the model. Robust and generalize model is better than accurate model as we have to apply it on unseen data.