

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:

Optimal Value for Ridge Regression Model : 10

Optimal Value for Lasso Regression Model : 0.001

Changes after doubling alpha:

There is no significant changes in the models only r^2_{score} is varying little by

0.9220000960179736

0.9026465663858932, when optimal value for ridge is 20

0.9292504648912956

0.9049835162192809, when optimal value for ridge is 10

0.900390552576016

0.9005187255993299, when optimal value for lasso is .002

0.9172391535152298

0.9081746360565274, when optimal value for lasso is .001

Top five variable after doubling the optimal value to 20:

1. For ridge:-

MSSubClass
BsmtFullBath
2ndFlrSF
OverallQual
Exterior1st_CBlock

2. For Lasso:-

MSSubClass
BsmtFullBath
Neighborhood_Timber
Exterior1st_CBlock
Neighborhood_Edwards

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?

I will choose Lasso regression model since it not only overcome the problem high value of coefficient $B(\beta)$ but also set it to zero if they are not relevant. Lasso model also eliminating the feature from huge number of features to significant features.

Alpha value for Lasso regression model is 0.001

Question 3

After building the model, you realized 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:

Top five variables will be these:

1. FireplaceQu_Ex
2. OverallQual
3. BsmtQual_Fa
4. Neighborhood_NridgHt
5. OverallCond

Question 4

How can you make sure that a model is robust and generalizable? What are the implications of the same for the accuracy of the model and why?

Answer:

To make sure of for a model being robust and generalizable

Bias and variance trade off can be used for model comparison

There are some other method like Repeated Holdout Validation test robustness.

Accuracy and Precision with the help of Bias Variance trade off.

Implications are the below methods to make a robust and generalized model,
Because

1. Add more data – more data will make us to predict better with training data
2. Treat missing and Outlier values- treating missing and outliers are very important
3. Feature transformation - Scaling or Standardizing gives better uniform dataset
4. Feature Selection- It's important to select features properly for better stable model
5. Multiple algorithms- We must try multiple algorithms to test the bias and variance of model
6. Algorithm Tuning- Algorithm tuning will help in making robust model
7. Cross Validation- This is very important for making robust model.

Question-2:

Repeat the above procedure.

