

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 regression is 2 and lasso regression is 0.01.

If we choose double the value of alpha, there is a bit change in values.

Lasso regression values initially

train data R2: 0.8854624158407248
test data R2: 0.8894603158029368
RMSE : 0.12573595366706636

Lasso regression values after doubled alpha

train data R2: 0.8810127660647746
test data R2: 0.0170177096441244
RMSE : 0.130451943811215

Ridge regression values initially

train data R2: 0.9364594823911135
test data R2: 0.9077597079466583
RMSE : 0.11485785595060986

Ridge regression values after doubled alpha

train data R2: 0.9344289827010624
test data R2: 0.9086565586717956
RMSE : 0.11429811158196207

Since alpha values are small, there is no major change in model after doubling the value of alpha.

What will be the most important predictor variables after the change is implemented?

Ridge regression

MSZoning_FV

MSZoning_RL

Neighborhood_Crawfor

MSZoning_RH

MSZoning_RH

lasso regression

GrLivArea

OverallQual

OverallCond

TotalBsmtSF

BsmtFinSF1

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 prefer to choose Lasso.

Reason:

Lasso- RMSE : 0.12573595366706636

Ridge- RMSE : 0.11485785595060986

Here RMSE for lasso and Ridge has no major difference.

Lasso performs feature selection, and it is easier to interpret models generated by lasso. Also lasso pushes some coefficients to be exact 0 so performs variable selection. Looking this, lasso is better over ridge.

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?

Five most important predictor variables in the lasso model

GrLivArea

OverallQual

OverallCond

TotalBsmtSF

BsmtFinSF1

After excluding these 5 variables and building new model, Five most important predictor variables now-

LotFrontage

LotArea

ScreenPorch

HouseStyle_2.5Fin

HouseStyle_2.5Unf

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?

To make model robust and generalisable it should be simple so that more bias and less variance should be there.

What are the implications of the same for the accuracy of the model and why?

For the accuracy - robust and generalisable model will perform well on both training and testing data resulting similar accuracy. So, accuracy of model can be maintained by keeping the balance between Bias and variance that minimize the total error.