

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?

Solution:

Alpha value for Ridge is **6**

Alpha value for Lasso is **100**

Doubled the Ridge value:

There is a drop of 1% in the r2 score post doubling the value for ridge

Doubled the Lasso value:

There is a drop of 1% in the r2 score post doubling the value for lasso

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?

Solution

Lasso regression would be best considering number of features are too many and Lasso does elimination of features . Hence I would chose Lasso for robust model.

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?

Solution

Top five important features

Neighborhood_NoRidge

Neighborhood_NridgHt

Neighborhood_Somerst

1stFlrSF

Neighborhood_Veenker

Above are dropped and built model again then below are the next top 5 features

2ndFlrSF

MSSubClass_190

MSSubClass_120

BsmtQual

HouseStyle_SLv1

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?

Solution

Below are some of the steps we can consider to build robustness of a model

- Add regularisation
- Bias variance trade off
- VIF values must be less than 5
- P values for predictors must be less 0.05

Above steps implies that the model accuracy is almost on the same lines for train and test data