

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 of Alpha for Ridge and Lasso is 510.

Ridge:

Before Double alpha: 510

('KitchenQual_TA', -3412.982), ('BsmtQual_Gd', -3399.439),
('TotalBsmtSF', 5985.811), ('OverallQual', 8272.63), ('GrLivArea', 8356.668)

After double alpha: 1020

('TotalBsmtSF', 5243.749), ('OverallQual', 6952.066), ('GrLivArea', 7142.03), ('ExterQual_TA', -2962.816),
('KitchenQual_TA', -2861.584)

Lasso:

Before Double alpha: 510

Top five predictor variables from Lasso model:

('BuitAge', -10418.838), ('KitchenQual_Gd', -10374.817), ('Neighborhood_NridgHt', 7891.611),
('OverallQual', 11927.839), ('GrLivArea', 25194.313)

After double alpha: 1020

('Neighborhood_NridgHt', 7852.762), ('OverallQual', 14155.049), ('GrLivArea', 26949.475),
('BuitAge', -8949.42), ('KitchenQual_Gd', -6303.696)

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 Lasso, since we have huge number of columns(features), the Lasso models do the process of RFE along with the finding best predictor columns. Reducing the no. of features improve our computation time and estimate the precise predictor variables.

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?

Ridge:

```
('ExterQual_TA', -4629.814), ('BsmtExposure_No', -3091.053),  
('Neighborhood_NridgHt', 6659.965), ('TotRmsAbvGrd', 7263.599),  
('1stFlrSF', 9601.747)
```

Lasso:

```
('BsmtQual_TA', -14564.136), ('BsmtQual_Gd', -14384.939), ('1stFlrSF',  
13798.698), ('TotalBsmtSF', 14551.889), ('2ndFlrSF', 21727.696)
```

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?

By making low bias and low variance, we can make the model more generic. Also this will give model without overfitting. We can use technique such as cross validation, to make sure the model gives more accurate with the new test and prod data, since these technique gives different set of data during training like unseen data.