

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:

From the created models, the optimal values for alpha are:

For Lasso: alpha=0.0001, train r2_score= 0.9203, test r2_score= 0.8301,

most imp predictor = 'GrLivArea'

For Ridge: alpha=1.0, train r2_score= 0.9260, test r2_score= 0.8301,

Most imp predictor = 'GrLivArea'

When doubled i.e., alpha for Lasso = 0.0002 and for Ridge = 2.0

For Lasso: alpha=0.0002, train r2_score= 0.9142, test r2_score= 0.8279,

Most imp predictor = 'GrLivArea'

For Ridge: alpha=2.0, train r2_score= 0.9232, test r2_score= 0.8335

Most imp predictor = 'GrLivArea'

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 the comparing the both models with optimal alpha values, we can see that there is no difference between the r2_scores between the models.

But the fact is that Lasso model removed 34 features which might not be necessary to explain the 'SalePrice'.

As a result, Lasso model is simpler compared to Ridge model and so I prefer to select Lasso model over Ridge model

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:

The best features after removing the 5 best features are :

- 1) 1stFlrSF
- 2) 2ndFlrSF
- 3) FullBath
- 3) LotShape_IR2
- 4) BsmtFinSF1

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?