

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 :

In my analysis, the alpha value came as 500 for Ridge regression.

0.01 for Lasso regression.

OverallCond remain as the most important predictor variable after doubling the alpha for Ridge regression.

2014

	Feaure	Coef
4	OverallCond	0.098024
15	BsmtFullBath	0.082643
58	Neighborhood_NridgHt	0.069177
13	2ndFlrSF	0.059941
59	Neighborhood_OldTown	0.059516
23	GarageArea	0.051801
12	1stFlrSF	0.051786
122	BsmtExposure_Mn	0.050058
14	GrLivArea	0.047481
20	Fireplaces	0.047306

After doubling the alpha for Lasso regression the BsmtFullBath become the important predictor variable.

0711

	Feature	Coef
15	BsmtFullBath	0.287332
4	OverallCond	0.214988
58	Neighborhood_NridgHt	0.113916
59	Neighborhood_OldTown	0.106490
23	GarageArea	0.087101
122	BsmtExposure_Mn	0.071399
6	YearRemodAdd	0.069838
49	Neighborhood_Edwards	0.061223
64	Neighborhood_StoneBr	0.053862
16	FullBath	0.052188

2. 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 : I will chose to apply the Lasso regression, since the R2 value is better compared to Ridge regression.

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?

Answer:

BsmtFullBath
OverallCond
Neighborhood_NridgHt
Neighborhood_OldTown
GarageArea

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?

Answer:

- a. Use cross validation to assess its performance on unseen data.
- b. Consider feature selection
- c. Carefully tune hyperparameters
- d. Understand the bias-variance trade off.