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 regression: 5
Optimal value of alpha for Lasso regression: 0.0004

If the alpha values for the Ridge and Lasso regression are doubled, the coefficients of the predictor variables would change.

The most important predictor variables after the change in Ridge regression would remain the same.

The most important predictor variables after the change in Lasso regression would change

Before doubling the alpha for Lasso regression:

	Features	rfe_support	rfe_ranking	Coefficient
11	MSZoning_RL	True	1	0.107075
5	GrLivArea	True	1	0.099862
12	MSZoning_RM	True	1	0.075820
1	OverallQual	True	1	0.069501
9	MSZoning_FV	True	1	0.068675
4	TotalBsmtSF	True	1	0.045775
2	OverallCond	True	1	0.045379
14	Foundation_PConc	True	1	0.041408
7	GarageCars	True	1	0.036892
3	BsmtFinSF1	True	1	0.032864

After doubling the alpha for Lasso regression:

	Features	rfe_support	rfe_ranking	Coefficient
5	GrLivArea	True	1	0.098867
11	MSZoning_RL	True	1	0.078572
1	OverallQual	True	1	0.072409
9	MSZoning_FV	True	1	0.053801
12	MSZoning_RM	True	1	0.048849
2	OverallCond	True	1	0.045713
4	TotalBsmtSF	True	1	0.045641
14	Foundation_PConc	True	1	0.037869
7	GarageCars	True	1	0.036670
3	BsmtFinSF1	True	1	0.033470

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:

Lasso regression would be a better option to choose. Lasso regression works well with less number of significant predictor variables. It produces a simpler model with more comprehensible interpretations.

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:

The 5 most important predictor variables now would be:

1. TotalBsmtSF
2. OverallCond
3. Foundation_PConc

4. GarageCars
5. GarageCars

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 model can be made robust and generalisable by ensuring that:

1. It’s not giving too weight to outlier data. This ensures that the accuracy of the predictions is high.
2. Its accuracy with the test data is not less than its accuracy with the training data
3. Confidence intervals are used. This makes more accurate predictions.