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 Ridge alpha = 2.3 and Lasso alpha = 0.001

After doubling Alpha the Model changes are as in the table below:

	R^2	RSS	MSE	RMSE
Ridge alpha = 2.3				
Train	0.93	2.27	0.003	0.05
Test	0.74	3.57	0.009	0.10
Ridge Double alpha = 4.6				
Train	0.92	2.57	0.003	0.05
Test	0.76	3.29	0.009	0.09
Lasso alpha = 0.001				
Train	0.87	3.99	0.004	0.07
Test	0.77	3.24	0.008	0.09
Lasso Double alpha = 0.002				
Train	0.84	5.23	0.006	0.08
Test	0.75	3.46	0.009	0.09

The models important predictor variables for optimal alpha and after doubling the alpha in descending order are;

Ridge for optimal alpha:

1) PosN   -0.196204 Near positive off-site feature--park, greenbelt, etc.

2) OverallQual: 0.164374 Rates the overall material and finish of the house

3) 1stFlrSF: 0.129407 First Floor square feet

4) GrLivArea: 0.126207 Above grade (ground) living area square feet

5) 2ndFlrSF: 0.114766 Second floor square feet

Ridge After doubling alpha:

1) OverallQual: 0.144309 Rates the overall material and finish of the house

2) PosN   -0.110288 Near positive off-site feature--park, greenbelt, etc.

3) 1stFlrSF: 0.109104 First Floor square feet

4) GrLivArea: 0.105249 Above grade (ground) living area square feet

5) 2ndFlrSF: 0.094119 Second floor square feet

Lasso for optimal alpha:

1) OverallQual: 0.263613 Rates the overall material and finish of the house

2) GrLivArea: 0.226042 Above grade (ground) living area square feet

3) TotalBsmtSF: 0.099747 Total square feet of basement area

4) GarageArea: 0.093736 Size of garage in square feet

5) TotRmsAbvGrd: 0.066573 Total rooms above grade (does not include bathrooms)

Lasso After doubling alpha:

1) OverallQual: 0.251120 Rates the overall material and finish of the house

2) GrLivArea: 0.240555 Above grade (ground) living area square feet

3) TotRmsAbvGrd: 0.079671 Total rooms above grade (does not include bathrooms)

4) GarageArea: 0.077119 Size of garage in square feet

5) GarageCars: 0.049733 Size of garage in car capacity

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:

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:

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: