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:

a) Alpha value for ridge: **6**

Alpha value for lasso: **0.0001**

Key Metrics with optimal alpha value

	Metric	Ridge Alpha	Ridge Regression	Lasso Alpha	Lasso Regression
0	R2 Score (Train)	6	0.904921	0.0001	0.91244
1	R2 Score (Test)		0.799586		0.825604
2	RSS (Train)		0.859902		0.7919
3	RSS (Test)		1.160661		1.009981
4	MSE (Train)		0.033118		0.031782
5	MSE (Test)		0.058774		0.054826

Coefficient of GrLivArea:

Ridge	Lasso
0.045807	0.235321

b) After doubling up the alpha values

Alpha value for ridge: **12**

Alpha value for lasso: **0.0002**

	Metric	Ridge Alpha	Ridge Regression	Lasso Alpha	Lasso Regression
0	R2 Score (Train)	12	0.890931	0.0002	0.900744
1	R2 Score (Test)		0.790665		0.821042
2	RSS (Train)		0.986425		0.897677
3	RSS (Test)		1.212322		1.0364
4	MSE (Train)		0.035471		0.033838
5	MSE (Test)		0.060067		0.055538

Coefficient of GrLivArea:

Ridge	Lasso
0.036138	0.19203

On doubling the alpha values:

- RSS increases for both train and test
- Coefficients (Beta values) decreases

c) For Ridge, before and after the change in alpha value,

- The most important variable stayed the same i.e. OverallQual

For Lasso, before and after the change in alpha value,

- The most important variable stayed the same i.e. 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:

I will choose **Lasso**

- Lower RSS values
- Difference between Test and Train R2 Square is lower
- Higher R2 Square value for both Train and Test

	Metric	Ridge Alpha	Ridge Regression	Lasso Alpha	Lasso Regression
0	R2 Score (Train)	6	0.904921	0.0001	0.91244
1	R2 Score (Test)		0.799586		0.825604
2	RSS (Train)		0.859902		0.7919
3	RSS (Test)		1.160661		1.009981
4	MSE (Train)		0.033118		0.031782

5	MSE (Test)		0.058774		0.054826

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:

Based on the Lasso Model, top five predictor variables are:

GrLivArea
PoolQC_Gd
OverallQual
RoofMatl_WdShngl
MasVnrArea

Now, Upon removal of these 5 variables, The top five predictor variables are:

1stFlrSF
2ndFlrSF
PoolQC_NP
Neighborhood_NoRidge
Neighborhood_StoneBr

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) The model can be made robust and generalizable by using regularization techniques by having a trade off between bias and variance and by considering the optimum complexity.

b) If the complexity increases, the bias decreases but the variance increases, the error also increases, there by impacting the accuracy. If the complexity decreases, the bias increases but the variance decreases, the error also increases, thereby impacting the accuracy. This means there has to be a proper trade-off between bias and variance so that an optimal complexity is achieved which has the low error level and higher accuracy