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:

The optimal value of alpha for ridge regression is **4.0** and for lasso regression is **0.0003.**

Lasso regression

Alpha value = 0.0003

Mean squared error on test dataset: 0.0174

R-squared value on test dataset: 0.8696

	Features	Coefficient
0	Neighborhood_StoneBr	0.144200
1	Neighborhood_Crawfor	0.099600
2	Neighborhood_NridgHt	0.095700
3	Exterior1st_BrkFace	0.075300
4	OverallQual	0.070800
5	2ndFlrSF	0.059700
6	TotalBsmtSF	0.057900
7	1stFlrSF	0.052100
8	OverallCond	0.047700
9	Neighborhood_Somerst	0.046200

Alpha value = 0.0006 (Double)

Mean squared error on test dataset: 0.017

R-squared value on test dataset: 0.8724

	Features	Coefficient
0	Neighborhood_StoneBr	0.110900
1	Neighborhood_Crawfor	0.088700
2	Neighborhood_NridgHt	0.077300
3	OverallQual	0.074600
4	Exterior1st_BrkFace	0.066600
5	TotalBsmtSF	0.052600
6	OverallCond	0.047800
7	GrLivArea	0.044000
8	2ndFlrSF	0.039600
9	Neighborhood_Somerst	0.039200

Ridge Regression

Alpha value = 4.0

Mean squared error on test dataset: 0.0174

R-squared value on test dataset: 0.8695

	Features	Coefficient
0	Neighborhood_StoneBr	0.121800
1	Neighborhood_Crawfor	0.087900
2	Neighborhood_NridgHt	0.087600
3	Exterior1st_BrkFace	0.078000
4	OverallQual	0.068800
5	TotalBsmtSF	0.059800
6	RoofStyle_Gambrel	0.053800
7	MSZoning_FV	0.050600
8	2ndFlrSF	0.049000
9	OverallCond	0.047800

Alpha value = 8.0 (Double)

Mean squared error on test dataset: 0.0172

R-squared value on test dataset: 0.871

	Features	Coefficient
0	Neighborhood_StoneBr	0.094300
1	Neighborhood_Crawfor	0.078000
2	Neighborhood_NridgHt	0.072600
3	OverallQual	0.070900
4	Exterior1st_BrkFace	0.063000
5	TotalBsmtSF	0.056700
6	OverallCond	0.047600
7	2ndFlrSF	0.043400
8	1stFlrSF	0.039400
9	GrLivArea	0.039200

Hence, after doubling the values of alpha value of R-Squared on test dataset has increased and mean squared error o n test dataset has slightly decreased which is not good for the model.

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:

We will choose Lasso Regression as it performs slightly better than the Ridge Regression. Moreover, Lasso helps in feature reduction by reducing the coefficients of redundant variables to 0.

Lasso trims down the coefficients of redundant variables to 0 thus, indirectly performs variables selection also. Wher eas, ridge regression reduces the coefficients to arbitrarily low values though not 0.

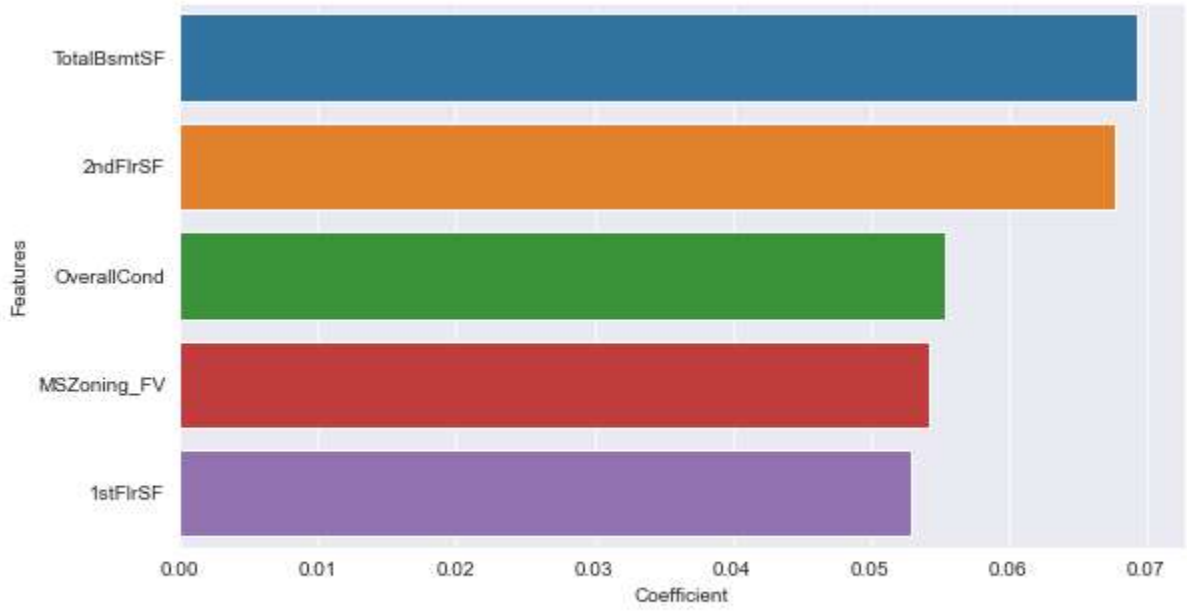
Thus, we can conclude that lasso regression is better option as it is trying to make the regression model simpler whil e balancing the “bias-variance” tradeoff.

Question 3

After building the model, you realised that the five most important predictor variables in the lasso model are not avai lable in the incoming data. You will now have to create another model excluding the five most important predictor v ariables. Which are the five most important predictor variables now?

Answer:
According to the given case after removing the five most important predictor variables in the lasso model from the g iven dataset we created another model.
These are the five most important predictors :

	Features	Coefficient
0	TotalBsmtSF	0.069300
1	2ndFlrSF	0.067800
2	OverallCond	0.055400
3	MSZoning_FV	0.054200
4	1stFlrSF	0.052900



- These are the five most important predictor variables for the required case.

Question 4

How can you make sure that a model is robust and generalizable? What are the implications of the same for the accu racy of the model and why?

Answer:
We can you make sure that a model is robust and generalizable it should not be over fitted.