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:

Ridge: Alpha = 3

Most important predictor variables in order:

	Features
5	OverallQual
10	GrLivArea
8	1stFlrSF
12	TotRmsAbvGrd
7	TotalBsmtSF

Lasso: Alpha=0.8

	Features
213	1stFlrSF
214	2ndFlrSF
192	SaleType_CWD
53	Neighborhood_StoneBr
205	LotArea

When alpha is doubled, coefficients reduce.

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 over Ridge because of better results as below:

	Metric	Linear Regression	Ridge Regression	Lasso Regression
0	R2 Score (Train)	8.786677e-01	8.751104e-01	9.461943e-01
1	R2 Score (Test)	8.754896e-01	8.749996e-01	9.037634e-01
2	RSS (Train)	6.257934e+11	6.441411e+11	2.775130e+11
3	RSS (Test)	2.696667e+11	2.707278e+11	2.084308e+11
4	MSE (Train)	2.623817e+04	2.662003e+04	1.747269e+04
5	MSE (Test)	2.629549e+04	2.634718e+04	2.311791e+04

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:

These are the top 5 features: '1stFlrSF', '2ndFlrSF', 'SaleType_CWD', 'Neighborhood_StoneBr', 'LotArea'

If they are removed, the new top 5 features are:

	Features	Coefficient
210	GrLivArea	176783.9856
14	MSZoning_FV	48142.8942
203	OverallQual	46875.2129
220	GarageArea	43897.6748
206	BsmtFinSF1	43843.3468

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:

There should not be substantial change in R-squared value.