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 alpha for ridge regression is 3.0

The alpha for lasso regression is 0.0001

The R2 Score (Train) for ridge and lasso will be lower.

The RSS (Train) for ridge and lasso will be higher.

The MSE (Train) for ridge and lasso will be slightly higher.

The R2 Score (Test) for ridge will be slightly lower.

The R2 Score (Test) for lasso will be slightly higher.

The RSS (Test) for ridge will be slightly higher.

The RSS (Test) for lasso will be lower.

The MSE (Test) for ridge will be slightly higher.

The MSE (Test) for lasso will be slightly lower.

	Metric	Linear Regression	Ridge Regression	Lasso Regression	Ridge Regression x 2	Lasso Regression x 2
0	R2 Score (Train)	9.522669e-01	0.941417	0.943455	0.933431	0.935952
1	R2 Score (Test)	-1.545299e+21	0.886887	0.886259	0.886730	0.889371
2	RSS (Train)	1.054026e+00	1.293600	1.248610	1.469964	1.414289
3	RSS (Test)	1.682703e+22	1.231703	1.238550	1.233413	1.204659
4	MSE (Train)	4.113217e-02	0.045568	0.044768	0.048575	0.047646
5	MSE (Test)	7.923851e+09	0.067793	0.067981	0.067840	0.067045

GrLivArea, TotalBsmtSF, OverallQual, Neighborhood, 1stFlrSF were the most important predictor old variables.

GrLivArea, TotalBsmtSF, OverallQual, Neighborhood, SaleCondition are the most important predictor variables after the change is implemented.

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:

The R2 Score is slightly for Lasso. The Lasso regression is chosen because it helps in reducing the features in the model, and create a simpler final model.

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:

GrLivArea, TotalBsmtSF, OverallQual, Neighborhood, 1stFlrSF were the most important predictor old variables.

BsmtFinSF1, BsmtUnfSF, MSZoning, 2ndFlrSF, SaleType are five most important predictor variables now

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 that is robust but not generalisable may have high accuracy on noisy or different data, but low accuracy on new data that is similar to the training data. This may indicate that the model is too complex or flexible and cannot capture the underlying patterns of the data.

A model that is generalisable but not robust may have high accuracy on new data that is similar to the training data, but low accuracy on noisy or different data. This may indicate that the model is too simple or rigid and cannot adapt to changes in the data distribution.

A model that is both robust and generalisable may have high accuracy on both new data that is similar to the training data and noisy or different data. This may indicate that the model has a good balance between complexity and flexibility and can learn the essential features of the data.