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:

- Best alpha for Ridge is 0.31622776601683794, and the best alpha for Lasso is 0.0016102620275609393.

Lasso Regression: In Lasso regression, the regularization penalty is stronger when alpha is doubled. More coefficients may become exactly zero as a result, which would successfully carry out feature selection. As the model gets more limited, less significant aspects might be left out.

Ridge Regression: In a similar manner, raising the alpha value by twofold in Ridge regression intensifies the regularization penalty without requiring coefficients to be precisely zero. Large coefficients are still penalized, which results in a more restricted model without completely eliminating any variables.

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: Lasso regression is a better option if we wish to undertake automatic feature selection and there is a view that many features are redundant or unnecessary. Lasso frequently drives some coefficients to zero, hence eliminating some features. Ridge regression may be a better option if the dataset has highly associated features. When presented with a set of associated features, Lasso often chooses one feature at random and zeros out the others. Ridge more fairly distributes the weights among linked attributes because of its penalty term.

Question 3

After building the model, you realized 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: Five most important predictor variables after dropping top 5:

- GrLivArea
- SalesType_New
- LandContour_HLS
- OverallQual
- BsmtExposure_Gd

Question 4

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

Answer:

- **Cross-Validation:** Use techniques like k-fold cross-validation to assess the model's performance across multiple subsets of the data.
- **Regularization:** Use regularization techniques like L1 (Lasso) and L2 (Ridge) to prevent overfitting by penalizing overly complex models. Regularization helps in producing simpler models that are more likely to generalize.
- **Hyperparameter Tuning:** Fine-tune hyperparameters using techniques like grid search or random search to optimize the model's performance on validation data.