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
If we choose double the value of alpha for both Ridge regression and Lasso regression, it would make the models more strict in reducing the coefficients and shrinking them towards zero.

With higher alpha values, the Ridge regression model would further constrain the coefficients, making them smaller. This means the model would focus more on reducing the impact of each predictor variable.

Similarly, the Lasso regression model would also increase the level of regularization with higher alpha values. This would result in more predictor variables being completely ignored (set to zero), leaving only the most important variables in the model.

After making these changes, the most important predictor variables would be the ones that still have a significant impact on predicting the target variable. These variables would have non-zero coefficients and contribute the most to the model's predictions.

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

Based on the optimal values determined, for controlling multicollinearity and obtaining stable coefficient estimates, Ridge regression with an optimal lambda value of 0.9 is preferred. However, if the goal is to perform feature selection and identify important predictors, Lasso regression with an optimal alpha value of 0.001 would be the choice. The decision depends on the specific objectives of the analysis and the trade-off between model simplicity and interpretability.

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

BsmtFinSF2, BsmtHalfBath, Neighborhood_IDOTRR, WoodDeckSF, LotShape

---

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

1. A robust and generalizable model can be achieved by using a diverse and representative dataset that captures real-world variability.
2. Splitting data into train and test sets helps evaluate the model's performance on unseen data, ensuring its ability to generalize.
3. Applying cross-validation techniques provides a more reliable estimate of the model's performance by using multiple subsets of data for training and testing.
4. Regularization methods like ridge and lasso regression prevent overfitting and promote generalization by adding a penalty term to the model's objective function.
5. Feature selection helps improve generalization by identifying and excluding irrelevant or redundant predictors, allowing the model to focus on the most informative features.
6. While enhancing generalization, the model's accuracy may slightly decrease as it aims to be more applicable in real-world scenarios.