Q1: 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:

- *What is the optimal value of alpha for ridge and lasso regression?*

  Optimal value of alpha for Ridge regression: 2.0

  Optimal value of alpha for Lasso regression: .0001

- *What will be the changes in the model if you choose double the value of alpha for both ridge and lasso?*

  The R2 score drops slightly post doubling alpha:

Regression Model	Alpha		Double Alpha	
	Train R2	Test R2	Train R2	Test R2
Ridge	0.9194	0.8806	0.9163	0.8796
Lasso	0.9214	0.8824	0.9186	0.8813

- *What will be the most important predictor variables after the change is implemented?*

  There is no change in predictor variables post doubling the alpha, only slight change in coefficients values

Regression Model	Alpha	Double Alpha
	Predictor Variables	
Ridge	SaleCondition_Partial SaleCondition_Normal SaleCondition_Family SaleCondition_Alloca SaleCondition_AdjLand	SaleCondition_Partial SaleCondition_Normal SaleCondition_Family SaleCondition_Alloca SaleCondition_AdjLand
Lasso	SaleCondition_Partial SaleCondition_Normal SaleCondition_Family SaleCondition_Alloca SaleCondition_AdjLand	SaleCondition_Partial SaleCondition_Normal SaleCondition_Family SaleCondition_Alloca SaleCondition_AdjLand

Q2: 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:

- Both Ridge and Lasso regression has approximately same values of R2 for test data.

Regression Model	Optimal Alpha	
	Train R2	Test R2
Ridge (alpha = 2.0)	0.9194	0.8806
Lasso (alpha = 0.0001)	0.9214	0.8824

- But , since Lasso Regression also takes into account the feature selection too, so it will be better to choose Lasso Regression

Q3: 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:

- New optimal alpha value remains same: .0001
- New predictor variable:

Lasso	GarageFinish_Unf GarageFinish_RFn GarageFinish_NA GarageType_NA GarageType_Detchd

Q4: 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?

Answers:

- We can focus on below strategies to enhance the robustness and generalizability of the model, focusing on R2 (coefficient of determination) and mean square error (MSE):
    - Cross-Validation:
        - Use cross-validation techniques, such as k-fold cross-validation, to assess the model's performance across multiple subsets of the data. This helps ensure that the model's performance is consistent and not overly dependent on a particular random split of the data.
    - Hyperparameter Tuning:
        - Tune the regularization strength hyperparameter (alpha) using cross-validation. This helps in finding the optimal trade-off between fitting the training data well and keeping the model coefficients small to prevent overfitting.
    - Feature Scaling:
        - Standardize or normalize the features to ensure that all variables contribute equally to the regularization term. This is important because regularization techniques like Ridge and Lasso are sensitive to the scale of the input features.
    - Outlier Handling:

        - Identify and handle outliers appropriately. Outliers can have a significant impact on the coefficients and performance metrics. Robust regression techniques may be considered to mitigate the influence of outliers.
    - Model Complexity:
        - Be cautious of overfitting by monitoring the model's complexity. Ridge and Lasso are regularization techniques designed to prevent overfitting, but it's essential to find a balance between simplicity and fitting the data.

- *Implications on the accuracy of the model and why?:*

    - Implication involve a trade-off between fitting the training data closely and preventing overfitting. The choice of regularization strength, feature selection properties, and the characteristics of the data all play a role in determining the overall impact on predictive accuracy.
    - Impact on Overfitting:
        - Ridge and Lasso are regularization techniques designed to prevent overfitting. By adding a regularization term to the loss function, these methods penalize large coefficients, encouraging the model to find a balance between fitting the training data well and keeping the coefficients small. This can lead to better generalization to new, unseen data.
    - Feature Selection:
        - Lasso regularization has the additional effect of encouraging sparsity in the coefficient estimates, effectively performing feature selection. Some coefficients may be exactly zero, meaning that the corresponding features are effectively ignored by the model. This can be beneficial if there are irrelevant or redundant features in the dataset.
    - Handling Multicollinearity:
        - Ridge regression is particularly useful when dealing with multicollinearity (high correlation between predictor variables). By penalizing the sum of squared coefficients, Ridge tends to distribute the impact of correlated variables more evenly. This can improve the stability of coefficient estimates.
    - Sensitivity to Hyperparameter Choice:
        - The performance of Ridge and Lasso models is sensitive to the choice of the regularization hyperparameter (alpha). Cross-validation is typically used to tune this hyperparameter, and the choice should be based on the specific characteristics of the data. Too much regularization (a high alpha) can lead to underfitting, while too little regularization (a low alpha) may result in overfitting.