Question1: 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?

Answer1: The optimal value of the regularization parameter for ridge and lasso regression models is typically determined using techniques such as cross-validation ,same has been done in the code with the given dataset and found that with ridge validation it is 100 and with lasso it is 100.When we doubled the value of alpha in Ridge and Lasso regression, it increases the regularization strength. Higher alpha values penalizes large coefficients more strongly, leading to simpler models with smaller coefficients. This helps to prevent over-fitting but might also under-fit the data if the regularization is very strong. The choice of the optimal alpha value depends on the specific dataset and problem given.

If you double the value of alpha, the models' coefficients will be reduced further, making the model even more conservative in its predictions. The effect will be more prominent in Lasso regression as it encourages sparsity (some coefficients become exactly zero), effectively performing feature selection.

To find the most important predictor variables after implementing this change, we can examine the coefficients of the trained models. Lower coefficients indicate less importance, and coefficients that are exactly zero (in Lasso) indicate that the corresponding variable is not contributing to the model.

Question2: 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?

Answer2: I will choose ridge regression because**:**

All features are relevant and you want to keep all of them in the model.

Have multi-collinearity issues in given dataset.

Primary goal is to improve predictive power while controlling overfitting.

Question3. 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

Answer3: As per the model after coding it is found the below are the five most important predictor variables:

Top 5 Most Important Predictor Variables in Lasso Regression:

Feature	Coefficient
Neighborhood_NridgHt	65304.675594
Neighborhood_Somerst	31202.037354
Neighborhood_CollgCr	30761.287583
LotConfig_FR2	15021.516843
MSZoning_RL	12827.265826

Question4: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?

Answer4: Ensuring that a model is robust and generalizable is crucial for its effectiveness in making predictions on unseen data. Here are several strategies and considerations to ensure the robustness and generalizability of a machine learning model:

Cross-Validation: Uses technique like k-fold cross-validation to assess the model's performance across multiple subsets of the dataset . Cross-validation provides better estimate of the model's performance on unseen data compared to single train-test split.

Holdout Validation Set: Split the data into three sets namely a training set, a validation set, and a test set. Use the training set for model training, validation set for hyper-parameter tuning and model selection, and test set to evaluate the final model's performance on the dataset. This approach helps prevent over-fitting to the test data.

Feature Selection: Carefully selects relevant features and avoid over-fitting by removing irrelevant or highly correlated features. Regularization techniques can be employed to penalise complex models and promote simpler, and more interpretable models.

Hyper-parameter Tuning: Tune the hyper-parameters of the model using techniques like grid search or randomized search to find the best combination of hyper-parameters. This ensures that the model is optimized in performance.

Data Pre-processing: Standardize or normalise the features to ensure that all features are on the same scale. Handles missing data appropriately (either by removing rows with missing data or imputing missing values) and encodes categorical variables properly.

Regularization: Use regularization techniques like (Lasso) and (Ridge) regularisation to prevent over-fitting by adding penalty terms to the loss function. Regularization discourages overly complex models.

Avoid Data Leakage: Ensure that there is no data leakage between the training and test data-sets. Data leakage occurs when information from the test set is used to train the model, leading to overly optimistic performance estimates.

Implications for Accuracy:

Training Accuracy vs. Test Accuracy: A robust and generalisable model should have almost same performance on both the training and test data. Large disparities between these accuracies may indicate over-fitting.

Consistent Performance: The model should perform consistently across different subsets of the dataset. If the performance varies significantly, the model may not be generalising well.

Stability: A robust model should provide stable predictions even when exposed to small variations in the input data. Small perturbations in the input should not result in drastically different predictions for the data.

By following these best practices, it can increase the likelihood that your model will generalize well to un-seen data and provide reliable predictions in real-world cases.