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?**

The optimal value of the regularization parameter (alpha) in Ridge and Lasso regression depends on the specific dataset and the problem at hand. There is no one-size-fits-all answer, and it is common practice to use techniques like cross-validation to find the best value for alpha.

<u>*Ridge Regression:*</u>

Ridge regression adds a penalty term to the linear regression cost function, which is proportional to the sum of the squares of the coefficients.

The Ridge regression cost function is given by:

$$J(\beta) = \text{MSE}\ (\beta) + \alpha \sum_{i=1}^{n} \beta_i^2$$

Where β= vector of regression coefficients, MSE (β) = mean squared error terms and α = regularization parameter

A smaller α allows the coefficients to vary more freely, while a larger α forces the coefficients to be smaller.

<u>*Lasso Regression:*</u>

Lasso regression adds a penalty term proportional to the absolute values of the coefficients

The Lasso regression cost function is given by:

$$J(\beta) = \text{MSE}\ (\beta) + \alpha \sum_{i=1}^{n} |\beta_i|$$

Where β= vector of regression coefficients, MSE (β) = mean squared error terms and α = regularization parameter

Lasso tends to produce sparse models, as it encourages some coefficients to become exactly zero. As α increases in Lasso, sparsity increases.

If we were to double the value of α for Ridge and Lasso:

In Ridge Regression:

The penalty for large coefficients would be stronger, leading to a more pronounced shrinkage of coefficients. Coefficients would be pushed closer to zero, but not necessarily to exactly zero.

In Lasso Regression:
The penalty for large coefficients would also increase, and the model would likely produce a sparser solution. More coefficients would be forced to exactly zero.

Generally, with higher alpha values, less important variables are more likely to have their coefficients pushed towards zero or become exactly zero. This can lead to a simpler and more interpretable model.

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?**
   Though the model performance by ridge regression was better in terms of R2 values of train and test, it is better to use Lasso, since it brings and assigns a zero value to insignificant features, enabling us to choose the predictive variables.

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?**
   Overallqual, GrLivArea, OverallCond, GarageArea, BsmtFullBath

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?**
   Ensuring a model's robustness and generalizability involves navigating the bias-variance tradeoff. Bias is when a model is too simple and misses important patterns, leading to mistakes. On the other hand, variance is when a model is too sensitive to the specific details of the training data, making it struggle with new data. Finding the right balance means picking a level of complexity that avoids both these issues. If a model is too simple, it might not be flexible enough, and if it's too complex, it might fit the training data too closely and not work well on new data. So, creating a good model means figuring out this tradeoff to get accurate predictions on different datasets without getting confused by noise in the training data. The goal is to find a sweet spot that makes the model reliable and useful in real-world situations.