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:

The optimal value of alpha in ridge and lasso regression influences the regularization strength in these models. It is determined through techniques like cross-validation, where different values of alpha are tested, and the one that provides the best model performance is chosen.

For ridge regression, alpha controls the amount of shrinkage applied to the coefficients. A higher alpha leads to greater shrinkage, which can help prevent overfitting by penalizing large coefficient values. However, choosing too high of an alpha may result in underfitting. In our case it was

For lasso regression, the role of alpha is similar, but it also introduces a feature selection aspect. Lasso tends to drive certain coefficients to exactly zero, effectively eliminating less relevant predictors from the model.

Changes in the Model with Double the Value of Alpha:

If we choose to double the value of alpha for both ridge and lasso regression, the coefficients of the predictor variables will shrink more, and in the case of lasso, some coefficients may be driven to exactly zero. This increased regularization can lead to a simpler model with fewer features having significant weights.

Most Important Predictor Variables After the Change:

The impact of doubling alpha on the model's predictor variables depends on the original alpha, the data, and the relationships between variables. Generally, with higher alpha values, the model tends to prioritize a subset of the most influential predictors.

For ridge regression, even with a higher alpha, all coefficients will be non-zero, but their magnitudes will be more constrained. The most important predictors will still have relatively higher weights.

In contrast, lasso regression tends to produce sparse models, and by doubling alpha, more coefficients are likely to be driven to zero. The most important predictors will be those that survive the increased regularization and maintain non-zero weights.

To determine the most important predictors after implementing the change, we would need to examine the coefficients of the updated model. Variables with non-zero coefficients in the updated model would be considered the most important predictors.

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:
The optimal value of lambda we got in case of Ridge and Lasso was:

- Ridge regression - 2
- Lasso regression - 0.01.

The corresponding r2 values we got in case of Ridge and Lasso was:

- Ridge regression - Train = 0.92, Test = 0.87, difference = -0.05
- Lasso regression - Train = 0.86, Test = 0.85, difference = -0.01.

The mean squared error in case of Ridge and Lasso was:

- Ridge regression - 0.14
- Lasso regression - 0.15

We can clearly observe that the Mean Squared Error of Lasso regression is slightly higher than that of Ridge regression. Also the difference of r2 between train and test is less in lasso as compared to ridge. Since, lasso helps in feature reduction (as the coefficient value of one of the lasso's features to be shrunk toward 0) and helps to increase model interpretation by taking the magnitude of the coefficients, thus *Lasso regression has a better edge over Ridge regression*.

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:
We dropped the top 5 most important predictor variables in the lasso regression model and again created the model and got the below five most important predictor variables:

1. TotalBsmtSF
2. TotRmsAbvGrd
3. OverallCond
4. Total_Bathrooms
5. LotArea

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:
To improve the model's accuracy, it is important to keep it as simple as possible while still retaining its generalizability. This can be accomplished by using a bias-variance trade-off. The simpler the model, the more bias it will have but less variance will be applicable throughout the population. However, if the model becomes too complex, this reduces its ability to be generalized and can even lead to overfitting. To avoid this, regularization methods can be used to shrink coefficients towards zero, which helps maintain the model's complexity while minimizing total error.