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?

Ans: The optimal value of LAMBDA we got in case of Ridge and Lasso is:

1. Ridge – 0.8

2. Lasso – 2.0

The changes in the model if we choose to double the value of alpha for both ridge and lasso are:

- When we double the alpha value in case of ridge and lasso for both, the r2 score on training data decreased but it increased on testing data.

Predictors are same but the coefficient of these predictor has changed:

- LotArea---------------Lot size in square feet
- OverallQual---------Rates the overall material and finish of the house
- OverallCond--------Rates the overall condition of the house
- YearBuilt-------------Original construction date
- BsmtFinSF1--------Type 1 finished square feet
- TotalBsmtSF------- Total square feet of basement area
- GrLivArea-----------Above grade (ground) living area square feet
- TotRmsAbvGrd----Total rooms above grade (does not include bathrooms)
- Street_Pave--------Pave road access to property
- RoofMatl_Metal----Roof material_Metal

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?

Ans: The optimal value of LAMBDA we got in case of Ridge and Lasso is:

1. Ridge – 0.8

2. Lasso – 2.0

The r2 value we got in case of Ridge and Lasso is:

• Ridge - Train = 0.880, Test = 0.860, difference – 0.020

• Lasso - Train = 0.882, Test = 0.853, difference – 0.029

The r2 score of lasso is slightly higher for the test dataset so we will choose lasso regression to solve this problem.

Given Lasso's ability to shrink coefficients towards zero and perform feature selection by setting some coefficients exactly to zero, it offers a more interpretable model by effectively identifying the most relevant features. Compared to Ridge regression, where coefficients are reduced but not necessarily set to zero, Lasso provides a clearer indication of which features have the most impact on the target variable, leading to enhanced model interpretability and potentially improved generalization performance.

Q3. 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?

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

LotArea, OverallQual, YearBuilt, BsmtFinSF1, TotalBsmtSF are the top 5 important predictors.

Q4. 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?

Ans: Ensuring a model's robustness and generalizability involves striking a balance between simplicity and accuracy. Simplifying the model increases its generality, even though this might lead to a decrease in accuracy. This principle aligns with the Bias-Variance trade-off, where simpler models exhibit higher bias but lower variance, making them more generalizable. Conversely, complex models tend to have lower bias but higher variance.

Challenges such as underfitting and overfitting arise when this balance is not maintained. Regularization techniques play a crucial role in managing model complexity by shrinking coefficients towards zero. This prevents the model from becoming overly complex and reduces the risk of overfitting.

By penalizing complexity, regularization ensures that the model remains optimally simple. It facilitates achieving the Bias-Variance trade-off by striking a balance between bias and variance, ultimately minimizing the total error. This equilibrium, known as the Optimum Model Complexity, represents the ideal compromise where the model is simple enough to be generalizable yet sufficiently complex to be robust.