Assignment Part II

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?

Ans:

1. The optimal value of LAMBDA we got in case of Ridge and Lasso is :
   a. Ridge - 2.0
   b. Lasso - 0.0001

2. The changes in the model if we choose to double the value of alpha for both ridge and lasso are:
   a. Ridge Regression
      i. Slight increase in mean squared error.
      ii. R2 value of train and test remains almost the same.
   b. Lasso Regression
      i. Slight increase in mean squared error.
      ii. R2 value of train slightly decreases.
      iii. Huge fall in R2 value of the test.
      iv. Model and prediction worsen.
   c. General Impact
      i. Increased alpha penalizes the model further.
      ii. More coefficients of a variable shrink towards zero.
3. The most important predictor variables after the change is implemented:
   a. For Ridge:
      i. Total_sqr_footage
      ii. OverallQual
      iii. GrLivArea
      iv. Neighborhood_StoneBr
      v. OverallCond
      vi. TotalBsmtSF
      vii. LotArea
      viii. YearBuilt
      ix. Neighborhood_Crawfor
      x. Fireplaces
   b. For Lasso:
      i. Total_sqr_footage
      ii. OverallQual
      iii. YearBuilt
      iv. GrLivArea
      v. Neighborhood_StoneBr
      vi. OverallCond
      vii. LotArea
      viii. Neighborhood_Crawfor
      ix. Neighborhood_NridgHt
      x. GarageCars

Assignment Part II

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?

Ans:

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

a. Ridge - **2.0**
b. Lasso - **0.0001**

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

a. Ridge - Train = **0.930**, Test = **0.896, difference** – **0.046**
b. Lasso - Train = **0.927**, Test = **0.902, difference** – **0.025**

The Mean Squared error in case of Ridge and Lasso is:

a. Ridge - **0.00297**
b. Lasso - **0.00280**

Here we can clearly see that Lasso is better than Ridge due to mse is lower,r2 is also less

Assignment Part II

Question 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?

Ans:

Model with the following Top 5 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?

Ans:

- **Model Simplicity and Generalization:**
  - Aim for a simple model; simplicity makes a model more 'generic.'
  - Simpler models, although less accurate, tend to be more robust.
- **Bias-Variance Trade-off:**
  - Bias increases with simplicity, while variance decreases.
  - Simple models are more generalizable due to a balanced bias-variance trade-off.
  - Complex models have high variance and low bias.
- **Underfitting and Overfitting:**
  - Underfitting and overfitting are common problems.
  - Strive for a balance between bias and variance to avoid these issues.
  - Achieving this balance is possible through regularization.
- **Role of Regularization:**
  - Regularization manages model complexity by shrinking coefficients toward zero.
  - Prevents the model from becoming overly complex and reduces the risk of overfitting.
  - Ensures an optimum level of simplicity by penalizing complexity.
- **Optimal Model Complexity:**
  - Regularization contributes to achieving an Optimum Model Complexity.
  - The model is sufficiently simple to be generalizable yet complex enough to be robust.
  - Balances the Bias-Variance trade-off by minimizing Total Error.
- **Regularization's Impact on Bias-Variance Trade-off:**
  - Leads to a compromise by increasing bias to a point where Total Error is minimized.

  - Striking a balance between simplicity and complexity.
- **Simplicity and the Bias-Variance Trade-off:**
  - Simplifying a model involves navigating the Bias-Variance trade-off.