Assignment Part-II

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: - Optimal Value of alpha for Ridge and lasso regression are 1 and 0.001 respectively.

Model	**Alpha values**
Ridge Regression	1
Lasso Regression	0.001

Changes in the model if we double the alpha values for both the model: -

For both Ridge and Lasso, when we double the alpha value, it has pushed the coefficient to push towards zero. In Lasso Regression, the coefficients did not change the order but in Ridge Regression the first three features did change in order.

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: - Ridge Regression works well when there are many variables and Lasso Regression works well when there are fewer variables. Although here, there are lot of variables being zero with Lasso Regression compared to Ridge Regression.

Also, the r2 score of Lasso is higher in test case compared to that for ridge regression. And speaking of mean squared errors, they are almost same for both regressions. So, I would choose Lasso over Ridge.

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?**

Ans: - Following are the five variables after the five most important predictor variables were not available:

- BsmtQual: Evaluates the height of the basement
- KitchenQual: Kitchen quality
- BsmtFinSF1: Type 1 finished square feet
- Foundation_PConc (Foundation: Type of foundation, PConc: Poured Contrete)
- 2ndFlrSF: Second floor square feet

4. **How can you make sure that a model is robust and generalizable? What are the implications of the same for the accuracy of the model and why?**

Ans: - A Model is said to be generalized when it's not overfitting or underfitting on the training data. It should not memorize the training data and it should perform well on the unseen data that is test data with some acceptable error. A model should perform accurately on unseen data even if there are changes on training data. Error should be minimal or say in acceptable range that's what a robust model should be like.

A model can be robust and generalizable if there is no overfitting and underfitting with higher accuracy. It should be a balance of all three. Regularization is the technique through in which the coefficients are more penalized that are large and so reduces the overfitting.

A simpler model is more robust and generalizable as we compromise the complexity of the model with accuracy and so accuracy might go down on training data but consistent on test data. A generalized model is when the error is lowest and variance and bias are almost the same. That is where we get the optimal model complexity.