**Subjective Assignment – Advanced Regression**

**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:

* In my final setting, the best alpha values for Ridge and Lasso are 50 and 10 respectively.
* The model's performance is unaffected by multiplying these numbers by two. Neighborhood StoneBr,GarageArea,Neighborhood NridgHt,TotalBsmtSF,GrLivArea, KitchenQual, and Neighborhood Names are now major predictor factors in Ridge. Neighborhood Edwards,BldgType TwnhsE,GarageFinish
* In Lasso, TotalBsmtSF and SaleType New are crucial predictor variables. Neighborhoods NAmes and OldTown, KitchenQual and SaleCondition Partial are all included in this property's MSZoning RM, GarageType Attached and GrLivArea.

**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:

In terms of results, we find that Ridge and Lasso are fairly close.

We decided to use the Lasso method even though the Ridge model performed marginally better on the testing data. A benefit of feature removal using Lasso is that it may assist identify the most relevant predictors in our dataset, which has around 130+ columns.

As a result, our overall result is Lasso, which has a r2 score of 88 on the training dataset and a r2 score of 84 on the testing dataset.

**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?

Answer:

Our Lasso model's five most critical predictor variables are: "GrLivArea," "GarageType Attchd," "MSZoning RM," "SaleType New," and "TotalBsmtSF." These are the five greatest critical predictor variables when we recreate the model without them: MasVnrArea, Neighborhood StoneBr, Neighborhood NridgHt, Fireplaces, and the GarageArea area.

**Question 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?

Answer:

We may increase the model's robustness and generalizability by not overfitting it and keeping it as basic as feasible. Overfitting the model will improve accuracy, but it will no longer be generalizable. To ensure the robustness of the model, it should be able to perform well on both training and testing datasets.