Surprise Housing Problem Statement - 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?

Answer:

The Current optimum value of alpha/lambda is:

> 1. Ridge Regression: 4.51
> 2. Lasso Regression: 0.0001

After doubling the optimum value, **very marginal change in the r-square** value was observed for both Ridge and Lasso regression.

All the top 5 predictors after doubling optimum value for Lasso regression.

> 1. **GrLivArea** - Living area is the strongest predictor
> 2. **Neighborhood_StoneBr** - Stone Brook area with Ames city is one of the strong predictor
> 3. **BsmtFinSF1** - Type 1 finished square feet is the second strongest predictor
> 4. **GarageArea** - Size of garage is one of the strong predictor
> 5. **KitchenQual** - Kitchen quality is one of the strong predictor

All the top 5 predictors after doubling optimum value for Ridge regression.

> 1. **GrLivArea** - Living area is the strongest predictor
> 2. **GarageArea** - Size of garage is one of the strong predictor
> 3. **KitchenQual** - Kitchen quality is one of the strong predictor
> 4. **BsmtFinSF1** - Type 1 finished square feet is the second strongest predictor
> 5. **ExterQual** - External quality is one of the strong predictorr

---

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:

Both Ridge Regression and Lasso Regression have done equally good job. However If I have to pick one out of these two then, I would pick **Lasso Regression model (Model4)** as it has **lesser number of Beta coefficients** to look into, therefore a **less complex model** without compromising on the model efficiency.

In [146…
```
final_metric
```

Out[146]:

	Metric	Model1: Linear Regression	Model2: RFE Regression	Model3: Ridge Regression	Model4: Lasso Regression
0	R2 Score (Train)	8.966597e-01	8.967519e-01	0.883060	0.885494
1	R2 Score (Test)	-9.862649e+21	-1.259083e+21	0.878199	0.878388
2	RSS (Train)	2.912265e+00	2.909667e+00	3.295524	3.226943
3	RSS (Test)	1.138765e+23	1.453767e+22	1.406349	1.404159
4	MSE (Train)	5.340754e-02	5.338371e-02	0.056813	0.056219
5	MSE (Test)	1.612427e+10	5.761166e+09	0.056664	0.056620

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:

The optimum value of alpha/lambda after excluding top 5 predictors is:

> 1. Ridge Regression: 1.02
> 2. Lasso Regression: 0.000059

All the top 5 predictors for Lasso regression.

> 1. **TotalBsmtSF**
> 2. **GarageQual**
> 3. **LotArea**
> 4. **Neighborhood_Edwards**
> 5. **ExterQual**

All the top 5 predictors for Ridge regression.

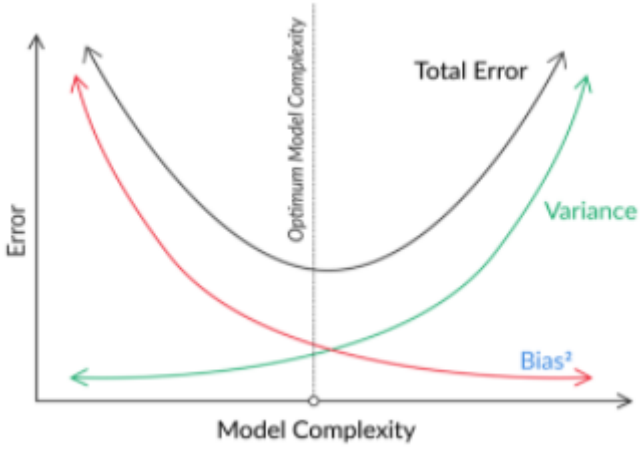
> 1. **TotalBsmtSF**
> 2. **LotArea**
> 3. **GarageQual**
> 4. **Neighborhood_Edwards**
> 5. **ExterQual**

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:

We can make sure that model is robust and generalisable by taking care of **bias - variance trade-off**. If the model is far too complex or overfit, then model will have low bias (on trining data) but will have high variance (on test data). On the contrary, if the model is far too simple or generalize then the model will have high bias and low variance.



Therefore we should go for the optimum model which will the least total error as shown in graph above. Here, to get the optimum model, Lasso and Ridge regularization techniques are very helpful. Using the hypermeter tuning, we can build the optimum model which will have least total error and also will have good accuracy.