Assignment- Part II

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**: Original Alpha Values:**

1. Optimal Value for Ridge: **0.0001**
2. Optimal Value for Lasso: **0.0001**

Alpha Values after doubling:

1. Optimal Value for Ridge: **0.0002**
2. Optimal Value for Lasso: **0.0002**

After double the value of alpha, there are no significant changes in both metrics & features as shown below.

	Linear	Lasso	Ridge
GrLivArea	1.139724	1.137593	1.139695
OverallQual	0.736756	0.740166	0.736759
LotArea	0.415474	0.372625	0.415456
GarageCars	0.361894	0.361095	0.361897
OverallCond	0.322477	0.315883	0.322471
BsmtFullBath	0.216168	0.220394	0.216166
SaleType_Oth	0.289904	0.185747	0.289873
Neighborhood_StoneBr	0.183744	0.177293	0.183744
Fireplaces	0.157507	0.159448	0.157511
Neighborhood_Crawfor	0.139153	0.135200	0.139153
BedroomAbvGr	0.138444	0.132308	0.138458
LandContour_HLS	0.131268	0.122248	0.131266
Neighborhood_NridgHt	0.116301	0.113971	0.116301
Neighborhood_ClearCr	0.111542	0.110183	0.111543
Exterior1st_BrkFace	0.107240	0.105928	0.107241
LandContour_Low	0.109288	0.101128	0.109288
Neighborhood_Veenker	0.093425	0.081798	0.093424
LandContour_Lvl	0.089515	0.080697	0.089514
SaleType_ConLD	0.082184	0.063401	0.082182
SaleCondition_Alloca	0.068985	0.047892	0.068984
HouseStyle_2.5Unf	0.049304	0.032217	0.049302
GarageQual_Po	0.132451	0.027948	0.132440
BsmtFinSF1	0.038560	0.027261	0.038571

** The solution is present in Jupyter Notebook.*

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: R2 Scores, RSS and MSE for Lasso are better than Ridge & Linear Regression. In Lasso, features having zero coefficients value can be removed from the model thus reducing model complexity.
Therefore, ***Lasso can be preferred over Ridge and Linear Regression.***

Q3: **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: * Top 5 features in Lasso Model:

['GrLivArea' , 'OverallQual' , 'LotArea' , 'GarageCars' , 'OverallCond']

**** After Removal of top 5:***

```
In [163]: df_lasso.sort_values(by = 'Lasso', ascending=False).head(5)
```

Out[163]:

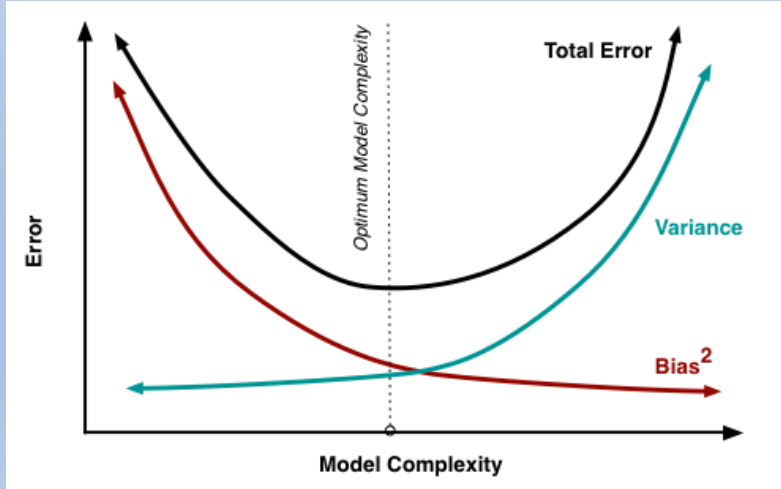
	Lasso
GrLivArea	1.131061
OverallQual	0.743851
GarageCars	0.360811
LotArea	0.330071
OverallCond	0.307435

Q4: 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: Model with testing error & training error has enough stability even after some noise. This means than an unprecedented change in one or more feature does not alter the value of predicted variables significantly.

We can control the tradeoff between model complexity and bias, to make the model more robust, there should be balance keeping the model simple and making simple model lead to bias variance tradeoff.

Accuracy of model can be maintained by keeping the balance bias and variance and minimize the total error.



Thank You!