Advanced Regression 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?

Answer: The Optimal Value of alpha for Ridge and Lasso Regression is mentioned below

- **Ridge – 8.0**
- **Lasso – 0.001**

If we double the value of alpha for both Ridge and Lasso Regression, I found that the R2 score is changed as follows

	R2 Score	
	Original	New
Lasso	89.9	88.3
Ridge	90.3	90.2

The Top 5 Predictors Post changing value of alpha from **0.001** to **0.002** for **Lasso Regression** are mentioned below

Top 5 Predictors				
	Original Value of Alpha = 0.001		**New Value of Alpha = 0.002**	
Lasso	SaleType_New	0.132506	SaleType_New	0.118491
	OverallQual_Excellent	0.126516	CentralAir	0.100391
	OverallQual_Very Excellent	0.117994	Functional_Typ	0.096175
	Neighborhood_Crawfor	0.107622	Neighborhood_Crawfor	0.082931
	Functional_Typ	0.100549	OverallQual_Excellent	0.061243

The Top 5 Predictors Post changing value of alpha from **8.0** to **16.0** for **Ridge Regression** are mentioned below

Top 5 Predictors				
	Original Value of Alpha = 8.0		**New Value of Alpha = 16.0**	
Ridge	Functional_Typ	0.092885	Functional_Typ	0.084971
	Neighborhood_Crawfor	0.092118	Neighborhood_Crawfor	0.078169
	OverallQual_Very Good	0.087309	CentralAir	0.074009
	Neighborhood_StoneBr	0.080271	OverallQual_Very Good	0.072914
	OverallCond_Excellent	0.077694	OverallQual_Excellent	0.060674

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: The Optimal Value of Lambda determines the amount of Regularization. The More Lambda you apply, the more regularization you achieve. Thereby it gives a simpler model in general.

The Choice to choose is totally dependent upon the model you desire. For Example, if you want a smaller number of variables it can be achieved with Lasso Regression as it offers reduction in features as it sets some of the coefficient values to 0.

On the comparison Ridge Regression offers more features as it does not set the coefficients to 0 but narrows them close to 0. Thereby preventing overfitting and avoids extreme values of coefficients.

My opinion is to go with Lasso in this case. Although it's giving me a slightly less R2 score as compared to Ridge as it makes eliminates more features and makes model simpler. That's why I choose it.

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 Top 5 predictors in my Lasso Model are

1. **SaleType_New**
2. **OverallQual_Excellent**
3. **OverallQual_Very Excellent**
4. **Neighborhood_Crawfor**
5. **Functional_Typ**

Post dropping the above 5 variables, the new top 5 predictors in my Lasso Model are

1. **SaleCondition_Partial,**
2. **CentralAir,**
3. **OverallCond_Excellent,**
4. **MSZoning_FV,**
5. **OverallQual_Very Good**

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: The model can be made more robust and generalizable by reducing Overfitting which can be achieved by Regularization by choosing optimal value of alpha. If we focus on more accuracy then there're high chances of overfitting thus the model will perform poorly on unseen test data and it won't be sustainable and generalisable. Thus, we need to ensure that our model is trained using

required features only by proper analysis of data and tuned and does not overfit thereby comparable accuracy is observed on training and test data.