Advanced Regression – Aasheesh Kumar Rai

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?

Ans :

The Optimal value of Alpha for Ridge and Lasso for the data set was :

Best alpha value for Lasso : 0.001

Best alpha value for Ridge : 7,0

When the Value of Alpha is Doubles the R2 Score is changes reasonably.

```
## Final Model.
lasso = Lasso(alpha=0.001)
lasso.fit(X_train,y_train)

y_train_pred = lasso.predict(X_train)
y_test_pred = lasso.predict(X_test)

print("R2 Score Train: "+str(round(r2_score(y_true=y_train,y_pred=y_train_pred),2)))
print("R2 Score Test: "+str(round(r2_score(y_true=y_test,y_pred=y_test_pred),2)))
```

```
R2 Score Train: 0.88
R2 Score Test: 0.86
```

```
## Final Model Doubled.
lasso = Lasso(alpha=0.002)
lasso.fit(X_train,y_train)

y_train_pred = lasso.predict(X_train)
y_test_pred = lasso.predict(X_test)

print("R2 Score Train: "+str(round(r2_score(y_true=y_train,y_pred=y_train_pred),2)))
print("R2 Score Test: "+str(round(r2_score(y_true=y_test,y_pred=y_test_pred),2)))
```

```
R2 Score Train: 0.86
R2 Score Test: 0.85
```

	Featuere	Coef
0	MSSubClass	7.380908
4	OverallCond	0.029930
55	Neighborhood_Edwards	0.025206
13	FullBath	0.020092
5	MasVnrArea	0.016736
20	GarageArea	0.015784
143	BsmtQual_2	0.012891
201	GarageType_Basment	0.009967
70	Neighborhood_StoneBr	0.008194
32	MSZoning_RM	0.006972

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?

Ans :

The Optimal value of Alpha for Ridge and Lasso for the data set was :

Best alpha value for Lasso : 0.001

Best alpha value for Ridge : 7,0

After creating model in both Ridge and Lasso we can see that the r2_scores are almost same for both of them but as lasso will penalize more on the dataset and can also help in feature elimination i am going to consider that as my final model.

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?

5	MasVnrArea	0.016736
20	GarageArea	0.015784
143	BsmtQual_2	0.012891
201	GarageType_Basment	0.009967
70	Neighborhood_StoneBr	0.008194
32	MSZoning_RM	0.006972

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?

Ans:

A model needs to be robust and generalizable so that they are not impacted by outliers in the training data. The model should be generalizable so that the test accuracy is not lesser than the training data. The model should perform well under all the data set and not just the data set. Outliers should not be given more importance as they might miss out on the key elements of the data. Confidence interval can we used to have standardizing.