

## Assignment: Advanced Regression – Subjective Questions

### 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 :

As per the model: Alpha for **Ridge** is 10 and **Lasso** is 0.001

- $r^2$  score for **Ridge** is 92% and 87% for test and train values
- $r^2$  score for **lasso** is 90% and 87% for test and train values respectively

As we increase the alpha, model flexibility decreases making it more variant to the change in Training Data.  $R^2$  Square for lasso decreases.

- $r^2$  score for **lasso** is 88% and 86% for test and train values respectively

Root Mean Square Error also increases from 0.0233 to 0.0253

### 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 :

As per the model: Alpha for **Ridge** is 10 and **Lasso** is 0.001.

I would choose 0.001 as optimal value because it is very close to 0 making it more regularized in terms of less variant.

### 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?

Top 5 variables as per **ridge** are :

118	OverallQual_9	0.087
66	Neighborhood_Crawfor	0.078
76	Neighborhood_NridgHt	0.077
82	Neighborhood_StoneBr	0.075
117	OverallQual_8	0.065

Top 5 variable as per **lasso** are :

118	OverallQual_9	0.129
66	Neighborhood_Crawfor	0.099
255	Functional_Typ	0.081
117	OverallQual_8	0.078
76	Neighborhood_NridgHt	0.073

After dropping them from the model the next variables are :

Top 5 variables as per **ridge** are :

252	Fireplaces_2	0.074
250	Functional_Typ	0.069
52	LandContour_HLS	0.066
122	OverallCond_9	0.066
300	SaleCondition_Alloca	0.065

Top 5 variable as per **lasso** are :

250	Functional_Typ	0.086
252	Fireplaces_2	0.084
270	GarageCars_3	0.078
137	Exterior1st_BrkFace	0.070
296	SaleType_New	0.060

#### 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?

- r2 score for **Ridge** is 92% and 87% for test and train values
- r2 score for **lasso** is 90% and 87% for test and train values respectively

This means that the selected features from models are able to explain 90% of data.

With a minimal change in alpha the model variables and its accuracy is not deviating much. Ex : by changing regularization factor to 0.002 in case of Lasso the test score changed to 86% from 87%. Hence it can be said to be more generalizable. By changing the alpha of ridge to 20 the r2 score dint change at all.

The Root square mean error of lasso is 0.0233 changed to 0.025 by increasing the alpha from 0.001 to 0.002. hence by changing the alpha a bit model test R2 decreases because of increase in variance towards the unseen data.