

## **ASSIGNMENT: ADVANCED REGRESSION**

### **Question 1:**

**a.** What is the optimal value of alpha for ridge and lasso regression?

**Ans a:**

The optimal value of alpha is as follows:

For Lasso: 0.001

For Ridge: 20

---

**b.** What will be the changes in the model if you choose double the value of alpha for both ridge and lasso?

**Ans b:**

If we increase or double the value of alpha, the model will tend to more generalize itself as follows:

For lasso the  $r^2$  score for train and test will get affected and will reduce to 0.887 for train and 0.886 for test.

While in case of ridge the  $r^2$  score for train is 0.90 and for test is 0.89

---

**c.** What will be the most important predictor variables after the change is implemented?

**Ans c:** The most predicted variable after making changes for lasso are

	<b>Featuere</b>	<b>Coef</b>
<b>0</b>	const	<b>11.948414</b>
<b>3</b>	OverallQual	<b>0.108351</b>
<b>9</b>	1stFlrSF	<b>0.066050</b>
<b>10</b>	2ndFlrSF	<b>0.064925</b>
<b>20</b>	GarageCars	<b>0.058755</b>
<b>102</b>	Neighborhood_Crawfor	<b>0.052425</b>
<b>68</b>	Condition1_Norm	<b>0.050852</b>
<b>134</b>	Exterior1st_BrkFace	<b>0.049522</b>
<b>218</b>	Foundation_PConc	<b>0.042089</b>
<b>4</b>	OverallCond	<b>0.040180</b>

For ridge are: -

	<b>Feaure</b>	<b>Coef</b>
<b>0</b>	constant	<b>11.884115</b>
<b>3</b>	OverallQual	<b>0.090281</b>
<b>10</b>	2ndFlrSF	<b>0.060810</b>
<b>9</b>	1stFlrSF	<b>0.051216</b>
<b>102</b>	Neighborhood_Crawfor	<b>0.050410</b>
<b>20</b>	GarageCars	<b>0.047984</b>
<b>134</b>	Exterior1st_BrkFace	<b>0.046082</b>
<b>68</b>	Condition1_Norm	<b>0.044461</b>
<b>118</b>	Neighborhood_StoneBr	<b>0.043488</b>
<b>177</b>	CentralAir_Y	<b>0.041</b>

**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.2**

As we know that lasso regression tends to eliminate some variables by making their coefficient equal to 0 so lasso regression proves to be better than ridge regression. Moreover, if we talk about their model prediction and accuracy, both tend to give almost same result.

Eliminating the variables gives lasso regression an edge over ridge regression as it makes the model simpler to interpret.

### Question 3

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:**

After removing the top five important predictor variables following are the variable important in Lasso are as follows:

	<b>Featuere</b>	<b>Coef</b>
<b>0</b>	const	10.700239
<b>107</b>	RoofMatl_WdShngl	1.286724
<b>106</b>	RoofMatl_WdShake	1.180355
<b>101</b>	RoofMatl_CompShg	1.157105
<b>105</b>	RoofMatl_Tar&Grv	1.085711
<b>104</b>	RoofMatl_Roll	1.065561

#### **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.**

- By using techniques of advance regression.
- Do not underfit or overfit the data.
- Maintaining tradeoff between variance and bias.
- Using cross validation with multiple folds.
- Using lasso and ridge regression techniques.
- Hyper parameter tuning and choosing optimal value of alpha.

The accuracy is not the only measure we should look on while model building because it can be biased towards our training data. We should always try to generalize our model to make it more robust.