

### Advanced regression Subjective Assignment

#### 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: In Ridge the optimal value for alpha(lambda) is 10 and for Lasso it is 0.001.

If we double the value of alpha for ridge there will be reduction in coefficient. With lasso there will be less importance feature turning coefficient to 0.

Most of the predictors variables would still be the same in case if we double the value.

#### 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: In this case my preference would be Ridge regression as its stats are much on the higher side. However if our primary goal is to zero in feature selection we would use the Lasso Regression

#### 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?

Ans: Here is the list of top predictors after dropping the top 5 predictors from the Lasso mode:

- BsmtFinSF1
- Functional\_Typ
- 1stFlrSF
- Dwelling since 2-STORY 1945 & OLDER
- Neighborhood\_Somerst

#### 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 made **robust** and **generalizable** so that they are not impacted by outliers in the training data. The model should be accurate for datasets other than the ones which were used during training. Too much weightage should not be given to the outliers so that the accuracy predicted by the model is high.

To ensure that is not the case, the outlier analysis needs to be done and only those which are relevant to the dataset need to be retained. Those outliers which it does not make sense to keep must be removed from the dataset.

In General, for making your model more **robust** to outliers following ways can be adopted:

- We can use a model that's resistant to outliers. Tree-based models are generally not affected by outliers, while regression-based models are. If we are performing a statistical test, try a non-parametric test instead of a parametric one.
- We can use a robust error metric: Switching from mean squared error to mean absolute difference reduces the influence of outliers
- Winsorize the data- Artificially cap the data at some threshold.
- Transform the data - If the data has a very pronounced right tail, try a log transformation.