

Assignment 4 - 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?

A:

- The optimal value of alpha for Lasso : {'alpha': 0.001}, and the optimal value of alpha for Ridge : {'alpha': 0.9}
- When these optimal values for alpha are doubled as 0.002 for Lasso and 1.8 for Ridge
- The features now regarded as best are -

```
#FullBath : Full bathrooms above grade  
#BsmtFinSF2 : Type 2 finished square feet  
#WoodDeckSF : Wood deck area in square feet  
#HeatingQC : Heating quality and condition  
#ExterQual : Evaluates the quality of the material on the exterior
```

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?

A:

After creating model in both Ridge and Lasso, and after determining the optimal value of the lambda, 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 too. The choice of lasso as regularized advanced regression model, makes it more robust.

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?

A:

```
#MiscVal      : Value of miscellaneous feature
#BsmtHalfBath : Basement half bathrooms
#LowQualFinSF : Low quality finished square feet (all floors)
#BsmtFullBath : Basement full bathrooms
#HalfBath     : Half baths above grade
```

When we build a model removing the above Best 5 features, and test again - The best features are changed and these columns above, take up the best - common features in both Ridge and Lasso regression overall.

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?

A:

A model is made more robust and general so that they are not impacted by an outlier present in the dataset. Analytically, we need a generalisable model to understand the model and draw predictions on the test set that it has not seen before, so the general model is able to have the test accuracy to be on par with the training accuracy. Also, too much weightage must not be given to the outliers present so that we obtain a higher accuracy on training. Those outliers which it does not make sense to keep must be removed from the dataset. This would help increase the accuracy of the predictions made by the model. Confidence intervals can be used (typically 3-5 standard deviations). Also, while considering regularization, we must learn to balance between the various models' accuracy

and how it is rightfully fitting the dataset given, so that even test set is as good as the train.