

## SUBJECTIVE QUESTION AND ANSWERS – ASSIGNMENT PART II

### 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 1

The optimal value of alpha for ridge and lasso regression for my model are :-

-Best alpha value for Lasso: {'alpha': 0.001}

-Best alpha value for Ridge: {'alpha': 0.9}

If we choose to double the value of alpha for both ridge and lasso, the R2 score did not changed much. Changed just by 1% for my model.

Below are the most important features, best explaining the dataset:

1. MiscVal : \$Value of miscellaneous feature
  2. BsmtHalfBath : Basement half bathrooms
  3. LowQualFinSF : Low quality finished square feet (all floors)
  4. BsmtFullBath : Basement full bathrooms
  5. HalfBath : Half baths above grade
- 

### 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 2

The optimal value of alpha for ridge and lasso regression for my model are :-

-Best alpha value for Lasso: {'alpha': 0.001}

-Best alpha value for Ridge: {'alpha': 0.9}

I would choose to use lasso as lasso will penalize more on the dataset and can also help in feature elimination.

---

### 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?

#### Answer 3

The five most important predictor variable after excluding the 5 most important predictor variables would be:-

1. Neighborhood\_MeadowV
2. Neighborhood\_Crawfor
3. Neighborhood\_ClearCr
4. PoolArea
5. CentralAir

---

### 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?

#### Answer 4

A model needs to be made robust and generalizable so that they are not impacted by outliers in the training data. The model should also be generalizable so that the test accuracy is not lesser than the training score. The model should be accurate for datasets other than the ones used during training set.

Pointers to be kept in mind:-

1. Outlier detection,
2. Confidence intervals typically between 3-5 standard deviations

The above pointers will help standardize the predictions made by the model and also increase the accuracy.

If the model is not robust, it cannot be trusted for predictive analysis.