

### 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

Ridge Alpha 1

lasso Alpha 10

2. In Ridge Regression, R2score on training data has decreased but it has increased on testing data
3. In Lasso Regression, R2score on training data has decreased but it has increased on testing data
  - LotArea-----Lot size in square feet
  - OverallQual-----Rates the overall material and finish of the house
  - OverallCond-----Rates the overall condition of the house
  - YearBuilt-----Original construction date
  - BsmtFinSF1-----Type 1 finished square feet
  - TotalBsmtSF----- Total square feet of basement area
  - GrLivArea-----Above grade (ground) living area square feet
  - TotRmsAbvGrd----Total rooms above grade (does not include bathrooms)
  - Street\_Pave-----Pave road access to property
  - RoofMatl\_Metal----Roof material\_Metal

Predictors are the same but the coefficient of these predictors has changed.

### 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 for and why?

Answer

The r2\_score of the lasso is slightly higher than the ridge for the test dataset so we will choose lasso regression to solve this problem.

### 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

The five most important predictor variables are:

- 1stFlrSF-----First Floor square feet

- GrLivArea-----Above grade (ground) living area square feet
- Street\_Pave-----Pave road access to property
- RoofMatl\_Metal-----Roof material\_Metal
- RoofStyle\_Shed-----Type of roof(Shed)

#### 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?

#### Answer

The model should be generalized so that the test accuracy is not lesser than the training score. The model should be accurate for datasets other than the ones which were used during training. Too much importance should not be given to the outliers so that the accuracy predicted by the model is high. To ensure that this is not the case, the outliers 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. If the model is not robust, It cannot be trusted for predictive analysis.