Questions And Answers

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:-

- Optimal value of lambda for Ridge regression is 0.2
- -Optimal value of lambda for Lasso regression is 0.0001
- Doubling the values of alpha for Ridge and Lasso, we get 0.4 and 0.0002 lambdas

respectively. When we double the lambda values, we get the following metrics

- RIDGE:
  Train R2 score: 0.8383721131401312
  Test R2 score: 0.8152767694815412

  Train MSE: 0.0027510022071777643
  Test MSE: 0.003058392640295085

  Train RMSE: 0.05244999720855821
  Test RMSE: 0.05530273628216858

- LASSO:
  Train R2 score: 0.8243624734971988
  Test R2 score: 0.8262284361668115

  Train MSE: 0.002989454557995706
  Test MSE: 0.002877070038393919

  Train RMSE: 0.05467590472955803
  Test RMSE: 0.053638326207982284

When the alpha values are doubled, we can see slight decrease in R2 values (Lasso model), no difference observed in Ridge model.

Following are the most important predictor variables

- GrLivArea
- Exterior1st_BrkComm
- TotalBsmtSF
- ExterQual_Fa
- KitchenAbvGr
- Functional_Maj2
- GarageCars
- LotArea
- Functional_Sev
- ExterQual_TA

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:-

- RIDGE:
  Train R2 score: 0.8383721131401312
  Test R2 score: 0.8152767694815412
- LASSO:
  Train R2 score: 0.8243624734971988
  Test R2 score: 0.8262284361668115
- The R2 Score of Ridge regression is greater than that of Lasso regression. Ridge regression decreases the coefficients nearly equal to 0, which indicates that all the 30 (RFE) features are included in the Ridge model.
- When the dataset has a greater number of features, it is time consuming to use Ridge regression
- As the number of features increase, the model becomes too complex. As the complexity increases, the variance also increases but the bias compromise is reduced. On the other hand, Lasso regression pushes some of the coefficients to exactly 0, which implies it does feature selection. Here, with the optimal value of lambda 0.0001, Lasso regression selected 21 significant features while the Ridge regression contains all the (30) features.
- As the number of features are mitigated, the model is not as complex as Ridge and the variance is also shrinked with reasonable bias compromise Hence, better to use Lasso regression as it does have feature selection and thereby reducing the complexity of the model.

- The Mean Squared Error of Lasso is slightly lower than that of Ridge.
- Also, since Lasso helps in feature reduction (as the coefficient value of one of the features becomes 0), Lasso has a better edge over Ridge.
- Therefore, the variables predicted by Lasso can be applied to choose significant variables for predicting the price of a house,

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?

Answers:-

The top five variable that are significant for Lambda 0.0001 in lasso regression are.

- GrLivArea
- Exterior1st_BrkComm
- TotalBsmtSF
- ExterQual_Fa
- KitchenAbvGr

If the predators are not present in incoming data will create another model using lasso for regression and Lambda 0.0001 and predict new significant features

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:-

- As Per, Occam's Razor -the two models that show similar 'performance' in the finite training or test data, we should pick the one that makes fewer on the test data due to following reasons:-
- Simpler models are usually more "generic' and are more widely applicable
- Simpler models require fewer training samples for effective training than the more complex ones and hence are easier to train.
- Simpler models are more robust. Complex models. tend to change wildly with changes in the training data set Simple models have low variance, high bias and complex models have law bias, high variance
- Simpler models make more errors in the training set. Complex models lead to overfitting. They work very well for the training samples, but fail miserably when applied to other test samples. Therefore to make the model more robust and generalizable, make the model simple but not simpler which will not be of any use. Regularization can be used to make the model simpler. Regularization helps to strike the delicate balance between keeping the model simple and not making it too naive to be of any use. For regression, regularization involves adding a regularization term to. the cost that adds up the absolute values or the squares of the parameters of the model
- A complex model will need to change for every little change in the dataset and hence is very unstable and extremely sensitive to any changes in the training data. A simpler model that abstracts out some pattern followed by the data points given is unlikely to change wildly even if more points are added or removed. Bias quantifies how accurate the model is likely to be on test data. A complex model can do an accurate job prediction provided there is enough training data. Models that are too naive, for e.g., one that gives same answer to all test inputs and makes no discrimination whatsoever has a very large bias as its expected error across all test inputs are very high. Variance refers to the degree of changes in the model itself with respect to changes in the training data. Thus accuracy of the model can be maintained by keeping the balance between Bias and Variance as it minimizes the total error as shown in the below graph