

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 regression is 0.0051 if we increase alpha value the model makes the co-efficient 0. But in ridge-regression I got the value of 15 it means the alpha value increased from 0, so train will be increasing and error term as well.

If we double the alpha value, In Ridge-regression it makes the model not to be useful for all the data sets. i.e., make the model unfit as it increases the alpha value, we may get more errors in testing and training the model.

But in lasso we got alpha value as 0.0005we double it will make more co-efficient of the variable and the variable will become closes to 0.

The most important predictor changes after the implementation of ridge-regression is

- GrLivArea
- SalePrice
- TotalBsmt
- FullBath
- YearBuilt

The most important predictor changes after the implementation of lasso regression is sale price

- FullBath
- yearBuilt
- GrLivArea
- TotalBSMTSF
- GarageCars

OverallQual

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) To Increase the accuracy in the model, we need to consider the co-efficients.

The two regression techniques has its own functionalities, But the lambda value increases in the Ridge-regression, the model is not going to fit for all, it will also decreases the variance and increase the bias. Because of penalty is lambda times of sum of co-efficients.

But in Lasso-Regression, the parameter lambda the penalty is square of co-efficients and it's been identified by the cross-validation. Due to this lasso decreases and it makes the value closure to the zero. As value increases, the lasso shrinkages and the variable is equals to or closure to zero value, And the model will get neglected.

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) The five most important predictor variable are

1. GrLivArea
2. FullBath
3. yearBuilt
4. OverallQual
5. TotalBSMTSF

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) As per my understanding, the models is totally depends upon the accuracy, It is all depends on the bias and variance, and to make the model more robust and generalisable, if the model is simpler, then the bias should be increased.

and the variance should be decreased and it works on all kinds of data. We need to maintain the bias and variance equally, otherwise sometimes the training data will be good but it will show impact on testing data error, it leads to the model weak.