

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: The optimal lambda value in case of Ridge and Lasso is as below:

Ridge - 10 Lasso - 0.01

When we double the value of alpha for our ridge regression to 20 from 10, the model will now apply more penalty on the curve and will try to make it more generalized which in turn makes the model more simpler and not eager to fit all the data. We now get more error for both test and train.

Similarly, when we increase the value of alpha for lasso from 0.01 to 0.02 the model tries to penalise more and more coefficients of the predictor variable will reduce to zero, and hence the value of R^2 also decreases.

Now after the changes:

The top five important predictor variable for Lasso are:

OverallQual, GrLivArea, TotalBsmtSF, OverallCond & BsmtFinSF1

If we double the value of lambda for Ridge:

The top five important predictor variable for Ridge are:

OverallQual, Neighborhood_Crawfor, GrLivArea, OverallCond, & SaleCondition_Normal

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: The optimal lambda value in case of Ridge and Lasso is as below:

Ridge - 10 Lasso - 0.01

Since Lasso helps in feature reduction (as the coefficient value of one of the feature became 0), Lasso has a better edge over Ridge.

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?

Answer:

Top Five variable are:

- 1stFlrSF
- 2ndFlrSF
- GarageArea
- Fireplaces
- BsmtFullBath

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: TO ensure your model is robust and generisable, keep these points in check:

1. Given two models which show similar performance in train and test data set, we should use the simpler model.
2. Regularization can be used to make the model simpler, IT helps to make the balance between the model being simple and too naïve to be of no use.
3. The accuracy of the model should be maintained by keeping the balance between Bias and Variance as it minimizes the total error.