Assignment Part-II

Q1. 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?

Ans:

Optimal value of alpha for ridge regression is 2.0 and Optimal value of alpha for lasso regression is 0.005.

By increasing the alpha value twice than before, we see there is a decrease in r2 score value and also more number of variables coeffecient became zero in case of lasso regression which makes the number of significant variables for the model decrease.

After increasing the alpha twice as before, the most important predictor variable for

Lasso is : "OverallQual"

Ridge is : "OverallQual"

Q2. 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?

Ans:

After applying ridge and lasso regression. If we compare r2 score metric, ridge regression has better r2 score than lasso.

Lasso:

R2 score on train set : 0.9206796193678233

R2 score on test set   : 0.8751925741208658

Ridge:

R2 score on train set : 0.8256542666910853

R2 score on test set   : 0.789799055962569

But we always prefer our model to be simpler with lesser number of significant predictor variables and also with better r2 score.

so keeping that as a condition, we know that lasso regression model will nullify some of the insignificant predictor variables making the model left with lesser number of predictor variables and also with good r2 score.

Therefore we choose Lasso over Ridge regression.

Q3. 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?

Ans:

As per the model that we obtain the 5 most important predictor variables are listed below.

Five most important predictor variables with original dataset.

    1.OverallQual

    2.GarageCars

    3.GrLivArea

    4.2ndFlrSF

    5.BsmtFullBath

As per the given condition, after removing above mentioned five important predictor variables, if we rebuild our model then below are the top five important predictor variables.

    1.TotalBsmtSF

    2.FullBath

    3.BedroomAbvGr

    4.BedroomAbvGr

5.Neighborhood_StoneBr

Q4.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?

Ans:

A simpler model will be more robust and generalisable. but model cannot be more simpler as it increases bias.

Therefore a good model will be having a good tradeoff between bais and variance. However making a model more robust and generalisable usually involves in compromising little bit on the accurracy on train data. But it will be fine as the model will give better results when applied on unseen data.

Its implication for the accuracy of model is that the model we obtain will give better results on unseen data which are as good as the results obtained from the training set