

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:

- Optimum value of lambda for Ridge Regression is 20. Optimum value of lambda for Lasso Regression is 0.0001.

- When we double the value of alpha for both ridge and lasso, regularization decreases which in turn reduces the score. It becomes simple.

- In Ridge, the train and test dataset r^2 score value decreases slightly and mean squared error increases.

- In Lasso, the train and test dataset r^2 score value decreases slightly and mean squared error doesn't change .

- The most important predictor variables after the change is implemented are:

- LotArea
- OverallCond
- GrLivArea
- OverallQual
- Neighborhood_Edwards

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:

- Optimum value of lambda for Ridge Regression is 20. Optimum value of lambda for Lasso Regression is 0.0001.
- The Mean Squared Error of Lasso marginally is not exactly that of Ridge. Likewise, since Lasso helps in include decrease (as the coefficient worth of a portion of the highlights have become 0), Lasso has a superior edge over Ridge despite the fact that r^2 for Lasso is somewhat not as much as Ridge.
- Thusly, the factors anticipated by Lasso can be applied to pick critical factors for foreseeing the cost of a house.
- In Training set mean squared error in ridge is 0.11253251518022964 and in lasso is 0.11163861736717866
- In Test set mean squared error in ridge is 0.0976644636234026 and in lasso is 0.09758277588747241

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:

Five most important predictor variables in the lasso model are not available in the incoming data are:

Neighborhood_Somerst, Neighborhood_Edwards, OverallQual, GrLivArea, MSSubClass_30.0

After creating another model excluding the five most important predictor variables and other 5 predictors changing alpha are:

Neighborhood_NridgHt, Neighborhood_StoneBr, MSZoning_FV , MSZoning_RL

MSZoning_RH

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:

The demonstrate ought to be generalized so that the test exactness isn't lesser than the preparing score. The demonstrate ought to be precise for datasets other than the ones which were utilized amid preparing. As well much significance ought to not given to the exceptions so that the exactness anticipated by the show is tall. To ensure that usually not the case, the exceptions investigation should be done and as it were those which are significant to the dataset must be held. Those exceptions which it does not make sense to keep must be evacuated from the dataset. On the off chance that the show isn't vigorous, it cannot be trusted for prescient examination.

As per Occam's Razor,

- 1.Simpler models are usually more 'generic' and are more widely applicable.
- 2.Simpler models require fewer training samples for effective training than the more complex ones and hence are easier to train.
- 3.Simpler models are more robust
- 4.Simpler models make more errors in the training set.