Name: Vaibhav Dandge Batch: MLc36

Assignment Part-II

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:

In Ridge Regression the alpha value is 0 And for lasso is 1.

Important predictor variables after change will be::

- LotArea----------Lot size in square feet
- OverallQual---------Rates the overall material and finish of the house
- TotalBsmtSF------- Total square feet of basement area
- GrLivArea----------Above grade (ground) living area square feet
- TotRmsAbvGrd----Total rooms above grade (does not include bathrooms)
- Street_Pave--------Pave road access to property
- RoofMatl_Metal----Roof material_Metal
- OverallCond--------Rates the overall condition of the house
- YearBuilt----------Original construction date
- BsmtFinSF1--------Type 1 finished square feet

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:

we will choose lasso regression to solve this problem, as the r2_score of lasso is slightly higher than lasso for the test dataset.

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:

The five most important predictor variables::

11stFlrSF----------First Floor square feet

GrLivArea----------Above grade (ground) living area square feet

Street_Pave---------Pave road access to property

RoofMatl_Metal------Roof material_Metal

RoofStyle_Shed------Type of roof(Shed)

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:

Test accuracy is not lesser than the training score for that the model should be generalized. The model should be accurate for datasets other than which were used during training. Do not give importance to the outliers so that the accuracy predicted by the model is high.

Always ensure that this is not the case, so the outliers analysis needs to be done and only those which are relevant to the dataset need to be retained.

Remove unwanted outliners dataset. It cannot be trusted for predictive analysis, If the model is not robust

Github link::

https://github.com/vaibhavdandge/Adv_Regression_Assignment_price-of-houses