

US Based Housing Company

Surprise Housing

Advanced Regression - Sandeep Roy

Question 1- Part A

What is the optimal value of alpha for ridge and lasso regression?

Ridge Alpha is 1

Lasso Alpha is 10

Question 1- Part B

What will be the changes in the model if you choose double the value of alpha for both ridge and lasso?

Alpha-1

R2score(train) - 0.9313905456057254

R2score(test) - 0.9183341082365928

Alpha-10

R2score(train) - 0.9311438423499492

R2score(test) - 0.9178118359931824

R2 Score on the training data has decreased and has increased on the testing data.

Question 1- Part C

What will be the most important predictor variables after the change is implemented?

YearBuilt - Original construction date

BsmtFinSF1 - Type 1 finished square feet

TotalBsmtSF - Total square feet of basement area

GrLivArea - Above grade (ground) living area square feet

Street_Pave - Pave road access to property

OverallQual - Rates the overall material and finish of the house

OverallCond - Rates the overall condition of the house

TotRmsAbvGrd - Total rooms above grade (does not include bathrooms)

LotArea - Lot size in square feet

RoofMatl_Metal - Roof material_Metal

Summary - The predictors are the same but the coefficients of these predictors has altered.

Question 2

You have determined the optimal value of λ for ridge and lasso regression during the assignment. Now, which one will you choose to apply and why?

Since lasso's r^2_{score} for the test dataset is marginally higher than ridge's, lasso regression will be used to address this issue.

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?

The top five most important predictor variables:

1. 1stFlrSF - First Floor square feet
2. GrLivArea - Above grade (ground) living area square feet
3. Street_Pave - Pave road access to property
4. RoofMatl_Metal - Roof material_Metal
5. 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?

In order to ensure that the test accuracy is equal to the training score, the model needs to be generalized. When applied to datasets other than the ones used for training, the model ought to yield reliable results. To ensure that the model's predicted accuracy is high, the outliers shouldn't be given undue weight. Only those outliers that are pertinent to the dataset should be kept after an outliers analysis has been completed to make sure this is not the case. The dataset has to have the outliers that don't make sense kept eliminated. A model cannot be relied upon for predictive analysis if it lacks robustness.