

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?

Ans: Optimal value of alpha for ridge is 1 and for lasso is 10.

When changed the optimal value of alpha for ridge as 2 and for lasso as 20, predictors are same but the coefficient of the predictors has changed.

LotArea

OverallQual

OverallCond

YearBuilt

BsmtFinSF1

TotalBsmtSF

GrLivArea

GarageArea

Please find Assignment Part II Question 1 under the attached Advanced Regression assignment - Housing company.ipynb for analysis using Python.

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?

Ans: The R2 score of Ridge is greater than Lasso, so will choose Ridge regression.

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?

Ans: After removing the five important predictors LotArea, OverallQual, OverallCond, YearBuilt, BsmtFinSF1

The new five important predictors will be

MasVnrArea, TotalBsmtSF, 1stFlrSF, 2ndFlrSF, GrLivArea

Please find Assignment Part II Question 3 under the attached Advanced Regression assignment - Housing company.ipynb for analysis using Python

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?

Ans: The model should be generalized so that the test accuracy is not lesser than the training accuracy. The model should be accurate for datasets other than the ones which were used in training. Too much importance should not be given to outliers. Those outliers which does not make sense can be removed from the dataset. If the model is not robust, it cannot be trusted for predictive analysis.