

Problem Statement Part-2

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?

ANS1. For ridge and lasso regression Ridge Alpha 1 lasso Alpha 10

R2score of training data has decrease and it has increase on testing data

R2score on training data has decreased but it has increased on testing data 3.

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 - The r2_score of lasso is slightly higher than ridge for the test dataset so we will choose lasso regression to solve this problem

Q3) 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?

ANS

- 1- 11stFlrSF-----First Floor square feet
- 2- 2- GrLivArea-----Above grade (ground) living area square feet
- 3- 3- Street_Pave-----Pave road access to property
- 4- 4- RoofMatl_Metal-----Roof material_Metal
- 5- 5- RoofStyle_Shed-----Type of roof(Shed)

Q4) How can you make sure that a model is robust and generalizable? What are the implications of the same for the accuracy of the model and why?

ANS The test accuracy is not lesser than the training score. The model should be accurate for datasets other than the ones which were used during training. Too much importance should not be given to the outliers so that the accuracy predicted by the model is high. Outliers need to be analyzed and only relevant are taken. Non relevant outliers need to be removed to make it robust in nature.