

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:

Best alpha value for Lasso : {'alpha': 0.001}

Best alpha value for Ridge : {'alpha': 0.9}

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?

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

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?

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?

Ans: The model should be generalized so that 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.. If the model is not robust, It cannot be trusted for predictive analysis.