

Assignment - Advanced Regression

Problem Statement - Part 2

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:

Output for `model_cv.best_params_` returned following for ridge and lasso

Ridge: 500 {'alpha': 500}

Lasso: 500 {'alpha': 500}

Even after changing the alpha value the most important predictor variables did not change with respect to Ridge or Lasso but coefficient of these predictor has changed

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:

R² Score of Lasso is higher compared to ridge regression. Hence, we would be choosing Lasso Regression. Also, Lasso is giving feature selection option which makes the model simple and accurate.

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 after removing the top five predictor variables from lasso are:

- MasVnrArea
- BsmtFinSF1
- BsmtFinSF2
- BsmtUnfSF
- TotalBsmtSF

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:

To ensure that the model is robust and generalisable we must

- Make sure model accuracy remains intact with changing dataset.
- The model should not be too simple and also too complex.
- Use cross-validation and regularization techniques during model training

These models may not give us accuracy with test data but over the time when exposed to unseen data models should be reliable and should perform efficiently in real-world scenarios.

The balance between achieved through these techniques enhances the model's accuracy.