

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?

The optimal value of alpha for Ridge regression is 10.

The optimal value of alpha for Lasso is 0.001.

The metrics with original hyper parameters are as follows:

The parameters after doubling both the hyper parameters is as follows:

We can observe that the r^2 score is nearly the same. We can observe a small increase in RSS train for ridge regression.

For lasso, we can see a small decline in R^2 score and a small increase in the RSS score

The new predictors of Ridge regression are ['OverallQual', 'GrLivArea', 'TotalBsmtSF', 'Neighborhood_Crawfor', 'Neighborhood_Somerst']

The new predictors of Lasso regression are ['OverallQual', 'GrLivArea', 'TotalBsmtSF', 'OverallCond', 'Foundation_PConc']

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 choice of the model depends on the use case of the solution. If our intent is to find the important input variables that influence the target variables, we can use Lasso as it aids us in the problem by helping in feature selection. If we intend to keep the coefficient values as low as possible then Ridge is a better solution.

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 5 important variables after dropping first top predictors in Lasso are ['Electrical_FuseP', 'GarageQual_Gd', 'Exterior1st_Stone', '2ndFlrSF', 'MasVnrType_Stone']

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?

To make the model more robust, we can take the following precautions:

1. identify and remove multi collinearity up front.
2. More diligent outlier treatment.
3. Do not limit the model evaluation to only one metric.

To make model more generalizable do the following:

1. Keep complexity as low as possible.
2. Identify optimal predictors for model building.

Accuracy and complexity are directly proportional. We need to trade off complexity and accuracy to an accepted minimum