

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 –

The optimum value of alpha for ridge is 0.3 and for lasso it is 0.0001

Double the value of alpha for ridge is 0.6 and for lasso it is 0.0002

Compared to the model built with optimum value of alpha the changes that are observed in the model built with double the value of alpha are

- Slight reduction in the R² – Score on the train data set for both Ridge and Lasso
- Slight increase in the R² – Score on the test data set for both Ridge and Lasso
- Slight increase in the RSS, MSE and RMSE value on the train data set for both Ridge and Lasso
- Slight decrease in the RSS, MSE and RMSE value on the test data set for both Ridge and Lasso

The most important variables after the changes implemented are

GrLivArea	OverallQual	LotArea	GarageCars
OverallCond	Exterior2nd_Other	Neighborhood_Crawfor	BsmtFullBath
Neighborhood_NridgHt	HouseAge	Exterior1st_BrkComm	PoolArea
MSSubClass	LotShape_IR3	LotFrontage	

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 –

I will choose the Ridge. The reasons behind it are

- The R² Score of Ridge is slightly higher than Lasso

- There is not much variation in the RSS, MSE and RMSE values between Ridge and Lasso

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 top 5 most important predictor variables identified by lasso to be get dropped are

'GrLivArea', 'OverallQual', 'LotArea', 'HouseAge', 'GarageCars'

After creating another model after dropping the above predictor variables the five most important predictor variables are

'TotalBsmtSF', '1stFlrSF', 'GarageArea', '2ndFlrSF', 'FullBath'

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 is said to be generalisable when the difference between train and test accuracy is very minimal. The accuracy of the model is it should be accurate on the unseen (new) data. A model can be made robust by making it provide consistent in accuracy of predicting on the unseen data, high R2 score, minimal RSS etc.