

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 value of alpha for Ridge : `alpha=100`

Best value of alpha for Lasso : `alpha=0.01`

Doubling the value of alpha for Ridge gave same almost same R2 score. But it may lead to underfitting and also the number of parameters/Features is more

Based on both the Ridge and Lasso Models, the key predictor variables for predicting the Sale Price of Houses in Australia are as follows:

Lasso Model

GrLivArea
ExterQual_Gd
FireplaceQu_OK
GarageType_Attchd
GarageArea
KitchenQual_Gd
OverallQual_7
BsmtQual_Gd
BsmtFullBath_1
SaleCondition_Normal

Ridge Model

ExterQual_Gd
ExterQual_TA
Exterior2nd_VinylSd
FireplaceQu_OK
GarageType_Attchd
OverallCond_7
BsmtQual_Gd
SaleCondition_Normal
GrLivArea

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 :

Lasso helps in feature reduction (as the coefficient value of one of the feature became 0), I will choose Lasso to get a simpler and efficient model

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?

After removing the first 5 important predictors, as per the notebook, below are 5 predictor variables

2ndFlrSF

1stFlrSF

FireplaceQu_OK

OverallQual_7

OverallCond_7

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?

