

Question1: 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 : As per the calculations done:

- Optimal value of Alpha for Ridge Regression is: 9.0
- Optimal value of Alpha for lasso Regression is: 0.0004

When we double the value of alpha for our ridge regression the model will apply more penalty on the curve and try to make the model more generalized that is making model more simpler and no thinking to fit every data of the data set . Similarly when we increase the value of alpha for lasso we try to penalize more our model and more coefficient of the variable will reduced to zero, when we increase the value of our r2 square also decreases.

The most important variable after the changes has been implemented for ridge regression are as follows:-

OverallQual
MSZoning_RL
TotalBsmtSF
MSZoning_FV
OverallCond
MSZoning_RM
Foundation_PConc
GarageCars
BsmtFinSF1

The most important variable after the changes has been implemented for lasso regression are as follows:-

MSZoning_RL
GrLivArea
MSZoning_RM
OverallQual
MSZoning_FV
TotalBsmtSF
OverallCond
Foundation_PConc
GarageCars
BsmtFinSF1

Question2: 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:

Looking at the Variable coefficients, I will prefer to use Lasso Regression Results because it makes multiple coefficients to Zero. And hence it automatically drops the non-important variables and let us know the important variables along with its relation which will help us in deciding the sales price of the house more effectively.

Question3: 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:

If now, top 5 variables are not available then I will make use of the next 5 important variables .

TotalBsmtSF
OverallCond
Foundation_PConc
GarageCars
BsmtFinSF1

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 make sure my model is robust and generalizable:

- I will test it in on both Train and Test dataset and make sure to choose that Hyperparameter for which the accuracy (R square) on both Test and Train is close.
- Make sure no overfitting is happening. If overfitting is there then we try to add some bias to it (using regularization techniques)
- Also, I will try to use lesser number of independent variables to predict the dependent variable. It can be achieved by VIF or applying Regularization (preferably Lasso). Implication of doing above is that → Accuracy of the model decreases which is not good to have