

Assignment Part-II

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 optimal value of alpha for Ridge is: 30

The Optimal value of alpha for Lasso is: 0.0002

If we double the value for both Ridge and Lasso i.e., 60 and 0.0004, the model will apply additional penalty and hence more generalization.

The mean square error will increase for both the train and the test set.

The important variables from the new Lasso and Ridge model will be:

Lasso_new		Ridge_new	
TotalBsmtSF	0.102673	OverallQual	0.084324
OverallQual	0.093567	TotalBsmtSF	0.070602
GrLivArea	0.070426	GarageCars	0.046440
MSZoning_RL	0.051344	GrLivArea	0.042232
GarageCars	0.047314	OverallCond	0.040110

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:

The optimal value of lambda for both our regression model is:

Ridge: 30

Lasso: 0.0002

The R squared value is slightly more, and the mean square error of Lasso is slightly less than that of Ridge.

Moreover, Lasso also helps us in feature reduction (selection by assigning the coefficient to zero), we can move forward and use Lasso for this business problem.

Question 3

After building the model, you realized 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:

After dropping the 5 important variables. We have built the new lasso model and the new lambda is 0.004.

The 5 most important variables in the new Lasso model will be:

1. **1stFlrSF**
2. **GarageCars**
3. **2ndFlrSF**
4. **BsmtQual**
5. **KitchenQual**

Question 4

How can you make sure that a model is robust and generalizable? What are the implications of the same for the accuracy of the model and why?

Answer:

The model will be robust and generalizable when the model is simple and performs better on unseen data. We can compromise on R squared or accuracy of the model in order to have low variance. We know the concept of Bias-Variance trade off and how it helps making model generalized.

Here, as we have implemented lasso and ridge regression to make a regularized regression model. We compromised on bias to reduce the variance, so our final model becomes simple, robust, and stable.