Advanced Regression

Graded Assignment – Part-2

Case Study : House Price Prediction

Submitted By : Sandeep Kumar

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?

Regularization method	Ridge	Lasso
Optimal Value of Alpha	2.0	0.0001
If we Choose Double the Value of Alpha	At Alpha 2.0 MSE = 0.00452 R2_Train = 0.926 R2_Test = 0.817 At Alpha 4.0 MSE = 0.00423 R2_Train = 0.923 R2_Test = 0.844	At Alpha 0.0001 MSE = 0.00385 R2_Train = 0.923 R2_Test = 0.844 At Alpha 0.0002 MSE = 0.00379 R2_Train = 0.919 R2_Test = 0.847
Most Important Factor after setting the Alpha as Double	• Total_Area_sqft • OverallQual • GrLivArea • OverallCond • Neighborhood_StoneBr	• Total_Area_sqft • OverallQual • YearBuilt • OverallCond • 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?

Regularization method	Ridge	Lasso
Optimal Value of Alpha	2.0	0.0001
MSE and R2 Comparison	At Alpha 2.0 MSE = 0.00452 R2_Train = 0.926 R2_Test = 0.817	At Alpha 0.0001 MSE = 0.00385 R2_Train = 0.923 R2_Test = 0.844
Conclusion: • Mean Squared Error of Lasso is lower than Mean Squared Error of Ridge. • Difference between R2 Value for Train and Test Data is minimal in Lasso In this trial, Lasso Lambda is winner so we will go with Lasso Lambda.		

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?

- After dropping the top most important predictor variable in the Lasso Model we again created the Model and below are the top 5 predictor this time.
- TotalBsmtSF
- SaleType_CWD
- Neighborhood_StoneBr
- TotRmsAbvGrd
- No_Of_Bathrooms

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?

A Model will be considered robust and generalized if it follow below rules.

- If a model has minimum R2 value difference for Train and Test data then it is considered as accurate and robust.
- Regularizing method (Ridge and/or Lasso) helps to manage the Model complexity which in turn help to create a robust and accurate model.
- Before Creating Model, we should do proper Data Cleaning, Scaling, Removal of Outliers so that these data specific issues should not impact the Model Accuracy.