Subjective Questions Advanced Regression

- Ashok Vashist

Agenda

Subjective Question and Answers

Question 1

- What is the optimal value of alpha for ridge and lasso regression?

Ans : Optimum value of Alpha :

  - Ridge -> 2.0
  - Lasso -> 0.0001

- What will be the changes in the model if you choose double the value of alpha for both ridge and lasso?

At Apha 2.0 for Ridge	At Alpha 0.0001 for Lasso
MSE = 0.00452	MSE = 0.00385
R2_Train = 0.926	R2_Train = 0.923
R2_Test = 0.817	R2_Test = 0.844
At Alpha 4.0	At Alpha 0.0002
MSE = 0.00423	MSE = 0.00379
R2_Train = 0.923	R2_Train = 0.919
R2_Test = 0.844	R2_Test = 0.847

- What will be the most important predictor variables after the change is implemented?

	Ridge	Lasso
The most important predictor variables after the change is implemented (Alpha is doubled)	Total_Area_sqft	Total_Area_sqft
	OverallQual	OverallQual
	GrLivArea	YearBuilt
	OverallCond	OverallCond
	Neighborhood_StoneBr	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	At Alpha 0.0001
	MSE = 0.00452	MSE = 0.00385
	R2_Train = 0.926	R2_Train = 0.923
	R2_Test = 0.817	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 experiment , Lasso Lambda is winner so we should use 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?

We dropped the top 5 most important predictor variables in the lasso model and again created again model and got the below five most important predictor variables:

- TotalBsmtSF
- SaleType_CWD
- Neighborhood_StoneBr
- TotRmsAbvGrd
- No_Of_Bathrooms

Question4

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.

- A model needs to be made robust and generalizable so that they are not impacted by outliers in the training data.
- Too much weightage should not be given to the outliers so that the accuracy predicted by the model is high
- The model is expected to be as simple as possible and simpler models are considered as more ‘generic’, though its accuracy will be decreased but it will be more robust.
- Underfitting and overfitting are the problems associated with the model. Hence, it is important to have balance in Bias and Variance to avoid such problems. This is possible with “Regularization”.
- Regularization helps in managing the model complexity by essentially shrinking the coefficients towards zero. This avoids the model becoming too complex, thus reducing the risk of overfitting.
- If a model has minimum R2 value difference for Train and Test data then it is considered as accurate and robust.
- Confidence intervals can be used (typically 3-5 standard deviations). This would help standardize the predictions made by the model. If the model is not robust , it cannot be trusted for predictive analysis
- 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.