Advance Regression

Subjective Assignment Part-II

Abstract

Subjective Q &A submitted as part of the assignment on Advance Regression Module. The responses to these questions are influenced based on the model development exercise in Part-1 of the assignment which is submitted along with this per the given evaluation criteria

MOHAMMED RAHAMATHULLA

iiitB and UpGrad Executive PG Programme

Machine Learning & Artificial Intelligence

mohammedrahmat@gmail.com

Subjective Assignment – Advanced Regression

Q1. 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: Optimal Value of alpha in Lasso and Ridge is 0.0001 and 5.0 respectively.
When we double these values, the model performance remains same in both the cases.

Ridge the important predictor variables are: Neighborhood_StoneBr, GarageArea, Neighborhood_NridgHt,TotalBsmtSF,GrLivArea , KitchenQual,Neighborhood_Names,Neighborhood_Edwards,BldgType_TwnhsE,GarageFinish

Lasso the important predictor variables are - TotalBsmtSF ,SaleType_New ,MSZoning_RM,GarageType_Attchd ,GrLivArea,Neighborhood_NAmes ,Neighborhood_OldTown,KitchenQual ,SaleCondition_Partial ,RoofStyle_Gable.

Q2: 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: We observe that Both Ridge and Lasso give very similar results in terms of performance. Although Ridge model performs slightly(1%) better than lasso on the test dataset, we still decide to choose the Lasso model to apply finally. Lasso helps with feature elimination and as our dataset has over 130+ columns , so feature elimination can be an advantage in realising the most important predictor variables. Hence, our final model is Lasso with r2 score of 92 on train and 91 on test datasets respectively.

Q3: 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:
The five most important predictor variables in our Lasso model are - 'GrLivArea','GarageType_Attchd','MSZoning_RM','SaleType_New','TotalBsmtSF'.
If we remove these and rebuild the model, the five most important predictor variables now are – MasVnrArea ,Neighborhood_StoneBr ,Neighborhood_NridgHt,Fireplaces ,GarageArea

Q4: 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: By ensuring that our model is not over fitting and it is simple as much as we can manage, we can be optimistic that our model is robust and is pretty generalized. A more focus on accuracy by over fitting the model will derail it from the scope of generalization. However, accuracy will be decently good on train and test data when we achieve generalization.