Advanced Regression Assignment

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 to 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 for ridge and lasso regression are:

• Optimal Value of lambda for Ridge: 10

• Optimal Value of lambda for Lasso: 0.001

If the optimal value of lambda is very high i.e. there is more regularization and model will underfitted. We basically want model which do not overfit or underfit the data but definitely identify pattern of data. If we choose to double the value of alpha In ridge it will lower the coefficients and in Lasso we there would be chance to important variable magnitude values will be turning zero and it will consider not important factor to determine Y. and

The most important predictor variable after the change is implemented are those which are significant to derive model effectively.

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: Optimal Value of alpha for ridge and lasso regression are:

• Optimal Value of lambda for ridge: 10

• Optimal Value of lambda for Lasso: 0.001

As we got good score for both the models so we can go with Lasso regression as it results in model parameters such that lesser important features coefficients become zero.

Ridge: Train: 90.9 Test: 87.4 and Lasso: Train: 89.8 Test: 86.

We know Lasso regression is also work as feature selection as it helps to lower the magnitude of unreliable features to zero. And help to build statistically efficient model.If we have too many variables and one of our primary goal is feature selection, then we will use Lasso.

Whereas in Ridge Regression, we can reduce the coefficient magnitude upto zero. this is our prime goals, then we will use Ridge Regression.

Question 3: 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: On running the same notebook and removing the top 5 significant variables:

We found below variables as next 5 significant

1. GarageType_BuiltIn'
2. GarageType_Detchd
3. GarageType_No Garage
4. GarageType_Others
5. GarageFinish_No Garage

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:

- We Know robust itself represent meaning of not affecting by any external force A model is robust when any variation in the data does not affect its performance much.
- A generalizable model is able to adapt properly to new, previously unseen data, drawn from the same distribution as the one used to create the model.
- It can be also understood using bias-variance trade-off. Simple model high bias but low variance and more generic model. Generalizes model will give high accuracy on training and test data set.
- To make sure a model is robust and generalizable, we have to take care of overfit issue. This is because an very high variance. At any smallest change in data affects the model prediction heavily. It work on training data , but fails if new data will share.
- the model should be simpler.not too complex.

- If we look at it from the perspective of Accuracy, A simpler model is usually more generic than a complex model. This becomes important because generic models are bound to perform better on unseen data sets