Problem Statement - 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?

Ans:

1)The optimal value of alpha for Ridge Regression=**5**
R2 score (train) : 0.882787

R2 score (test): 0.857034

The optimal value of alpha for LassoRegression=**0.0001**
R2 score (train) : 0.879379

R2 score (test) : 0.859336

2) As the optimum values of alpha are doubled accuracy of the model is decreased in both Ridge and Lasso Regressions for both test and train data

3)Top Features: 'GrLivArea', 'OverallQual', 'Neighborhood_NoRidge', '2ndFlrSF' , 'LotArea'

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

Ans:

We'll go with Lasso because it has a feature selection option. It removed unwanted features from the model without affecting its accuracy. As a result, our model is generalised, simple, and accurate.

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?

Ans:

'GrLivArea,' 'OverallQual,' 'Neighborhood NoRidge,' '2ndFlrSF,' and 'LotArea' are the top five features. Model accuracy was dropped alittle bit after they were removed. Following the removal of the top five predictors, the top five features are: 1stFlrSF,FullBath,MasVnrArea,GarageArea,Neighborhood Edwards

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?

Ans:

The Occam Rajor principle must be followed, and the model should be as simple as possible. Outliers should have no effect on the model. Too much emphasis on outliers when developing

When using test data, the model can have an impact and fail. It must be ensured that the model's training and test data are accurate.The model must be comparable. Otherwise, the model may underperform in untested data.High training r2 and low test r2 indicate overfitting, which is not universal. The regularisation method should be used to keep the model as simple as possible. It is punishing if the model becomes more complex.

The regularisation method aids in the achievement of the Bias-Variance trade off. It strikes a balance by increasing bias until the Total Error is as low as possible. This is also known as Optimum Model Complexity, and it occurs when the model is both simple enough to be generalizable and complex enough to be robust.