Assignment (Regularization)

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?

Sol: Best alpha value for Lasso : {'alpha': 0.001}
Best alpha value for Ridge : {'alpha': 2.0}

So if we double the value for Lasso :{'alpha: 0.002} and for ridge :{'alpha': 4.0}

In Lasso:

r_2 value has increased from 0.85 to 0.86 and important predictor value is MiscVal in both the cases but coef value has decreased from 1.35 to 1.07

In Ridge:

r_2 value has increased from 0.85 to 0.86 and important predictor value is MiscVal in both the cases but coef value has decreased from 1.09 to 0.80

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?

Sol: I'll choose to go with Ridge regression as the r2_value in ridge regression model is better compared with lasso regression model and the chance that the important variables may be deleted in the case of lasso regression

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?

Sol: LotArea, YearRemoveAdd_Old, BsmtFinType2, OverallCond, Landslope

Q4: 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?

Sol: A model is considered to be robust if the model is stable, i.e does not change drastically with the change in the training set.

A model is generalisable if it does not overfits the training data and works well with the new data.

It's implication in terms of accuracy is that a robust and generalisable model will perform equally well on both training and test data i.e the accuracy does not change much for training and test data.