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?

Answer1:

Optimal value of alpha for ridge: **10**

Optimal value of alpha for ridge: **100**

After make the double alpha for ridge and lasso i.e. **20 and 200**

For Ridge: ***Coeff values are increasing as alpha will increase.r2_score of train data is also drop from .84 to 0.82***

For Lasso: ***Coeff values are increasing as alpha will increase.r2_score of train data is also drop from 0.84 to 0.82***

Top Features: Neighborhood_NoRidge, Neighborhood_NridgHt, 2ndFlrSF, Neighborhood_Crawfor , 1stFlrSF

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 2:

We will choose **Lasso** as its giving **feature selection** option also. It has removed unwanted features from model without affecting the model accuracy. Which makes are model generalized and simple and accurate.

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 3:

Top 5 features are **Neighborhood_NoRidge, Neighborhood_NridgHt, 2ndFlrSF, OverallQual, Neighborhood_Veenker**. After dropping them model accuracy reduced from 84 and 84% to 77% and 77%. Now top most features are: **Next top 5 features** after droping 5 main predictors 1stFlrSF, MSSubClass_90, MSSubClass_120, TotalBsmtSF, HouseStyle_1Story

<u>Question 4:</u>

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?

<u>Answer 4:</u>

To make model robust and generalisable 3 features are required:

1. **Model accuracy** should be > 70-75%: I our case its coming 84%(Train) and 84%(Test) which is correct.
2. **P-value** of all the features is < 0.05
3. **VIF** of all the features are < 5

Thus we are sure that model is robust and generalisable.