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 and lasso regression are

- 10
- 100

If we double the value for alpha for ridge and lasso. The changes will be as follows: 20 and 200

For Ridge: Coeff values will increase as alpha increases. r2_score of train data
dropped from .81 to 0.65

For Lasso: As we increased the alpha value more features were removed from the model.
r2score dropped a bit in both test and train data

Top Features:

1. Neighborhood_NoRidge
2. Neighborhood_NridgHt
3. OverallQual
4. 2ndFlrSF
5. Neighborhood_Veenkar

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 it is giving a feature selection option.In addition It has removed unwanted features from the model without affecting the model accuracy. This helps in making the model simple, generalized and increases accuracy.

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?

Answer 3: Top 5 features are

1. Neighborhood_NoRidge
2. Neighborhood_NridgHt
3. OverallQual
4. 2ndFlrSF
5. Neighborhood_Veenkar

After dropping them model accuracy reduced from .81 to .65.

New top most features after dropping 5 main predictors are

1. BsmtUnfSF',
2. 'TotalBsmtSF',
3. '1stFlrSF',
4. 'GarageYrBlt',
5. 'MSSubClass_45'

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?
Answer 4:
To make model robust and generalisable 3 features are required:

1. Model accuracy should be > 70-75%: I our case it's coming around 80%
2. P-value of all the features is < 0.05
3. 3. VIF of all the features are < 5

Thus we can conclude that the model is robust and generalisable.