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 is 0.05 .
Optimal value of alpha for lasso is 20.
After doubling alpha for ridge and lasso i.e. 0.1 and 40 respectively.
For Ridge: Coeff values are increasing and r2_score of train data drops
For Lasso: As alpha value increased more features removed from model and r2score dropped both for test and train data
Top Features: Neighborhood_NoRidge, LotArea, TotalBsmtSF, Property_age

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: Next top 5 features after droping 5 main predictors 1stFlrSF, MSSubClass_90, MSSubClass_120, TotalBsmtSF, HouseStyle_1Story

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 Model accuracy should be > 70
In our case its coming 89%(Train) and 79%(Test) which is correct.