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?
Answer:
Ridge: 20
Lasso: 0.001
On doubling, coefficients decreased but r2 dropped.
Important Predictor variables after change:
OverallQual, Neighborhood_NoRidge, **Neighborhood_NridgHt,**

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:
Lasso can be selected as this is more generic. This helps us with feature selection option.

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:
Five most important predictors now:
HouseStyle_1Story, TotalBsmtSF, 1stFlrSF, 2ndFlrSF, TotalBsmtSF

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:
For a model to be robust and generalisable, the model should be as simple as possible.
Also Bias-Variance trade off can be helpful. Simple model will give more bias but less variance.
More accurate model will have equal accuracy for both train and test data.