## 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 1:**

Ridge:
Lambda = 2
R2 score (train): 0.917904
R2 score (test): 0.879608
MSE (train): 0.113599
MSE (test): 0.1407

Lasso:
Lambda = 0.001
R2 score (train): 0.9169
R2 score (test): 0.8779
MSE (train): 0.1142
MSE (test): 0.1417

After make the double alpha for ridge and lasso i.e., 4 and 0.002

For Ridge: Coefficient values of features have decreased as alpha increased. r2_score of train data is also drop from .917 to 0.915.
For Lasso: As alpha value increased more features got their coefficient 0 and were removed from model. Coefficient values of features have decreased as alpha increased. R2_score dropped from 0.9169 to 0.873 for train data.
Top Features: OverallQual_Poor, BsmtQual_No Basement, OverallQual_Excellent, Neighborhood_MeadowV, OverallCond_fair

**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:**

After creating model in both Ridge & lasso, we can see that the r2_score is almost same for both of them, but as Lasso will penalize more on data set & can help in feature elimination. There fore we are going to consider that our final model.

**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 Features: OverallQual_Poor, BsmtQual_No Basement, OverallQual_Excellent, Neighborhood_MeadowV, OverallCond_fair

After dropping them r2 score on train set reduced from 0.91 to 0.889 and on test set from 0.87 to 0.85.

Now topmost features are: Neighborhood_NridgHt, MSSubClass_1-STORY 1945 & OLDER, GrLivArea, Neighborhood_Edwards, Neighborhood_Crawfor.

**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:**

We should ensure that the model is robust and generalisable by regularizing the model and using a regularisation term with the RSS because the hyper parameter will ensure to strike the right balance between the model being too simple or too complex. Making the model more generalisable may take a toll on accuracy up to some extent but we can also have a look at the precision and recall of the model because sensitivity and specificity also play an important role in the model

evaluation criteria. Together if all three are above average we may accept the model. A very accurate model may have a chance of getting overfitted.