## Assignment-based Subjective Questions

**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:**

Below are the optimal values of alpha -

- Ridge regression = 0.1
- Lasso regression = 0.0001

After doubling the value of the alpha of Ridge regression -

- alpha = 0.2
    - R Squared on train set = 95%
    - R Squared on test set = 78%
    - RMSE = 0.035

After doubling the value of the alpha of Lasso regression -

- alpha = 0.0002
    - R Squared on train set = 95%
    - R Squared on test set = 69%
    - RMSE = 0.049

The most important predictor variables after the change are implemented -

- MSSubClass

**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:**

R Squared and RMSE on the test set is a little better than on Ridge regression so I will go with the Ridge regression model. Also, it works well when the number of predictors is very large like we have in this assignment.

Result from Ridge Regression Model:

- alpha: 0.1
- R Squared on train set = 95%
- R Squared on test set = 74%
- RMSE = 0.041

Lasso Regression Model:

- alpha: 0.0001
- R Squared on train set = 95%
- R Squared on test set = 65%
- RMSE = 0.056

**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:**

After building the model, five most important predictor variables in the Lasso model were:

- MSSubClass
- MSZoning_RL
- MSZoning_FV
- MSZoning_RH
- MSZoning_RM

After excluding the above five most important predictor variables, the new five most important predictor variables now:

- LotFrontage
- SaleType_ConLI
- Neighborhood_Edwards
- Neighborhood_Timber
- FireplaceQu_Fa

**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:**

The model should be as simple as possible. Accuracy may decrease of the model but it will be more robust and generalisable and it will also perform good on both train and tets data.