# Subjective Assignment – Advanced Regression

## 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:**

• In my final model, The optimal value of alpha for Ridge is 2 and for Lasso is 0.0001. • After doubling these values, the model performance remains same in both the cases.

 • In Ridge the important predictor variables are –

- MSZoning_RL   ,MSZoning_RM ,GrLivArea ,  OverallQual , MSZoning_FV

• In Lasso, the important predictor variables are

-  MSZoning_RL , GrLivArea , MSZoning_RM , OverallQual , MSZoning_FV ,

## 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:**

- The R2Score of Lasso (89.1%) is slightly higher than that of Ridge(88.8%).

- Also, since Lasso helps in feature reduction (as the coefficient value of one of the feature became 0), Lasso has a better edge over Ridge.

- Hence based on Lasso, the factors that generally affect the price are as follows.

  - MSZoning_RL
  - GrLivArea
  - MSZoning_RM
  - OverallQual
  - MSZoning_FV

# 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:**

The five most important predictor variables in our Lasso model are -

MSZoning_RL , GrLivArea , MSZoning_RM , OverallQual , MSZoning_FV

If we remove these and rebuild the model, the five most important predictor variables now are -

TotalBsmtSF , FullBath , HalfBath , Foundation_PConc , OverallCond

# Question 4

**How can you make sure that a model is robust and generalizable? What are the implications**

**of the same for the accuracy of the model and why?**

**Answer**:

The built model should over fitting and is as simple as possible, we are ensuring that it is robust and generalizable. We can achieve this by using hyper parameters and penalized methods like adjusted r2_score . The accuracy of the model will go up if we try to over fit the model but that no longer makes it generalizable. When the model is generalized the accuracy should be pretty good on both the training and the testing dataset making the model robust. Here in the above built model the accuracy is around 89 percent in the test data set making it a very good model to interpret.