

## Problem Statement 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?

**Ans:**

Optimal values of alpha for ridge is :50.0 and Lasso is 0.001.

-----  
Ridge:

-----  
Before:

R2 score (test) : 0.9067569209782779

RMSE (test) : 0.12654259850754085

After changing alpha to 100 for ridge:

R2 score (test) : 0.9052295739298476

RMSE (test) : 0.12757478990931834

-----  
Lasso:

Before:

-----  
R2 score (test) : 0.9040824949239074

RMSE (test) : 0.12834453537568824

After changing alpha to 0.002 for for lasso:

R2 score (test) : 0.9073016854877393

RMSE (test) : 0.1261723999869297

### 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?

**Ans:**

Here ridge has perform lillte better than Lasso. But We don't see much signficance diff. Will use Lasso model as final model becuase as it marks insignificance features to zero which we have already analyse as a part of EDA anlysis and also helps to keep our model simple.

### 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?

**Ans:**

The top five features are:

MSZoning\_RL-> 0.073488

2ndFlrSF-> 0.072756

LotArea-> 0.062613

MSZoning\_RM-> 0.059765

GarageCars-> 0.058023

### 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?

**Ans:**

To make sure our models are robust. We need to make sure out model doesn't overfit. Model overfits means it perform well and train test data but doesn't perform well on test data. A overfit models have high valience and small change in data affects the model peformance a lot. If a model using two many features means that models has remebers all the data along with noiseness and this kind of models isn called complex model. To protect form overfitting the model we add some kind biasness to our model. We Generally use Regularization technique using Lasso/Ridge and need to tune hypepareamaetr to get the optimal values of alpha.