

### 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:**

- Optimal value of alpha for ridge : 2.0, for Lasso : 0.0001
- If we double the alpha values of both ridge and lasso, the  $r^2$  value will change slightly, it also penalizes the model and a greater number of coefficients shrinks towards zero
- Most important predictor variables are :

### 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 optimal value of Lambda :
  - Ridge : 20
  - Lasso : 0.0001
- $R^2$  value
  - Ridge : Train : 0.83, Test : 0.84
  - Lasso Train : 0.832 , Test : 0.843
- Lasso helps in feature reduction and coefficient values of some of the variables shrink to zero so Lasso is better than Ridge

### 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 :

	Feature	Coef	mod
0	MSZoning	2.302743e+06	2.302743e+06
11	RoofStyle	1.494010e+04	1.494010e+04
32	GarageArea	1.259945e+04	1.259945e+04
28	TotRmsAbvGrd	-1.016340e+04	1.016340e+04
16	ExterCond	-8.092778e+03	8.092778e+03
9	HouseStyle	-7.474450e+03	7.474450e+03
30	GarageFinish	4.978562e+03	4.978562e+03
39	Total_Porch	4.919235e+03	4.919235e+03
5	LotConfig	4.657008e+03	4.657008e+03
20	BsmtFinType1	-4.026339e+03	4.026339e+03

### 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 :

- Cross validation : Split data into training and test and make more splits for validation of the model

- Regularization : Use ridge and Lasso to prevent overfitting
- Feature selection : Select and make more meaningful features from the data provided and make selection
- Hyperparameter Tuning : optimize hyperparameters using grid search or random search
- Unseen data: Test on unseen data for more robustness
-