

## Question 1

What is the optimal value of alpha for ridge and lasso regression?

**Alpha for Lasso is 0.0001 and Alpha for ridge is 4.0**

What will be the changes in the model if you choose double the value of alpha for both ridge and lasso?

Doubling alpha value is impacting coefficient and  $r^2$  score train,  $r^2$  score test, rss value train, rss value test, mse value train and mse value of test.

**Fore Ridge when alpha = 4.0, then**

r2 score train=0.9416391454684006
r2 score test=0.9132665845149273
rss value train=0.9390451579920435
rss value test=0.581594644781799
mse value train=0.0009504505647692748
mse value of test=0.001371685482975941

**Fore Ridge when alpha = 8.0, then**

r2 score train=0.9343140390179773
r2 score test=0.9105899025173947
rss value train=1.0569085066227908
rss value test=0.5995432509429006
mse value train=0.0010697454520473592
mse value of test=0.001414017101280426

**Fore Lasso when alpha = 0.0002, then**

r2 score train=0.9343707276049392
r2 score test=0.9089162353523307
rss value train=1.0559963687946639
rss value test=0.610766098041732
mse value train=0.0010688222356221293
mse value of test=0.0014404860802871037

**Fore Lasso when alpha = 0.0001, then**

r2 score train=0.9403499026334465
r2 score test=0.9092163239090469
rss value train=0.9597894951227478
rss value test=0.6087538413287865
mse value train=0.000971446857411688
mse value of test=0.0014357401918131758

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?

What will be the most important predictor variables after the change is implemented?

Ridge	Lasso
GrLivArea	GrLivArea
1stFlrSF	OverallQual_10
TotalBsmtSF	TotalBsmtSF
OverallQual_10	OverallQual_9
BsmtFinSF1	LotArea

## 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?

Ans2:

Lasso model is better because it makes features selection and makes 212 coeff as 0. it removes those features which does not make much impact in prediction.

## 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?

Ans3:

1stFlrSF	0.217290
2ndFlrSF	0.166643
BsmtFinSF1	0.126167
BsmtFinSF2	0.067566
BsmtUnfSF	0.063220

#### 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 4:

As per Occam's Razor simple model is robust and generalizable. But need to take care a model should be as simple as necessary but not simpler than that. It is basically trade-off between bias and Variance. We need to select correct lambda value which will give optimum bias and variance. When complexity increases, bias reduces and variance increases and when model is simple vise versa.