

### 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 –

Ridge alpha – 2

Lasso alpha – 0.01

When we increase the value of alpha we are penalizing our model. R2 score will be reduced and more number of coefficient will be reduced to zero.

Ridge – when alpha is 4 , below are the important parameters.

Variable	Coeff
constant	11.715
MSZoning_RH	0.138
MSZoning_FV	0.121
MSZoning_RL	0.121
Neighborhood_StoneBr	0.108

Lasso – when alpha is 0.02

The most important variable after the changes has been implemented for lasso regression are as follows:-

Variable	Coeff
constant	12.019
OverallQual	0.138
GrLivArea	0.109
GarageArea	0.045
OverallCond	0.034
Fireplaces	0.026
BsmtFullBath	0.021
TotalBsmtSF	0.017
FullBath	0.014

### 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 -

We saw that Ridge gives better performance than Lasso , but when we use Ridge, model become complex while Lasso have tendency to zero the coefficients . As we increase the value of lambda the variance in model is dropped and bias remains constant. Ridge regression includes all variables in final model but in Lasso Regression some of the coefficient becomes zero and model becomes simpler.

**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?

- GrLivArea
- OverallQual
- OverallCond
- TotalBsmtSF
- GarageArea

**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?

Robust and generalized model are simple and can perform better on unseen data than complex model. However accuracy will be decreased but they are robust. Simpler model are more biased and less variance. When bias is high model perform poor on training and testing dataset. When Variance is high .i.e. model performs very well on training data but poor on test data which is case of overfitting. Thus it is important to have balance between bias and variance to get optimal model.