

## 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 1:** As per the model that I have created, the optimal value of alpha for Ridge is 10 and for Lasso its 0.001.

When I double the value of alpha for Ridge, more penalty is being applied to the coefficients of the model which lowers the magnitude of them. In ridge regression, the coefficients will be lowered very close to 0 but will not be zero. However in Lasso, when the alpha is increased more and more feature's coefficients will become 0.

For Ridge regression, the most important predictors after doubling the alpha are:

LotFrontage	12.033395
BsmtFullBath	0.105813
OverallCond	0.075648
Neighborhood_Edwards	0.072950
MasVnrArea	0.063457

For Lasso regression, the most important predictors after doubling the alpha are:

LotFrontage	11.999633
BsmtFullBath	0.128506
OverallCond	0.093354
Neighborhood_Edwards	0.071142
MasVnrArea	0.063839

## 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 2:** I will choose the Lasso regression model, and the reason for choosing Lasso is because it gives higher penalty to the features which are not contributing to the target variable prediction thereby working as a feature reduction technique.

Since after lasso regression, I am left with lesser number of features, they are easy to analyze and the client can implement better measures on these limited features in order to achieve their goals.

### 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 3:** The five most important predictors after dropping the previous 5 most important predictors are as follows:

LotArea	12.040651
BsmtHalfBath	0.127536
BsmtFinSF1	0.109778
SaleCondition_Others	0.061745
WoodDeckSF	0.049018

### 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 4:**

To be sure of model being robust and generalisable, we need to follow following steps/pointers

1. The data used for training should represent the underlying insights and patterns better.
2. Normalization of parameters helps in making the coefficients better interpretable and helps in generalizing the model.
3. Applying regularization techniques helps prevent overfitting and improves generalization ability of the models.
4. Cross-validation helps in evaluating the model's performance on unseen data to make sure that the model generalize well.

The implications of making more generalisable and robust model is that it lowers the accuracy of the model at the cost of explainability.