

### 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:**

The optimal value for alpha for Ridge regression is: 20.0

The optimal value for alpha for Lasso regression is :0.001

Lasso Regression is a type of regression that uses shrinkage and is useful for Feature selection as it makes some of the coefficients to absolute zero. It encourages simple and sparse models

When doubling the value for alpha for both Ridge and Lasso, model complexity will have a greater contribution to the cost. The hypothesis selected when there is minimum cost, higher lambda will bias the selection towards models with lower complexity.

Doubling the values of alpha for both Lasso and Ridge does not significantly decrease the r-squared value and the important predictors still remain the same though their coefficients have changed.

### 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:**

Lasso Regression is a better option as we are predicting the important predictors affecting the SalesPrice.

Lasso makes it easier for the coefficients to be zero and therefore easier to eliminate some of the input variables not contributing to predict the output

Lasso can produce many solutions to the same problem, ridge can only produce one output

### 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:**

It might not be appropriate to build a model based on mere guess. How we define 'most important' is based on our goals and subject expertise. While statistics can help predict the most important variables, applying subject matter expertise to all aspects wouldn't be sufficient always as the data depends on the collection of samples whether biased or unbiased.

**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:**

A model needs to be made robust and generalisable so that they are not impacted by outliers in the training data. The model should also be generalizable so that the testing accuracy is not lesser than training score, leading to overfitting of the data. Too much weightage should not be given to the outliers so that accuracy predicted is very high. To ensure that this doesn't happen, outlier treatment such as capping the values, dropping the values higher than the IQR should be followed to have relevant data in place.