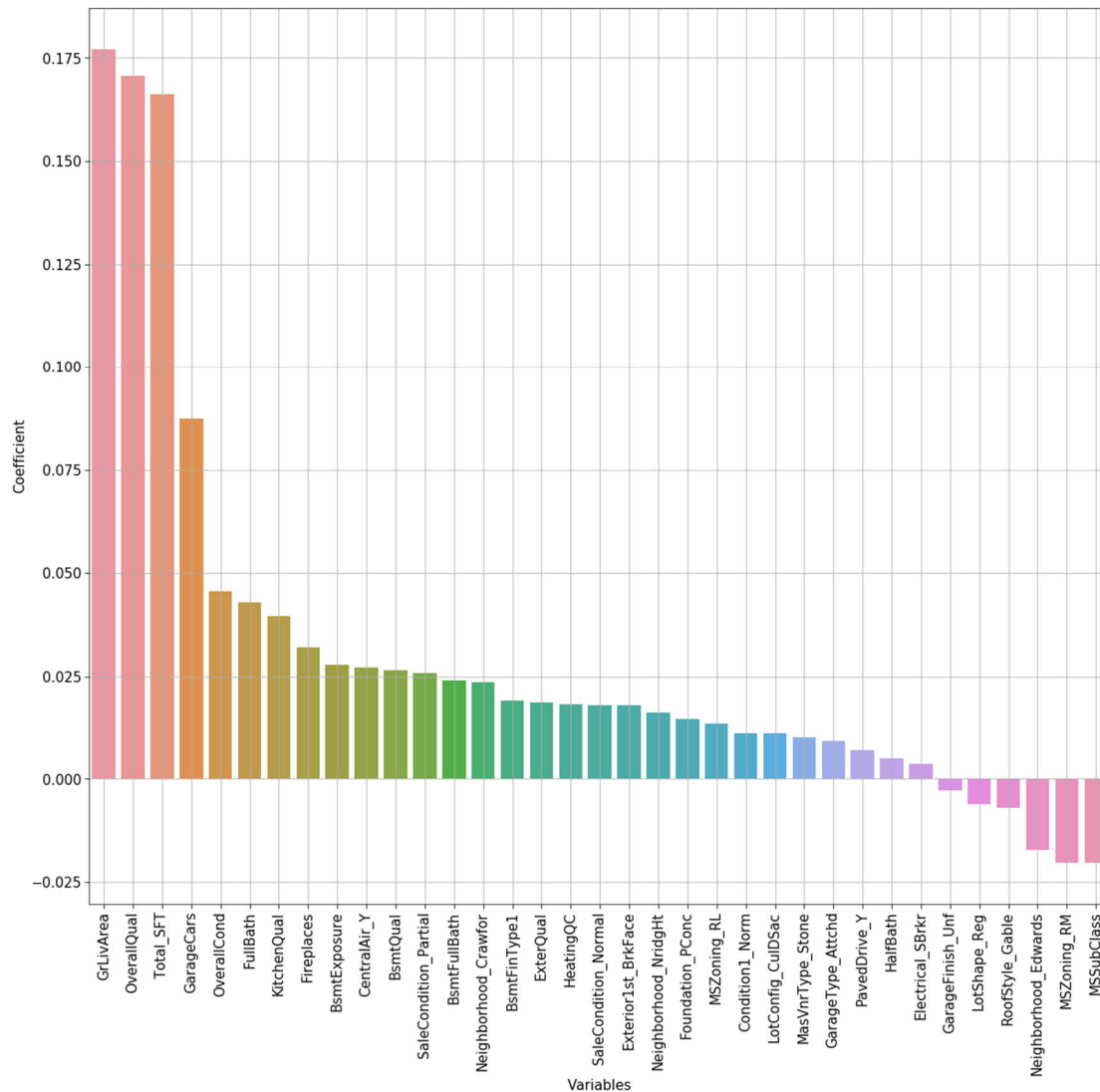


Question 1:

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:

- The optimal value of Alpha in Lasso is 0.001 with 35 features in the assignment.
- The optimal value of Alpha in Ridge is 1 which came in the assignment .
- If we choose double the value of alpha in Ridge and Lasso, more coefficients of variables will be moved towards zero.
- More alpha value penalize the coefficient of the variables.
- After the change is implemented, most important variables are GrLivArea and OverallQual.
- Most important predictor variables are based on absolute value of coefficients



## 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:

- I will go for Lasso because Lasso can make coefficient zero if the corresponding variable is not significant.
- Lasso helps in parameter shrinkage as well as variable selections.
- Ridge regression does not help in variable selection. It includes all variables in the model .

### Question 3

After building the model, you realized 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

- The five most important variables are - whose absolute value of coefficients are maximum.
- We need check also the change in R-Squared , because it explains the change in variance an variable explains.
- Sometimes statistical measure cannot show the importance of different predictor variables.
- At that moment, we need to consider subject area knowledge .

### Question 4

How can you make sure that a model is robust and generalizable? What are the implications of the same for the accuracy of the model and why?

Answer:

- A model will be robust when there is no impact of outliers to the model.
- During model building Sometime training score is high, but the test score becomes low. In that case the model does not perform well on unseen data. During training the model memorize the data points with noise. It is called overfitting.
- Usage of outlier detection technique like IQR and treat the outliers.
- We need to use robust scaling for to scale data.
- For skewed data and outlier, we can use log transformation.
- Usage of cross validation .
- For robust model we can use adjusted R-squared instead of R-squared. For classification we can use Sensitivity, specificity, precision and recall instead of accuracy.