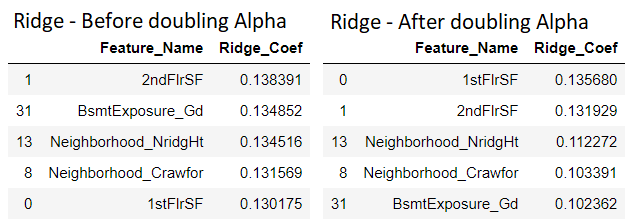
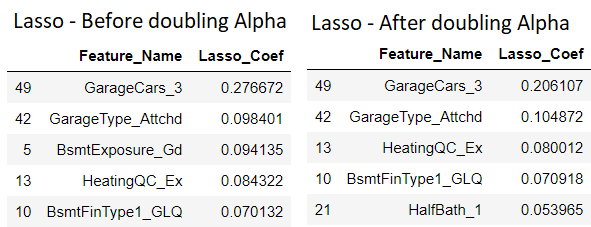
**Question 1**

**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?**

* From Ridge and Lasso plots of mean\_train\_error and mean\_test\_error it is clear that the optimal Alpha would be at around 12 and 0.005 respectively. At these Alpha values, the difference between train error and test error is minimum. As we increase the Alpha the difference is almost constant and the two curves are almost parallel to each other.
* After doubling the Alpha value, even though there is a change in the order, the top 5 predictor variables from Ridge remains same. We can also notice the drop in the coefficients of all variables except for “1stFlrSF” variable. Even though there is a hike in the coefficient of “1stFlrSF” the change is not significant with respect to the scale of coefficients.





* From above Lasso tables also, we can witness the drop in coefficients for majority of the variables and the top 5 predictor variables after doubling the Alpha value are almost same except for “HalfBath\_1” which is a new entry into top 5 after changing the Alpha value.
* So, it is clear that as we increase the Alpha value the cost function would increase and it will force the model to be simpler than we would like to, which may lead to underfitting of data.

**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?**

Lasso regression has multiple advantages. It will give good prediction accuracy because it shrinks the coefficients and it can also bring down the coefficients to absolute 0 which helps in reducing the variance of model without any substantial increase in Bias and performs feature selection efficiently. This is extremely useful when we have large number of features and small number of observations.

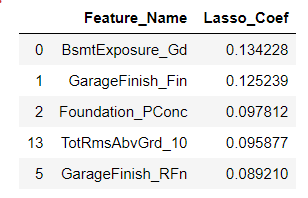
But on our data set, Ridge has performed better and it has very high accuracy and at the same time the Ridge model has 50 predictor variables which is extremely difficult for the business to act on such high number of variables, where as the Lasso model has only 20 predictor variables and it is easy for the business to take action items accordingly. The other 30 variables are zeroed out as they are insignificant with respect to Target variable.

Using “Oscam’s Razor” principle, which tells us that the model should be as simple as possible, I choose to pick the model given by Lasso Regression as it is simple and hence it is more generalizable and at the same time it has done relatively well on the dataset.

**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?**

After rebuilding the model by excluding the first 5 most predictor variables, below are the updated set of top 5 predictor variables along with their coefficients.



From above features it is quite intuitive that when we have a property with a greater number of Rooms and also furnished/semi furnished in our case garage with good basement exposure and with solid foundation then obviously they will lead to higher price tag of the property.

**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?**

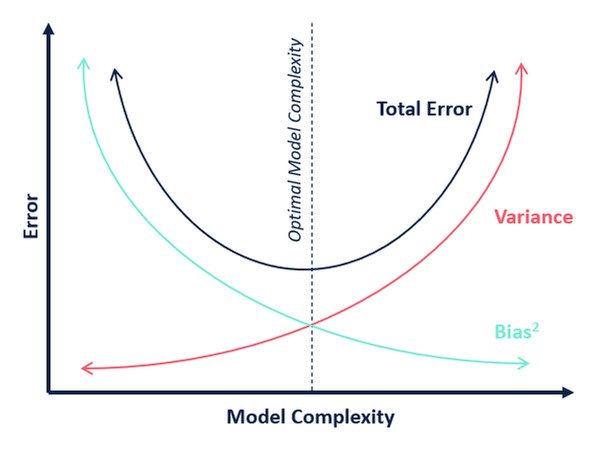
Ideally the model should be simple enough so that it performs well on unseen data and complex enough to not make too many mistakes. Then the model is robust and generalizable.

Simple models are usually more generic as they understand the underlying principles very well. If the model is too simple then it will only identify the dominant patterns in the data and may end up giving high bias.

Complex models will make far too many assumptions in training and they perform well on training data and gives high accuracy, but when we test the model on unseen data most of the assumptions made in training will fail on test data and will give poor accuracy on test data. This phenomenon is called Overfitting.

Complex model are Variance models which will swing wildly with small changes in the training data.

So complex models will have high variance but will give high accuracy in training but poor accuracy while validating the model on unseen data. Simple models have high bias and we may get relatively low accuracy with Simple models.



In the above figure, when the model complexity is low i.e the model is simple so the model will have low variance but it has high bias due to which the error is high and we may end up with low accuracies.

On the other hand, when the model has high complexity then it will have low bias but the model will swing wildly to small changes in training data as the variance of the model is high due to which the error is also high. So, we may end up with low accuracy on unseen data even though the model performs well on training data.

In order to make sure that the model is not too simple and not too complex and to minimise the error we need to strike balance between Bias and Variance at which the test error is also low and then the model starts performing well on both training data and unseen data which in turn leads to high accuracies on both training and unseen data.

When we strike the right balance between the Bias and variance, then the model is complex enough to understand the tricky relationships between variables and simple enough so that the model is generalizable.