**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 to 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:**

Identified optimal value of alpha for Ridge and Lasso regression are 20 and 0.001 respectively.

If we choose to double the value of alpha for both Ridge and Lasso, we may get a complex model that is likely to overfit. When both the alphas are double than optimal value, bias will be increased i.e., little far from origin, where the error contours are low, but the regularization contours might be very high. Such models are likely to overfit the data, as the error is quite low.

The most important predictor variables after the change (double the alpha value) are as follows:

**Ridge**

Most important predictor variables for Ride after double the alpha value (taken top 10 and bottom 10 of sorted coefficients as Ridge makes unimportant predictor variable’s coefficient to near to 0) are:

['Neighborhood\_Edwards', 'Neighborhood\_IDOTRR', 'BsmtFinType1\_Unf',

'BldgType\_Twnhs', 'Neighborhood\_Gilbert', 'MSSubClass', 'LotShape\_IR3',

'BsmtExposure\_No', 'GarageFinish\_Unf', 'HeatingQC\_TA', 'OverallCond',

'MSZoning\_RL', 'YearBuilt', 'Condition1\_Norm', 'MSZoning\_FV',

'GarageCars', 'Neighborhood\_NridgHt', 'Neighborhood\_Crawfor',

'OverallQual', 'GrLivArea']

**Lasso**

Most important predictor variables for Lasso after double the alpha value (taken top 10 and bottom 10 of sorted coefficients as Lasso makes unimportant predictor variable’s coefficient to 0) are:

['Neighborhood\_Edwards', 'BsmtFinType1\_Unf', 'MSSubClass', 'PoolArea',

'HeatingQC\_TA', 'BsmtExposure\_No', 'BldgType\_Twnhs', 'GarageFinish\_Unf',

'BsmtQual\_Gd', 'YrSold', 'OverallCond', 'Condition1\_Norm',

'MSZoning\_RL', 'GarageCars', 'MSZoning\_FV', 'Neighborhood\_NridgHt',

'YearBuilt', 'Neighborhood\_Crawfor', 'OverallQual', 'GrLivArea']

**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?**

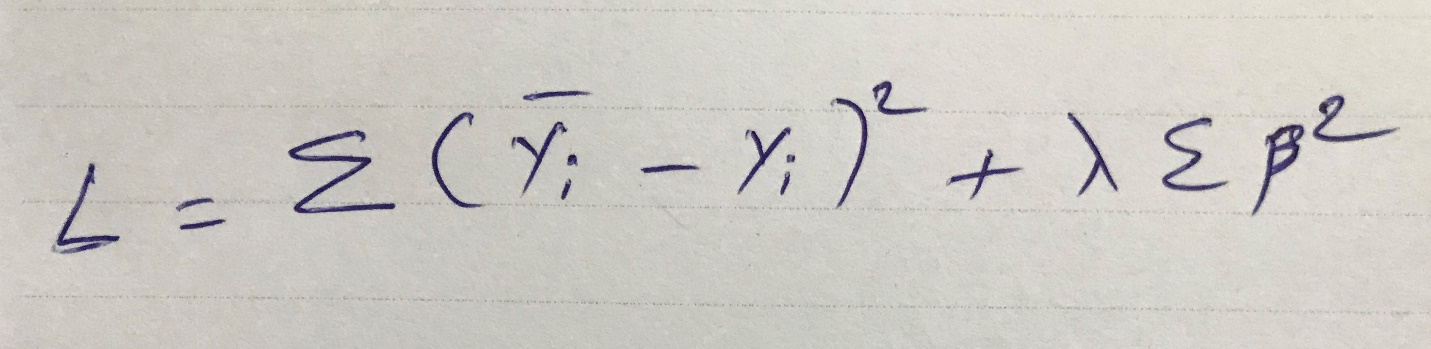
**Answer-2:**

I will choose Lasso regression. Because, Ridge regression makes irrelevant beta coefficients to near to zero but not zero. This way, Ridge keeps all coefficients in model.

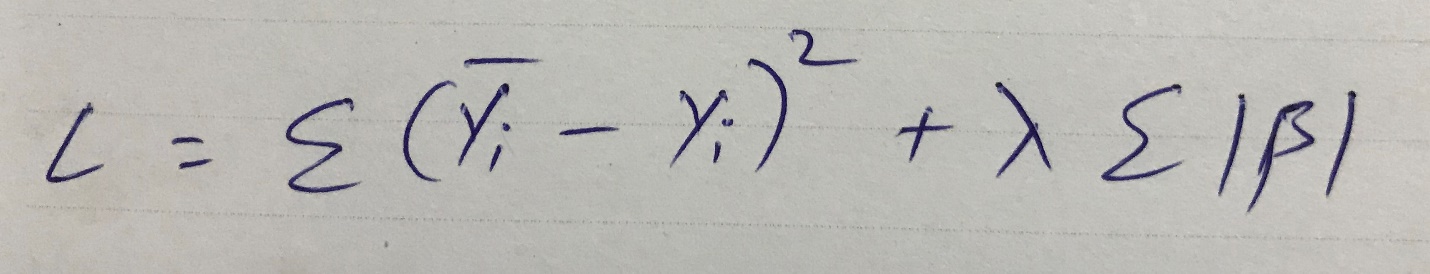
Whereas Lasso Regression makes the irrelevant beta coefficients to zero when lambda is large enough while ridge does not. Additionally, Lasso performs variable selection automatically.

Lambda parameter in Ridge is residual sum of squares of the coefficients whereas in Lasso it is sum of the absolute values of the coefficients. This difference has a huge impact on the bias-variance trade-off.

**Ridge Loss Function**



**Lasso Loss Function**



Like the best subset selection method, lasso performs variable selection. The tuning parameter lambda is chosen by cross validation. When lambda is small, the result is essentially the least squares estimates. As lambda increases, shrinkage occurs so that variables that are at zero can be thrown away.

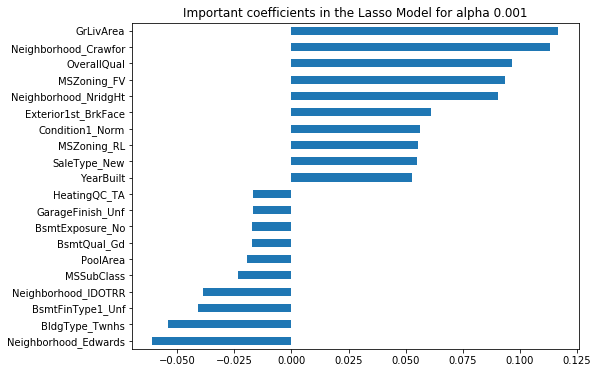
So, a major advantage of lasso is that it is a combination of both shrinkage (makes the irrelevant beta coefficients to zero) and selection of variables. In cases with very large number of features, lasso allow us to efficiently find the sparse model that involve a small subset of the features.

**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?**

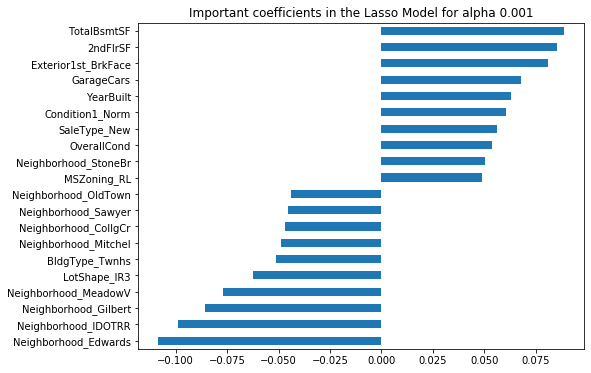
**Answer-3:**

The top five most important predictor variables returned by the final lasso model are:

**['GrLivArea', 'Neighborhood\_Crawfor', 'OverallQual', 'MSZoning\_FV', 'Neighborhood\_NridgHt']**

****

After removing them from train data set and recreating another model, returned next top 5 important variables by lasso model now are as follows:

**['TotalBsmtSF', '2ndFlrSF', 'Exterior1st\_BrkFace', 'GarageCars’, ‘YearBuilt']**

**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-4:**

A predictive model must be as simple as possible, but not simpler. There is an important relationship between the complexity of a model and its usefulness in a learning context because of the following reasons:

* Simpler models are usually more generic and are more widely applicable (i.e., they are generalizable).
* Simpler models require fewer training samples for effective training than the more complex ones.

Regularization is a process used to create an optimally complex model, i.e., a model that is as simple as possible while performing well on the training data. Through regularization, one tries to strike the delicate balance between keeping the model simple and yet not making it too naive for any use i.e., not underfitting as well as not overfitting too.

Any model can be said as robust and generalizable when the model gives us the least test error. Hence, the test error, and not only the training error, needs to be estimated in order to say that the model is robust and generalizable. Such model can be determined as a model having better accuracy.

This can be done in the following two ways:

1. Use metrics that consider both the model fit and its simplicity. They penalize the model for being too complex (i.e., for overfitting) and, consequently, more representative of the unseen ‘test error’. Some examples of such metrics are adjusted R2, AIC and BIC.
2. Estimate the test error via a validation set or a cross-validation approach. In the validation set approach, we find the test error by training the model on a training set and fitting on an unseen validation set, whereas in n-fold cross-validation approach, we take the mean of errors generated by training the model on all folds except the kth fold and testing the model on the kth fold, where k varies from 1 to n.

-End-