Problem Statement - Part II

- By Gargi Singh

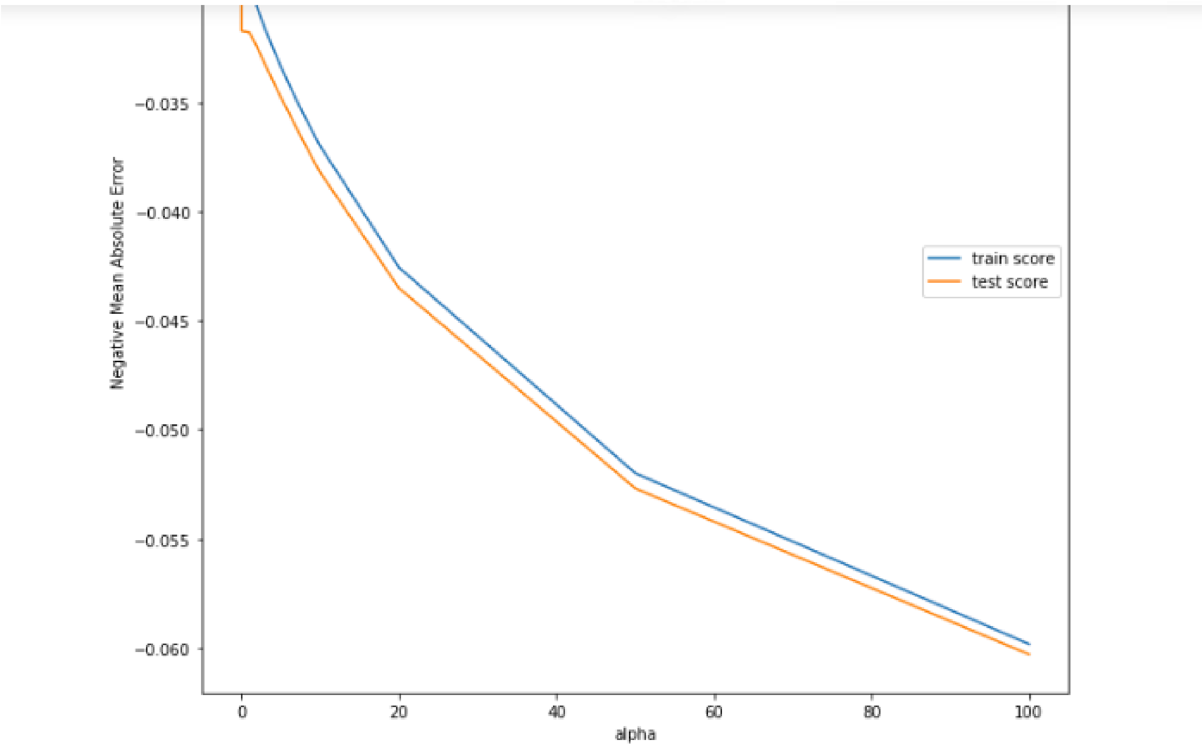
Question 1:

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 for ridge and lasso regression are:

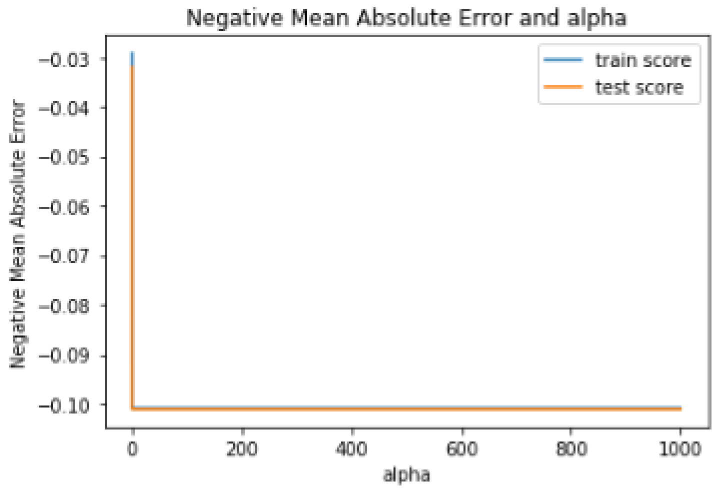
Ridge regression:0.01




```
[531]: print(model_cv.best_params_)

{'alpha': 0.01}
```

And the Lasso regression is:0.0001




```
print(model_cv.best_params_)
```

```
{'alpha': 0.0001}
```

Often we want conduct a process called regularization, wherein we penalize the number of features in a model in order to only keep the most important features.

if we double the value of alpha :

Ridge regression – With the increase in alpha leads to the decrease in the flexibility . As the alpha increases the flexibility of the ridge regression fit decreases leading to decreased variance but increased bias. In other words, the coefficients of the predictors decrease then their value in the model decreases. That is their effect decreases that is the flexibility of the model decreases.

This case happens in general when the regularization is done on the model.

Lasso regression:

When the lambda is small the result is essentially the least square estimates. As lambda increases shrinkage occurs so that variables that are zero can be thrown away.

That is, when alpha is 0, Lasso regression produces the same coefficients as a linear regression. When alpha is very very large, all coefficients are zero.

The most important predictor variables after the change is implemented is :

Ridge regression:

```
print(model_cv.best_params_)
```

```
{'alpha': 0.01}
```

```
alpha = 0.02
ridge = Ridge(alpha=alpha)

ridge.fit(X_train_rfe, y_train)
ridge_param=ridge.coef_
ridge_param
```

and the important predicted variable is

Ridge regression:

```
RoofMatl_Metal          Score: 1.0034695895255759
RoofMatl_Tar&Grv        Score: 0.9757516128069734
RoofMatl_WdShngl        Score: 0.9731342748236612
RoofMatl_CompShg        Score: 0.9579815530287603
RoofMatl_Roll           Score: 0.9557982761559901
```

RoofMatl_Metal is the most important variable.

Lasso regression

```
OverallQual                 Score: 0.26021127531178073
```

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:

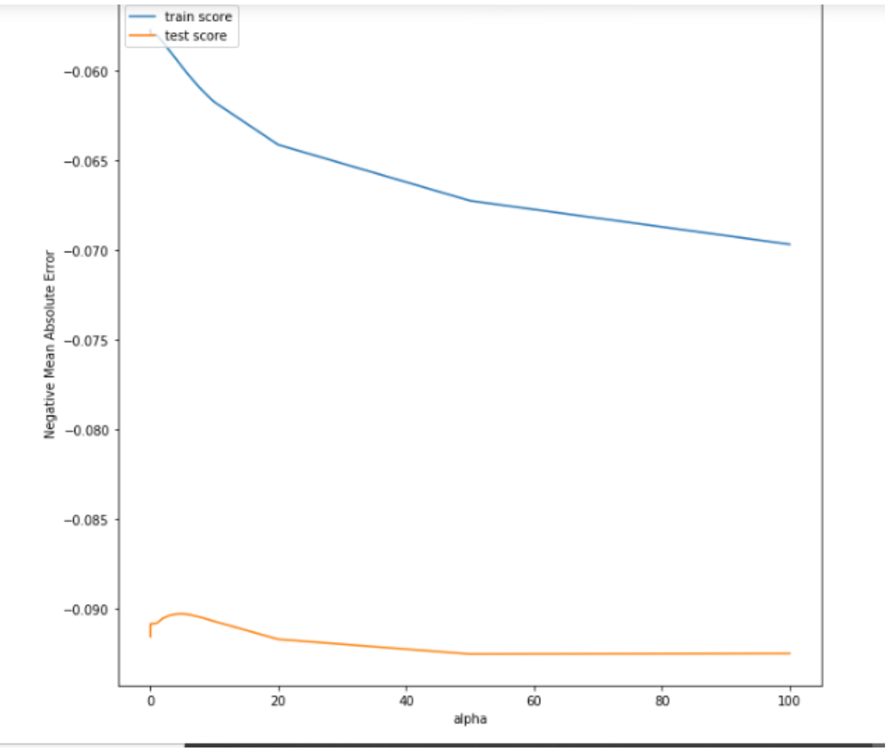
Regularization is the process which prevents overfitting by discouraging developers learning a more complex or flexible model, and finally, which regularizes or shrinks the coefficients towards zero. The basic idea is to penalize the complex models by adding a complexity term in such a way that it tends to give a bigger loss for evaluating complex models.

Ridge regression and Lasso regression. Both these methods are used to make the regression model simpler while balancing the 'bias-variance' tradeoff.

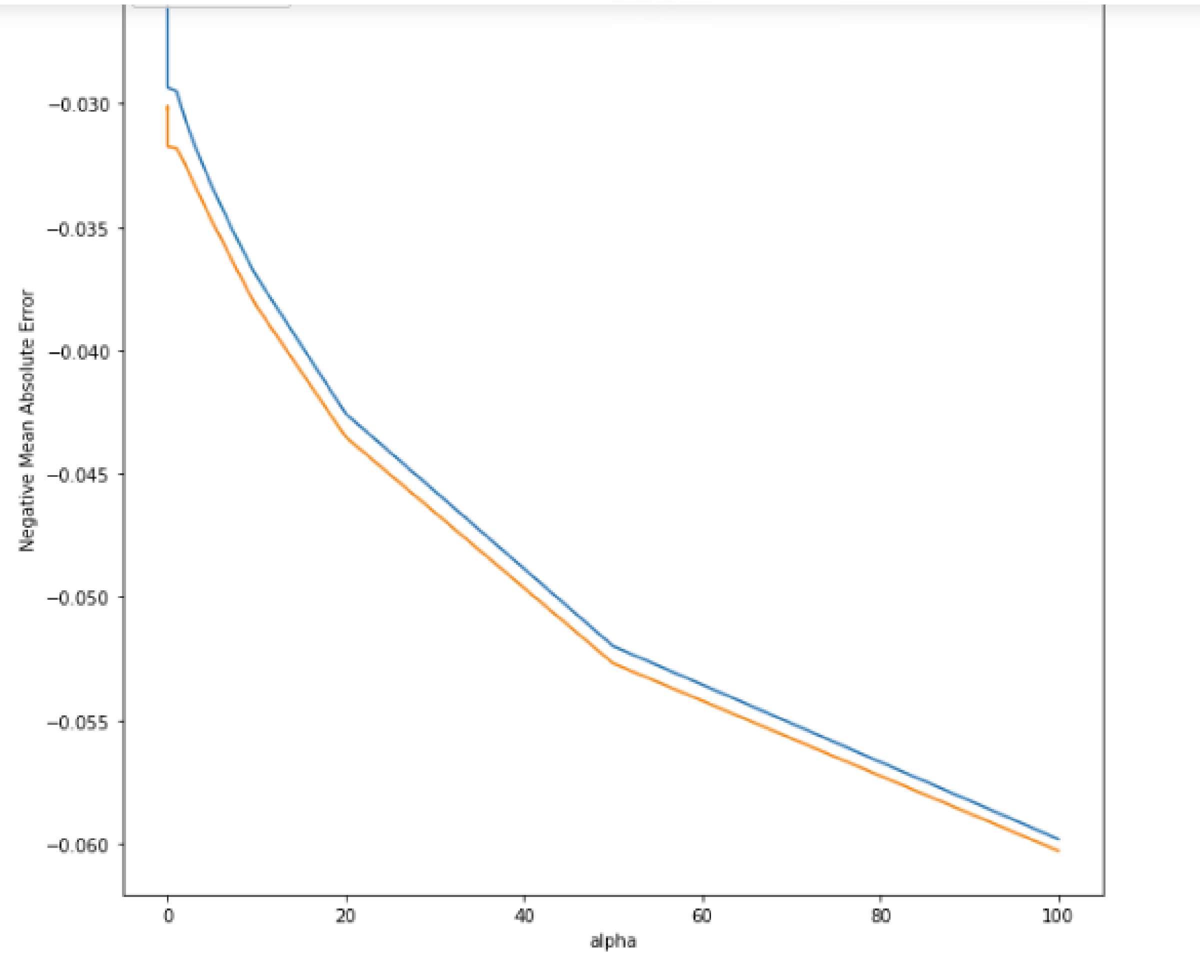
Ridge regression: an additional term of "sum of the squares of the coefficients" is added to the cost function along with the error term,

Lasso Regression: whereas in case of lasso regression, a regularisation term of "sum of the absolute value of the coefficients" is added.

We see from the plot that the test negative mean absolute error first increases and then decreases forming a bell curve. But the training negative mean absolute error keeps on decreasing as we increase the value of hyperparameter, which is in accordance with the bias-variance trade-off.



The above plot was obtained when the features without the rfe selection was passed and thus ridge regression was performed and thus by checking the plot we can see that the negative mean absolute error is initially increasing and thus after a certain alpha is decreasing thus we can keep the value where we get the really good high value of this metric and thus we will keep trader where best score for both the test score and the train score hence the alpha = 5.



Above plot is the ridge regression plot after we passed the rfe selected features the test and the train score is almost similar though but we can use another parameter to determine the best alpha

```
print(model_cv.best_params_)
```

```
{'alpha': 0.01}
```

So ridge regression puts constraint on the coefficients (w). The penalty term (lambda) regularizes the coefficients such that if the coefficients take large values the optimization function is penalized. So, ridge regression shrinks the coefficients and it helps to reduce the model complexity and multi-collinearity.

Note : setting λ to 0 is the same as using the OLS, while the larger its value, the stronger is the coefficients' size penalized.

Selecting lambda value that produces the lowest test mean squared error (MSE)

Lasso is the better option as it also does the feature selection which makes the model more robust.

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:

With the drop in the five variables selected by the lasso model . I build the mod el with the left over features and noticed that the alpha value in the ridge mod el has increased . Also the preference in the lasso model for the features has changed the feature in descending order are:

- 1stFlrSF
- TotalBsmtSF
- MSZoning_FV
- LotArea
- MSZoning_RL

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:

The model is said to be robust when an unseen data is been tested and the model performs its best .When a model is trained and once the model is tested on the unseen data the model should provide its best results and perform effectively.

Simpler models are usually more ’generic’. Simpler models are more robust.

So it is generally better for a model to be not too sensitive to the specifics of

the data set on which it has been trained. Complex models tend to change wildly with changes in the training data set.

Generalised model: In the generalised regression, you attempt to model the data using 'derived features', which could be linear or non-linear, rather than the original raw attributes. machine learning algorithms are used to fit a model to dataset we receive Then we go ahead with the training of the model. Like examples are presented to the model and the model tweaks its internal parameters to better understand the data. Once training is over, the model is performed upon new data and then uses what it has learned to explain that data.

Now here's where problems can rise . If you overtrain the model on the training data, then it will be able to identify all the relevant information in the training data, but will fail miserably when presented with the new data. We then say that the model is incapable of generalizing, or that it is overfitting the training data.

Affect of robustness on the accuracy of the model:

- One of the advantages of simplicity or generalised model are generalisability, robustness, making few assumptions and less data required for learning
- **Bias measures how accurately a model can describe the actual task at hand**:
  we aim to minimize error or increase accuracy. High bias means It is simply your model is less trained or less complex enough to predict any unseen example. So we cannot keep the model too simple too as it may fail to predict the unseen data resulting in underfitting .