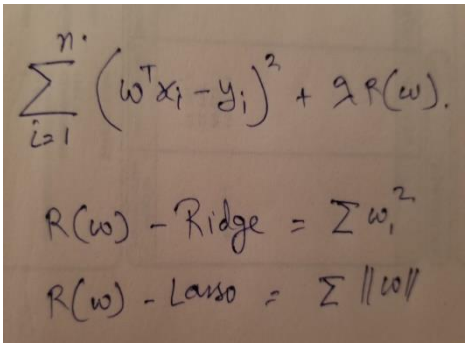
## Question 1

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?

The purpose of regularization is to ensure that model is not overtly complex. For ridge and lasso regression we penalize the model for its complexity. Regularized regression has 2 objective error term and regularized term.




Lambda is the coefficient for the regularization term R(w).

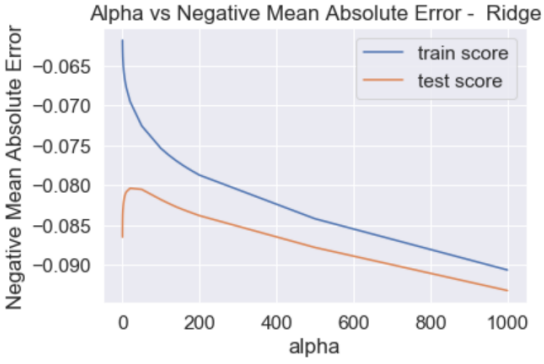
- Ridge uses sum of squared coefficients
- Lasso uses sum of absolute value of coefficients

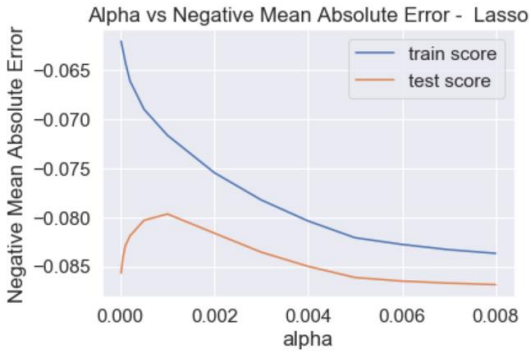
*Note – lambda discussed in theory above is indicated by alpha in scikit-learn package.*

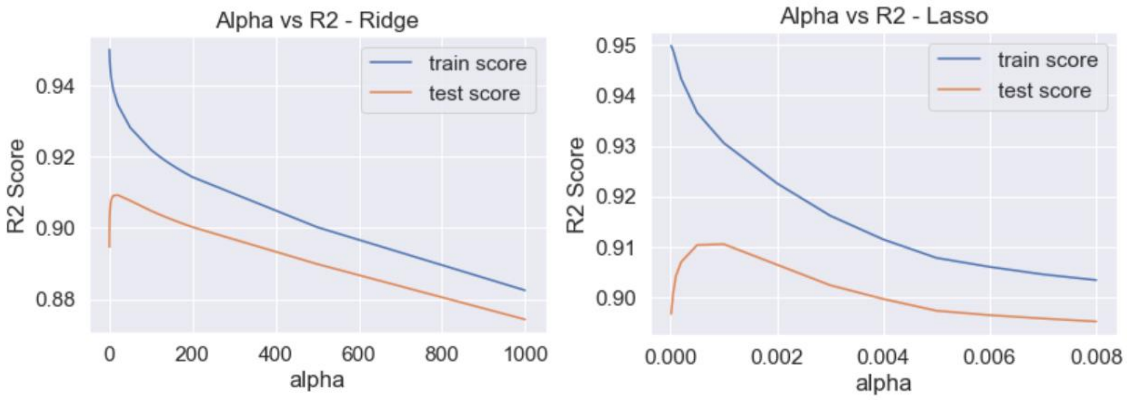
Optimal value of alpha for our models are as below:

- Ridge regression – 20
- Lasso regression – 0.001

As we increase the value of hyperparameter then premium on regularized term increases. At optimal value of hyperparameter alpha, the value of error term is minimum, and the score is maximum. Let's see graphical representation of impact alpha on error term.






As the value of alpha is doubled (2 times optimal value), for both ridge and lasso regression,

- the negative mean absolute error decreases
- the r2 decreases

Top 5 predictor variables for ridge regression when alpha is doubled are as below.

| Top 5 predictors with optimal alpha | | Top 5 predictors with double the optimal alpha | |
|---|---|---|---|
| **Feature** | **Value** | **Feature** | **Value** |
| Neighborhood_Crawfor | 0.077 | OverallQual | 0.062 |
| GrLivArea | 0.061 | GrLivArea | 0.058 |
| OverallQual | 0.061 | Neighborhood_Crawfor | 0.055 |
| AgeWhenSold | 0.051 | OverallCond | 0.048 |
| SaleCondition_Normal | 0.05 | 1stFlrSF | 0.039 |

Top 5 predictor variables for lasso regression when alpha is doubled are as below.

| Top 5 predictors with optimal alpha | | Top 5 predictors with double the optimal alpha | |
|---|---|---|---|
| **Feature** | **Value** | **Feature** | **Value** |
| Neighborhood_Crawfor | 0.119 | GrLivArea | 0.111 |
| GrLivArea | 0.109 | Neighborhood_Crawfor | 0.088 |
| SaleCondition_Partial | 0.097 | OverallQual | 0.074 |
| OverallQual | 0.068 | SaleCondition_Partial | 0.07 |
| Neighborhood_Somerst | 0.067 | OverallCond | 0.05 |

## 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?

As we add more predictors to the model, the training r2 score increases whereas the test r2 score does not increase. This is because of overfitting of the model i.e. model has memorized the training data. To overcome this issue, we use regularized regression. Regularized regression penalizes the model for its complexity.

Both ridge and lasso regression does feature selection. Ridge uses L2 regularization technique regression model whereas Lasso uses L1 regularization technique.

Optimal value of alpha for our models are as below:

- Ridge regression – 20
- Lasso regression – 0.001

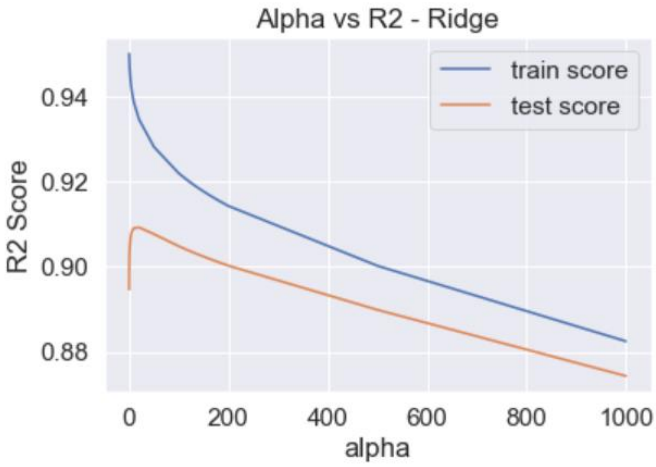
For optimal value of alpha in ridge and lasso regression, r2 score for both training and test data was calculated.

R2 score for ridge and lasso regression are as below.

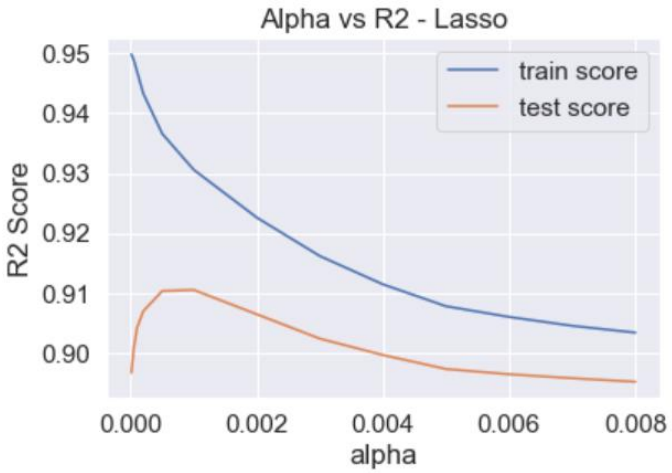
- Ridge regression train r2:  0.9331
- Ridge regression test r2:  0.9133

- Lasso regression train r2:  0.9281
- Lasso regression test r2:  0.9122

R2 score using Ridge regression



R2 score using Lasso regression




We will use lasso regression because:

1. Scores by lasso are more consistent across training and test dataset
2. Lasso regression has produced a simpler model since the number of predictors are less.

## 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?

Top 5 predictors for ridge and lasso regression are mentioned in question1. Since top 5 predictors are not available, all dummy variables if any associated with that variables are also removed. A new model is created after removing the top 5 predictors.

Since we are removing the top predictors the predictive power of the model has also dropped. This can be validated using r2 score.

Top predictors for lasso regression after removing the top5 predictors are:

| Feature | Value |
|---|---|
| 1stFlrSF | 0.100 |
| 2ndFlrSF | 0.092 |
| MSZoning_FV | 0.091 |
| Functional_Typ | 0.071 |
| PavedDrive_Y | 0.055 |

Change in r2 score after removing the top predictors are as below.

| | Before removing predictors | After removing predictors |
|---|---|---|
| Train | 0.9281 | 0.9094 |
| Test | 0.9122 | 0.8903 |

Top predictors for ridge regression after removing the top5 predictors are:

| Feature | Value |
|---|---|
| 1stFlrSF | 0.084 |
| 2ndFlrSF | 0.081 |
| MSZoning_FV | 0.061 |
| Functional_Typ | 0.056 |
| PavedDrive_Y | 0.051 |

Change in r2 score after removing the top predictors are as below.

| | Before removing predictors | After removing predictors |
|---|---|---|
| Train | 0.9331 | 0.9143 |
| Test | 0.9133 | 0.8865 |

## 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?

Per Occam's razor – model should be as simple as necessary. Advantages of simple model are as below:

- Generalizability
- Robustness
- Making few assumptions
- Less data is required for learning

If the r2 for trainings set is very high and r2 for test set is low, this is indication of overfitting. Such a model is not generalized because of which it is performing poor on unseen data. Robust model is not sensitive to training data. Robut model have low variance and high bias.

- Variance = How sensitive is model to the training data. This refer to consistency of the model.
- Bias = Accuracy of the data on unseen future data.

Model is always trained on training set and evaluated on unseen data (test set). Adding to many predictor variables in the model may lead to complex model. Complex model deteriorates the performance of the model (r2 score). Complex model introduces problem of overfitting where model memorized the data and is not generalized. When such model is evaluated against the unseen data the performance is very poor.