

## Assignment Part-II – Advanced Regression (RIDGE and LASSO)

By : Dhananjaya BL

### 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?

### ANS:

ALPHA VALUES:

For Ridge:     alpha = 7.4

For Lasso     alpha = 200

### DUOBLING THE ALPHA VALUE

When alpha doubled for the Ridge Regression;

- Training AdjR2 reduced and Testing AdjR2 increase,
- Overall variance decreased

When alpha doubled for the Lasso Regression;

- Both Training AdjR2 and Testing AdjR2 decreased,
- Overall variance increase

Top 10 Features before doubling the alpha were;

- Total Bath Rooms
- Neighborhood belonging to - StoneBr, NridgHt, NoRidge
- Total SFT of the house
- Total rooms above grade
- Number of fireplaces
- GarageArea
- Neighborhood belonging to - Veenker, Timberland, Somerset
- Wood deck area in square feet
- Basement Type 1 finished square feet
- Overall Quality

Top 10 Features after doubling the alpha are;

- Neighborhood belonging to - StoneBr, NridgHt, NoRidge
- Total Bath Rooms
- Number of fireplaces
- Total rooms above grade
- Total SFT of the house
- GarageArea
- Neighborhood belonging to - Veenker, Timberland, Somerset
- Overall Quality
- Wood deck area in square feet
- BedroomAbvGr

There is a shuffle in the top 10 feature, their positions have changed based on the changes to the 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?

### ANS:

Comparing both the model Ridge with an alpha value of 7.4 produced,  
Training AdjR2 of 84.5 and Test AdjR2 of 82.8  
And

Lasso with an alpha value of 200 produced,  
Training AdjR2 of 84.0 and Test AdjR2 of 82.7

With very large alpha which means more regularization, Lasso has produced almost close results as Ridge with smaller alpha, which indicates Ridge has performed better on this data set.

Will chose Ridge over Lasso for this 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?

### ANS:

The Next best predictor variables for Lasso are;

- GarageArea
- Neighborhood belonging to - Veenker, Timberland, Somerset
- Overall Quality
- Wood deck area in square feet
- BedroomAbvGr

## 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?

### ANS:

For the model to perform better on the real-world data when implemented, it is most important that model is more generalized than being very accurate one. There is a trade of between accuracy and generalizability of the model.

During model development we ensure that the variance in the model that is the performance metrics (AdjustedR2, RMSE, F1Score, etc...) between the Training and Test Data are close, or with in the acceptable limits. If there is a wider gap this tells us that model is not generalized or having higher variance. Model Tuning goal is to reduce this variance during model development.

At the same time we also need to ensure that the model is not underfitting, that is the model training performance is to the acceptable level or against any determined benchmark or against the human parity. Underfitted model might have lower variance but of no use for making accurate predictions.

A very complex model would produce a high level of training accuracy but may fail to produce same level of accuracy on unseen data, that means model is overfitting.

So it is required to develop a model that has acceptable level of accuracy and low level of variance, which is, high generalizability.