

# Assignment Part-II

## 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

Optimal value of alpha for ridge regression = 2.0

Optimal value of alpha for lasso regression = 0.0001

Double the value of alpha for ridge = 4.0

Double the value of alpha for lasso = 0.0002

Changes in the model after doubling the alpha values,

	Ridge	Lasso
<b>R2 score train</b>	0.894541961222369	0.8866375591632907
<b>R2 score test</b>	0.8602941483537443	0.8588896759609219
<b>RSS score train</b>	1.7903225154655833	1.9245126553704224
<b>RSS score test</b>	1.0695529468800853	1.080305235126096
<b>MSE score train</b>	0.0017534990357155565	0.0018849291433598652
<b>MSE score test</b>	0.002436339286742791	0.0024608319706744784

Most important predictor variables for Ridge and Lasso (after the changes are implemented),

Ridge	Lasso
OverallQual	OverallQual
FullBath	FullBath
BedroomAbvGr	BedroomAbvGr
OverallCond	OverallCond
BsmtFinSF1	BsmtFullBath
BsmtFullBath	BsmtFinSF1
NoRidge - Neighborhood	HalfBath
BsmtUnfSF	NoRidge - Neighborhood
Crawfor - Neighborhood	Crawfor - Neighborhood
LotArea	NridgHt - Neighborhood

From the above table we can conclude that the below are the most important predictor variables,

***OverallQual, FullBath, BedroomAbvGr, OverallCond, BsmtFinSF1, BsmtFullBath, NoRidge - Neighborhood and Crawfor - Neighborhood***

\*\*\*\* Refer the python notebook for the implementation \*\*\*\*

## 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

Optimal value of alpha for ridge regression = 2.0

Optimal value of alpha for lasso regression = 0.0001

- Lasso regression (0.864071) has slightly higher value for R2 Score than the Ridge regression (0.862928).
- Lasso regression (0.048687) has slightly lower value for Mean square error than the Ridge regression (0.048892).
- Lasso regression will reduce the coefficients to 0 which Ridge will not do.
- Lasso can be building the simple model with fewer features.

So, I will choose Lasso regression model to predict the more accurate values for the sale price for house.

## 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

5 most important predictor variables for lasso are,

OverallQual, BsmtFinSF1, BedroomAbvGr, FullBath and LotArea

Dropping the 5 most important predictor variables in the lasso model and building model again.

After building it again, 5 most important predictor variables are,

**OverallCond, BsmtUnfSF, NoRidge - [Neighborhood], RH - [MSZoning] and BsmtQual**

\*\*\*\* Refer the python notebook for the implementation \*\*\*\*

## 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

Model is robust and generalize when,

- Performance of the model is more accurate on the unseen test data and performance should not affect even after some variation in the test data.
- Complexity of the model should be low, and model should be built on the less predictor variables, which implies model should be simple.
- Multicollinearity should be avoided in the predictor variables.
- Regularization technique (Ridge, Lasso) should be used to avoid the overfitting and underfitting.
- Bias – variance tradeoffs can be used to get the good accuracy in the model.
- Accuracy is how much correct prediction the model done on the unseen test data.