

# Advanced Regression Assignment

Monday, June 6, 2022 1:50 PM

## Problem Statement - Part II

### 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 double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?**

The Optimal value of alpha for ridge is 100 and lasso is 500.

Ridge	Alpha=100	Alpha=200
	r2_train_lr = 0.9201710244637384 r2_test_lr = 0.8888762982137223 rss1_train_lr = 282478935584.0551 rss2_test_lr = 158369374167.9046 mse_train_lr = 314214611.3282037 mse_test_lr = 410283352.7665922	r2_train_lr = 0.934536630911627 r2_test_lr = 0.8985947367633689 rss1_train_lr = 231645498337.00323 rss2_test_lr = 144519016357.13885 mse_train_lr = 257670187.24916935 mse_test_lr = 374401596.7801525
	Important predictor = 'OverallQual' 6822.341448	Important predictor = : 'OverallQual', 8309.588099
Lasso	Alpha=500	Alpha=1000
	r2_train_lr = 0.9329593994080164 r2_test_lr = 0.9061018200665976 rss1_train_lr = 237226613130.43018 rss2_test_lr = 133820199944.01015 mse_train_lr = 263878323.83807585 mse_test_lr = 346684455.8134978	r2_train_lr = 0.9195009834279172 r2_test_lr = 0.8947923732141403 rss1_train_lr = 284849910249.89514 rss2_test_lr = 149938003719.60413 mse_train_lr = 316851958.00878215 mse_test_lr = 388440424.14405215
	Important predictor = 'GrLivArea' 21704.8727	Important predictor = 'GrLivArea' 21434.90664

#### 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 : I will choose Lasso with Alpha=500 as the R2 for train and test data is very close. There is no case of overfitting.

### 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?

Lasso features	Lasso coeffs
'YearRemodAdd'	4477.469797
'LotArea'	4360.061042
'OverallCond'	3877.979481
'Neighborhood_StoneBr'	3803.049397
'GarageArea'	3739.792975

### 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?

Very complex models have high variance (ie, are less robust) but have low bias(error)  
However they cause overfitting.

Less complex models have low variance (ie, error terms don't change much) but cause underfitting ie, they have high bias (error)

Hence we need to find an optimal value of bias and variance to get a more generalized and robust models.

