# AI/ML - CS 337 Lab - Lab Assignment 1 - Part 2

B.Nikhil 170050099

September 1, 2019

## Task 4: Lasso regression

### a) Parameters used

Parameters used for this task are,

$$\eta = \frac{1}{38000}$$

$$\lambda = 10000$$

$$max\,number\,of\,iterations = 31000$$

$$lr = 0.00002872$$

## Task 5: Observations

### a) Important features for Ridge regression

The five most important features correspond to the indices of W with maximum value of weight. For the given test case the important features for ridge-regression correspond to columns with indices [87 211 158 66 221]
See Figure 1 for features vs weight(W[i]). The top 5 highest peaks correspond to indices above

### b) Important features for Lasso regression

The five most important features correspond to the indices of W with maximum value of weight. For the given test case the important features for lasso-regression correspond to columns with indices [73 275 256 97 40]
See Figure 2 for features vs weight(W[i]). The top 5 highest peaks correspond to indices above

## Task 6: Learning rate vs execution time and error

### a) Learning rate vs execution time

Increasing learning rate will make error convergence to minima quickly but there are chances that the gradient might overshoot minima. In case of ridge-regression increasing learning rate will lead to decrease in number of iterations but for the case of lasso it almost runs max-iterations times so the execution time not always depends on rate of error convergence
See Figure 3 for Learning rate vs execution time for ridge grad descent. See Figure 4 for Learning rate vs Error for coord grad descent
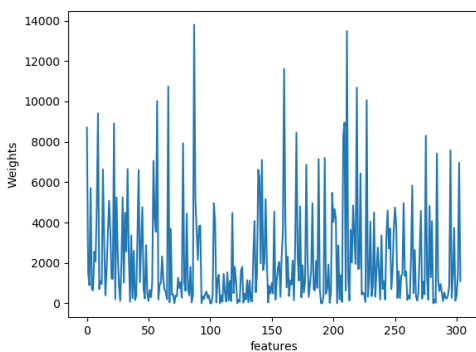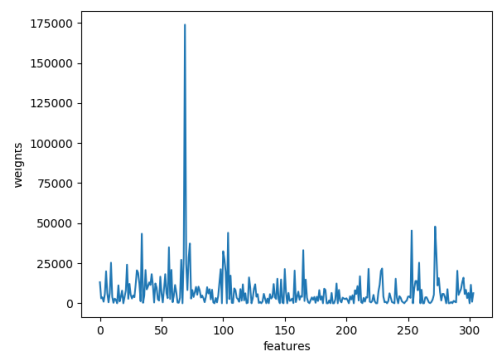


Figure 1: Ridge Important features.



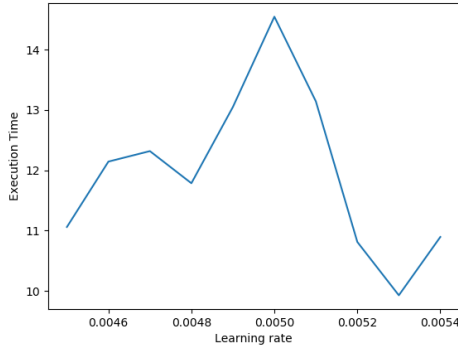Figure 2: Lasso Important features.
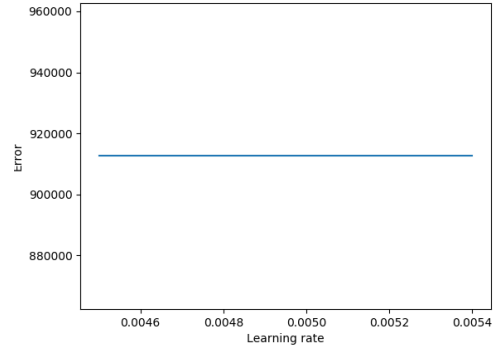
Figure 3: Ridge Execution time vs Learning rate.


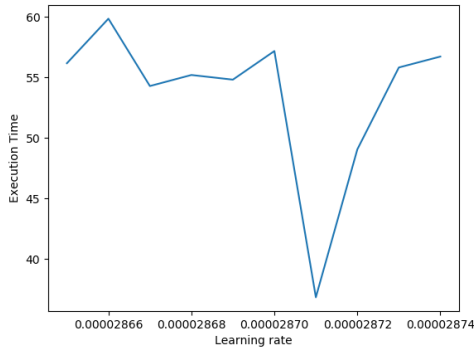
Figure 4: Ridge Observed Error vs Learning rate.
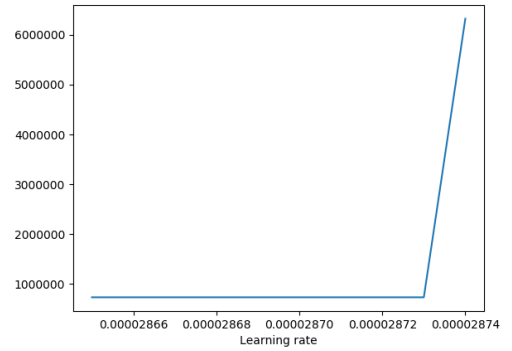


Figure 5: Lasso Execution time vs Learning rate.



Figure 6: Lasso Observed Error vs Learning rate.

## b) Learning rate vs observed error

If the error converges then the observered error is independent of Learning rate. But if the error doesn't converge and reaches max-iterations then increasing learning rate decreases observed error.

See figure 5 for Learning rate vs observed-error for ridge regression See figure 6 for Learning rate vs observed-error for lasso