ELL409 - Assignment Report - 1

Ashi Veerman, 2021MT10241

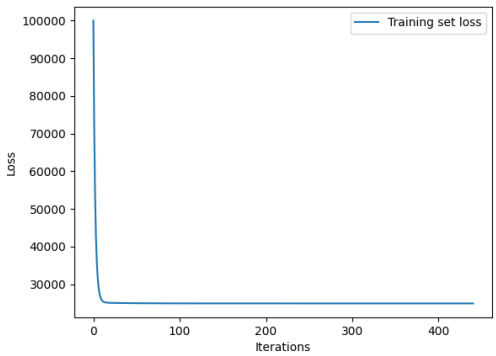
August 2024

1 Gradient Descent

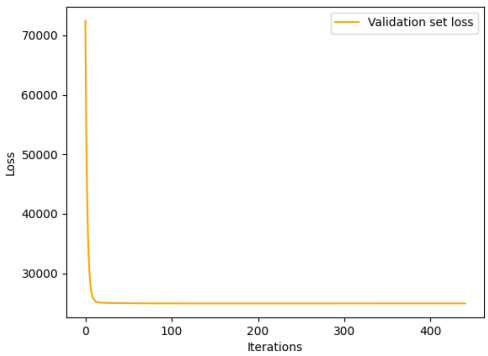
Learning rate	Number of iterations to converge
0.1	441
0.01	3676
0.001	29279

Table 1: Learning rate and number of interations to converge

1.1 Training and validation loss plots

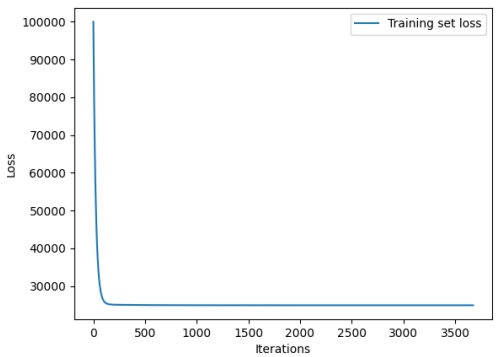


(a) Training loss

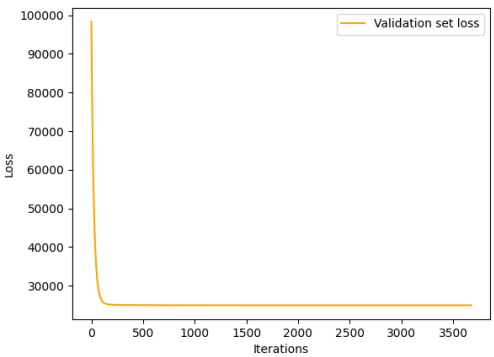


(b) Validation loss

Figure 1: Training and validation loss plots for learning rate 0.1

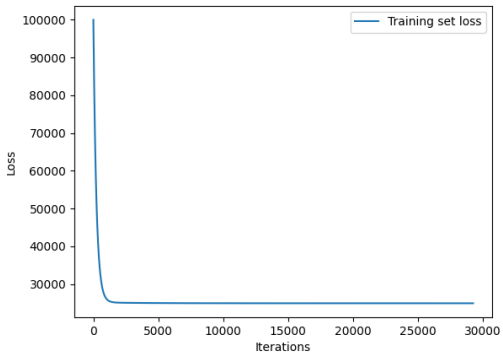


(a) Training loss

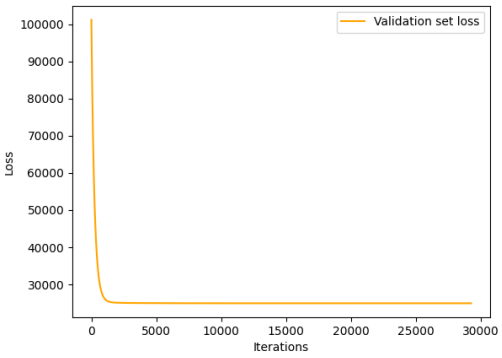


(b) Validation loss

Figure 2: Training and validation loss plots for learning rate 0.01



(a) Training loss



(b) Validation loss

Figure 3: Training and validation loss plots for learning rate 0.001

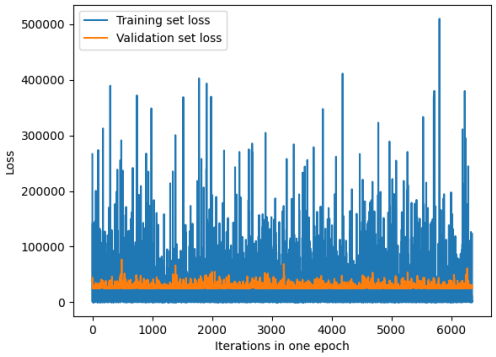
2 Stochastic Gradient Descent

Batch size	Time taken for convergence (s)
1	145.5390169620514
10	10.794893026351929
100	1.5699028968811035
1000	0.4364297389984131

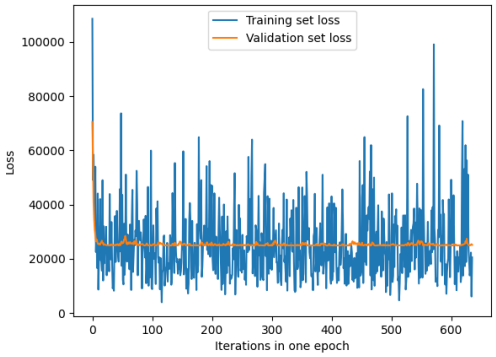
Table 2: Batch size and time taken for one epoch

2.1 Training and validation loss plots

Number of epochs set to 10,000
The below graphs don't have the learning rate decay technique applied. The graphs have error value plotted for one epoch.

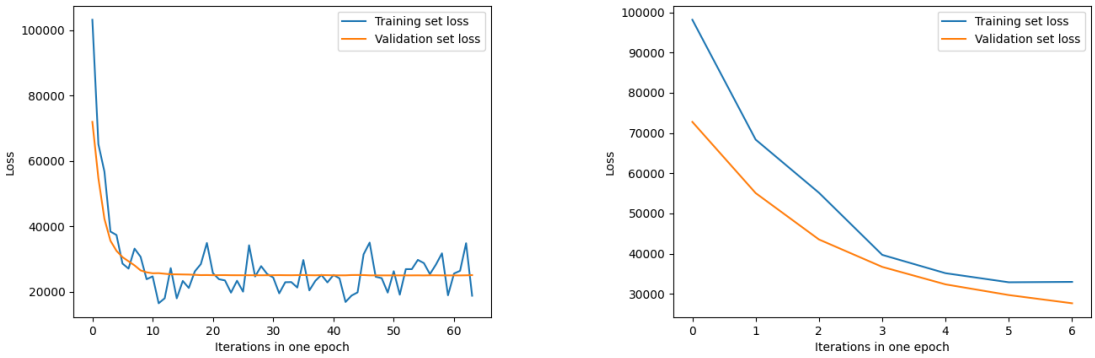


(a) Batch size 1



(b) Batch size 10

Figure 4: Training and validation loss plots for different batch sizes for one epoch



(a) Batch size 100

(b) Batch size 1000

Figure 5: Training and validation loss plots for different batch sizes for one epoch

3 Observations

- As learning rate is decreased more iterations are required for convergence
- As batch size is increased the time taken for one epoch reduces significantly
- For stochastic gradient descent time for convergence is very high as the error values oscillate a lot
- Used learning rate decay approach so that the stochastic gradient method converges