Linear MNIST Classifier

- Bhagyesh Gaikwad.

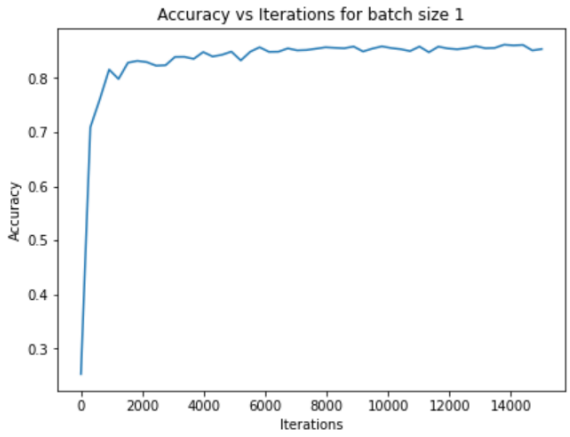
Stats for different batch size

Choosing the number of iterations to be 20,000 which should be sufficient for convergence.

Learning rates tried for all the batch sizes are 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}

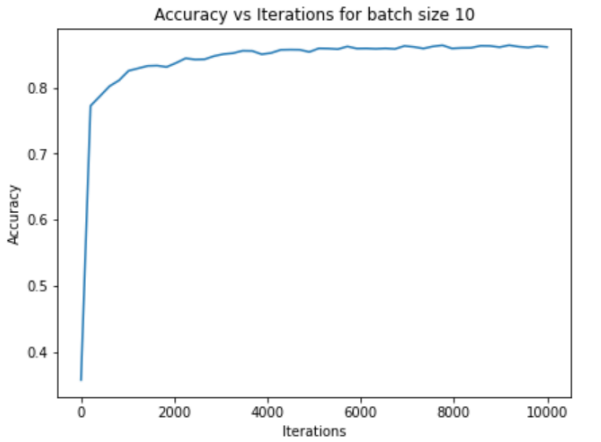
- Batch size 1:
  - Training Time : 2.79 seconds
  - Best Learning Rate : 0.001
  - Accuracy : 85.31%

  Plots :



- Batch size 10:
  - Training Time : 3.30 seconds
  - Best Learning Rate : 0.001
  - Accuracy : 86.22%

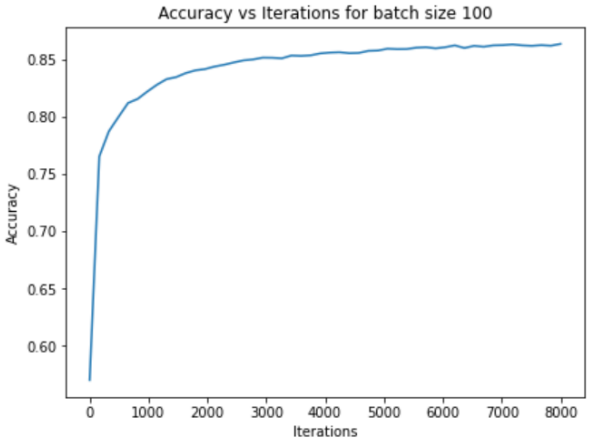
  Plots :



.ipynb#

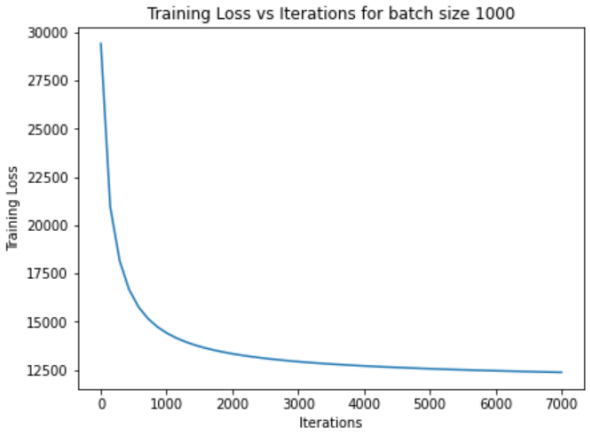
- Batch size 100:
  - Training Time : 15.29 seconds
  - Best Learning Rate : 0.001
  - Accuracy : 86.36%

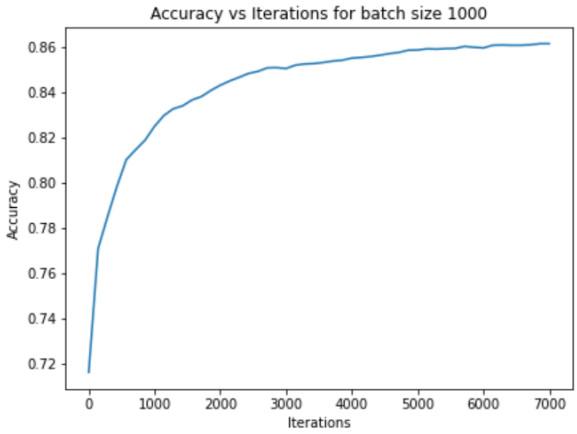
Plots :



- Batch size 1000:
  - Training Time : 3.30 seconds
  - Best Learning Rate : 0.001
  - Accuracy : 86.54%

Plots :



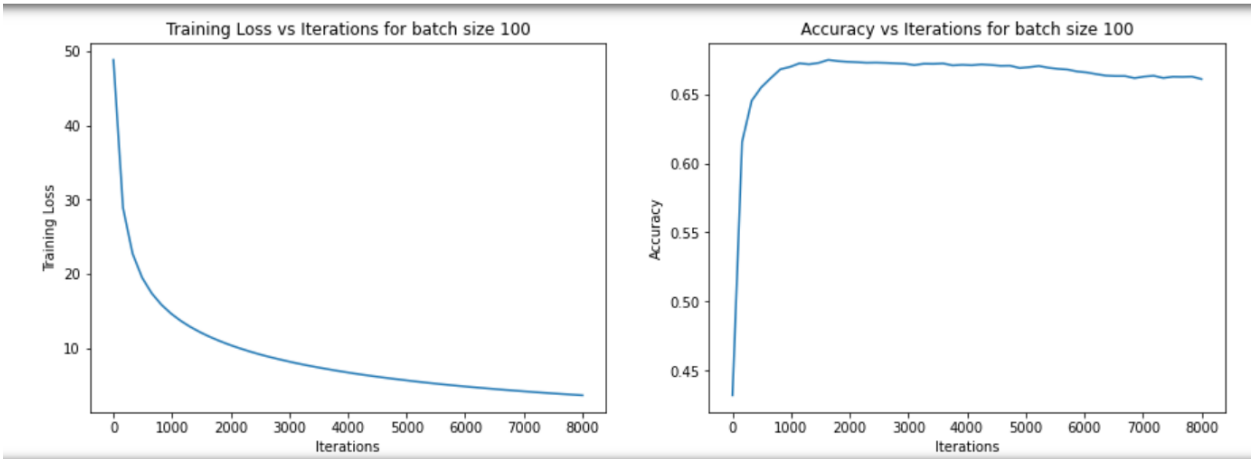


- Role of Batch sizes :
    - Here we have kept the number of iterations same for all the sizes, which in real sense should not be the case, we can keep more number of iterations for a smaller batch size and vice versa.
    - The reason for the above observation is smaller batch sizes take less training time but tend to be error prone during training.
    - Larger batch sizes on the other hand have more training time with more accurate estimates, this can be seen in the stats above, as the batch size increases the accuracy increases given the same number of iterations.
    - In practice the suitable batch size can be the one which gives the amount of accuracy needed without being computationally expensive, i.e after increasing the size from the chosen one there is no substantial improvement in accuracy, can be a good choice.

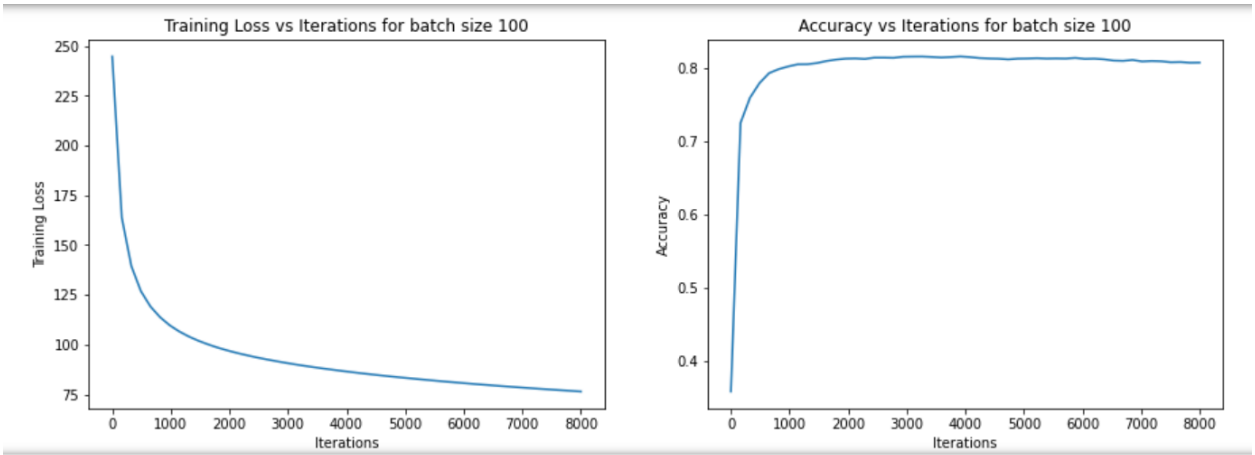
Below are the accuracy stats and plots for dataset sizes 100; 500; 1000; 10000; 50000; 60000
Note : I have added 50,000 and 60,000 for observation purposes:

```
max accuracy for data size  100  is:  0.6748
max accuracy for data size  500  is:  0.8157
max accuracy for data size  1000  is:  0.8403
max accuracy for data size  10000  is:  0.8602
max accuracy for data size  50000  is:  0.8624
max accuracy for data size  60000  is:  0.8631
```

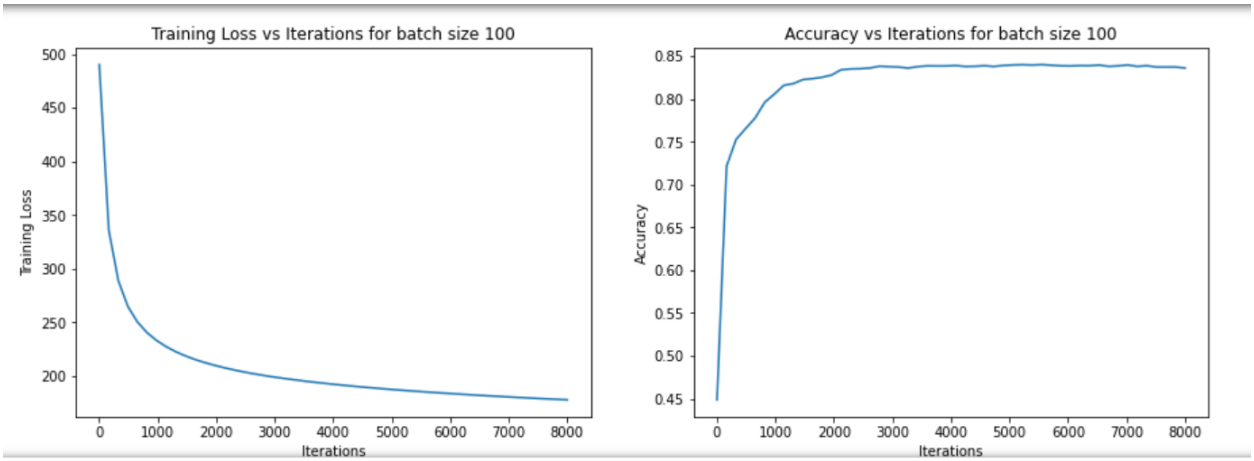
- Dataset size 100:



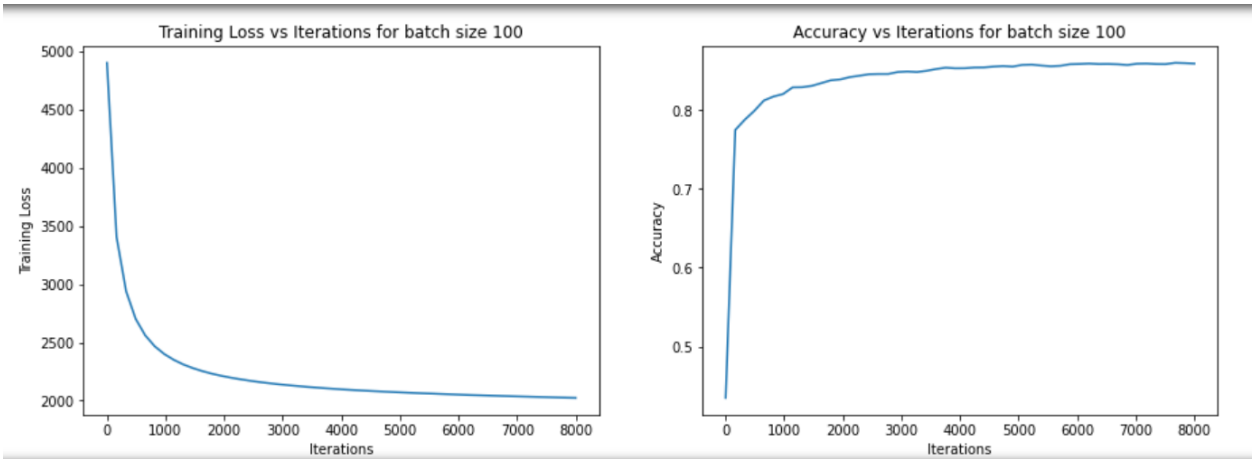
- Dataset size 500:



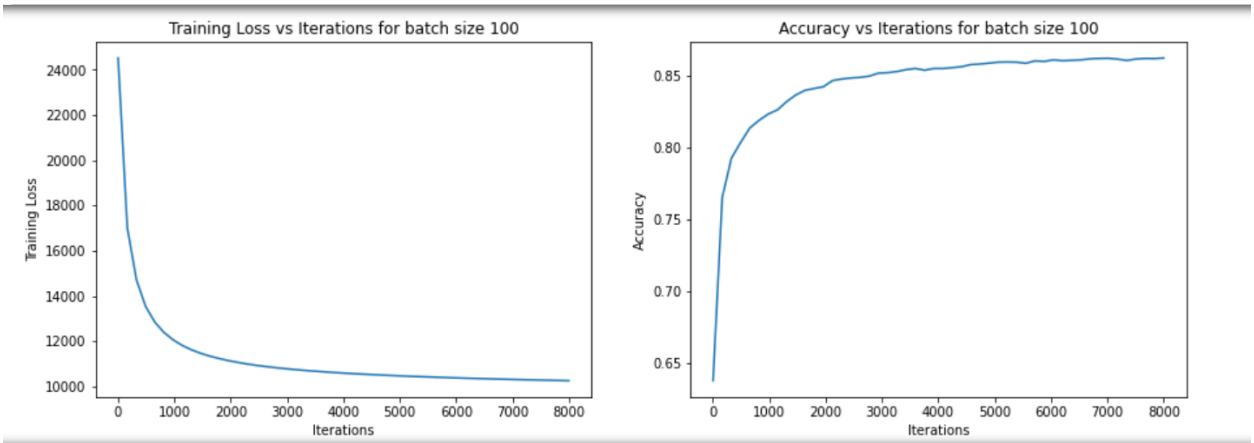
- Dataset size 1000:



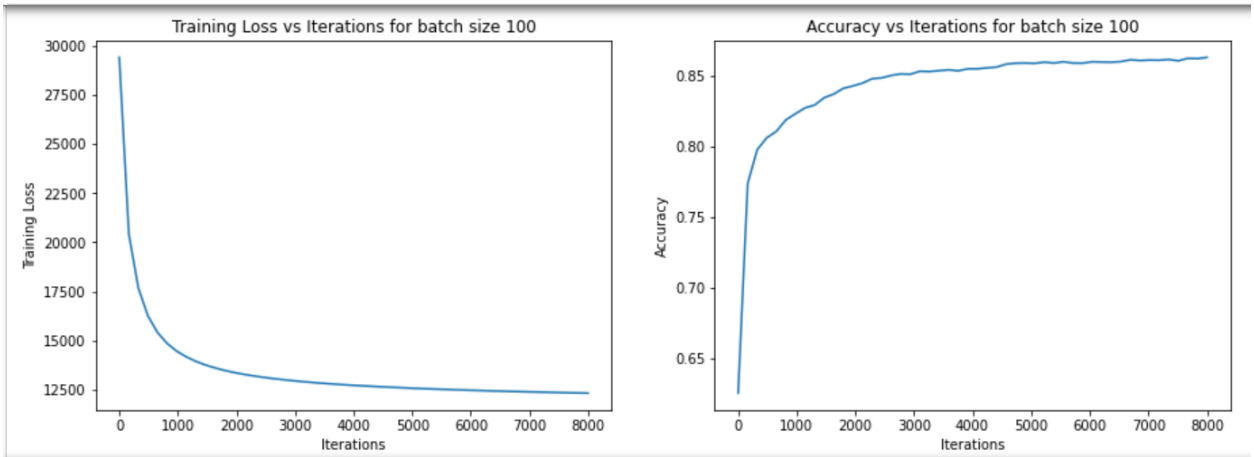
- Dataset size 10000:



- Dataset size 50000:



- Dataset size 60000:



- Role on training dataset size:

- We can observe that smaller dataset sizes have lower accuracies since the observed data is less models aren't trained effectively and thus do not perform well on test data.
- Although it should be noted that having too much data also does not significantly improve accuracy e.g. here 10,000 size also gives accuracy of 86.02 which is really close to the max accuracy achieved.
- In practice it is best not to train the model on a full dataset which makes it prone to overfitting. Instead a validation set can be used with different combinations of hyperparameters and see what works best.