

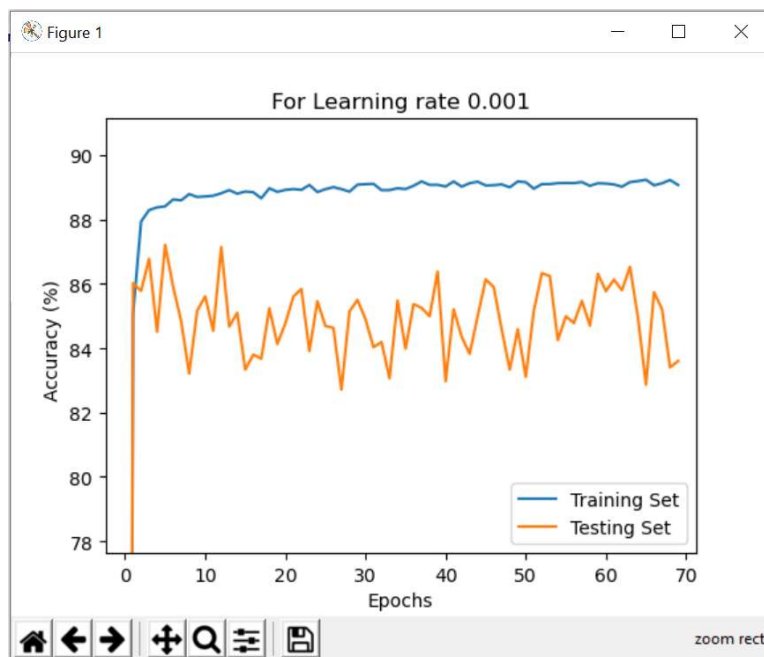
Perceptron Training and testing, with learning rates of $\eta = 0.001$, 0.01 and 0.1**Description of the program:**

This program involves a training of a single perceptron, with 785 inputs and 10 outputs, using MNIST dataset, with three different learning rates (0.001, 0.01 and 0.1). this program runs up to 70 epochs for each training (60,000 X 785) and testing (10,000 X 785) datasets. The data is normalized before starting to input the data, and the initial weights are randomly chosen between the interval $[-0.05, 0.05]$ to create a weights nd-array of (10 X 785).

The program cycle's through the training data, changing the weights according to the perceptron learning rule. After the weights are calculated for one epoch, the testing data is passed. Hence after each epoch the accuracy of testing and training data are calculated and stored in a file. Later they are plotted on an accuracy graph.

i. Learning Rate of 0.001

- a. The plot of accuracy between training set and testing set is as follows.



- From the graph we can see that the accuracy of testing set is lower than training set. We can say that the model did not learn it yet. It is an underfit case.

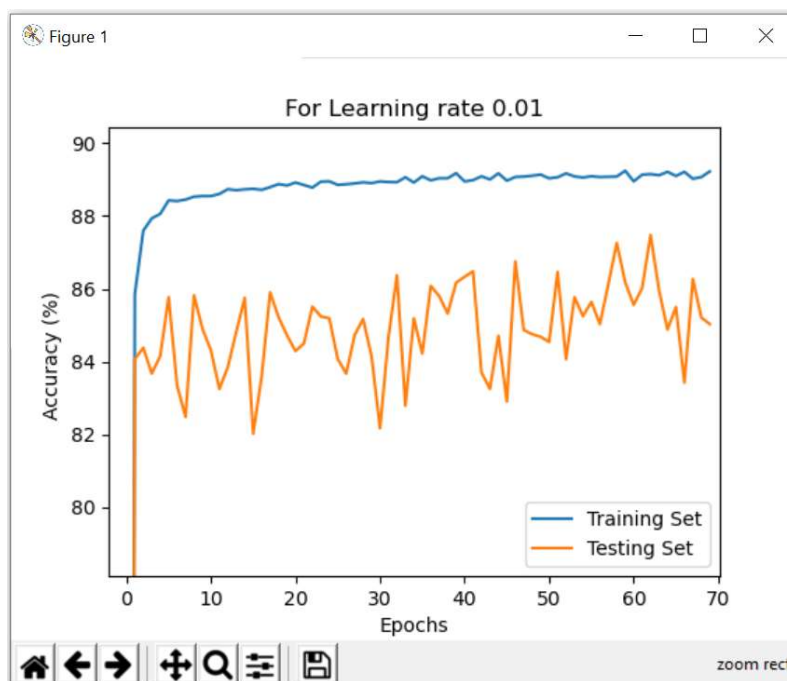
b. The confusion matrix of testing dataset after 70 epochs is as follows

[[924	0	2	6	0	5	18	3	12	10]
[0	1114	0	5	1	2	3	0	9	1]
[3	60	713	141	6	8	24	10	57	10]
[2	2	5	944	1	6	4	7	12	27]
[1	1	3	1	683	0	28	1	10	254]
[5	2	1	116	8	591	33	5	69	62]
[4	5	2	3	2	12	925	0	5	0]
[2	15	20	19	7	3	1	798	8	155]
[4	40	3	77	13	14	22	6	707	88]
[3	8	1	16	9	3	2	5	1	961]]

- After 70 epochs the model has 89.07% accuracy on training set and 83.6% accuracy on testing set.

ii. Learning Rate of 0.01

a. The plot of accuracy between training set and testing set is as follows.



- From the graph we can see that the accuracy of testing set is lower than training set. It is an underfit case.

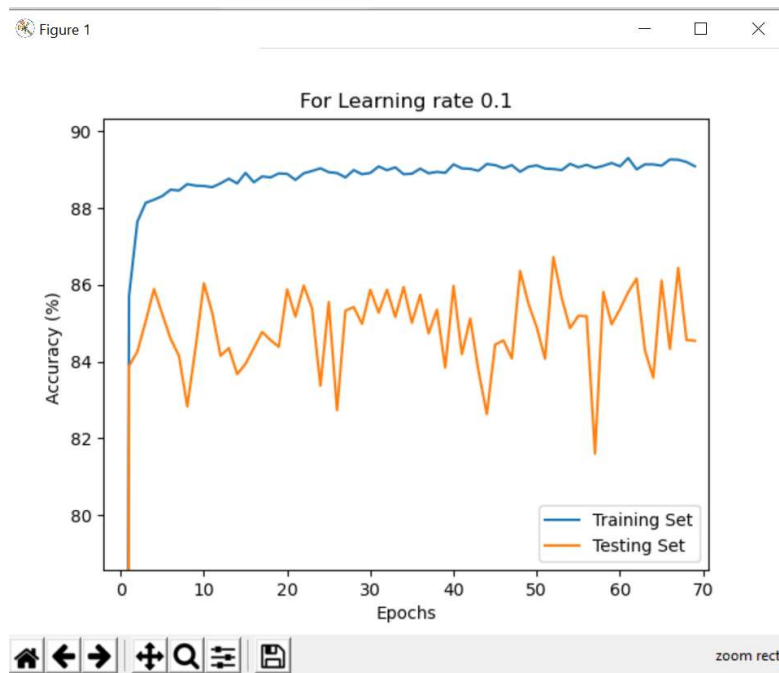
b. The confusion matrix of testing dataset after 70 epochs is as follows

```
[[ 940    0    3    2    0    5   17    3    7    3]
 [    0 1112    0    7    0    0    3    1   12    0]
 [    6   28  736  166    4   10   18    5   55    4]
 [    3    3   12  948    0    8    4    5   14   13]
 [    1    3    9    6  750    0   26    1   12  174]
 [    9    4    3  115   12  601   36    4   77   31]
 [    3    5    3    2    2   12  922    0    9    0]
 [    4   15   22   29   11    2    2  805   13  125]
 [    7   23    6  120   10    6   17    3  742   40]
 [    6    8    2   21   11    1    2    5    6  947]]
```

- After 70 epochs the model has 89.23% accuracy on training set and 85.03% accuracy on testing set.

iii. For learning rate 0.1

a. The plot of accuracy between training set and testing set is as follows.



- From the graph we can see that the accuracy of testing set is lower than training set. We can say that the model did not learn it yet. It is an underfit case.

b. The confusion matrix of testing dataset after 70 epochs is as follows

```
[[ 959    0    3    2    0    3    6    2    3    2]
 [   0 1115    1    6    0    1    3    1    8    0]
 [  10   41  783  124    4    6   21    6   29    8]
 [   3    2   10  962    1    5    1    8    4   14]
 [   3    3    4    7  692    0   29    1   13  230]
 [  12    3    1  166    6  586   30    8   39   41]
 [   9    5    5    3    1   18  915    0    2    0]
 [   5   14   19   15    6    1    1  858    6  103]
 [  10   36    7  165   11   10   23    7  641   64]
 [   8    6    2   27    4    3    3    9    4  943]]
```

- After 70 epochs the model has 89.08% accuracy on training set and 84.5% accuracy on testing set.