Pratik Walawalkar

667624808

Q3) Report:

Code is written in Python (attached at the end of report)

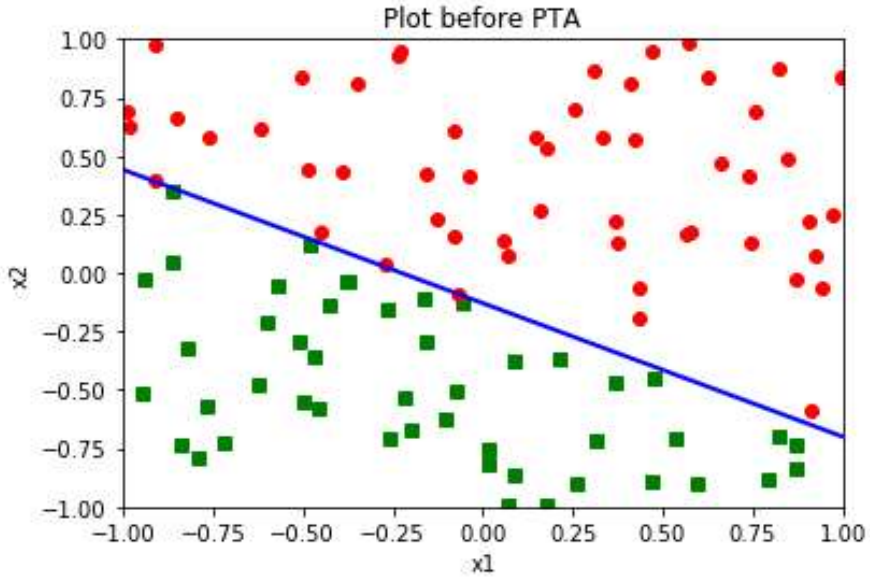
We observe following things while implementing Perceptron Training Algorithm:

1) For 100 samples (i.e. S collection)

Optimal weights before PTA i.e**. w: [-0.053561683916629854, -0.23711318193007447, -0.41519609106873356]**

Updated weights for carrying out PTA i.e**. w': [-0.4031964925425351, -0.6104564412967706, -0.24484741310837865]**

Total vectors in S: 100, Total vectors in S0: 53, Total vectors in S1: 47



Index-> Red circles= Class S0; Green squares = Class S1

Eta is nothing but our training parameter:

- **For eta = 1:**

Total number of epochs required for convergence: **10**

Final weights: **[-0.8830797924695428, -3.3133576319780977, -5.561407680702544]**

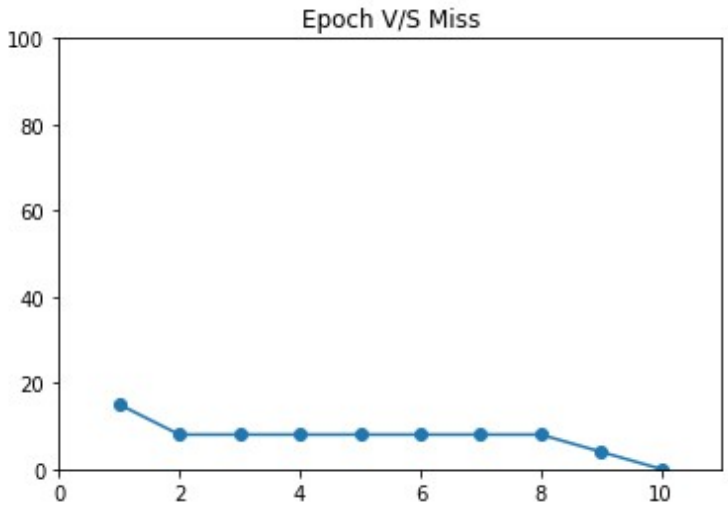


Figure 1

- For **eta = 0.1:**

Total number of epochs required for convergence: **8**

Final weights: [**-0.08307979246954283, -0.35125590759269676, -0.6160964223720395**]

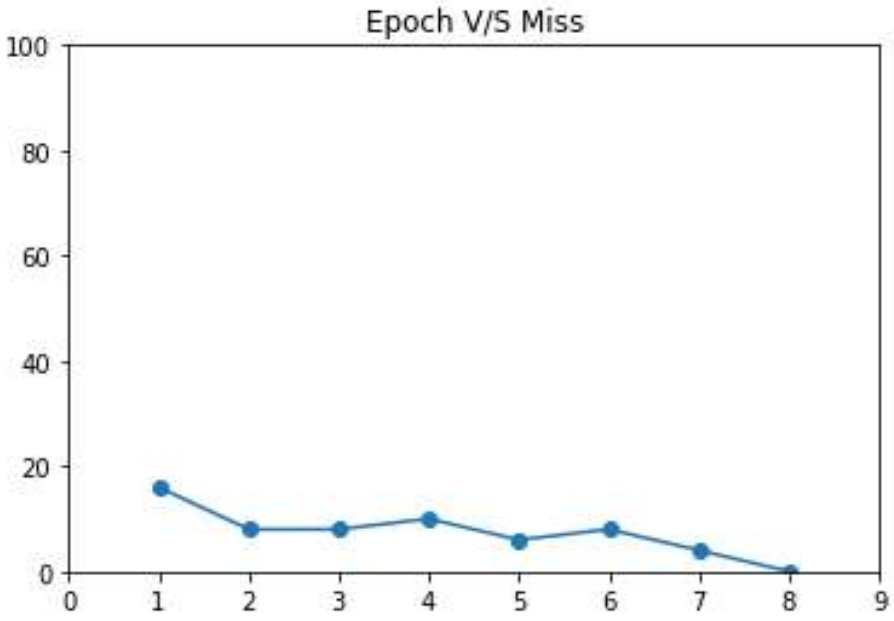


Figure 2

- For eta = **10**:

Total number of epochs required for convergence: **16**

Final weights: [**-9.883079792469545, -41.36321279289809, -71.67697180703308**]

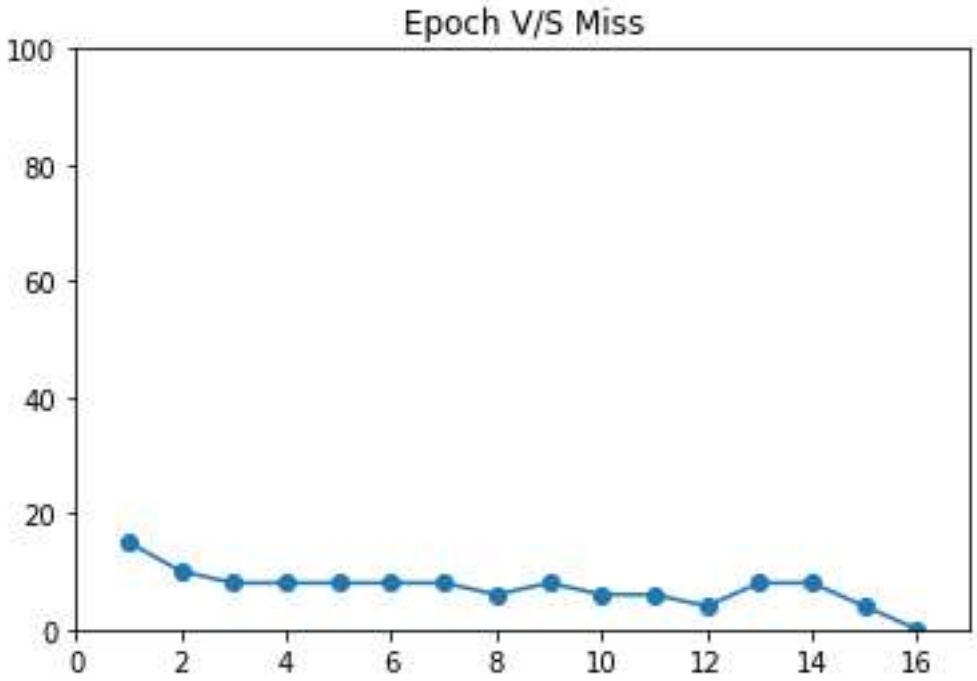
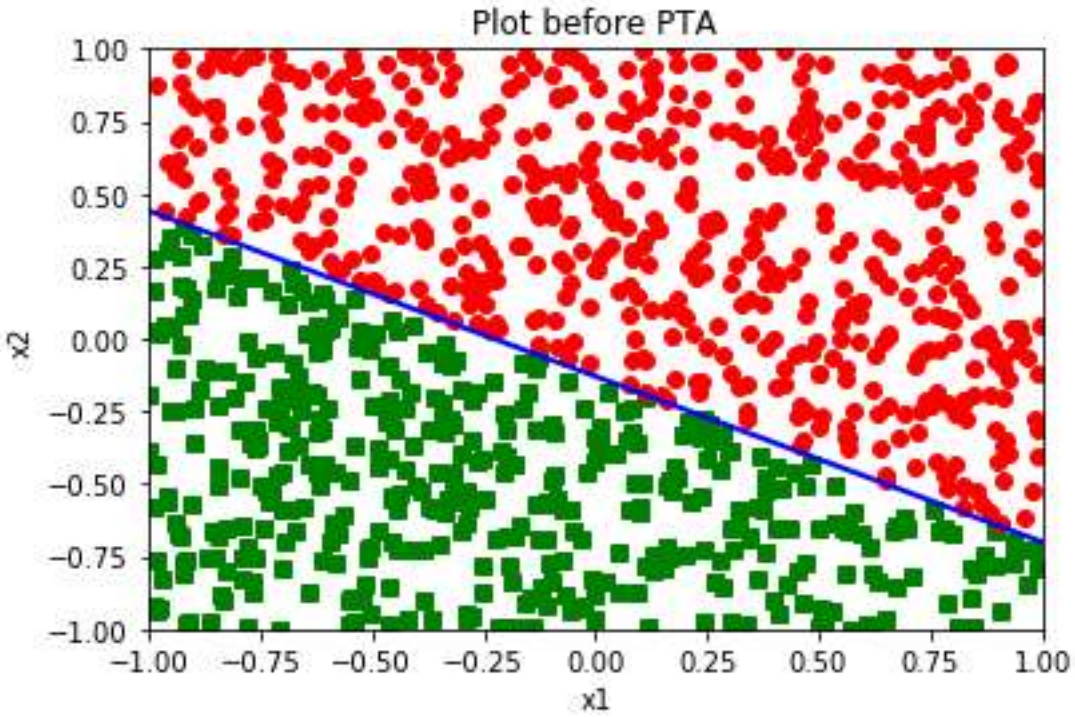


Figure 3

2) For 1000 samples (i.e. S collection)

Optimal weights before PTA i.e. **w: [-0.053561683916629854, -0.23711318193007447, -0.41519609106873356]**

Total vectors in S: 1000, Total vectors in S0: 571, Total vectors in S1: 429



Index-> Red circles = Class S0 ; Green squares = Class S1

- **For eta = 1:**

Total number of epochs required for convergence: **19**

Final weights: **[-1.8830797924695428, -8.122010921691878, -14.212969723127763]**

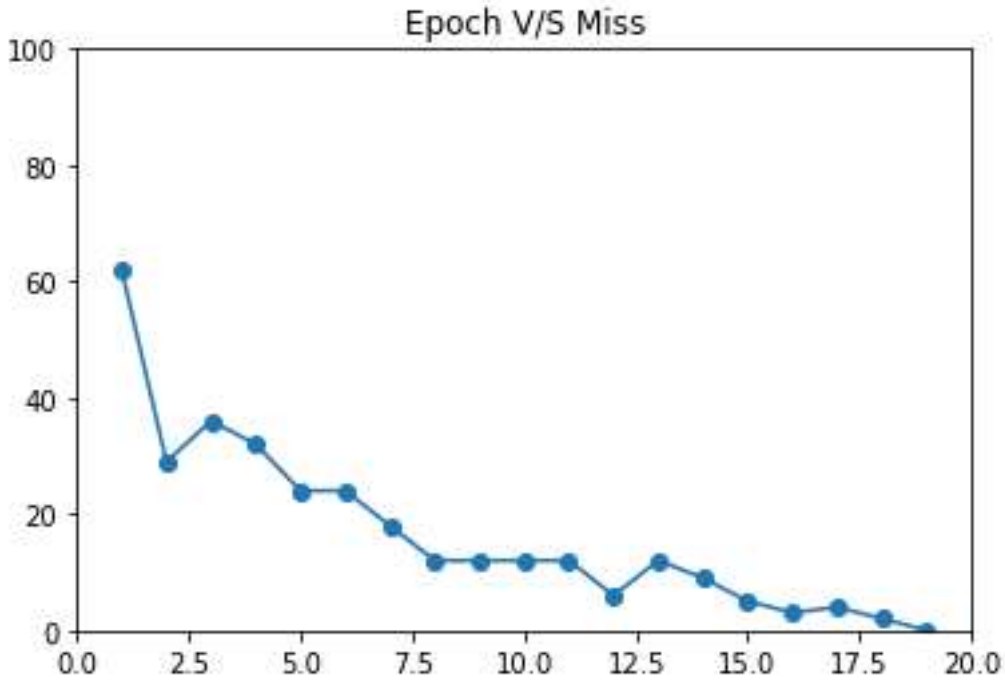


Figure 4

- **For eta = 0.1:**

Total number of epochs required for convergence: **18**

Final weights: **[-0.18307979246954284, -0.7890092585782806, -1.3819038661791228]**

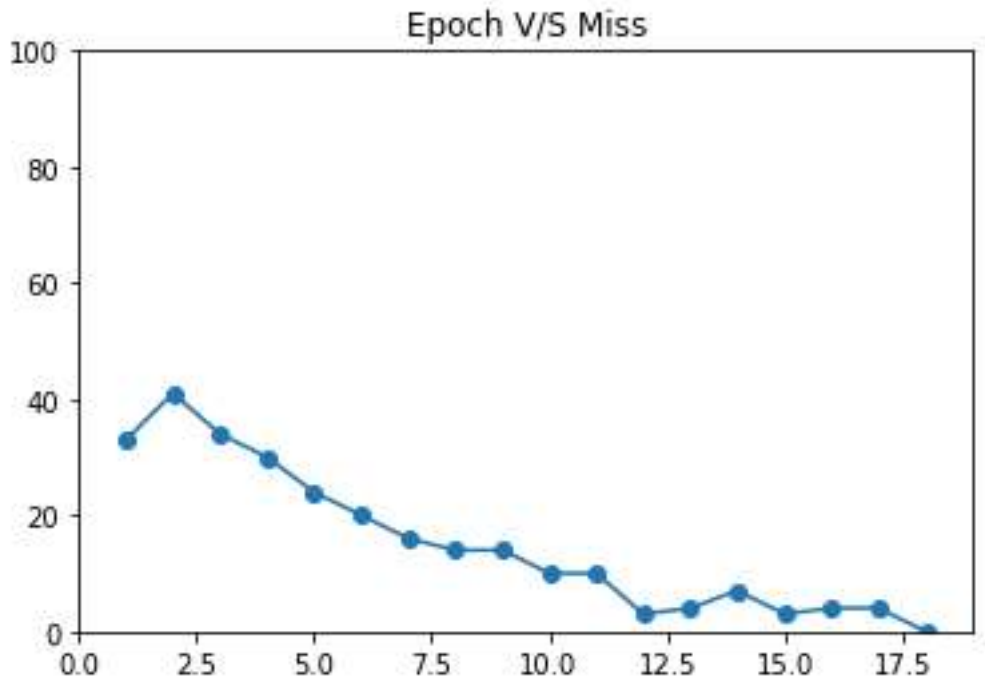


Figure 5

- **For eta = 10:**

Total number of epochs required for convergence: **19**

Final weights: **[-19.883079792469545, -84.81459721405383, -148.39275132913838]**

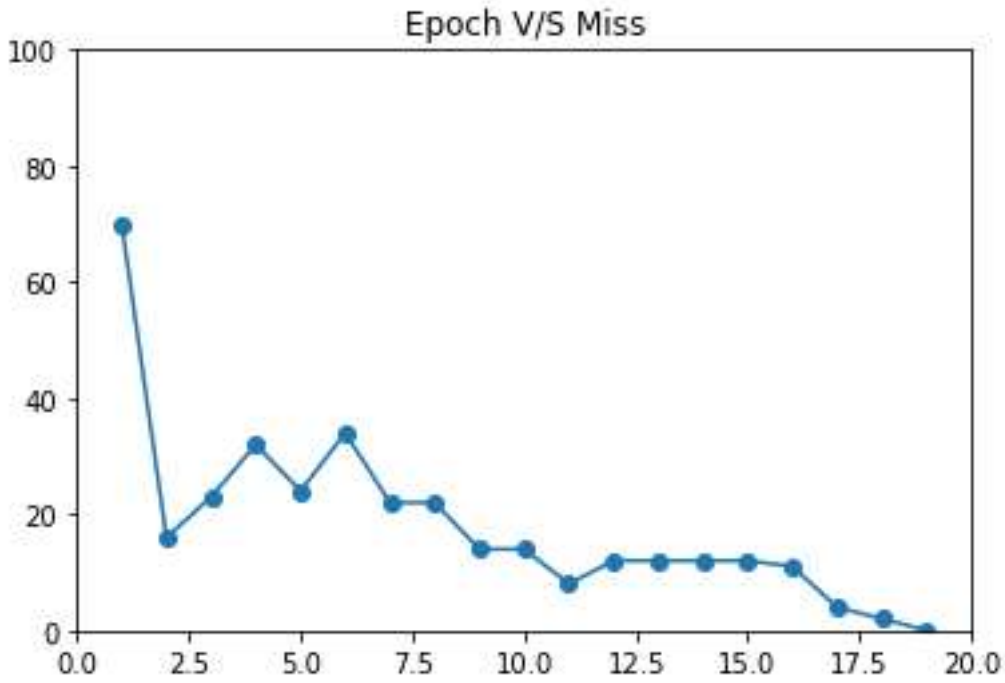


Figure 6

Solutions Q3 by observing above results:

(e) Randomly generated optimal weights are

w = [w0 w1 w2] = [-0.053561683916629854, -0.23711318193007447, -0.41519609106873356]

Note: We have taken same optimal weights for samples = 100, 1000 and for each training parameters i.e. 1, 0.1, 10

(i)

Graphs are plotted for 100 and 1000 samples as above. Red circles denote S1 collection of (x1, x2) vectors whereas green squares denote S2 collection of (x1, x2). Blue line denotes line 1.w0+w1.x1+w2.x2 = 0

(j)

ii. Randomly updated weights for carrying out PTA i.e**.**

w' = [w0' w1' w2'] = [-0.4031964925425351, -0.6104564412967706, -0.24484741310837865]

Note: We have taken same optimal weights for samples = 100, 1000 and for each training parameters i.e. 1, 0.1, 10

vii Final weights

- For 100 samples and training parameter = 1
  Final weights: **[-0.8830797924695428, -3.3133576319780977, -5.561407680702544]**
  Comparison with optimal weights: final weights decreases as compared to each optimal weights. With these weights, the convergence is achieved (misclassification=0).

(k) Graph that shows the epoch number vs the number of misclassifications for 100 samples and 1 training parameter: refer figure1.

(l) Graph that shows the epoch number vs the number of misclassifications for 100 samples and 0.1 training parameter: refer figure2.

(m) Graph that shows the epoch number vs the number of misclassifications for 100 samples and 10 training parameter: refer figure3.

Observation from all the graphs of epoch v/s misclassification for 100 samples:

- Initially misclassification rate is high but as the input (x1,x2) are fed into PTA, the rate damps and decreases for every incrementing epoch iteration. At the point of convergence, misclassification is zero. Lower the number of epoch, faster is the execution of PTA

(n) **Observation for changes in training parameters for 100 samples:**

Training parameters	Epochs required
0.1	8
1	10
10	16

From above table, we can say that as we increase no. of training parameters with the factor of 10, no. of epochs to achieve convergence also increases. But this is not true always as it also depends on the effect of change of initial weights taken randomly before and during implementing PTA

(o) To answer this question different weights were taken randomly while executing the experiment separately. Those were: w: [-0.08204609663199319, 0.6071159950347611, 0.14343841783349665]

Training parameters	Epochs required
0.1	25
1	2
10	6

From above table, we can say that for separate set of random weights, epoch is higher for eta=0.1 but decreases when eta increases. This is not the case with above question. Hence the PTA algorithm depends on how close the initial weights are chosen to that of optimal weights.

(p)

For 1000 samples

Training parameters	Epochs required
0.1	18
1	19
10	19

Observation as compared with 100 sample spaces:

- Number of epochs remains almost same with change in training parameters
- There is lot of damping oscillations for epoch v/s misclassification graph.
- Initial misclassification is much higher due to more number of input vectors
- More number of epochs are required
- Final weights are much greater than optimal weights. This is not the case with 100 samples