MLFA LAB
Assignment - 1

Name: Aryan Satpathy

Roll: 20EC10014

Question 1

I tried the **four values of K and results** are as following:

Fold	Accuracy	Precision	Recall	F1
1	1.0	1.0	1.0	1.5
2	1.0	1.0	1.0	1.5
3	1.0	1.0	1.0	1.5

Sub-Table 1: K = 3

Fold	Accuracy	Precision	Recall	F1
1	1.0	1.0	1.0	1.5
2	1.0	1.0	1.0	1.5
3	1.0	1.0	1.0	1.5
4	1.0	1.0	1.0	1.5
5	1.0	1.0	1.0	1.5

Sub-Table 2: K = 5

Fold	Accuracy	Precision	Recall	F1
1	1.0	1.0	1.0	1.5
2	1.0	1.0	1.0	1.5
3	1.0	1.0	1.0	1.5
4	1.0	1.0	1.0	1.5
5	1.0	1.0	1.0	1.5
6	1.0	1.0	1.0	1.5
7	1.0	1.0	1.0	1.5
8	1.0	1.0	1.0	1.5

Sub-Table 3: K = 8

K	Accuracy	Precision	Recall	F1
3	1.0	1.0	1.0	1.5
5	1.0	1.0	1.0	1.5
8	1.0	1.0	1.0	1.5
10	1.0	1.0	1.0	1.5
5	1.0	1.0	1.0	1.5
8	1.0	1.0	1.0	1.5
10	1.0	1.0	1.0	1.5
5	1.0	1.0	1.0	1.5
8	1.0	1.0	1.0	1.5
10	1.0	1.0	1.0	1.5

Sub-Table 4: K = 10

Table 1: Values of Performance Metrics over different K

Mean and Variances for the different values of K are as follows:

K	Accuracy	Precision	Recall	F1
3	1.0	1.0	1.0	1.5
5	1.0	1.0	1.0	1.5
8	1.0	1.0	1.0	1.5
10	1.0	1.0	1.0	1.5

Table 2: Mean & Variance for each K

Plot of Misclassification v/s Iterations on 80:20 split:

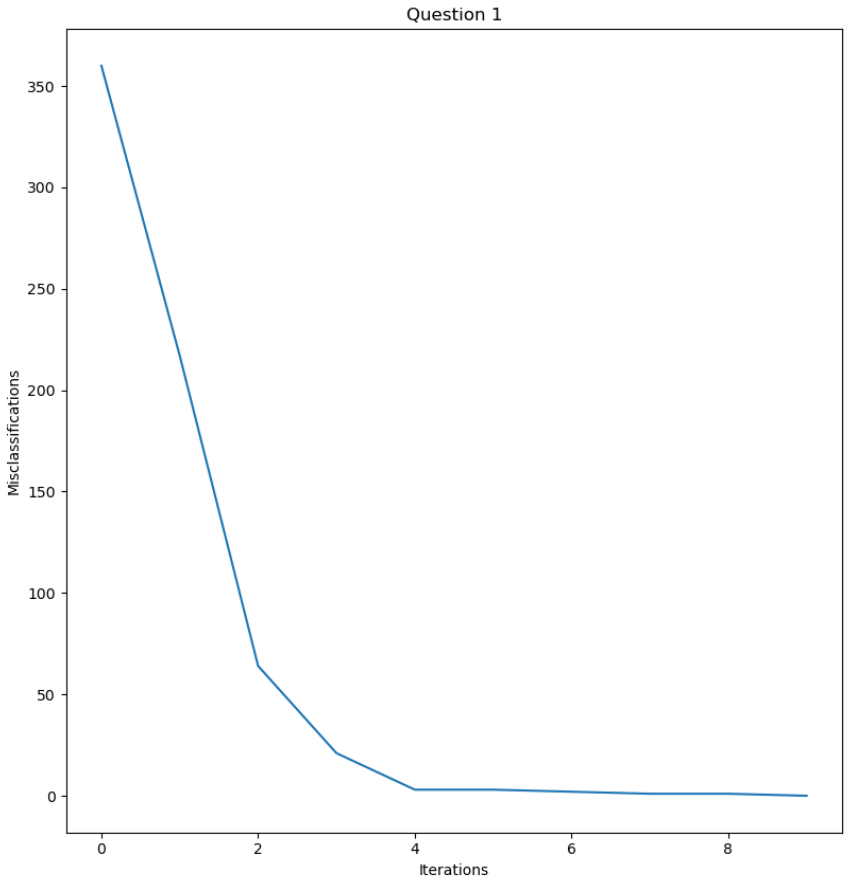

Figure 1: Misclassification v/s Iterations

Observation and remarks about the dataset:

1. Data is for certain **Linearly Separable** as all the **performance metrics** are at their **perfect values**, even with K Fold Cross Validation.
2. **PLA is very fast to converge**. With 1000 data points, it only takes 10 iterations to converge(**when data is linearly separable**).

Question 2

Plot of Misclassification v/s Iterations on 80:20 split:

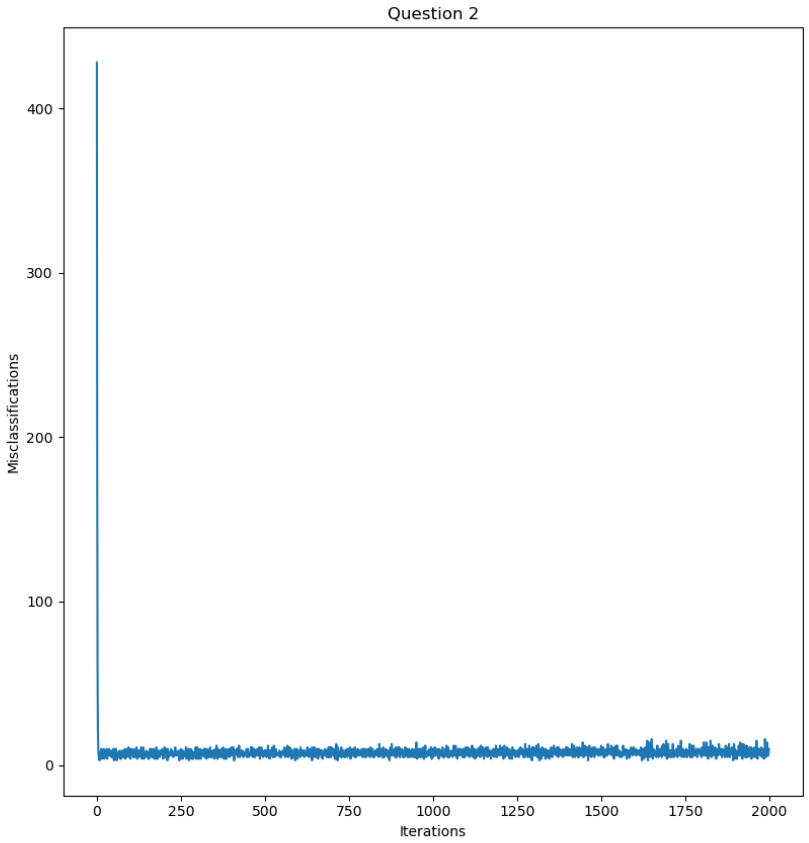

Figure 2: Misclassification v/s Iterations

Performance Metrics on the training are as follows:

Accuracy	Precision	Recall	F1
0.97	0.97	0.97	1.5

Table 3: Performance Metrics on Dataset 2

Observation and remarks about the dataset:

1. Data is for **Mostly Linearly Separable** as PLA fails to converge but hovers around a very small number of Misclassifications. It is also evident in the Performance Metrics.
2. Yet again, **PLA is very fast to converge when data is mostly linearly separable**.

Question 3

Plot of Misclassification v/s Iterations on 80:20 split:

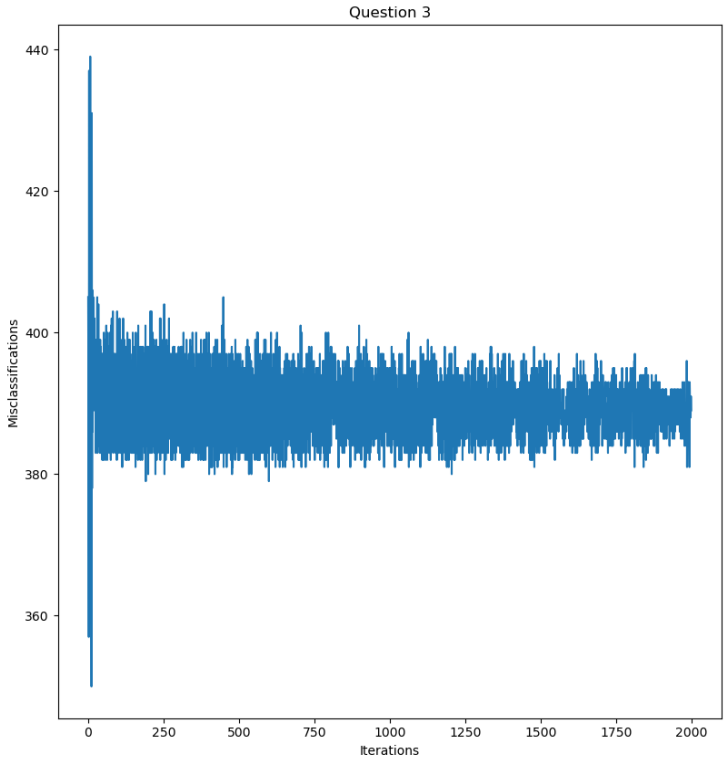

Figure 3: Misclassification v/s Iterations

Performance Metrics on the training are as follows:

Accuracy	Precision	Recall	F1
0.54	0.55	0.61	1.48

Table 4: Performance Metrics on Dataset 3

Observation and remarks about the dataset:

1. Data is for **Not Linearly Separable** as PLA fails to converge and oscillates around a very large number of Misclassifications(almost half the dataset). It is also evident in the Performance Metrics.
2. **PLA fails to converge when data is not linearly separable**.