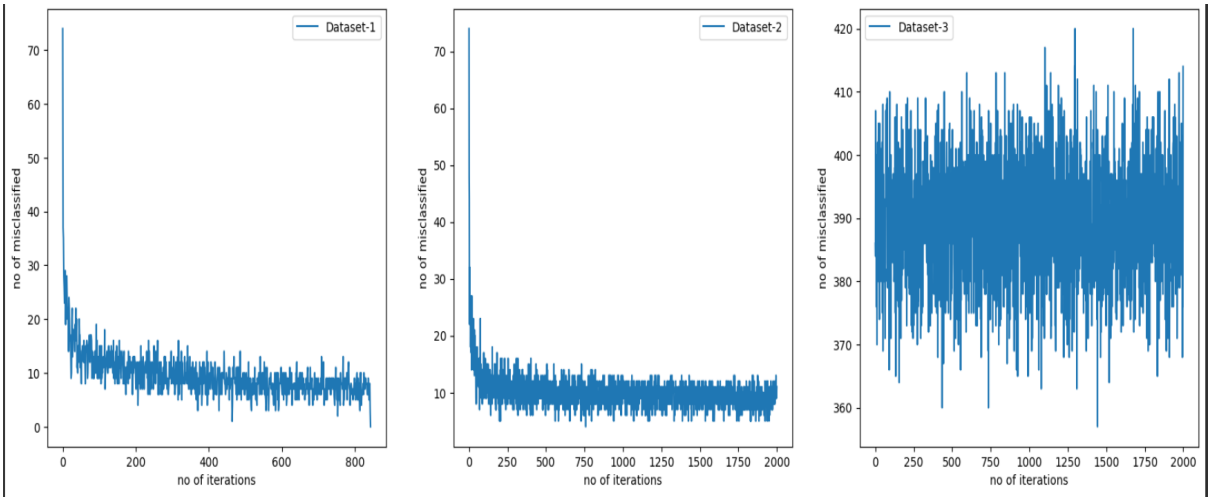
MLFA Assignment-1

Divyanshu Vaibhav
21BT10014

For Dataset-1				
	Table 1			
Folds:		Metrics		
	Accuracy	Precision score	Recall score	F1 score
Fold=2	0.998	1	0.99565217	0.99782135
	0.98	0.98418972	0.97647059	0.98031496
Mean:	0.989	0.99209486	0.98606138	0.98906816
Variance:	1.62E-10	9.67E-06	3.47E-06	5.47E-07
Fold=3	0.99101796	0.99375	0.98757764	0.99065421
	0.99099099	0.99310345	0.98630137	0.98969072
	0.99099099	1	0.98314607	0.99150142
Mean:	0.99099998	0.99561782	0.98567503	0.99061545
Variance:	1.62E-10	9.67E-06	3.47E-06	5.47E-07
Fold=5	0.995	1	0.99009901	0.99502488
	0.99	0.97802198	1	0.98888889
	0.99	0.97619048	1	0.98795181
	0.985	0.99038462	0.98095238	0.98564593
	0.995	1	0.99074074	0.99534884
Mean:	0.991	0.98891941	0.99235843	0.99057207
Variance:	1.40E-05	1.06E-04	5.09E-05	1.53E-05
Fold=10	1	1	1	1
	1	1	1	1
	1	1	1	1
	1	1	1	1
	0.99	0.97560976	1	0.98765432
	0.99	0.97674419	1	0.98823529
	0.99	1	0.98214286	0.99099099
	1	1	1	1

	0.99	1	0.98181818	0.99082569
	1	1	1	1
Mean:	0.996	0.99523539	0.9963961	0.99577063
Variance:	2.40E-05	9.09E-05	5.20E-05	2.77E-05
	Table 2	(Done for test data)	On 80:20 split train:test	
For 2nd Dataset				
	Accuracy	Precision score	Recall score	F1 score
	0.97	0.9555555556	0.9772727273	0.9662921348
For 3rd Dataset				
	0.57	0.5739130435	0.640776699	0.6055045872



Conclusions:

- Dataset-1 is linearly separable as the number of misclassified got down to zero after around ~800 iterations.
- Dataset-2 is semi-linearly separable as the number of misclassified dropped finally to around 10 on reaching max iterations, so PLA couldn't converge for it completely because of some noise in each data.
- Dataset-3 is non-linearly separable as the number of misclassified is very high and averaging around 390 till 2000 iterations(our MAX limit) .

Also we have got the metrics for all datasets, the metrics being very low for Dataset-3 also proves that it's non-linearly separable.

Output for my code:

```
For 1st dataset
 For K fold = 2
Metrics:
 [[0.998      1.         0.99565217 0.99782135]
 [0.98       0.98418972 0.97647059 0.98031496]]
Mean Metrics: [0.989      0.99209486 0.98606138 0.98906816]
Variance Metrics: [8.10000000e-05 6.24912122e-05 9.19833073e-05 7.66184239e-05]

 For K fold = 3
Metrics:
 [[0.99101796 0.99375    0.98757764 0.99065421]
 [0.99099099 0.99310345 0.98630137 0.98969072]
 [0.99099099 1.         0.98314607 0.99150142]]
Mean Metrics: [0.99099998 0.99561782 0.98567503 0.99061545]
Variance Metrics: [1.61677131e-10 9.67143942e-06 3.46929241e-06 5.47187012e-07]

 For K fold = 5
Metrics:
 [[0.995      1.         0.99009901 0.99502488]
 [0.99       0.97802198 1.         0.98888889]
 [0.99       0.97619048 1.         0.98795181]
 [0.985      0.99038462 0.98095238 0.98564593]
 [0.995      1.         0.99074074 0.99534884]]
Mean Metrics: [0.991      0.98891941 0.99235843 0.99057207]
Variance Metrics: [1.40000000e-05 1.05697111e-04 5.09214073e-05 1.53221368e-05]

 For K fold = 10
Metrics:
 [[1.         1.         1.         1.        ]
 [1.         1.         1.         1.        ]
 [1.         1.         1.         1.        ]
 [1.         1.         1.         1.        ]
 [0.99       0.97560976 1.         0.98765432]
 [0.99       0.97674419 1.         0.98823529]
 [0.99       1.         0.98214286 0.99099099]
 [1.         1.         1.         1.        ]
 [0.99       1.         0.98181818 0.99082569]
 [1.         1.         1.         1.        ]]
Mean Metrics: [0.996    0.99523539 0.9963961  0.99577063]
Variance Metrics: [2.40000000e-05 9.08702197e-05 5.19575392e-05 2.77278582e-05]
```

Metrics for 2nd dataset on test data: [0.97, 0.9555555555555556, 0.9772727272727273, 0.9662921348314608]

Metrics for 3rd dataset on test data: [0.57, 0.5739130434782609, 0.6407766990291263, 0.6055045871559634]