MLFA Assignment 1 - REPORT

Name : Shiva Ganesh Reddy Lakkasani
Roll Number : 20EE10069

A.RESULT TABLES :

1. TABLE - 1 :

Values of the performance metrics and their variances for different values of K :

a. For K = 2 :

S.No	Accuracy	Precision	Recall	F1 Score
1	0.996	0.9913793103448276	1.0	0.9956709956709957
2	0.986	0.9881889763779528	0.984313725490196	0.9862475442043221
Mean of the metrics	0.991	0.9897841433613902	0.9921568627450981	0.9909592699376589
Variance of the metrics	1.9659870489776835e-05	1.344217021371054e-06	5.546273584274039e-05	1.7162867078117587e-05

b. For K = 4 :

S.No	Accuracy	Precision	Recall	F1 Score
1	1.0	1.0	1.0	1.0
2	0.992	0.9809523809523809	1.0	0.9903846153846154
3	0.988	0.9919354838709677	0.984	0.9879518072289156
4	0.988	1.0	0.9769230769230769	0.9883268482490273
Mean of the metrics	0.992	0.9932219662058372	0.9902307692307692	0.9916658177156396
Variance of the metrics	2.097693807336304e-05	7.1726842161322e-05	9.152815834965111e-05	2.209823614688751e-05

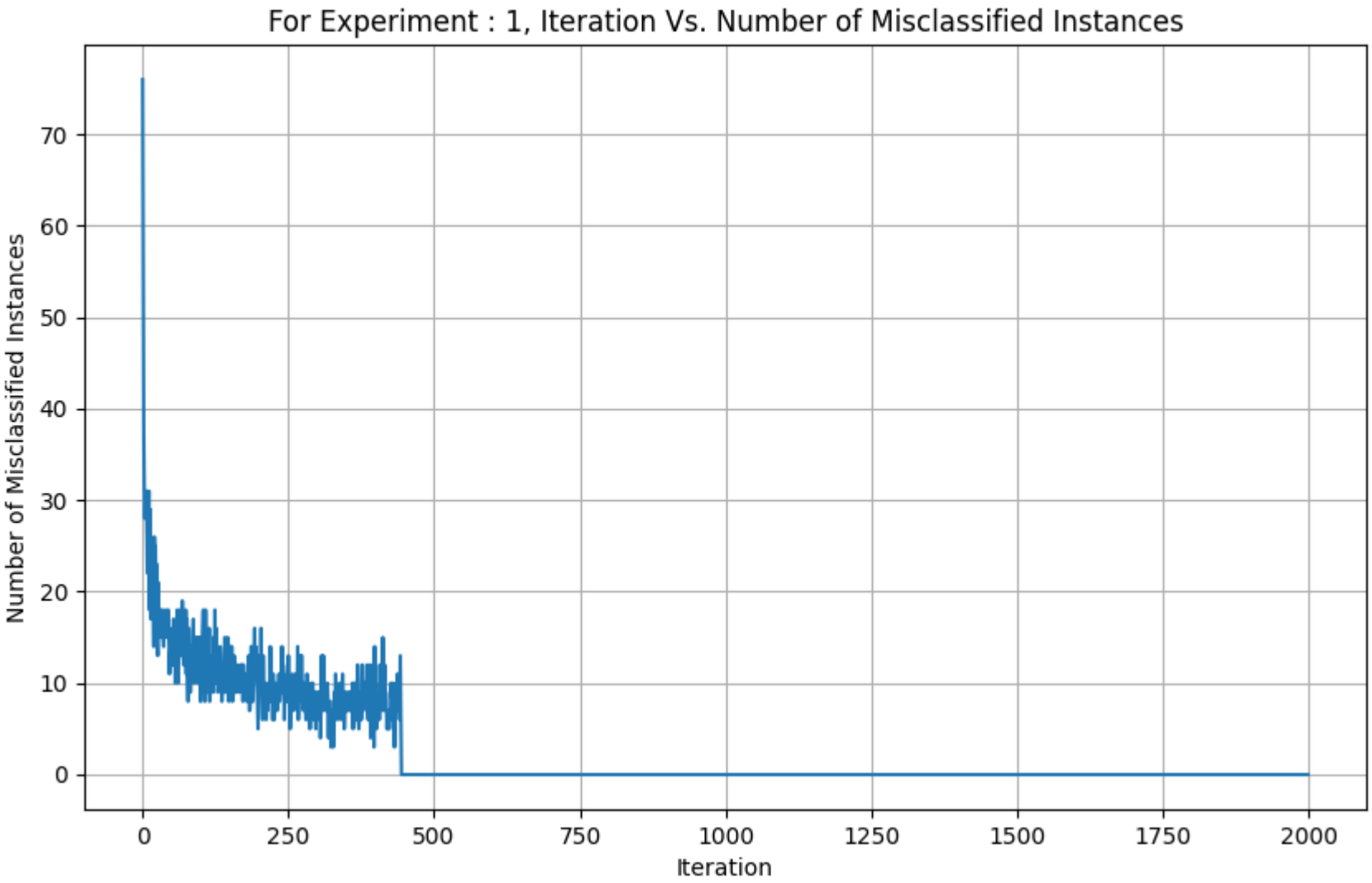
c. For K = 7 :

S.No	Accuracy	Precision	Recall	F1 Score
1	0.993006993006993	1.0	0.9863013698630136	0.993103448275862
2	1.0	1.0	1.0	1.0
3	0.986013986013986	0.9836065573770492	0.9836065573770492	0.9836065573770492
4	0.986013986013986	0.966666666666666	1.0	0.9830508474576272
5	0.972027972027972	0.9863013698630136	0.96	0.972972972972973
6	1.0	1.0	1.0	1.0
7	1.0	1.0	1.0	1.0
Mean of the metrics	0.991008991008991	0.9909392277009614	0.9899868467485804	0.9903905465833588
Variance of the metrics	1.7289644496130665e-05	0.00010949520115679022	3.390128270573564e-05	2.586623165978767e-05

d. For K = 8 :

S.No	Accuracy	Precision	Recall	F1 Score
1	1.0	1.0	1.0	1.0
2	1.0	1.0	1.0	1.0
3	0.992	0.9795918367346939	1.0	0.9896907216494846
4	0.992	0.9821428571428571	1.0	0.990990990990991
5	0.984	0.9821428571428571	0.9821428571428571	0.9821428571428571
6	0.992	0.9857142857142858	1.0	0.9928057553956835
7	1.0	1.0	1.0	1.0
8	0.992	1.0	0.9852941176470589	0.9925925925925926
Mean of the metrics	0.994	0.9911989795918368	0.9959296218487395	0.9935278647214512
Variance of the metrics	1.640428233825639e-05	6.374994752778497e-05	2.1489309197875152e-05	1.985219553068339e-05

Plot of Iteration Vs. Number of Misclassified Instances for Dataset - 1



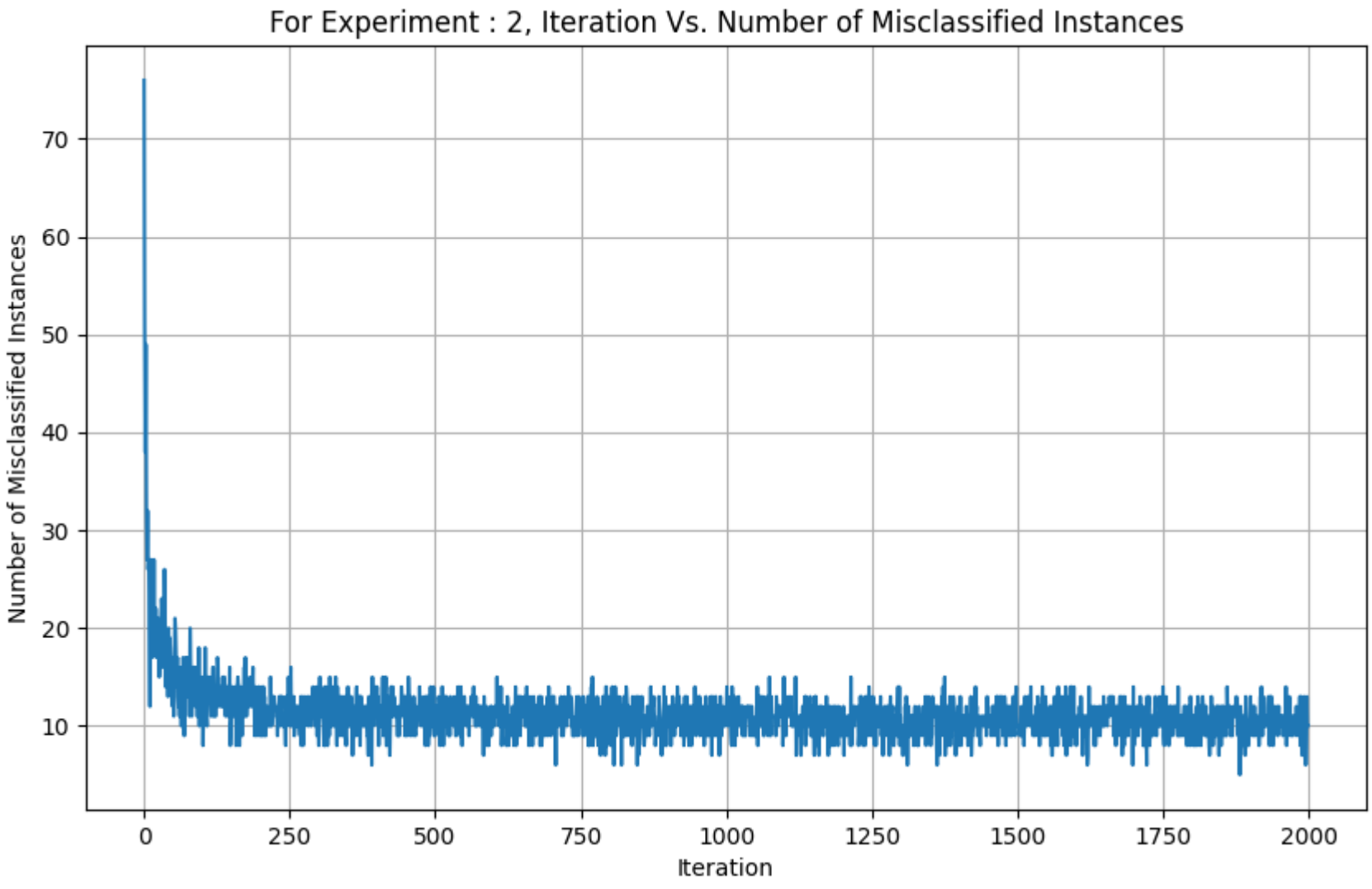
A. ii) Report training and test accuracy for experiment 2 and 3 :

TABLE - 2 :

Experiment No.	Training Accuracy	Test Accuracy
Experiment - 2	0.99	0.975
Experiment - 3	0.50875	0.49

EXPERIMENT - 2 :

Plot of Iteration Vs. Number of Misclassified Instances for Dataset - 2



OBSERVATIONS ON TRAINING FOR EXPERIMENT - 2 :

1. **Fluctuation in Misclassified Instances:**

   As the PLA iteratively adjusts the weights based on the misclassified instances, we can observe that the number of misclassified instances varies over iterations.

   This fluctuation continues until the algorithm finds a hyperplane that separates the data or until the maximum number of iterations is reached.

2. **Convergence Issues:**

   From the Iteration Vs Number of Misclassified Instances, we can conclude that the Dataset - 2 is **semi linearly separable data.**

   As we are training on **semi-linearly separable data**, the PLA might still find a hyperplane that separates the instances with a small number of misclassifications, but it won't be a perfect separation.

   This is because the Perceptron Learning Algorithm is designed to find a linear separating hyperplane, and semi-linearly separable data does not have a perfect linear separation.

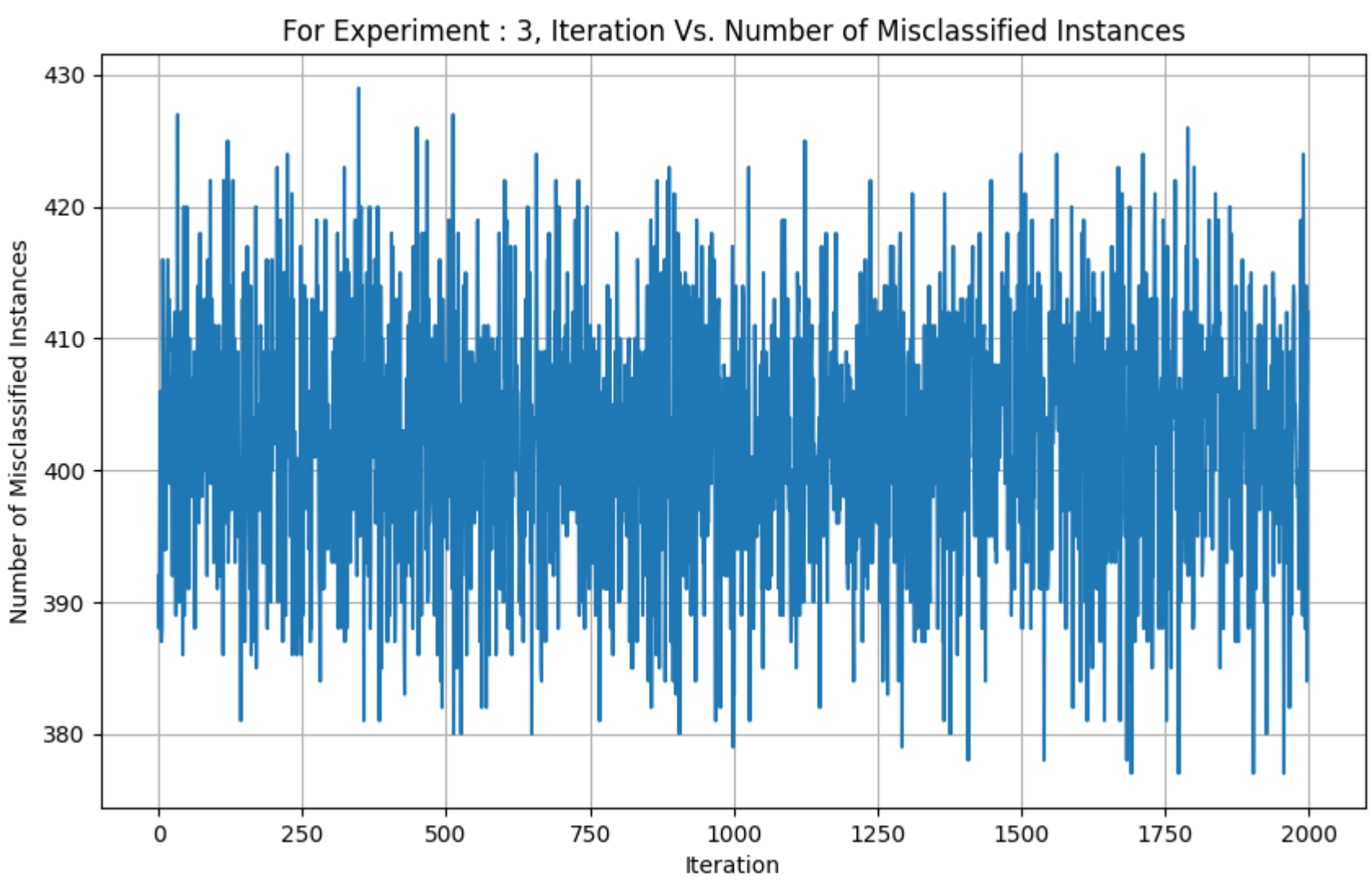
   The fluctuations are eventually stabilizing, but a s**mall number of misclassified instances are being remained over the iterations,** suggesting that the dataset is **semi linearly separable** as we can observe from the above plot.

PERFORMANCE METRICS FOR TEST DATA IN EXPERIMENT - 2 :

Experiment No.	Accuracy	Precision	Recall	F1 Score
Experiment - 2	0.975	0.9565217391304348	0.9887640449438202	0.9723756906077348

EXPERIMENT - 3 :

Plot of Iteration Vs. Number of Misclassified Instances for Dataset - 3



OBSERVATIONS ON TRAINING FOR EXPERIMENT - 3 :

1. From the Iteration Vs Number of Misclassified Instances, we can conclude that the Dataset - 3 is **Non-linearly separable data.**
2. For non-linearly separable data, the PLA will struggle to find a hyperplane that separates the instances of different classes, as no such linear hyperplane exists for this type of data.
3. Despite numerous iterations, the algorithm failed to stabilize at a **low number of misclassified instances**, suggesting that the dataset is **not linearly separable**. This can be observed from the above plot.

PERFORMANCE METRICS FOR TEST DATA IN EXPERIMENT - 3 :

Experiment No.	Accuracy	Precision	Recall	F1 Score
Experiment - 3	0.49	0.5108695652173914	0.4519230769230769	0.4795918367346939

CONNECT EXPERIMENT 1, EXPERIMENT 2 AND EXPERIMENT 3 BY COMMENTING ON THE NATURE OF THE DATASETS :

By Observing the **Iteration Vs. Number of Misclassified Instances Plots**, we can conclude that the :

1. **Dataset - 1 is “Linearly Separable Data”.**
2. **Dataset - 2 is “Semi - Linearly Separable Data”.**
3. **Dataset - 3 is “Not - Linearly Separable Data”.**

CONCLUSION :

i.) What can you say about the datasets used in three experiments?

Ans :

By Observing the **Iteration Vs. Number of Misclassified Instances Plots**, that we have plotted on the “train data”, we can conclude that the :

1. **Dataset - 1 is “Linearly Separable Data”.**
2. **Dataset - 2 is “Semi - Linearly Separable Data”.**
3. **Dataset - 3 is “Not - Linearly Separable Data”.**

ii.) How do you connect the nature of the datasets with the experimental results?

Ans :

A. For Dataset - 1 :

For a linearly separable dataset, the PLA will eventually find a hyperplane that perfectly separates the two classes, i.e., it will classify all instances correctly.

In the **"Iteration vs Number of Misclassified Instances" plot**, we can observe that the number of misclassified instances **decreases to zero and stays at zero**. The point where the number of misclassified instances reaches zero is **where the algorithm has found a separating hyperplane.**

The algorithm will converge, and once it does, **it won't change the weights of the hyperplane**, as there will be no more misclassified instances to adjust the weights.

B. For Dataset - 2 :

For semi-linearly separable dataset, the PLA might still find a hyperplane that separates the instances with a small number of misclassifications, but it won't be a perfect separation.

In the **"Iteration vs Number of Misclassified Instances" plot**, we can observe a fluctuating pattern, where the number of misclassified instances decreases and then possibly increases slightly, **showing small oscillations** but not reaching zero.

The fluctuations may eventually stabilize, **but a small number of misclassified instances will likely remain**.

C. For Dataset - 3 :

For non-linearly separable data, the PLA struggles to find a hyperplane that adequately separates the instances of different classes, as no such linear hyperplane exists for this type of data.

In the **"Iteration vs Number of Misclassified Instances" plot**, you will observe large fluctuations in the number of misclassified instances, and **the number of misclassified instances does not stabilize to a low value.**

The algorithm might continue to adjust weights without showing a clear trend towards decreasing misclassified instances, even as the number of iterations becomes large.

THANK YOU !