MLFA Assignment 3 - REPORT

Name : Shiva Ganesh Reddy Lakkasani
Roll Number : 20EE10069

DATA SET INFORMATION :

[150 rows x 5 columns]
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 5 columns):

#	Column	Non-Null Count	Dtype
0	SepalLengthCm	150 non-null	float64
1	SepalWidthCm	150 non-null	float64
2	PetalLengthCm	150 non-null	float64
3	PetalWidthCm	150 non-null	float64
4	Species	150 non-null	object

dtypes: float64(4), object(1)
memory usage: 6.0+ KB

Dataset sizes after splitting in 60:20:20 ratio :

Training dataset: (90, 4) (90,)
Validation dataset: (30, 4) (30,)
Testing dataset: (30, 4) (30,)

EXPERIMENT - 1 :

[EXPT - 1] Performance Metrics of the model WITH FEATURE SCALING for learning rate = 1e-05 :

Accuracy: 0.62
Precision: 0.3353
Recall: 0.4333
F1 Score: 0.3572

[EXPT - 1] Performance Metrics of the model WITH FEATURE SCALING for learning rate = 0.0001 :

Accuracy: 0.58
Precision: 0.1932

Recall: 0.3667
F1 Score: 0.2454

[EXPT - 1] Performance Metrics of the model WITH FEATURE SCALING for learning rate = 0.001 :

Accuracy: 0.80
Precision: 0.7981
Recall: 0.7000
F1 Score: 0.6069

[EXPT - 1] Performance Metrics of the model WITH FEATURE SCALING for learning rate = 0.01 :

Accuracy: 0.93
Precision: 0.9024
Recall: 0.9000
F1 Score: 0.8997

[EXPT - 1] Performance Metrics of the model WITH FEATURE SCALING for learning rate = 0.05 :

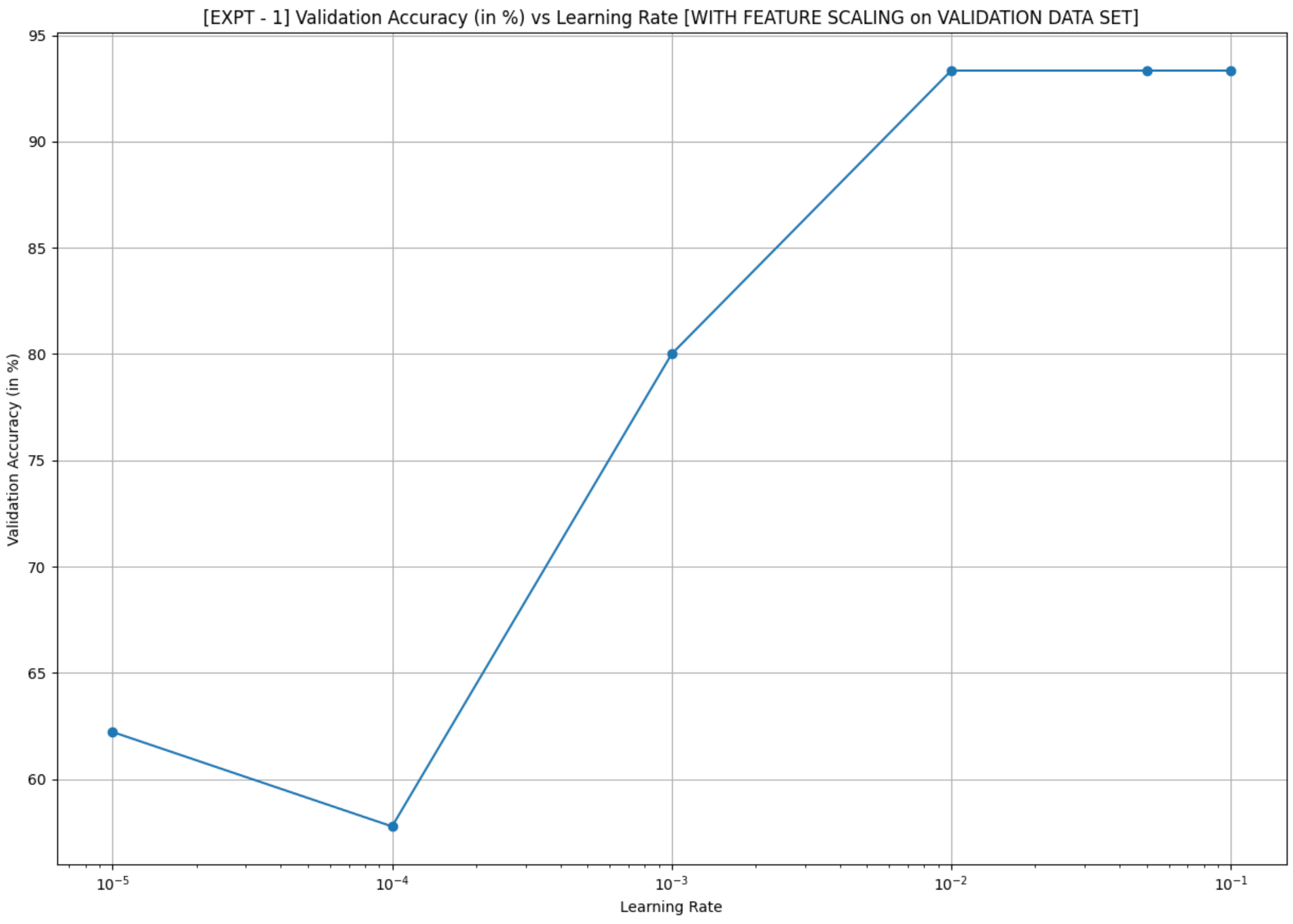
Accuracy: 0.93
Precision: 0.9024
Recall: 0.9000
F1 Score: 0.8997

[EXPT - 1] Performance Metrics of the model WITH FEATURE SCALING for learning rate = 0.1 :

Accuracy: 0.93
Precision: 0.9024
Recall: 0.9000
F1 Score: 0.8997

EXPERIMENT - 2 :

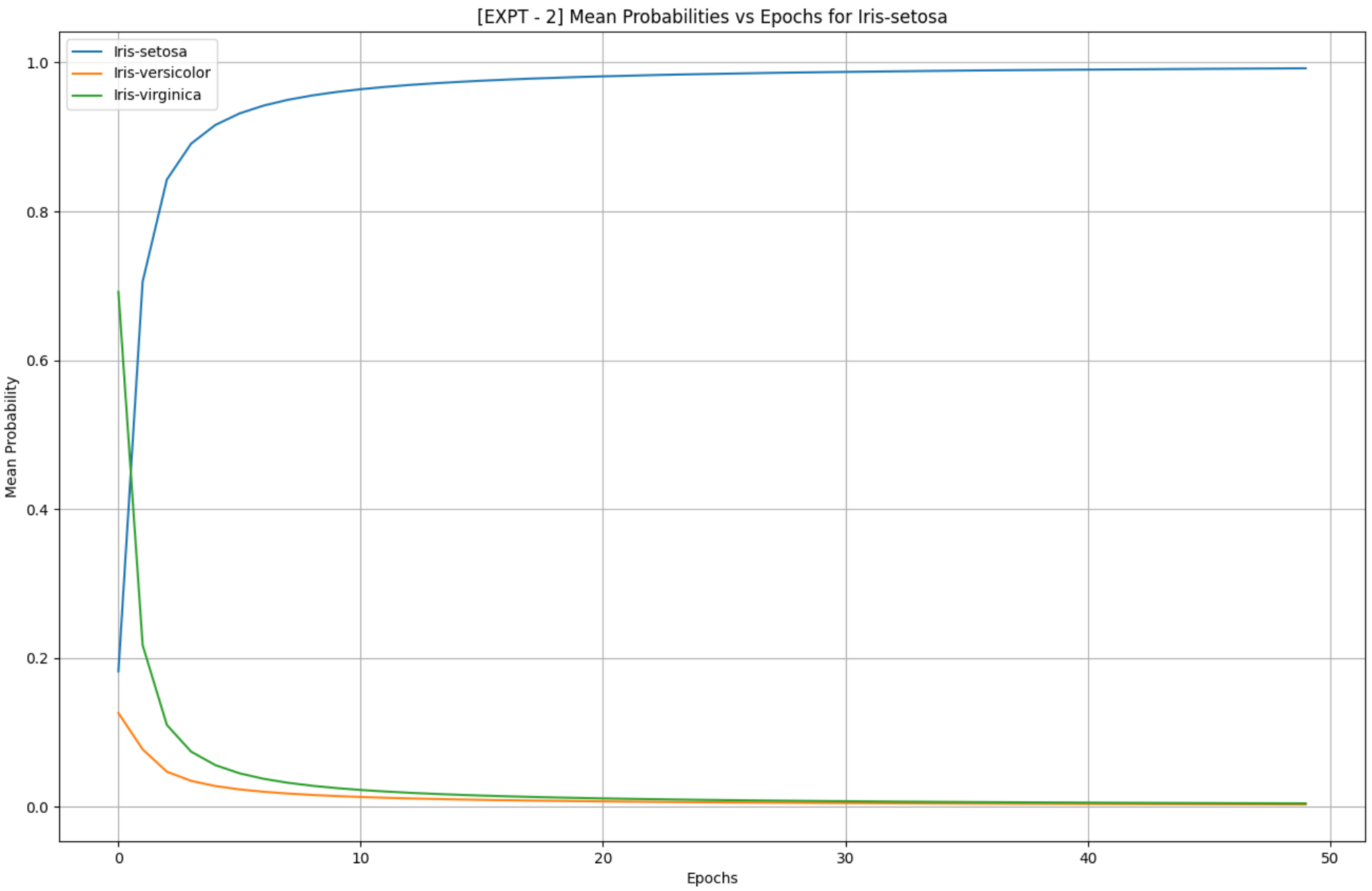
Percentage Accuracy (in %) vs Learning rate plot :



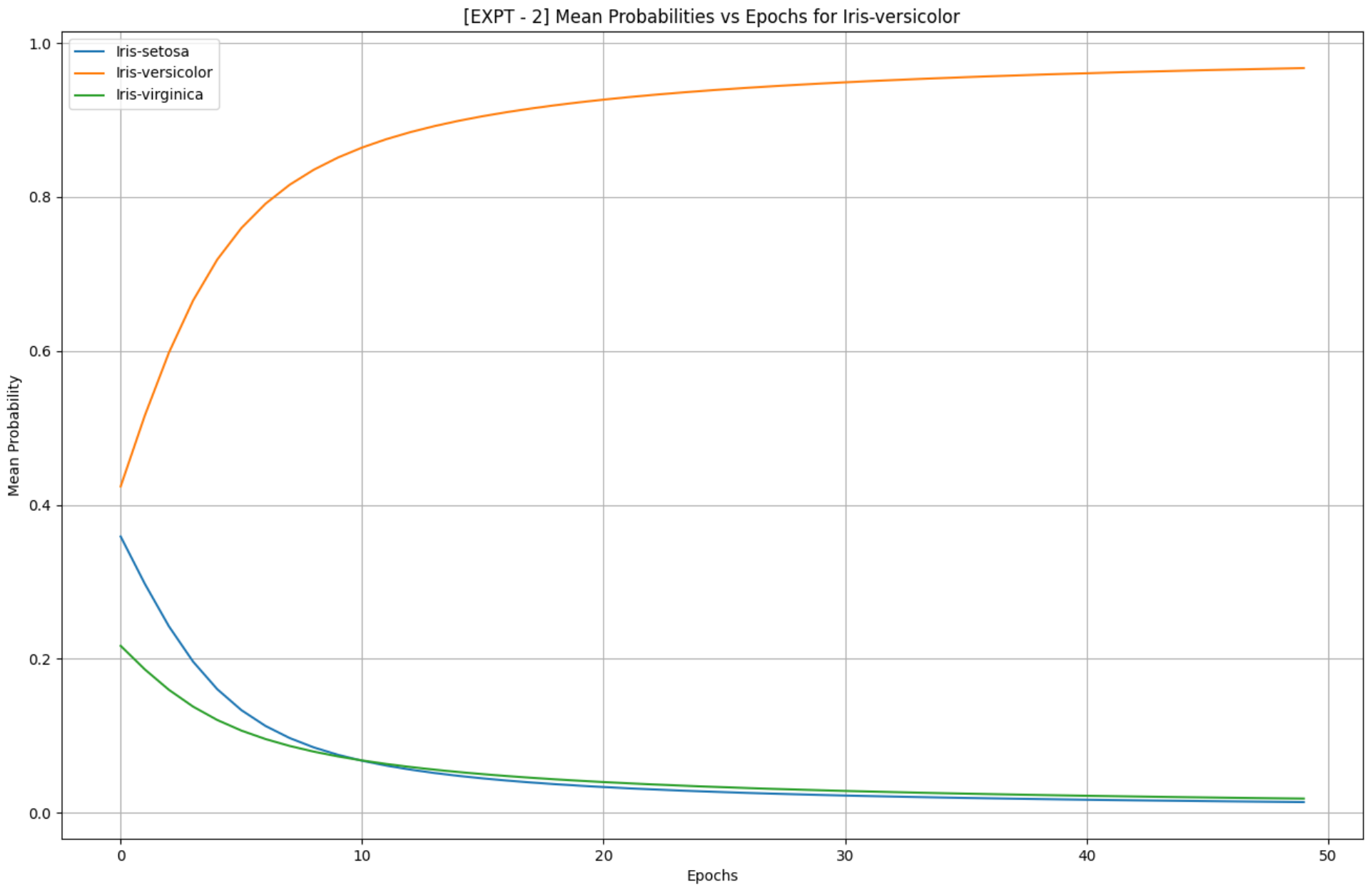
Best value of Hyperparameter Learning Rate WITH FEATURE SCALING on VALIDATION DATA SET : **0.01**

EXPERIMENT - 3 :

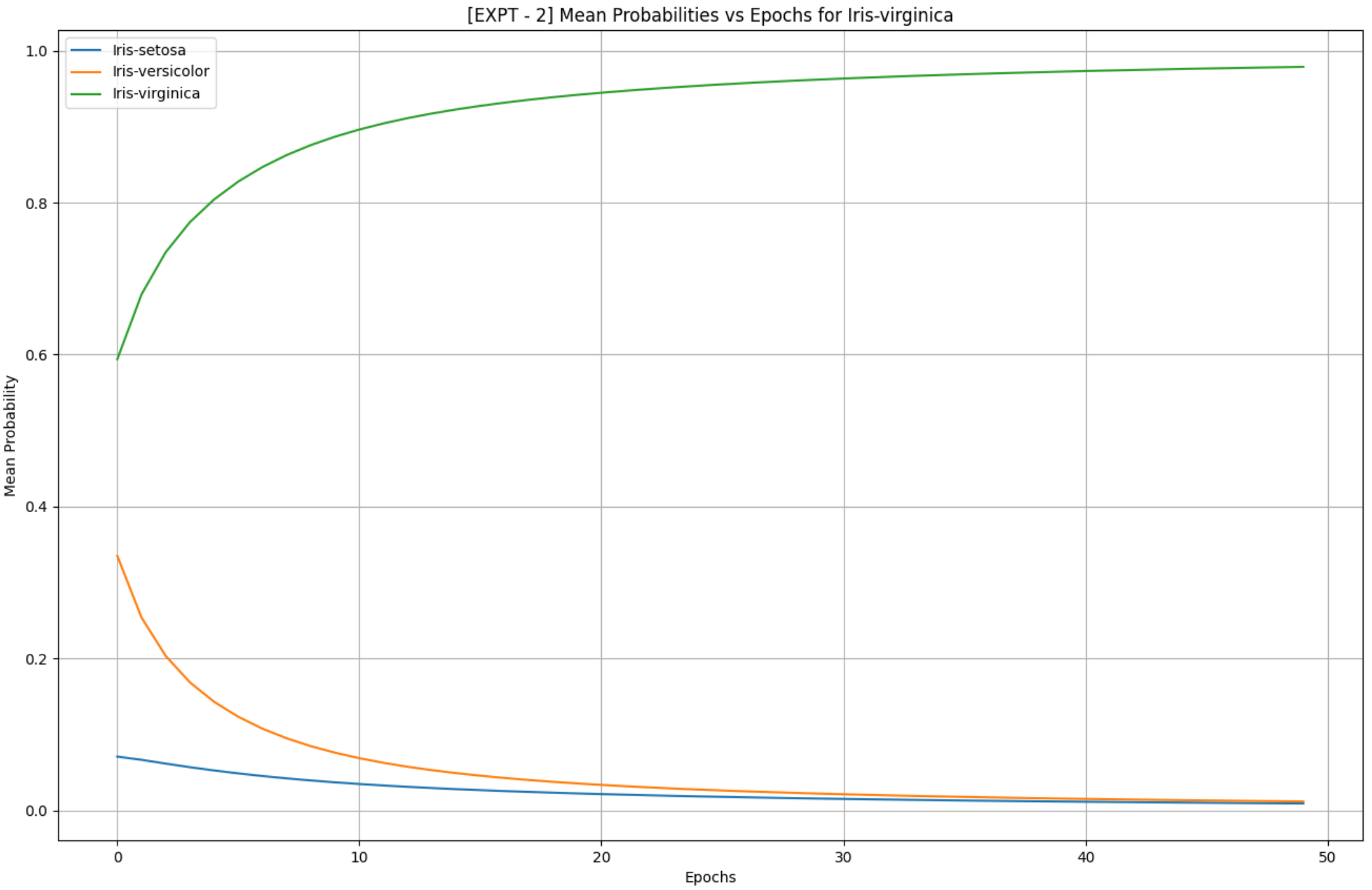
1. Mean Probabilities vs Epochs for Iris-setosa :



2. Mean Probabilities vs Epochs for Iris-versicolor :



3. Mean Probabilities vs Epochs for Iris-virginica :



[EXPT - 3] OBSERVATIONS :

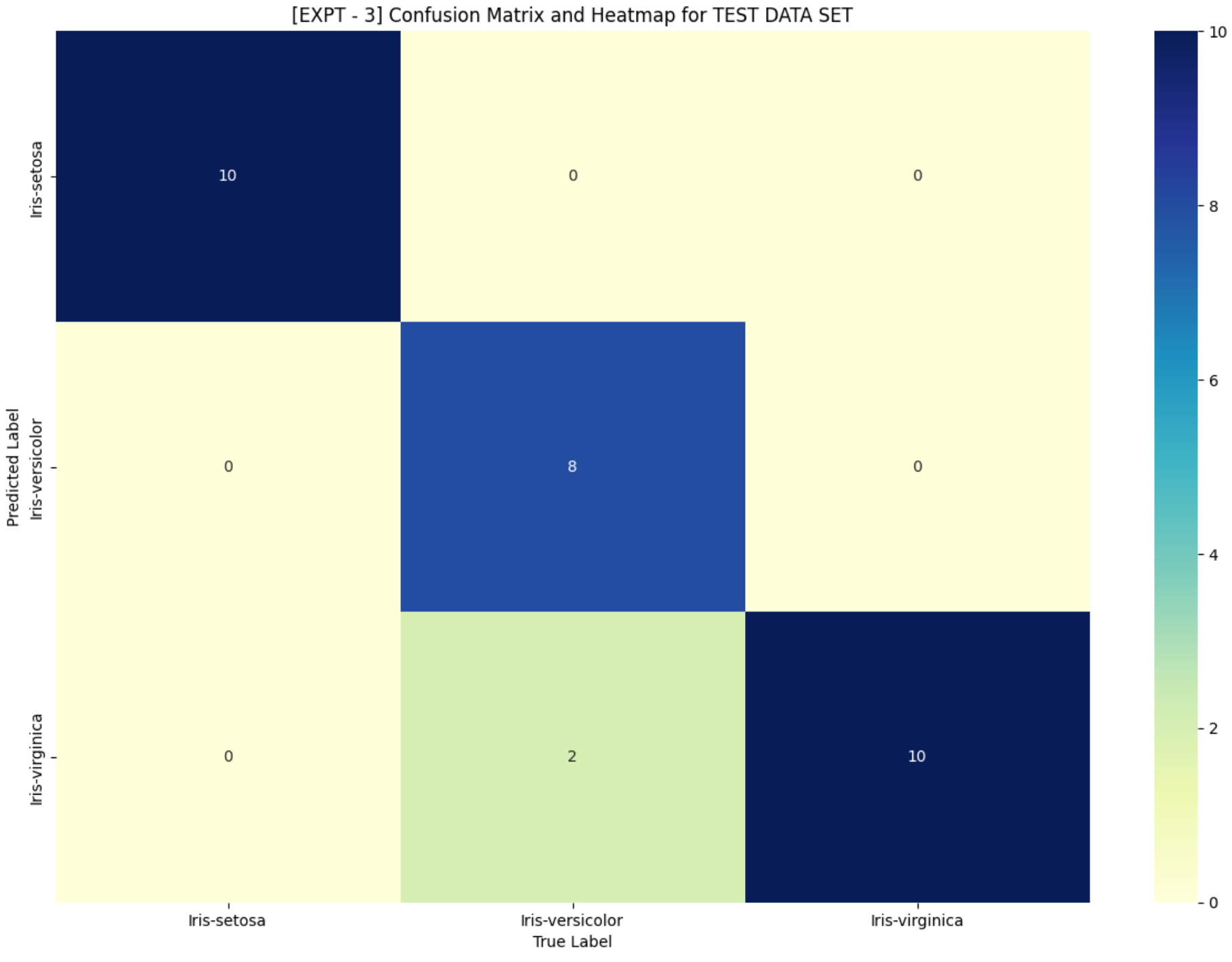
The observations of from the plots of **Mean Probabilities vs Epochs** for different classes of the softmax regression trained on the Iris dataset, are:

1. **For correct classes** : The mean probabilities for true class labels gradually rise towards 1.0 as the classifier becomes more confident in its predictions for that class.
2. **For incorrect classes** : The mean probabilities for other class labels gradually fall towards 0.0, indicating that the classifier is increasingly confident that those samples do not belong to those other classes.

EXPERIMENT - 4 :

A. [EXPT - 4] For the TEST DATA :

1. Confusion Matrix and the Heat Map :



2. Accuracy, Precision, Recall and f1-score for individual class :

Performance metrics for Class in TEST DATA SET **: Iris-setosa**

Accuracy: 1.0000
Precision: 1.0000
Recall: 1.0000
F1 Score: 1.0000

Performance metrics for Class in TEST DATA SET **: Iris-versicolor**

Accuracy: 0.9333
Precision: 1.0000
Recall: 0.8000
F1 Score: 0.8889

Performance metrics for Class in TEST DATA SET **: Iris-virginica**

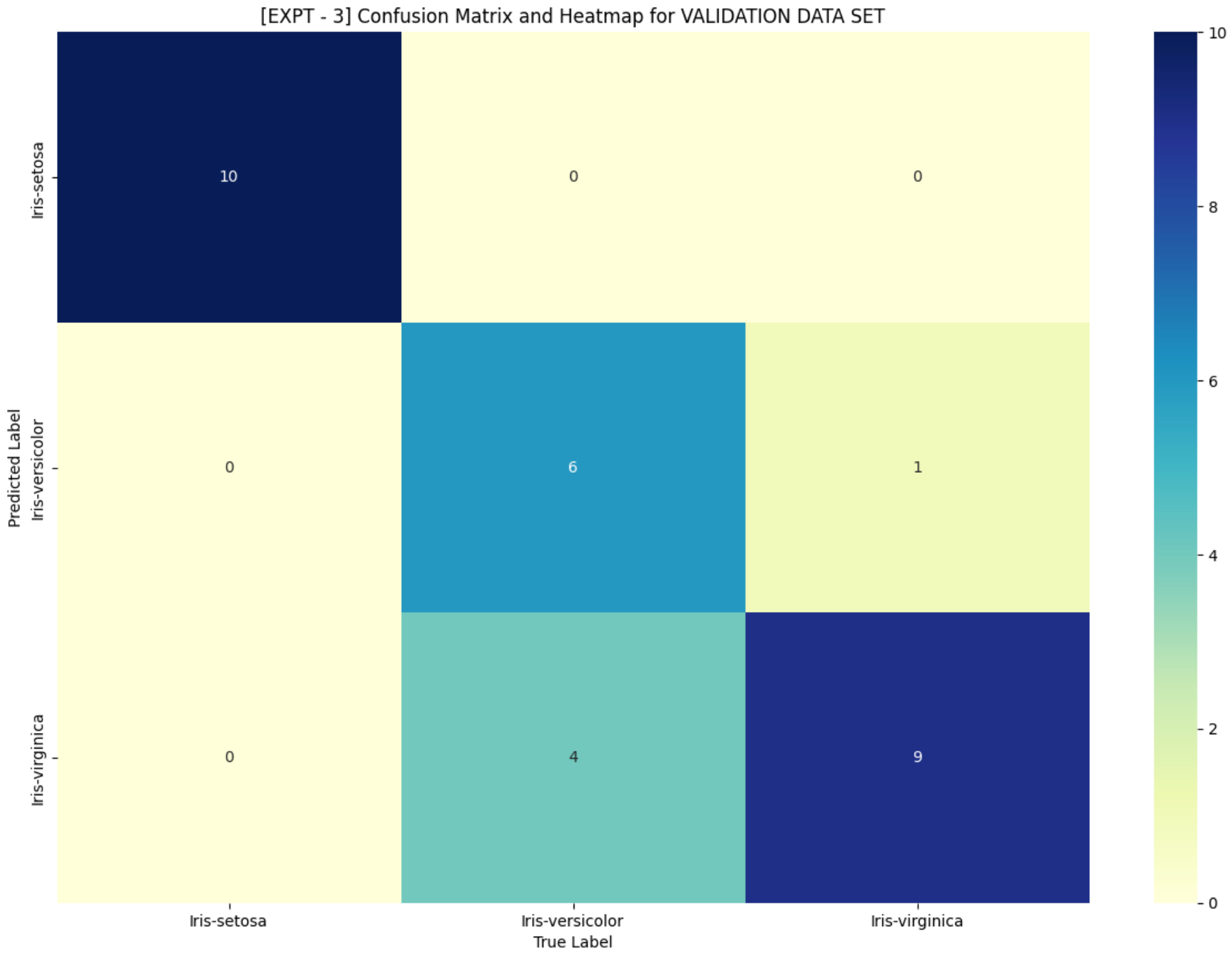
Accuracy: 0.9333
Precision: 0.8333
Recall: 1.0000
F1 Score: 0.9091

[EXPT - 4] TABLE for TEST DATA :

	Precision	**Recall**	**F1 Score**
Iris-setosa	1.0000	1.0000	1.0000
Iris-versicolor	1.0000	0.8000	0.8889
Iris-virginica	0.8333	1.000	0.9091

B. [EXPT - 4] For the VALIDATION DATA :

1. Confusion Matrix and the Heat Map :



Performance metrics for Class in VALIDATION DATA SET **: Iris-setosa**

Accuracy: 1.0000
Precision: 1.0000
Recall: 1.0000
F1 Score: 1.0000

Performance metrics for Class in VALIDATION DATA SET **: Iris-versicolor**

Accuracy: 0.8333
Precision: 0.8571
Recall: 0.6000
F1 Score: 0.7059

Performance metrics for Class in VALIDATION DATA SET **: Iris-virginica**

Accuracy: 0.8333
Precision: 0.6923
Recall: 0.9000
F1 Score: 0.7826

[EXPT - 4] TABLE for VALIDATION DATA :

	Precision	Recall	F1 Score
Iris-setosa	1.0000	1.0000	1.0000
Iris-versicolor	0.8571	0.6000	0.7059
Iris-virginica	0.6923	0.9000	0.7826

[EXPT - 4] OBSERVATIONS :

1. **Diagonal Dominance :** From the confusion matrix and Heat Map, we can observe most of the values are present on the diagonal of the matrix, it indicates that the model is performing well, as it is predicting many **true positives (TP)** and **true negatives (TN)**.

2. **Off-Diagonal Values** : The values outside the diagonal indicate misclassification. The row represents the true class, and the column represents the predicted class. We can observe that the off-diagonal values are very less bright in Heat Map, which indicates that the model is performing well.

- **Precision [TP / (TP + FP)]**: Precision reflects the accuracy of positive predictions for each class. High precision indicates a low rate of false positives.

- **Recall [TP / (TP + FN)]**: Recall represents the model's ability to identify all instances of a class. High recall means most positive samples were correctly identified.

- **F1-Score [2 * (Precision * Recall) / (Precision + Recall)]**: The harmonic mean of precision and recall, is used to measure overall model accuracy, especially in imbalanced datasets.

3. **Class-wise Evaluation**: Classes with high precision but low recall might indicate the model is very selective when predicting that class but misses several actual instances. Conversely, high recall but low precision could mean the model frequently misclassifies other samples as this class.

4. **Imbalanced Dataset**: A class with fewer samples might have high precision but low recall, signifying the model rarely predicts this class and thus has few false positives, but misses many actual instances.

5. **Noisy Data**: If the dataset for a particular class is noisy or overlaps with other classes, both precision and recall may be reduced, resulting in a low F1-score.