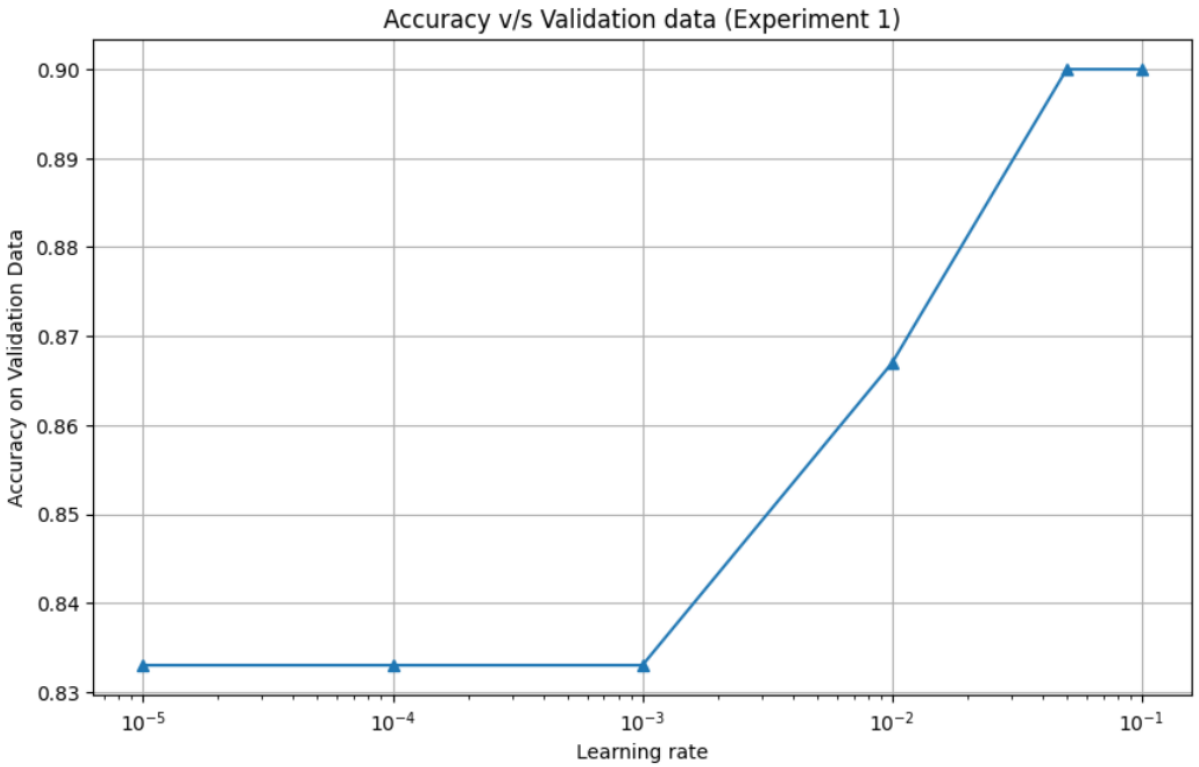
Assignment 3

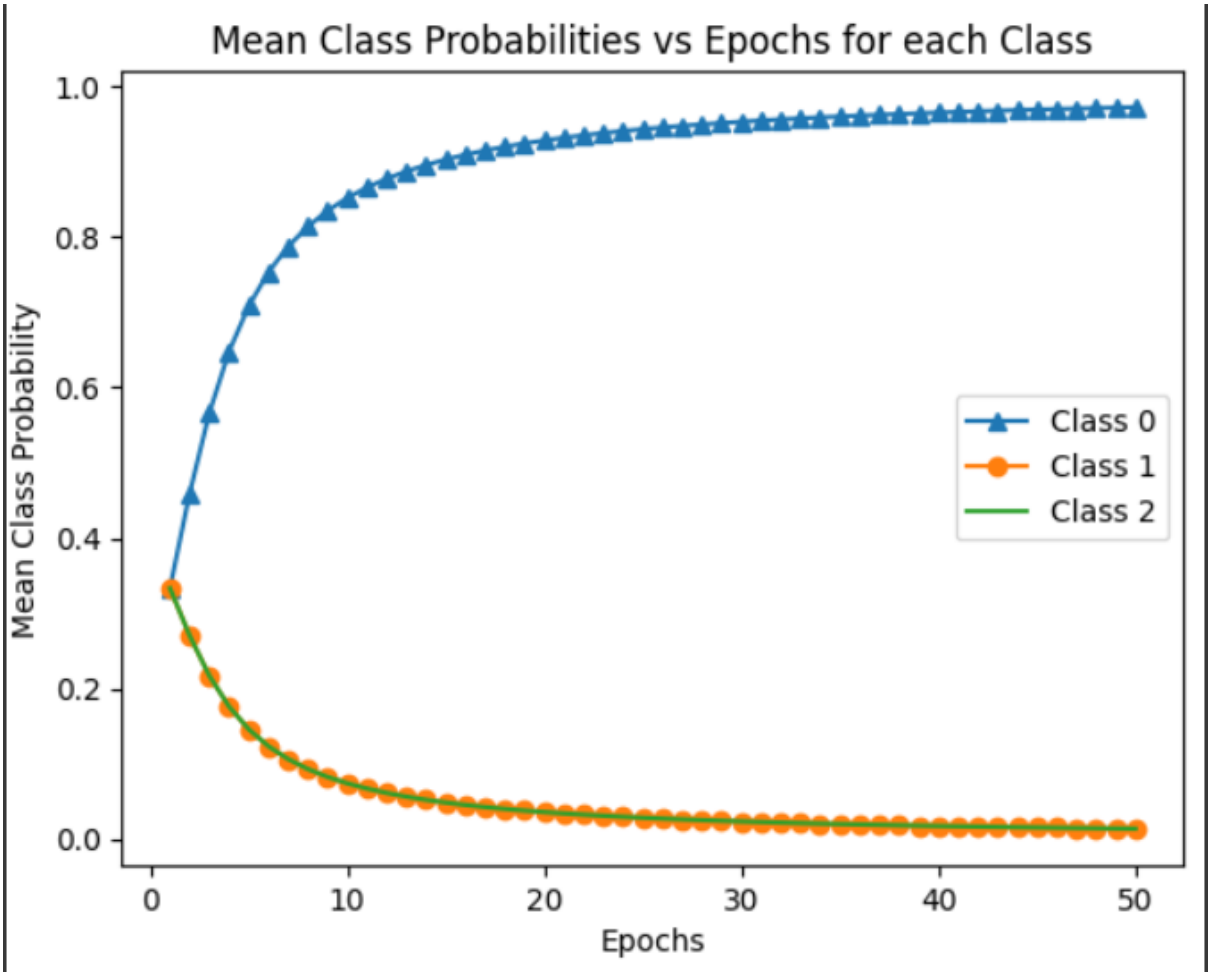
Divyanshu Vaibhav
21BT10014

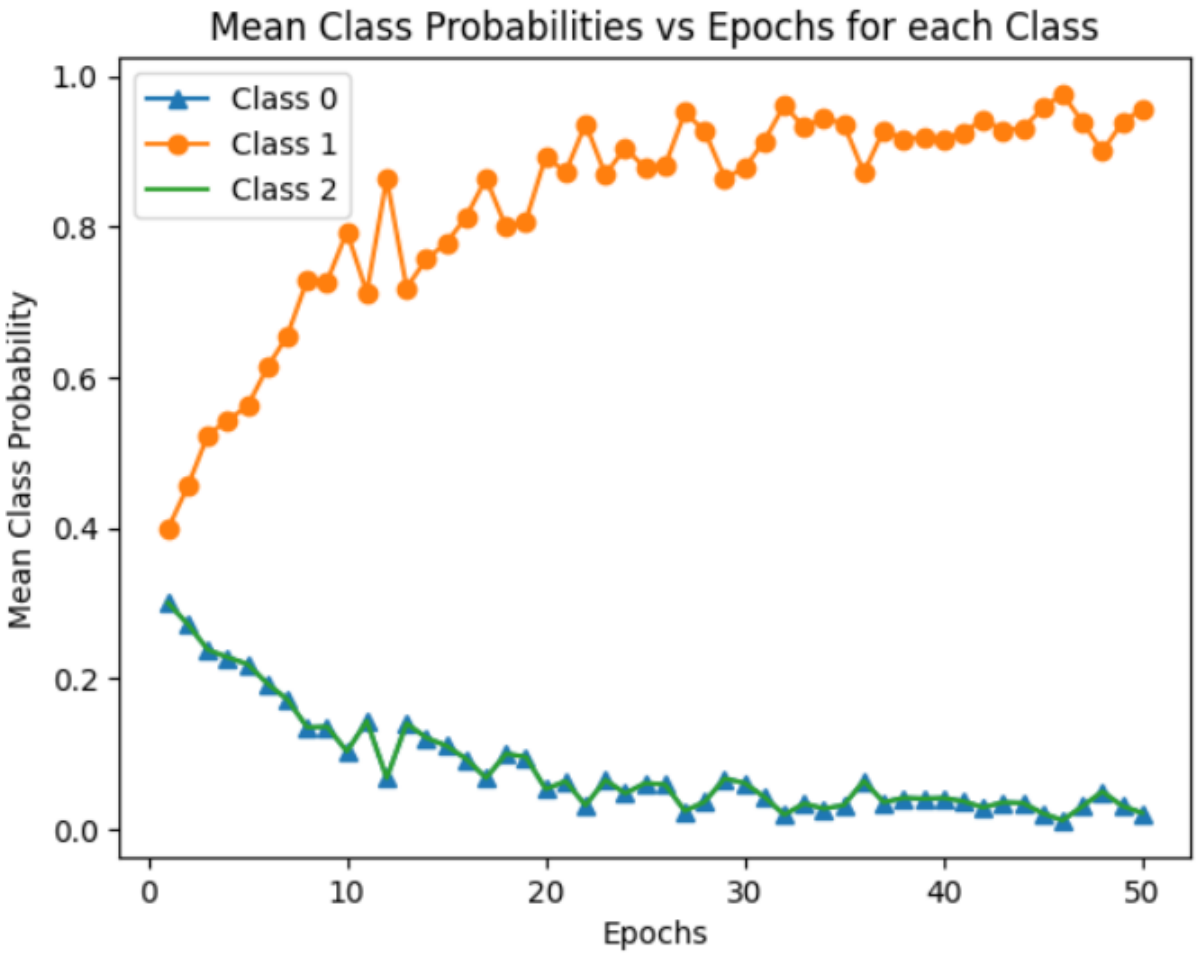
Experiment 1

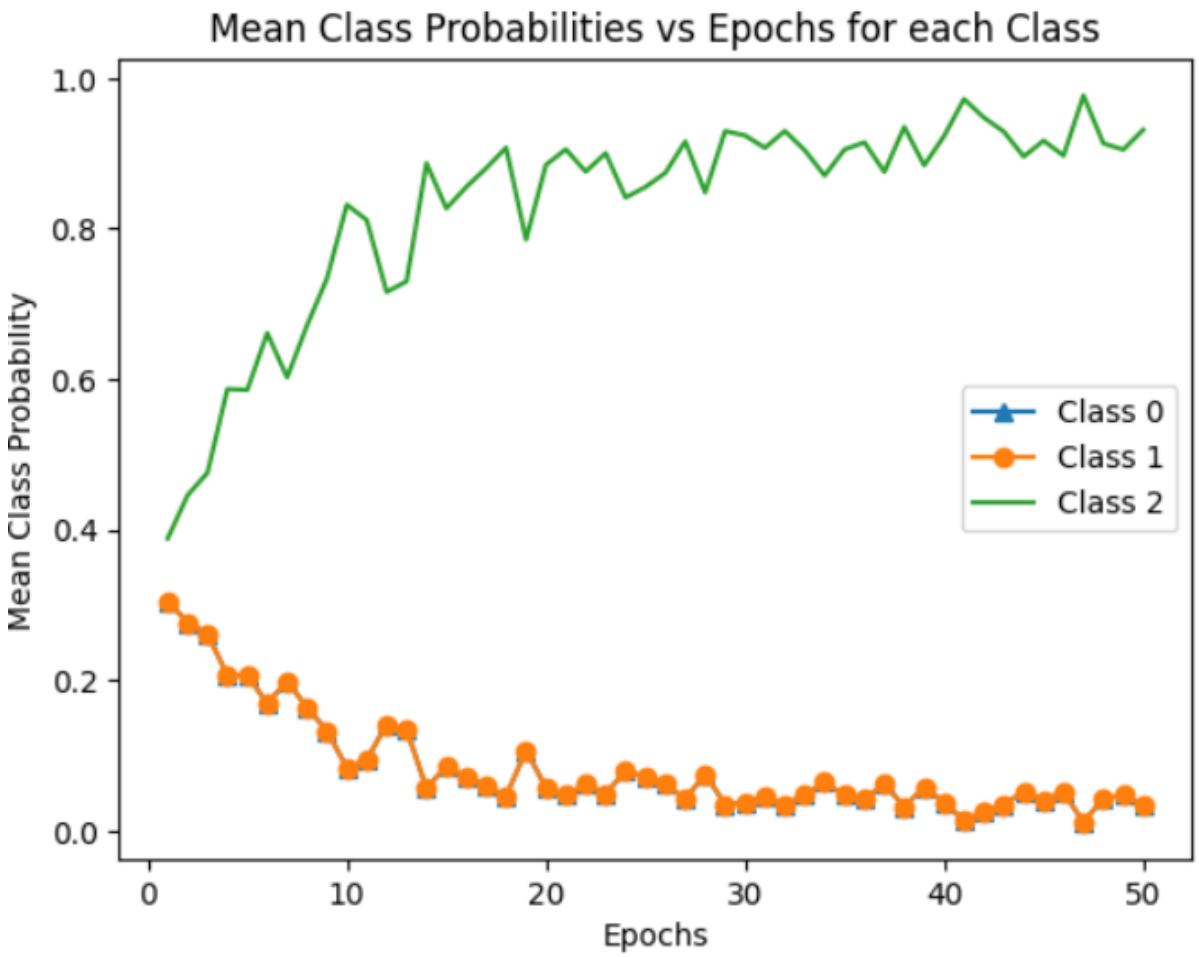


- Best learning rate found was to be 0.1 for which accuracy was 90%
- The confusion matrix also suggests that 0.1 is best when tested on test data.

Experiment 2

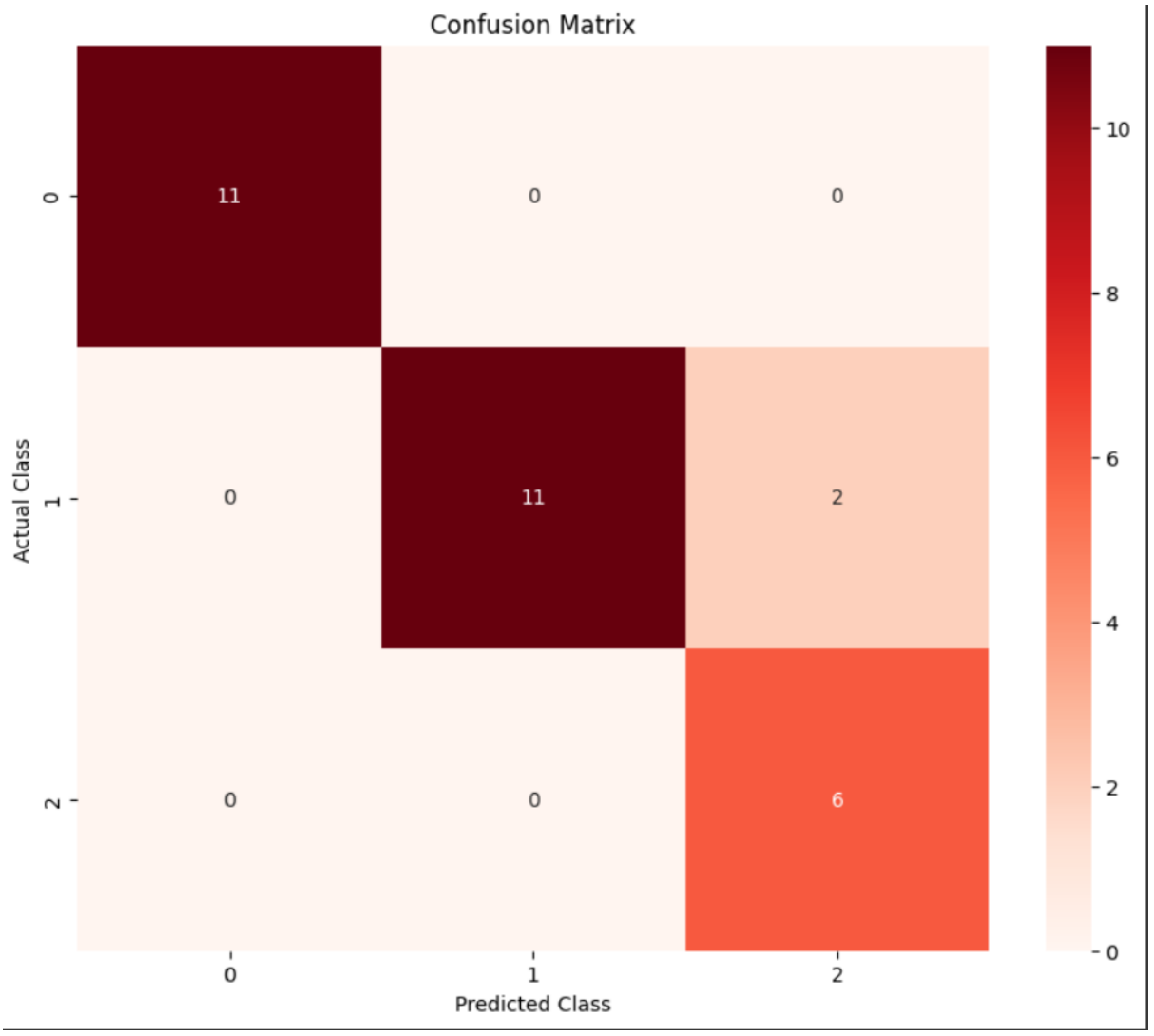



Mean Class Probabilities vs Epochs for each Class
Class 0
Class 1
Class 2
Mean Class Probability
Epochs
1.0
0.8
0.6
0.4
0.2
0.0
0
10
20
30
40
50


Mean Class Probabilities vs Epochs for each Class
Mean Class Probability
Epochs
1.0
0.8
0.6
0.4
0.2
0.0
0
10
20
30
40
50
Class 0
Class 1
Class 2

Experiment 3

Class	1	2	3
1	11	0	0
2	0	11	2
3	0	0	6



Class ⬇️	Precision	Recall	F1
1	1	1	1
2	1	0.85	0.92
3	0.75	1	0.86

Accuracy	0.93		
Macro Avg	0.92	0.95	0.92
Weighted Avg	0.95	0.93	0.94

Observations-

• The accuracy observed was 93%

• The jagged appearance of the mean class probability curve for a class 1 and 2 in logistic regression is likely due to the following reasons:

1). Stochasticity in Mini-Batch Gradient Descent: In the code provided earlier, we are using mini-batch gradient descent for training the logistic regression model. Mini-batch gradient descent updates the model's weights and biases based on a subset (mini-batch) of the training data in each iteration. This introduces stochasticity or randomness into the optimization process. As a result, the model's parameters (weights and biases) may not change smoothly in a single direction, leading to jagged curves in the mean class probability.

2).Noise in Data: Even if we are training on a dataset consisting purely of data from one class, there can still be some inherent noise or variability in the data. This noise can lead to fluctuations in the model's predictions and, consequently, in the mean class probability curve.

3).Learning Rate and Optimization Parameters: The choice of learning rate and other hyperparameters can affect the convergence of the optimization algorithm.