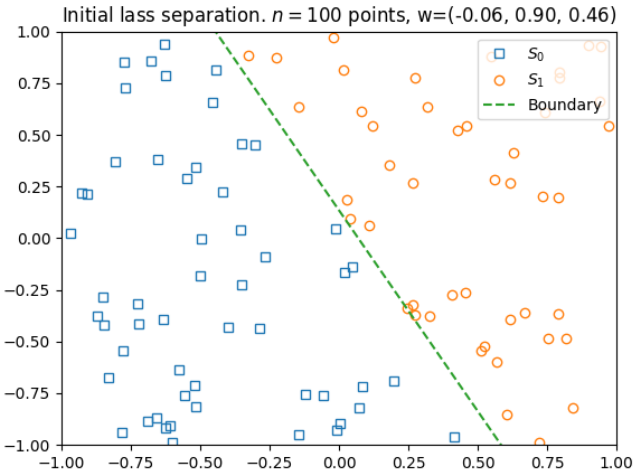
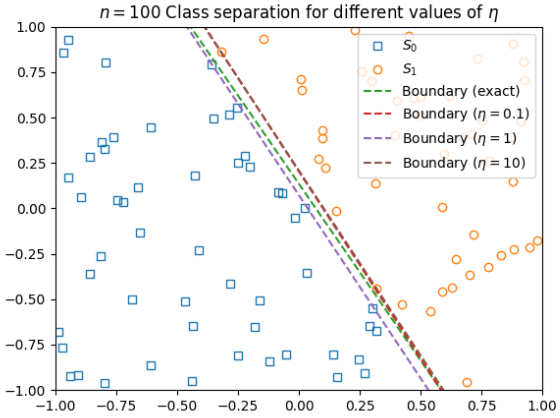
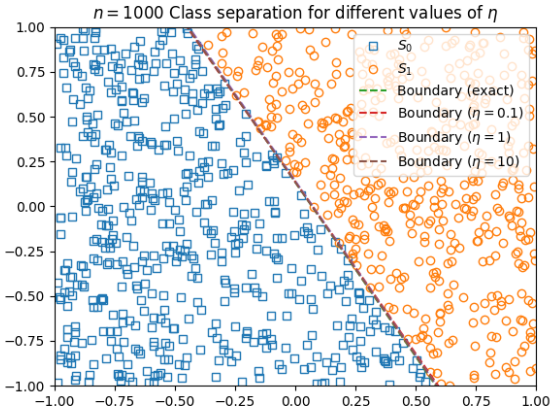
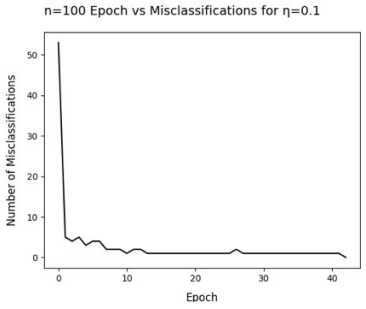
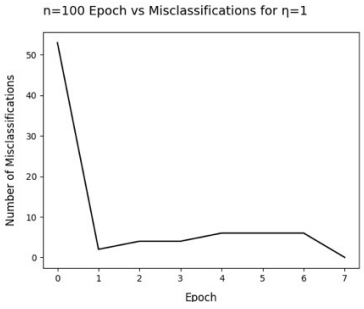
HW2 Yaroslav Popryho
ypopry2@uic.edu

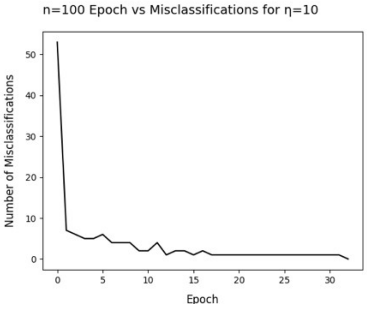


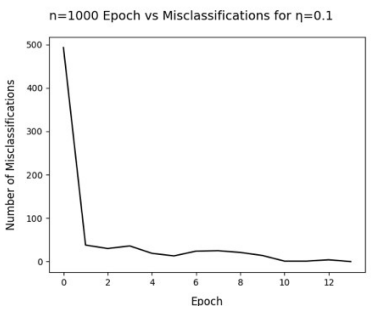


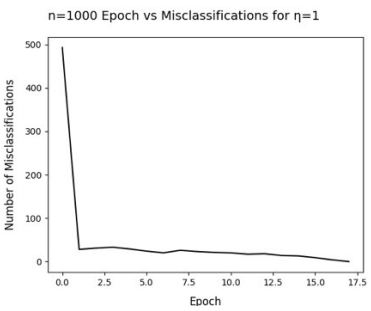


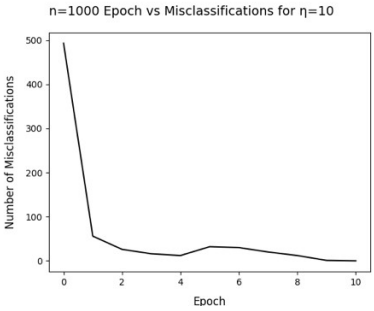












n=100 eta=0.1 epochs[eta]=42
w = [-0.06272994 0.90142861 0.46398788]
w' = [0.79710838 0.21285812 -0.9816059]
w'' = [-0.10289162 1.04000926 0.50522601]

n=100 eta=1 epochs[eta]=7
w = [-0.06272994 0.90142861 0.46398788]
w' = [0.79710838 0.21285812 -0.9816059]
w'' = [-0.20289162 5.62563639 2.79336622]

n=100 eta=10 epochs[eta]=32
w = [-0.06272994 0.90142861 0.46398788]
w' = [0.79710838 0.21285812 -0.9816059]
w'' = [-9.20289162 89.88405489 44.19493733]

n=1000 eta=0.1 epochs[eta]=13
w = [-0.06272994 0.90142861 0.46398788]
w' = [0.79710838 0.21285812 -0.9816059]
w'' = [-0.10289162 1.43011695 0.74644818]

n=1000 eta=1 epochs[eta]=17
w = [-0.06272994 0.90142861 0.46398788]
w' = [0.79710838 0.21285812 -0.9816059]
w'' = [-1.20289162 16.52181938 8.51457458]

n=1000 eta=10 epochs[eta]=10
w = [-0.06272994 0.90142861 0.46398788]
w' = [0.79710838 0.21285812 -0.9816059]
w'' = [-9.20289162 132.03619302 68.3104778]

vii. Write down the final weights you obtain in your report. How does these weights compare to the "optimal" weights [w0, w1, w2]?

The optimal and final weight vectors are nearly collinear but differ in length, indicating close alignment but different magnitudes.

(l) Comment on how the changes in η effect the number of epochs needed until convergence

A smaller η increases the number of epochs needed for convergence due to gradual weight adjustments, potentially leading to precise convergence. In contrast, a larger η decreases the epoch number but may cause fluctuations and potential overshooting, possibly extending convergence time.

(m) Comment on whether we would get the exact same results (in terms of the effects of η on training performance) if we had started with different w0, w1, w2, S, w′0 , w′1 , w′2 .

Altering the initial values [w0, w1, w2, S, w′0, w′1, w′2] could result in varied training performances at different η values. These changes influence the learning trajectory, potentially affecting the convergence speed and solution quality, thereby causing inconsistency in results across different initializations.

(n) Do the same experiments with n = 1000 samples. Comment on the differences compared to n = 100.

Using n = 1000 samples instead of n = 100 tends to yield more stable and accurate results by better representing the underlying distribution and possibly reducing the influence of noise or outliers found in the smaller dataset. However, it might necessitate greater computational resources and time.