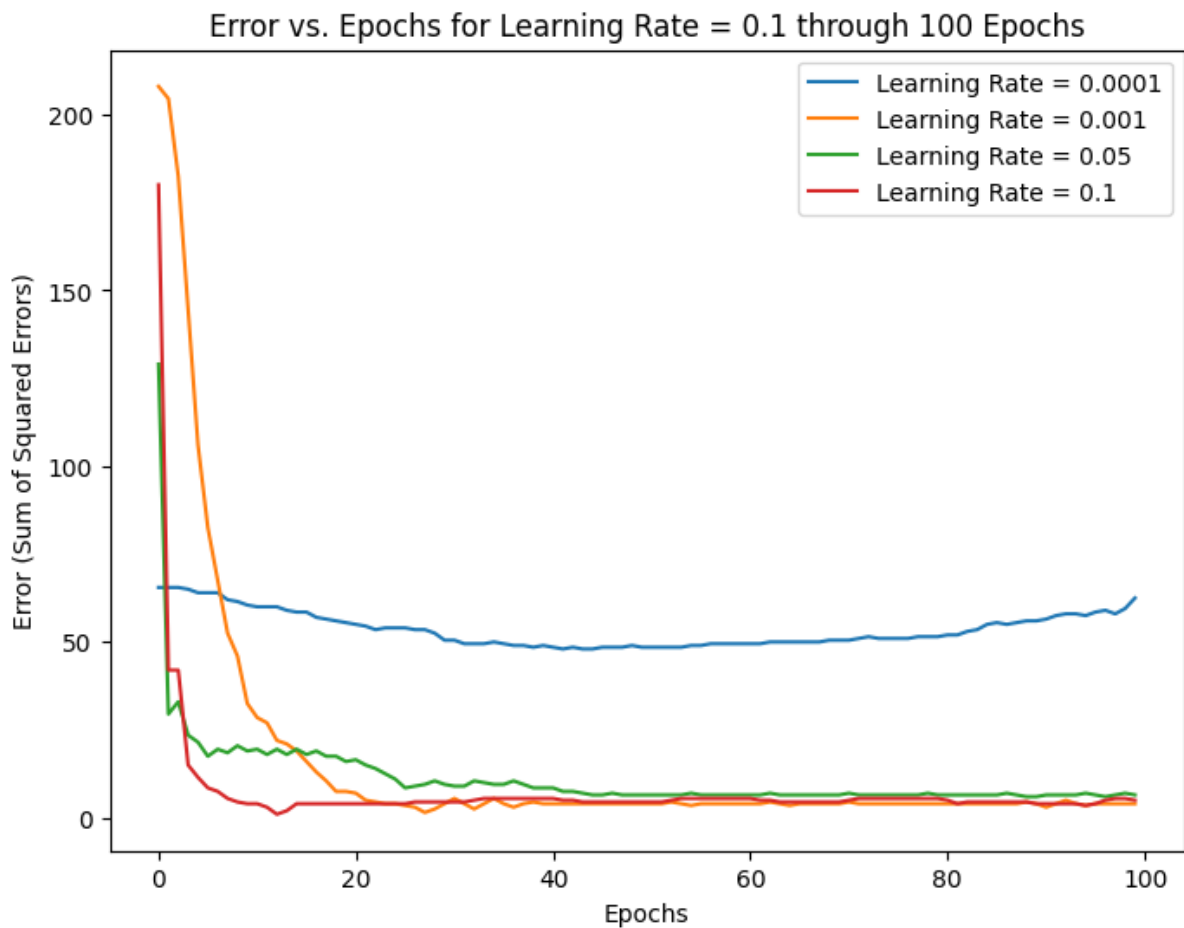


Machine Learning HW4

Du Nguyen and Thanh Nguyen have equal contributions to the assignment.

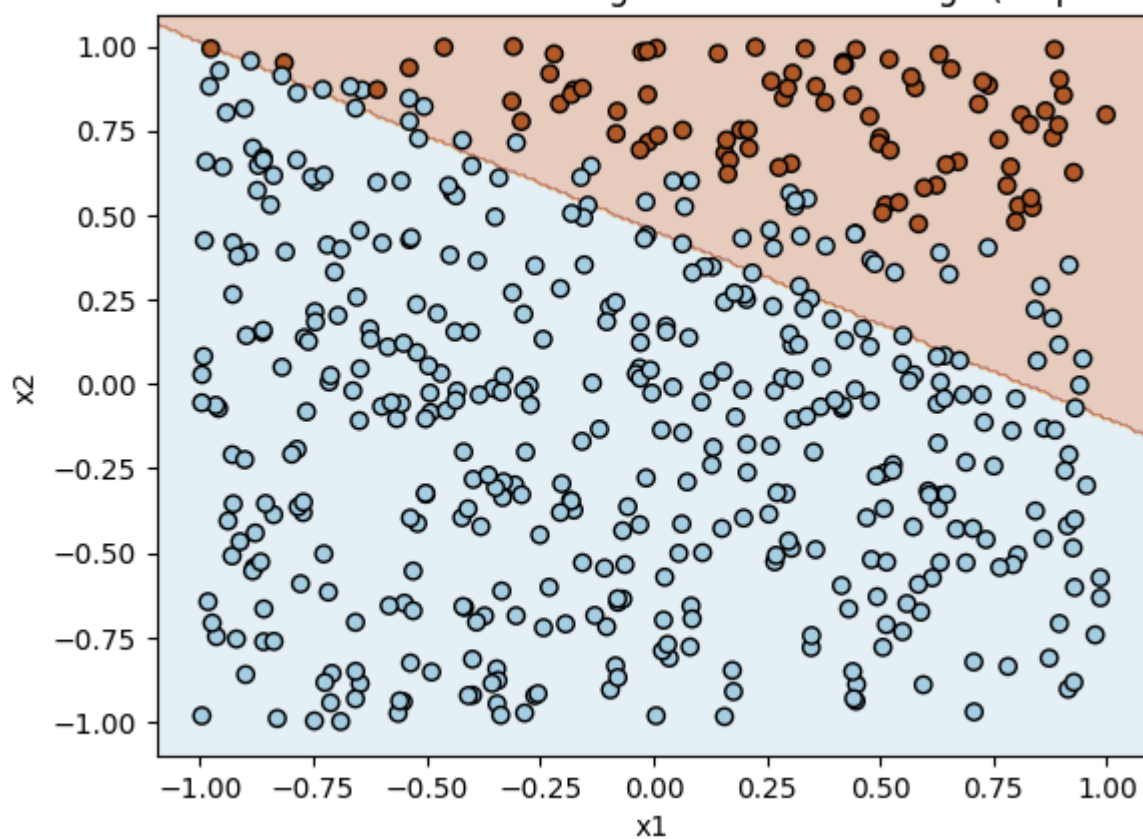
Problem 1:

a)

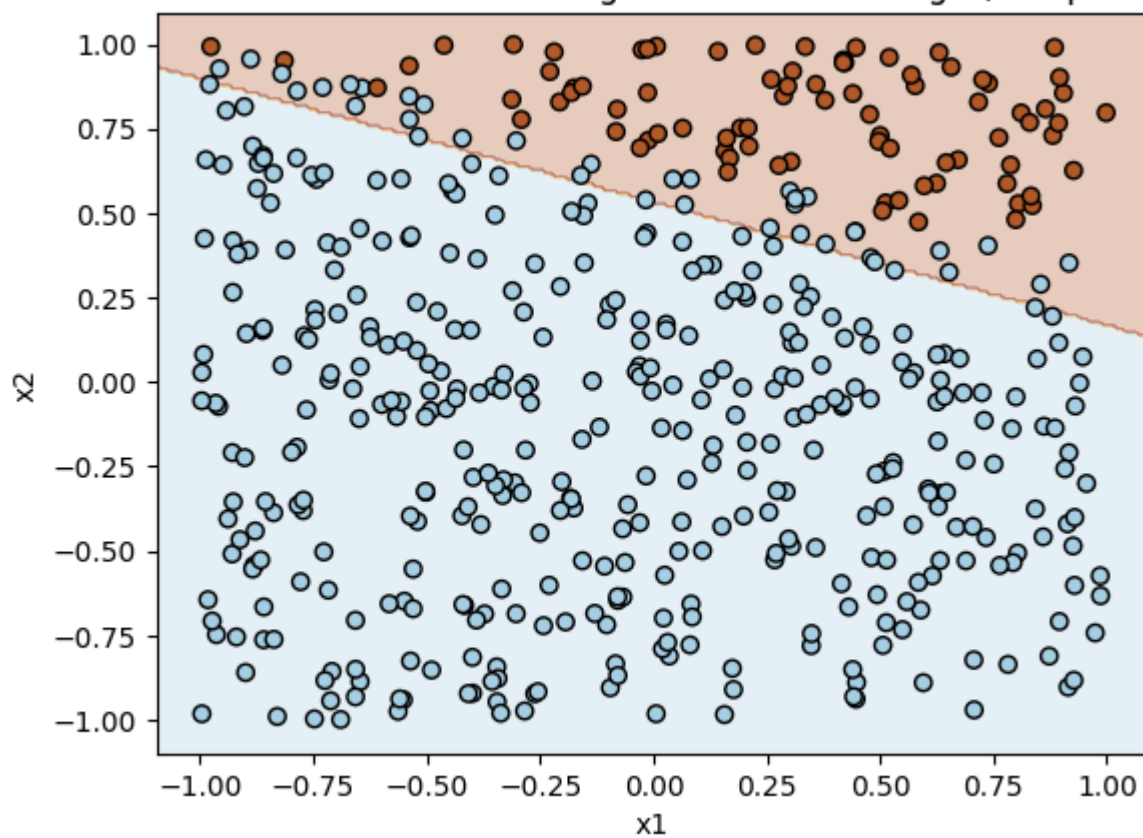


b).

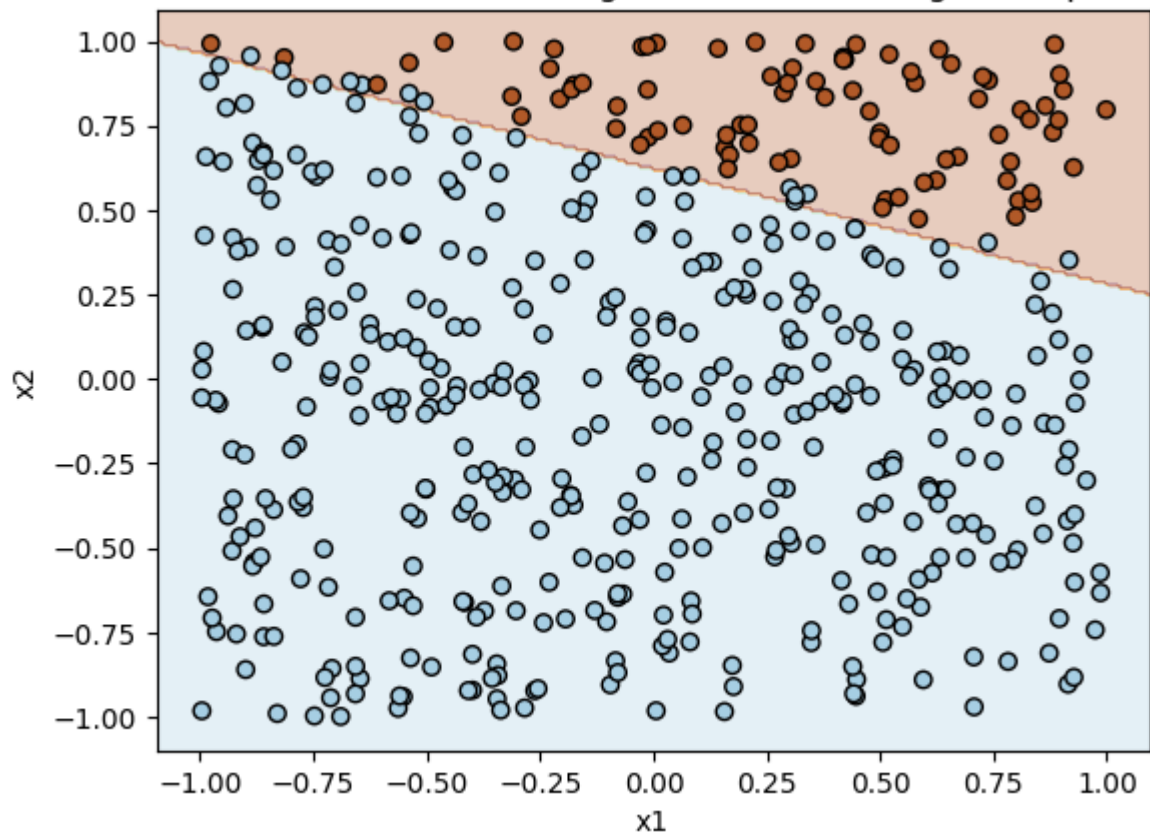
Decision Surface for Learning Rate = 0.01 through (5 Epochs)



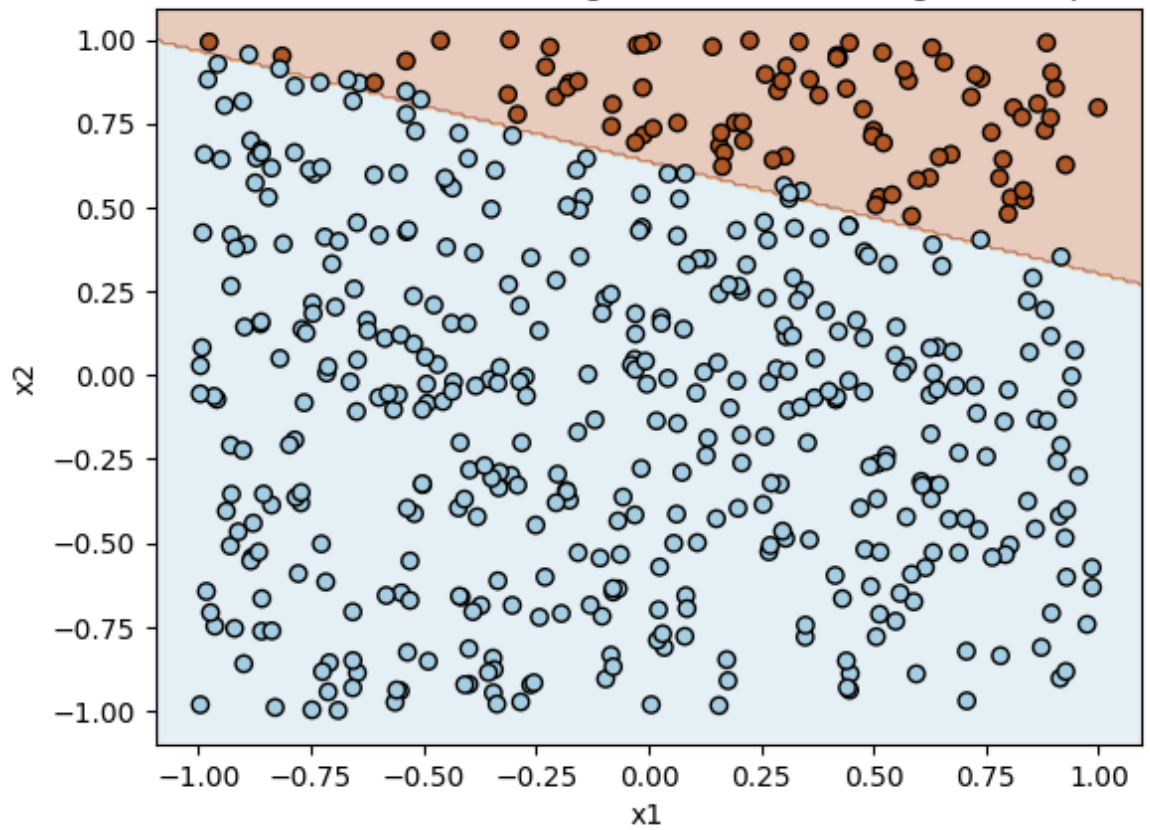
Decision Surface for Learning Rate = 0.01 through (10 Epochs)



Decision Surface for Learning Rate = 0.01 through (50 Epochs)



Decision Surface for Learning Rate = 0.01 through (100 Epochs)

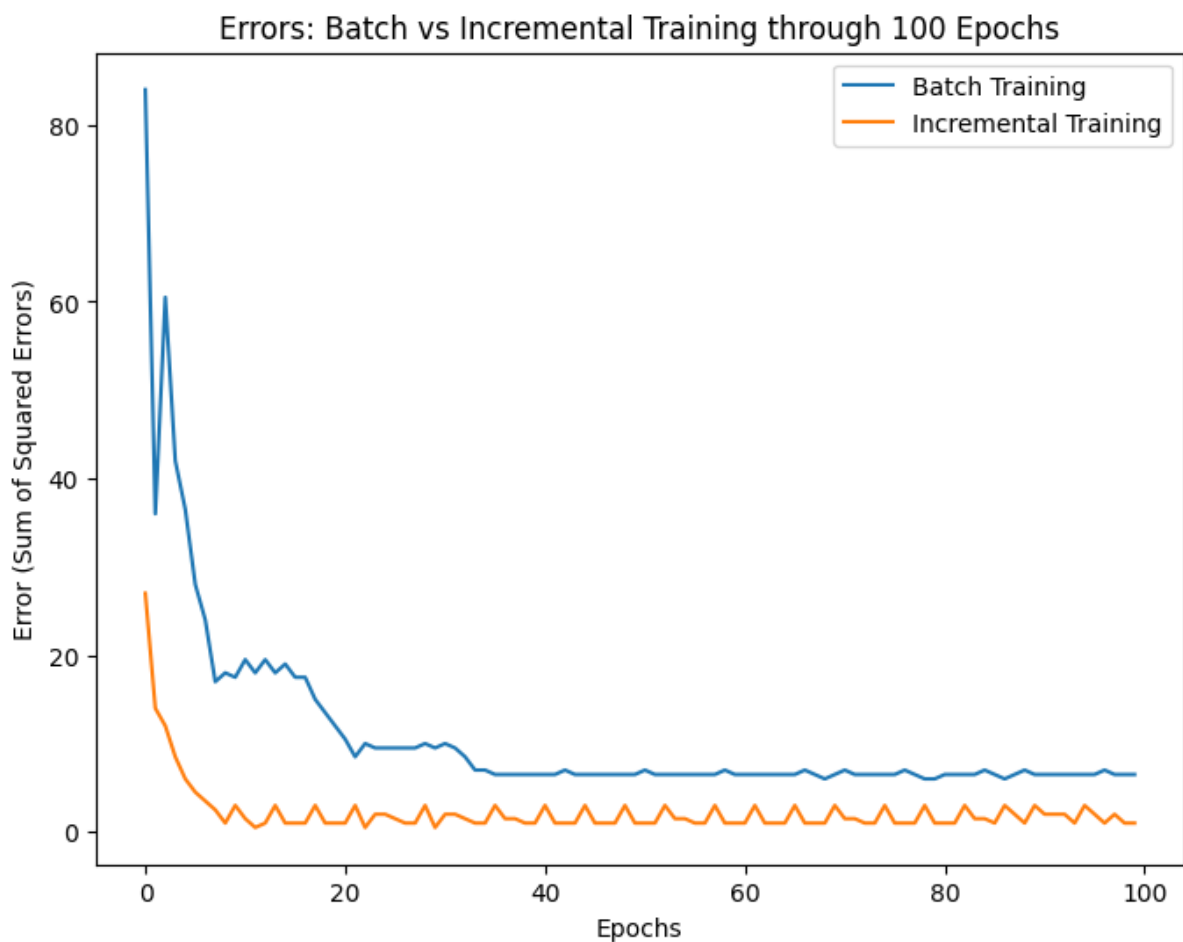


c)

```
Analysis of learning rates:  
Learning Rate: 0.0001, Final Error: 55.5  
Learning Rate: 0.001, Final Error: 0.0  
Learning Rate: 0.05, Final Error: 4.0  
Learning Rate: 0.1, Final Error: 4.0
```

As we can see, the learning rate **0.001 work the best** since it achieved the perfect convergence while the other fails to do so. Rates below 0.001 make updates too small to converge, or it learn too slow and not effective, which results in higher loss (55.5 for 0.0001). And rates above 0.001 causing weight oscillations (shown in 1a), thus it gets stuck in the suboptimal region (4.0)

d)



Batch Training Time: 0.003195 seconds

Batch Training Weight Updates: 100

Incremental Training Time: 0.413930 seconds

Incremental Training Weight Updates: 50000

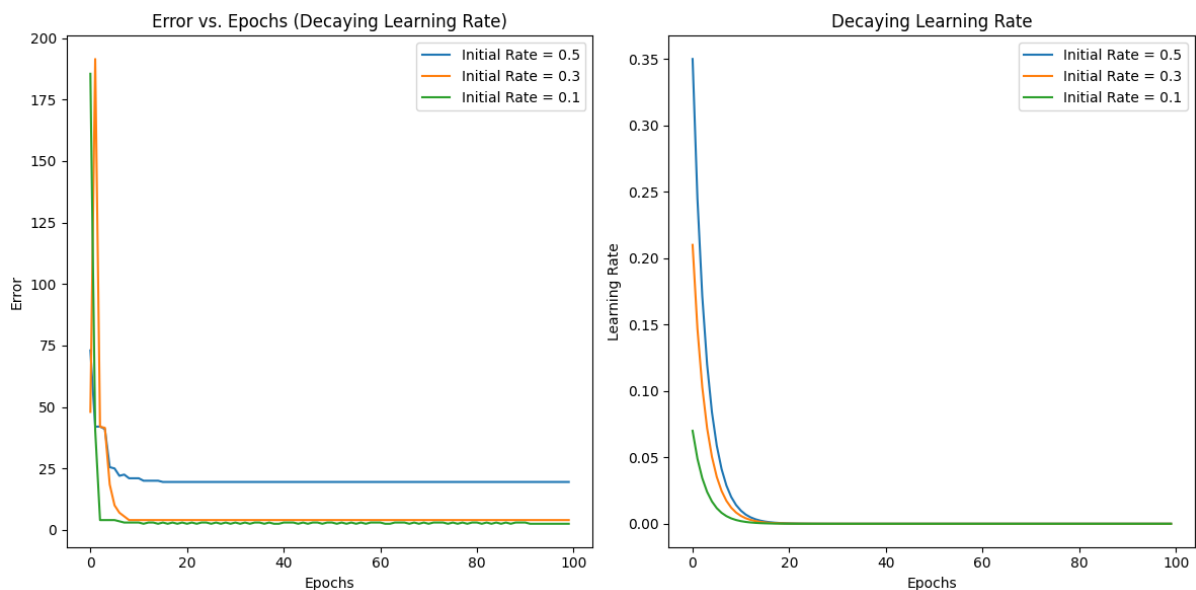
Analyzing:

The Batch Training time is significantly faster than that in Incremental training time (about 140 times faster) while the batch training weights updates is only 100 updates, whose incremental training updates 50000, that is 500 times greater.

This gives us an insight about pros of each method. Batch training is effective for small to medium dataset and fast computation while with larger dataset, incremental training is a more fit choice, and also if we want the model to adapt continuously.

Problem 2:

a)



Discussion:

The plots shown that using a decaying learning rate help the perceptron learning fast as first and then stabilize as the rate decreases. Higher initial rate such as 0.5, 0.3 lead to faster early learning and also have oscillation as a side effect, while smaller initial learning rate produces a more stable convergence with lower final error.

b)

Errors with Adaptive Learning Rate through 100 Epochs

