**Group #1**

**Balaji Arcot Kulasekaran**

**Harsh Kantilal Sharma**

**Rudrashis Poddar**

**Rabie Ahmed M Alsaihati**

**University of Cincinnati**

**College of Applied Science and Engineering**

**Machine learning**

**Professor Anca Ralescu**

**Homework four**

# Problem 1

## Accuracy output and training error VS number of iterations (25)

A picture containing chart

Description automatically generated

Figure 1 shows the training error (y axis) vs number of iterations (x axis)

1. Decision surface graphs

1) 5 iterations

Chart

Description automatically generated

Figure 2 shows the decision surface for 5 iterations

2) 10 iterations

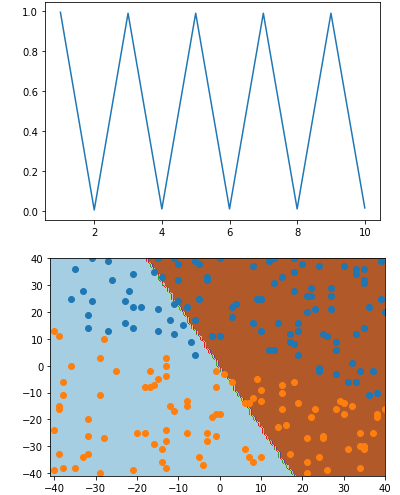


Figure 3 shows the decision surface for 10 iterations

At 50 iterations

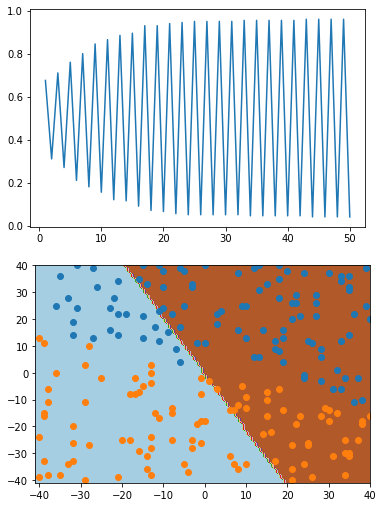


Figure 4 shows the decision surface for 50 iterations

At 100 iterations

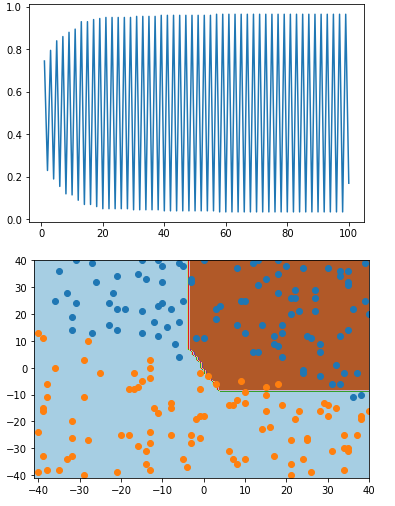


Figure 5 shows the decision surface for 100 iterations

## Different learning rates (at 50 epochs)

1. 0.1

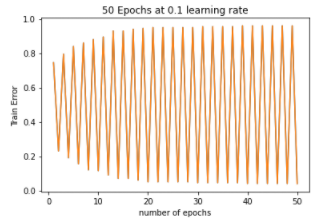


Figure 6 shows train error against 50 iterations at 0.1 lr

1. 0.01

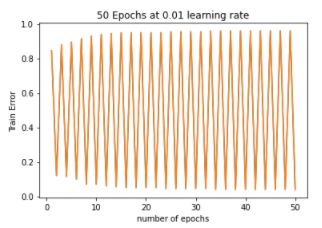


Figure 7 shows train error against 50 iterations at 0.01 lr

1. 0.001

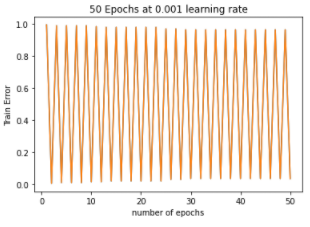


Figure 8 shows train error against 50 iterations at 0.001 lr

1. 0.0001

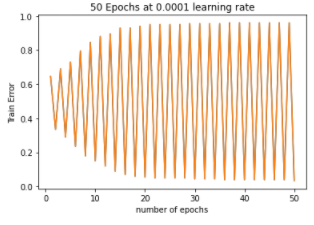


Figure 9 shows train error against 50 iterations at 0.0001 lr

## d) For stochastic gradient descent where the weight is updated for every training instance in X, we have chosen 0.01 for learning rate.

SGD number of iterations VS Epochs (50)

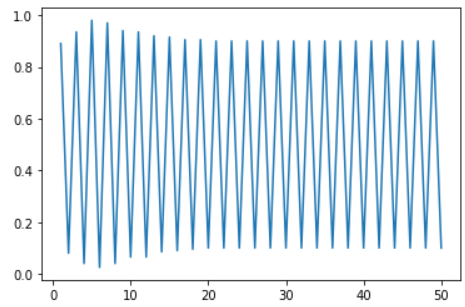


Figure 10 shows train error(X-axis) against 50 iterations(yaxis) at 0.01 lr using SGD

We can see that the median error for this method is much less than gradient descent even though the output is still fluctuating. Therefore, we can say that Stochastic GD is more efficient than GD.

# Problem 2

## a) Here we have chosen the random number between 0 and 1 to be 0.8. For nth iteration the learning rate becomes eta \*(0.8)n

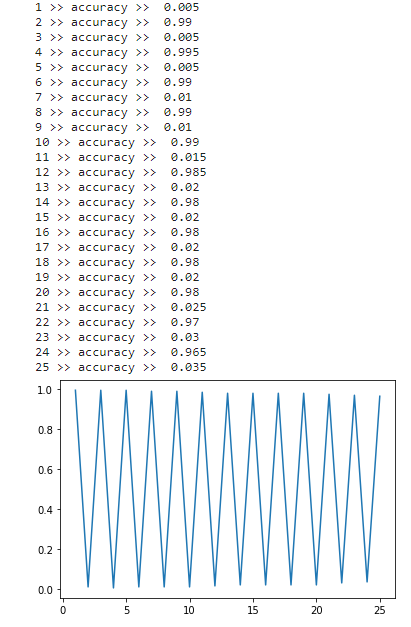


Figure 11

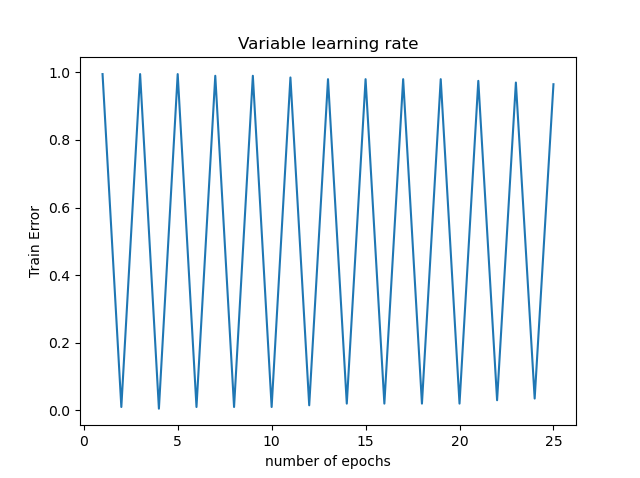


Figure 12 shows the same graph as above

Though there were great fluctuation depending on if the nth epoch is odd or even for all learning rates, we can see that the median error for learning rate=0.01 was the least. That is why we have chosen learning rate=0.01 If we compare figure 6 with figure 1 above it could be inferred that the decaying rates approach is better than the constant rates.

2B) Here we have taken eta=0.5, t=0.03, d=0.9, and D=1.02 like given in the example and the graph is as follows.

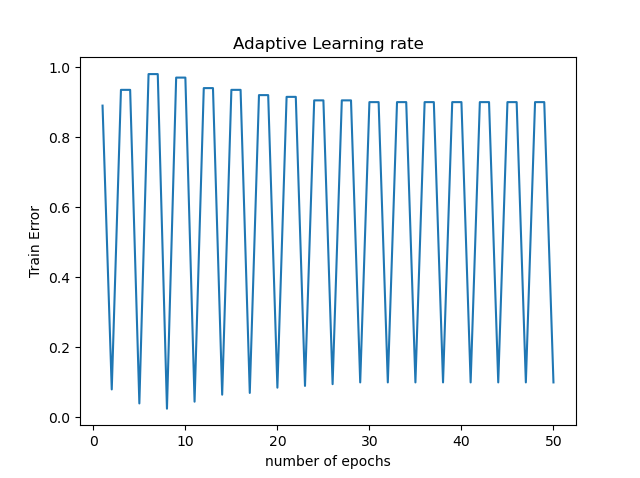


Figure 13 shows the same graph but with adaptive learning procedure implemented

We can see that the graph has flattened even more for adaptive learning rate than for decaying learning rate and has even lower error median than decaying learning rate

We can tell from the graph that the adaptive learning rate has more efficiency than decaying or constant learning rate as the value of the median error v epoch is the least for adaptive.