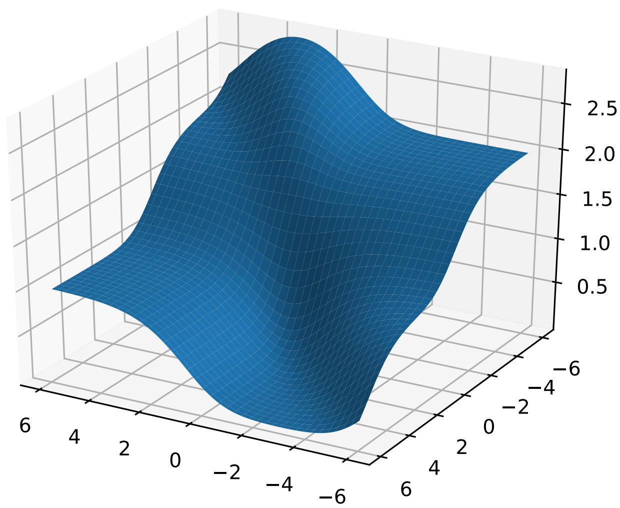
I

1

Figure 1: Plot of the loss-function



w is minimized at $w_1 = 5.8$ and $w_2 = -2.9$ at this point $L_{simple} = 0.00498$.

2

To see the calculation see images 7 and 8 at the end of the document.

The resulting gradient was

$$\nabla_w L_{simple}(w) = \left[\frac{\partial}{\partial w_1} L_{simple}, \frac{\partial}{\partial w_2} L_{simple}\right]$$

$$\frac{\partial}{\partial w_1} L_{simple} = 2(\sigma(w, [1,0] - 1)e^{-w_1} \cdot \sigma^2 w, [1,0] + 2(\sigma(w, [1,1]) - 1)e^{-w_1-w_2} \cdot \sigma^2(w, [1,1])$$

$$\frac{\partial}{\partial w_2} L_{simple} = 2(\sigma(w, [0,1])e^{-w_2} \cdot \sigma^{(}2w, [0,1]) + 2(\sigma(w, [1,1]) - 1)e^{-w_1-w_2} \cdot \sigma^2(w, [1,1])$$

As no stop condition was suggested or implemented I let the algorithm run for 1000 steps.

For values $\eta \leq 0.1$ the algorithm will converge to L_{simple} close to zero. However it is very slow. The larger one chooses η, the lower L_{simple} will become, but for values over 100 no real change in L_{simple} is seen anymore.

Figure 2: Gradient Descent $\eta = 0.1$

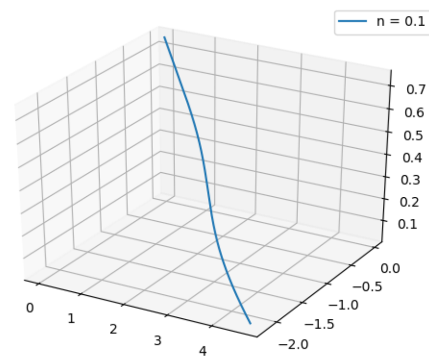


Figure 3: Gradient Descent $\eta = 1$

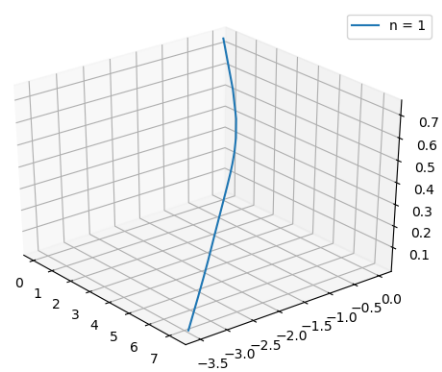


Figure 4: Gradient Descent $\eta = 10$

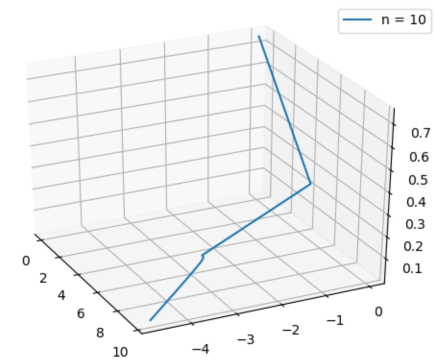
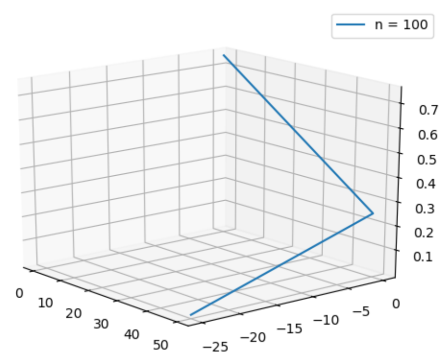


Figure 5: Gradient Descent $\eta = 100$



II

2

The gradient was a follows:

3

Gradient

The gradient was calculated to:

$$\frac{\partial L_n}{\partial w_i} = (\sigma(w, x_n) - y_n) \cdot x_i \cdot e^{-w^T x} \cdot \sigma(w, x_n)^2$$

Stochastic training

For any number of iterations, the speed is almost instant, however the accuracy is depending highly on a large number of iterations. This is so for both the big and small seperable sets.

For the non separable set, the error, regardless of iteration size stops at around 0.17. Increasing the iteration size beyond this does nothing, because the set is non separable and some errors cannot be avoided.

Batch training

Either, something is wrong with my implementation, or this is clearly the worse algorithm at least when it comes to computation time. Even for the separable set, with only 10 iterations, the execution time was about 4 seconds. The algorithm then achieved an error-rate of 0.5, which is crap.

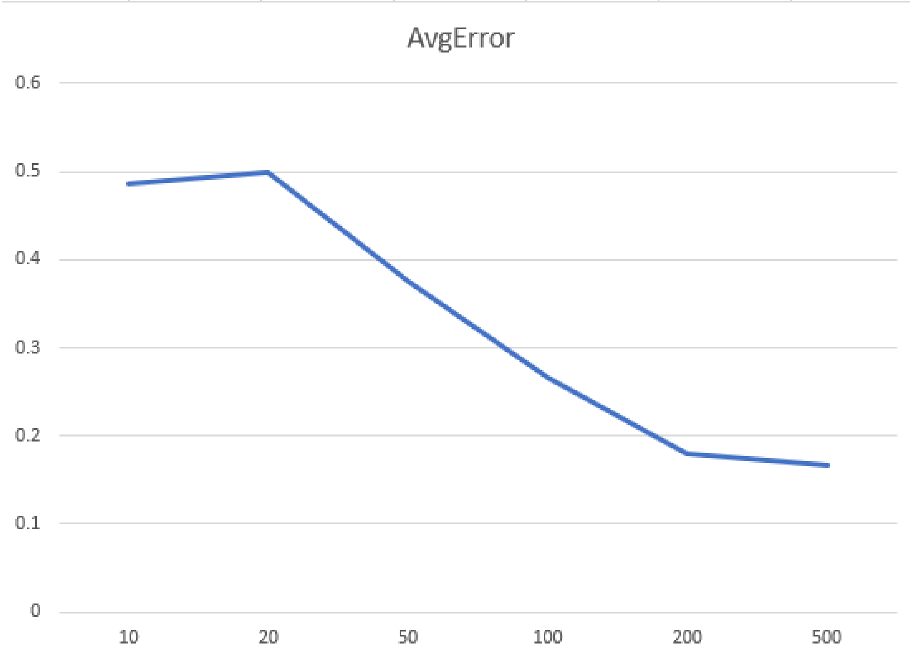
Running for a 100 iterations takes around 45 seconds, and yields an error-rate of 0.458, which is still pretty bad.

Running the non separable set for 100 iterations yields an error of 0.435.

I don't have time to run for longer.

Error vs iterations

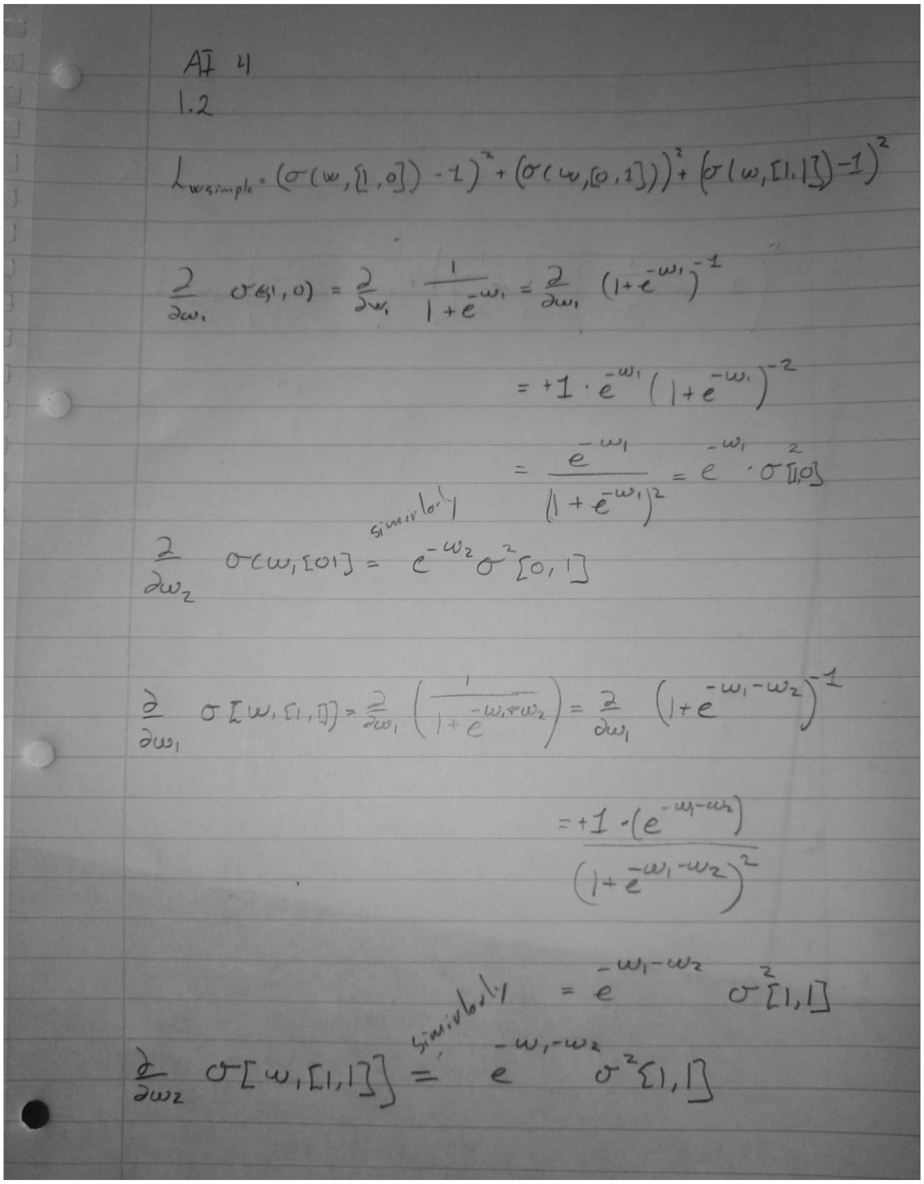
Figure 6: Comparison of avg error and iterations for stochastic



As seen in figure 6 the avg error falls slowly with more iterations. As more data is added the more close to the actual extremities of the two sets the perceptron will lie and thus make fewer errors as it ”knows” more about the two sets and their edges.

The execution time is pretty much constant at around 0.3 seconds. This shows that the execution time of the algorithm itself is very low compared to other python related overhead.

Figure 7: Calculation $\nabla_w L$



AI 4

1.2

$$L_{w\,simple} = (\sigma(w,[1,0]) - 1)^2 + (\sigma(w,[0,1]))^2 + (\sigma(w,[1,1]) - 1)^2$$

$$\frac{\partial}{\partial w_1}\sigma([1,0]) = \frac{\partial}{\partial w_1}\frac{1}{1+e^{-w_1}} = \frac{\partial}{\partial w_1}(1+e^{-w_1})^{-1}$$

$$= +1 \cdot e^{-w_1}(1+e^{-w_1})^{-2}$$

$$= \frac{e^{-w_1}}{(1+e^{-w_1})^2} = e^{-w_1} \cdot \sigma^2[1,0]$$

similarly

$$\frac{\partial}{\partial w_2}\sigma(w,[0\,1]) = e^{-w_2}\sigma^2[0,1]$$

$$\frac{\partial}{\partial w_1}\sigma[w,[1,1]] = \frac{\partial}{\partial w_1}\left(\frac{1}{1+e^{-w_1-w_2}}\right) = \frac{\partial}{\partial w_1}(1+e^{-w_1-w_2})^{-1}$$

$$= \frac{+1 \cdot (e^{-w_1-w_2})}{(1+e^{-w_1-w_2})^2}$$

$$= e^{-w_1-w_2}\,\sigma^2[1,1]$$

similarly

$$\frac{\partial}{\partial w_2}\sigma[w,[1,1]] = e^{-w_1-w_2}\,\sigma^2[1,1]$$

Figure 8: Calculation $\nabla_w L$

$$\frac{\partial L_{simple}(w)}{\partial w_1} = 2(\sigma[1,0]-1)\cdot\frac{\partial}{\partial w}\sigma[1,0] + 0$$

$$+ 2(\sigma[1,1]-1)\cdot\frac{\partial}{\partial w_1}\sigma[1,1]$$

$$= 2(\sigma[1,0]-1)\, e^{-w_1}\,\sigma^2[1,0]$$

$$+ 2(\sigma[1,1]-1)\, e^{-w_1-w_2}\,\sigma^2[1,1]$$

$$\frac{\partial L_{simple}(w)}{\partial w_2} = 0 + 2\sigma[0,1]\cdot\frac{\partial}{\partial w_2}\sigma[0,1] +$$

$$2(\sigma[1,1]-1)\cdot\frac{\partial}{\partial w_2}\sigma[1,1]$$

$$= 2\sigma[0,1]\, e^{-w_2}\,\sigma^2[0,1] +$$

$$2(\sigma[1,1]-1)\cdot e^{-w_1-w_2}\,\sigma^2[1,1]$$

$$\nabla_w L_{simple}(w) = \left[\frac{\partial L_{simple}}{\partial w_1}, \frac{\partial L_{simple}}{\partial w_2}\right]$$