Mathew Varughese

CS1699 – Deep Learning

**1.**

*Calculations done via Google Sheets -* [*Link*](https://docs.google.com/spreadsheets/d/1dNRrzIWn67y9aBVLq4GO1qpwQcuoKeKAMl_jtFPNrFU/edit?usp=sharing)

x = [1, 1, 1]

y = [0, 0]

w(1)10 = w(1)20 = w(1)11 = w(1)21 = w(1)12 = w(1)22 = w(2)10 = w(2)20 = w(2)11 = w(2)21 = w(2)12 = w(2)22 = 0.05

**Activations**

z0 = 1

z1 = σ(Σj w(1)1j x) = σ (0.05\*1 + 0.05 \*1 + 0.05 \* 1) = 0.53743

z2 = σ(Σj w(1)2j x) = σ (0.05\*1 + 0.05 \*1 + 0.05 \* 1) = 0.53743

y1 = σ(Σk w(2)1k z) = σ (0.05\*1 + 0.05 \*0.53743 + 0.05 \* 0.53743) = 0.52591

y2 = σ(Σk w(2)2k z) = σ (0.05\*1 + 0.05 \*0.53743 + 0.05 \* 0.53743) = 0.52591

**Errors**

δy1 = y1 \* (1 - y1) \* (y1 – ytrue) = 0.5259 \* (1-.5259) \* (0.525 -0) = 0.13112

δy2 = y2 \* (1 – y2) \* (y2 – ytrue) = 0.5259 \* (1-.5259) \* (0.5259 -0) = 0.13112

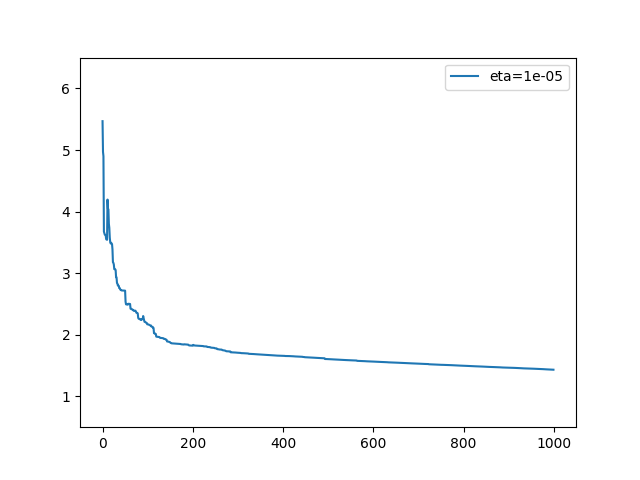
δz1 = z1 \* (1 – z1) \* Σk δkwk1 = (0.53743) \* (1-0.53743) \* (0.13112 \* 0.05 + 0.13112 \* 0.05) = 0.00326

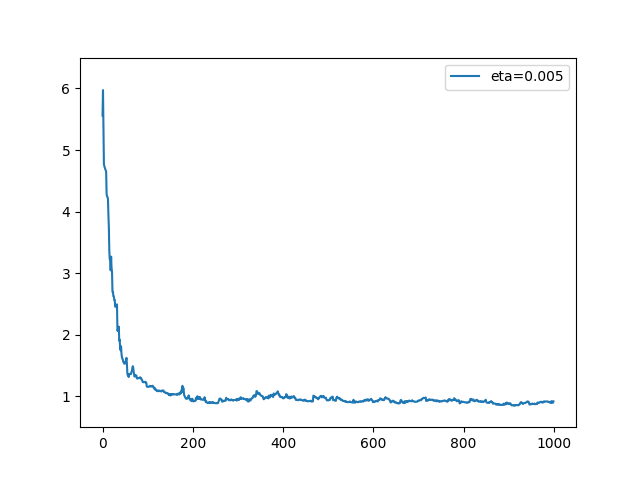
δz2 = z2 \* (1 – z2) \* Σk δkwk2 = (0.53743) \* (1-0.53743) \* (0.13112 \* 0.05 + 0.13112 \* 0.05) = 0.00326

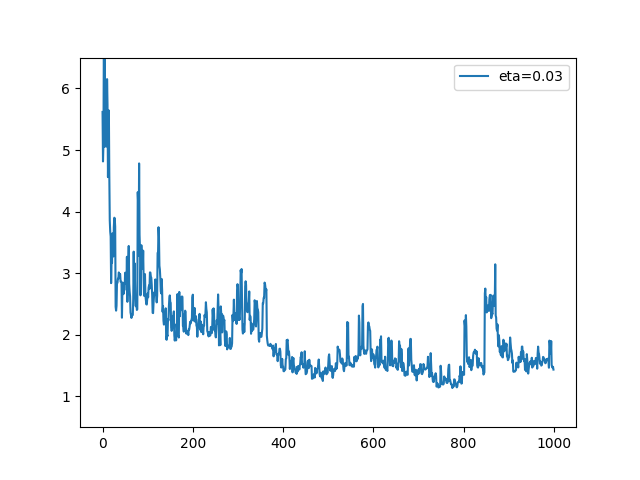
**Update Weights**

* w(2)10= w(2)10 – 0.3 \* δy1 \* z0 = 0.05 – 0.3 \* 0.13112 \* 1 = 0.0106625
* w(2)20= w(2)20 – 0.3 \* δy2 \* z0 = 0.05 – 0.3 \* 0.13112 \* 1 = 0.0106625
* w(2)11= w(2)11 – 0.3 \* δy1 \* z1 = 0.05 – 0.3 \* 0.13112 \* 0.5374 = 0.0288608
* w(2)21= w(2)21 – 0.3 \* δy2 \* z1 = 0.05 – 0.3 \* 0.13112 \* 0.5374 = 0.0288608
* w(2)12= w(2)12 – 0.3 \* δy1 \* z2 = 0.05 – 0.3 \* 0.13112 \* 0.5374 = 0.0288608
* w(2)22= w(2)22 – 0.3 \* δy2 \* z2 = 0.05 – 0.3 \* 0.13112 \* 0.5374 = 0.0288608
* w(1)10= w(1)10 – 0.3 \* δz1 \* x0 = 0.05 – 0.3 \* 0.004889445 \* 1 = 0.048533
* w(1)20= w(1)20 – 0.3 \* δz2 \* x0 = 0.05 – 0.3 \* 0.004889445 \* 1 = 0.048533
* w(1)11= w(1)11 – 0.3 \* δz1 \* x1 = 0.05 – 0.3 \* 0.004889445 \* 1 = 0.048533
* w(1)21= w(1)21 – 0.3 \* δz2 \* x1 = 0.05 – 0.3 \* 0.004889445 \* 1 = 0.048533
* w(1)12= w(1)12 – 0.3 \* δz1 \* x2 = 0.05 - 0.3 \* 0.004889445 \* 1 = 0.048533
* w(1)22= w(1)22 – 0.3 \* δz2 \* x2 = 0.05 – 0.3 \* 0.004889445 \* 1 = 0.048533

**RMS Error on Validation Set** = 0.9510375399011316 (Learning Rate was 0.005)

Weights initialized randomly with mean of 0 and stddev of 0.15





The plots of the learning rate line up with what I would expect. A learning rate of 0.03 is a little high, which causes fluctuations. Conceptually, this might be because it overshoots the minimum. A small learning rate of 0.00005 makes a very smooth curve. This makes sense, since only tiny incremental changes are made to the weights. A learning rate of 0.005 seems ideal, because it has less fluctuations, learns somewhat fast, but still is able to find a smaller error.