**NUMERICAL OF LSTM IN DEEP LEARNING…**

# The Example

Let us begin by defining out internal weights:

Wa=[0.45,0.25], Ua=[0.15], Ba=[0.2]

Wi=[0.95,0.8] , Ui=[0.8] , Bi=[0.65]

Wf=[0.7,0.45] , Uf=[0.1] , Bf=[0.15]

Wo=[0.6,0.4] , Uo=[0.25] , Bo=[0.1]

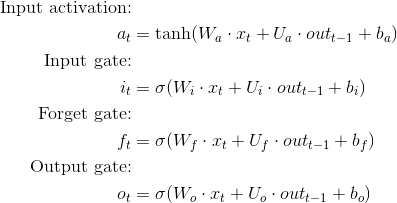
And now input data:

X0=[1,2] with label 0.5

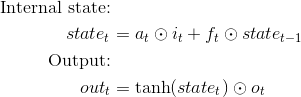
X1=[0.5,3] with label 1.25

**Forward components:**

The gates are defined as/ Formulas:



Which leads to:



**Apply forward for time step 0:**

* a0= tanh(Wa.X0+Ua.t-1+Ba)

putt values in activation 0:

tanh([0.45 o.25] [1 2] + [0.15][0] + [0.2]) = 0.81775

* i0= sigmoid(Wi . X0 +Ui . t-1 +Bi)

putt values in input 0:

sigmoid([0.95 0.8] [1 2] + [0.8] [0] +[0.65]) = 0.96083

* f0 = sigmoid(Wf . X0 + Uf . t-1 +Bf)

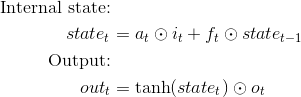
now, putt values in forget cell:

sigmoid([0.7 0.45] [1 2] + [0.1] [0] + [0.15]) = 0.85195

* O0 = sigmoid(W0 . x0 + U0 . t-1 +B0)

Putt values in output cell

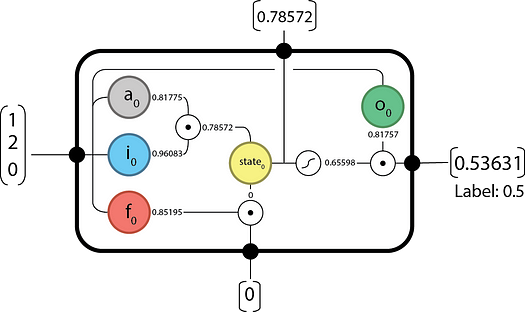
Sig([0.6 0.4] [1 2] +[0.25] [0] +[0.1])= 0.81757



Putt values in these states:

State0= 0.81775 \* 0.96083 + 0.8195 \* 0 = 0.78572

Output0= tanh(0.78572) \* 0.81757 = 0.53631



**Apply forward for time step 1:**

* a1= tanh(Wa.X1+Ua.t-1+Ba)

putt values in activation 1:

tanh([0.45 o.25] [0.5 3] + [0.15][0.53631] + [0.2]) = 0.84980

* i1= sigmoid(Wi . X1 +Ui . t-1 +Bi)

putt values in input 1:

sigmoid([0.95 0.8] [0.5 3] + [0.8] [0.53631] +[0.65]) = 0.98118

* f1 = sigmoid(Wf . X1 + Uf . t-1 +Bf)

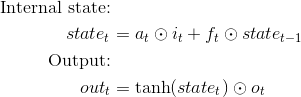
now, putt values in forget cell:

sigmoid([0.7 0.45] [0.5 3] + [0.1] [0.53631] + [0.15]) = 0.87030

* O1 = sigmoid(W0 . x1 + Uo . t-1 +B0)

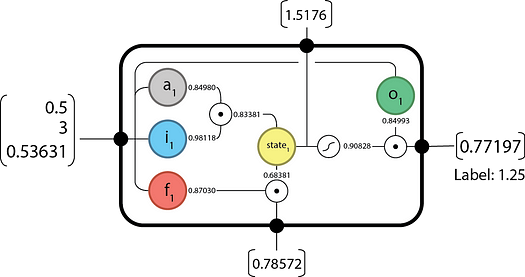
Putt values in output cell

Sig([0.6 0.4] [0.5 3] +[0.25] [0.53631] +[0.1])= 0.84993



State1 = 0.84980 \* 0.98118 + 0.87030 \* 0.78572 = 1.5176

Output1 = tanh (1.5176) \* 0.84993 = 0.77197



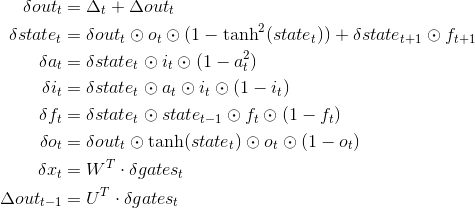
**Backward components:**

The gates are defined as/ Formulas:

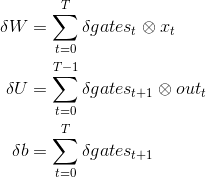
* ΔT the output difference as computed by any subsequent layers (i.e. the rest of your network), and;
* Δout the output difference as computed by the next time-step LSTM (the equation for t-1 is below).

Find:

Delta 1 = output of forward – label



The final updates to the internal parameters is computed as:



**Apply backward for time step 1:**

Delta 1 = output of forward 1 – label(0) = 0.77197 – 1.25 = -0.47803

**Delta out1 = 0 because there are no future time step**

Out1 = -0.47803 + 0 = -0.47803

State 1 = -0.47803 \* 0.84993 \* (1- tanh^2(1.5176) + 0\*0 = -0.07111

a1= -0.07111 \* 0.98118 \* (1- 0.84980^2) = -0.01938

i1= -0.07111 \* 0.84980 \* 0.98118 \* (1- 0.98118) = -0.00112

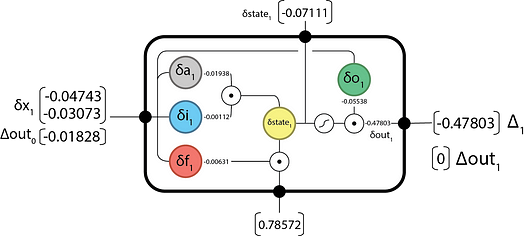
f1= -0.07111 \* 0.78572 \* 0.87030 \* (1- 0.87030) = -0.00631

o1= 09,47803 \* tanh(1.5176) \* 0.84993 \* (1- 0.84993) = -0.05538

X1= Wt . gates1= [0.45 0.95 0.70 0.60]

delta x 1 =W^ T \* delta gates\_{1} =( [ 0.45 0.95 0.70 0.60\\ 0.25 0.80 0.45 0.40 ]) \* ([ -0.0138\\ [-0.00112]\\[-0.00631]\\ [-0.05538]] = [[- 0.04743]\\ [- 0.03073]]

Souto = [0.15\\ 0.8 \\ 0.1 \\ 0.25] \* [[- 0.01938]\\ [- 0.00112]\\ [- 0.000631]\\ [- 0.05538]] = - 0.01828



**Apply backward for time step 0:**

Delta 0 = output of forward 0 – label (0) = 0.53631 – 0.5 = 0.03631

**Out0 = -0.01828 passed back from T= 1**

Out0 = 0.03631 + - 0..1828 = 0.01803

State 0 = 0.01803 \* 0.81757 \* (1- tanh^2(0.78572) + -0.07111\*0.87030 = -0.05349

a0= -0.05349 \* 0.96083 \* (1- 0.81775^2) = -0.01703

i0= -0.05349 \* 0.81775 \* 0.96083 \* (1- 0.96083) = -0.00165

f0= -0.05349 \* 0 \* 0.85195 \* (1- 0.85195) = 0

o0= 0.01803 \* tanh(0.78572) \* 0.81757 \* (1- 0.81757) = 0.00176

delta x 0 =W^ T \* delta gates\_{0}

[0.45 0.95 0.70 0.60 \\ 0.25 0.80 0.45 0.40]\*[-0.01703\\ -0.00165 \\ 0\\ -0.00176]= [0.00817 \\ -0.00487]

Delta out-1 = UT. delta gates\_1

= = [0.15 0.80 0.10 0.25] \* [-0.01703\\ -0.00165\\ 0\\ 0.00176] = -0.00343

