Problem 1 Assignment 2

$$V_t = \beta_1 V_{t-1} + (1-\beta_1)\frac{\partial J}{\partial W}$$

$$S_t = \beta_2 S_{t-1} + (1-\beta_2)\left(\frac{\partial J}{\partial W}\right)^2$$

$$V_{corr} = \frac{V_t}{1-\beta_1^t} \quad W_t = W_{t-1} - \alpha \frac{1}{V_{corr} \sqrt{S_{corr} + \epsilon}}$$

$$S_{corr} = \frac{S_t}{1-\beta_2^t}$$

$\epsilon = 10^{-8}$
$\beta_1 = 0.9$
$\beta_2 = 0.999$
$\alpha = 0.1$

$$W_1 = \begin{bmatrix} -0.22 & 0.33 & 0.67 \\ 0.49 & -0.86 & -0.90 \end{bmatrix}$$

$$\frac{\partial J}{\partial W} = \begin{bmatrix} 0.78 & 0.02 & 0.04 \\ -0.88 & -0.14 & 1.01 \end{bmatrix}$$

$$V_1 = \begin{bmatrix} -0.27 & 0.27 & -0.07 \\ -0.12 & -0.02 & 0.12 \end{bmatrix}$$

$$S_1 = \begin{bmatrix} 0.01 & 0.07 & 0.11 \\ 0.14 & 0.17 & 0.02 \end{bmatrix}$$

a. $V_2 = \beta_1 V_1 + (1-\beta_1)\frac{\partial J}{\partial W}$

$$V_2 = 0.9\begin{bmatrix} 0.01 & 0.07 & 0.11 \\ 0.14 & 0.17 & 0.02 \end{bmatrix} + (1-0.9)\begin{bmatrix} 0.78 & 0.02 & 0.04 \\ -0.88 & -0.14 & 1.01 \end{bmatrix}$$

$$V_2 = \begin{bmatrix} 0.686 & 0.043 & 0.030 \\ -0.812 & -0.129 & 0.930 \end{bmatrix}$$

b. $S_2 = B_2 S_1 + (1 - B_2)\left(\frac{\partial J}{\partial W}\right)^2$

$$S_2 = 0.999\begin{bmatrix} 0.01 & 0.07 & 0.11 \\ 0.14 & 0.17 & 0.02 \end{bmatrix} + (1 - 0.999)\begin{bmatrix} 0.78 & 0.02 & 0.04 \\ -0.88 & -0.14 & 1.01 \end{bmatrix}^2$$

$$S_2 = \begin{bmatrix} 0.011 & 0.699 & 0.110 \\ 0.141 & 0.170 & 0.021 \end{bmatrix}$$

c. $W_2 = W_1 - \alpha \dfrac{1}{V_{corr} + \sqrt{S_{corr}} + \epsilon}$

$$V_{corr} = \frac{V_2}{1 - B_1^2} = \begin{bmatrix} 0.753 & 0.047 & 0.331 \\ -0.892 & -0.142 & 1.022 \end{bmatrix}$$

$$S_{corr} = \frac{S_2}{1 - B_2^2} = \begin{bmatrix} 10.598 & 69.930 & 109.892 \\ 140.634 & 169.850 & 21.000 \end{bmatrix}$$

$$W_2 = \begin{bmatrix} -0.22 & 0.33 & 0.67 \\ 0.49 & -0.86 & -0.90 \end{bmatrix} - 0.1 \frac{1}{\begin{bmatrix} .753 & .047 & .331 \\ -.892 & -.142 & 1.022 \end{bmatrix} + \sqrt{\begin{bmatrix} 10.598 & 69.93 & 109.892 \\ 140.634 & 169.85 & 21 \end{bmatrix}} + 10^{-8}}$$

$$W_2 = \begin{bmatrix} -0.261 & 0.074 & 0.382 \\ 0.499 & -0.806 & -0.921 \end{bmatrix}$$

# Problem 2

a. $\phi(z) = \sigma(z) - 0.5$

$\sigma(x) = \frac{1}{1+e^{-x}}$

$z_t$ input of activation function

$\sigma'$ for derivative of $\sigma$

$\nabla h_t = f(\nabla h_{t+1})$

$h_m = f(W_h h_{t-1} + W_x x_m)$

$$\underbrace{\frac{\partial h_m}{\partial h_{m-1}}}_{\downarrow} = W_h^T \operatorname{diag}(f'(W_h h_{m-1} + W_x x_m))$$

$\nabla h_t = W_h^T \operatorname{diag}(\sigma'(z_{t-1}))$

single $h_{t-1}$

$$\boxed{\nabla h_t = \prod_{m=t+1}^{j} W_h^T \operatorname{diag}(\sigma'(z_{t-1}))}$$

$\sigma'(0) = \frac{1}{4}$

Scanned with CamScanner

# Problem 2

6. $h\nabla_t = \prod_{m=t+1}^{j} W_h^T \left(\frac{1}{4}\right)$

$$0 \approx \left[W_h^T\left(\tfrac{1}{4}\right)\right]\left[W_h^T\left(\tfrac{1}{4}\right)\right]\left[W_h^T\left(\tfrac{1}{4}\right)\right]\ldots$$

$$(0.0001)^{\frac{1}{100}} = \left[\left[W\left(\tfrac{1}{4}\right)\right]^{\wedge}100\right]^{\frac{1}{100}}$$

$$4 \cdot 1 = W\tfrac{1}{4} \cdot 4$$

$$4 = W \qquad \alpha = 4$$

if $W < 4$ then vanishing gradient

if $W > 4$ then exploding gradient

# Problem 3

a. if $x_t$ = 0 vector

then $h_t = h_{t-1}$

Answer: False

Because: the bias terms of $b_f$ & $b_i$ can still change the output

b. if $f_t \approx 0$

then: the back propigation will not reach earlier time steps

answer: False

Reason: an output of 0 can still carry an error of the intended output

c. $f_t$, $i_t$, & $o_t$ are non negative

Answer: False

Reason: $f_t$, $i_t$, and $o_t$ are a probability distribution whose sum must equal 1

Problem 3

a. False

- Even if $x_t$ vector is 0 the value $h_t$ can change from $b_f$ bias term

b. False

Although a $f_t$ near 0 results in $f_t \odot c_{t-1}$ to be near 0 the summation of $f_t \odot c_{t-1} + i_t \odot c_t$ results in backpropigation to earlier time steps

c. True

$f_t$, $i_t$, and $o_t$ cannot be negative because they are all activated by the sigmoid function to have values between 0 and 1

d. False

Although $f_t$, $i_t$, and $o_t$ cannot be negative and are between 0 and 1 the summation of their values can be greater then 1

# Problem 4

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f)$$

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i)$$

$$\tilde{c}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o)$$

$$h_t = o_t \odot \tanh(c_t)$$

$$\sigma(x) = \frac{1}{1+e^{-x}} \qquad \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

$$\underset{1\times 2}{W_f = [1 \ \ 2]}, \ U_f = [0.5], \ b_f = [0.2]$$

$$W_i = [-1 \ \ 0], \ U_i = [2], \ b_i = [-0.1]$$

$$W_c = [1 \ \ 2], \ U_c = [1.5], \ b_c = [0.5]$$

$$W_o = [3 \ \ 0], \ U_o = [-1], \ b_o = [0.8]$$

$$x_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}^{2\times 1} \quad y_1 = 0.5$$

$$x_2 = \begin{bmatrix} 0.5 \\ -1 \end{bmatrix} \quad y_2 = 0.8$$

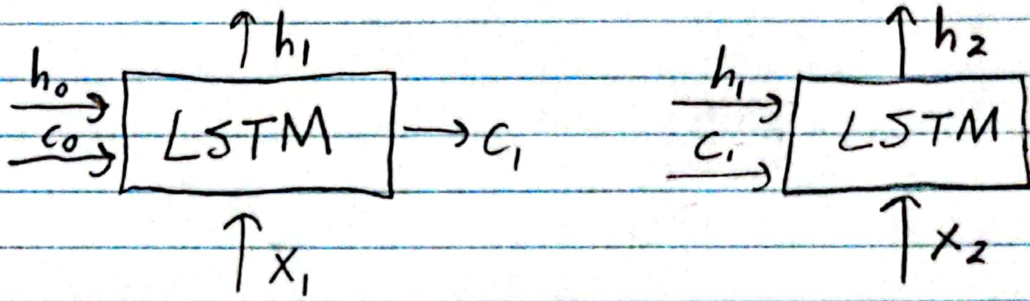
$$h_0 = [0] \qquad c_0 = [0]$$

# Problem 4

a. $f_t, i_t, o_t, h_t$ have dimension of 1.

b. $h_0 = 0$ vector    Calculate $h_1, h_2$

$c_0 = 0$ vector




$f_t = \sigma\left([1\ 2]\begin{bmatrix}1\\0\end{bmatrix} + [0.5][0] + [0.2]\right)$

$f_{t_1} = \sigma(1.2)$

$f_{t_1} = 0.231$

$i_{t_1} = \sigma\left([-1\ 0]\begin{bmatrix}1\\0\end{bmatrix} + [2][0] + [-0.1]\right)$

$i_{t_1} = \sigma(-1.1)$

$i_{t_1} = 0.750$

$o_{t_1} = \sigma\left([3\ 0]\begin{bmatrix}1\\0\end{bmatrix} + [-1][0] + [0.8]\right)$

$o_{t_1} = \sigma(3.8)$

$o_{t_1} = 0.022$

# Problem 4

$$\tilde{c}_{t_1} = \tanh(W_c x_t + U_c h_{t-1} + b_c)$$

$$\tilde{c}_{t_1} = \tanh(1.5)$$

$$\tilde{c}_{t_1} = 0.905$$

$$c_{t_1} = 0.231(0) + 0.750(0.022)$$

$$c_1 = 0.679$$

$$h_{t_1} = o_{t_1} \odot \tanh(c_{t_1})$$

$$h_{t_1} = 0.022 \odot \tanh(0.679)$$

$$h_1 = 0.013$$

$$f_{t_2} = \sigma\left(\begin{bmatrix} 1 & 2 \end{bmatrix}\begin{bmatrix} 0.5 \\ -1 \end{bmatrix} + \begin{bmatrix} 0.5 \end{bmatrix}\begin{bmatrix} 0.013 \end{bmatrix} + \begin{bmatrix} 0.2 \end{bmatrix}\right)$$

$$f_{t_2} = \sigma(-1.294)$$

$$f_{t_2} = 0.785$$

$$i_{t_2} = \sigma\left(\begin{bmatrix} -1 & 0 \end{bmatrix}\begin{bmatrix} 0.5 \\ -1 \end{bmatrix} + \begin{bmatrix} 2 \end{bmatrix}\begin{bmatrix} 0.013 \end{bmatrix} + \begin{bmatrix} -0.1 \end{bmatrix}\right)$$

$$i_{t_2} = \sigma(-0.574)$$

$$i_{t_2} = 0.640$$

$$o_{t_2} = \sigma\left(\begin{bmatrix} 3 & 0 \end{bmatrix}\begin{bmatrix} 0.5 \\ -1 \end{bmatrix} + \begin{bmatrix} -1 \end{bmatrix}\begin{bmatrix} 0.013 \end{bmatrix} + \begin{bmatrix} 0.8 \end{bmatrix}\right)$$

$$o_{t_2} = \sigma(2.287)$$

$$o_{t_2} = 0.092$$

# Problem 4

$\tilde{c}_{t_2} = \tanh(W_c x_t + U_c h_t + b_c)$

$\tilde{c}_{t_2} = \tanh\left([1\ 2]\begin{bmatrix}0.5\\-1\end{bmatrix} + [1.5][0.013] + [0.5]\right)$

$\tilde{c}_{t_2} = \tanh(-0.981)$

$\tilde{c}_{t_2} = -0.753$

$c_{t_2} = f_t \odot c_t + i_t \odot \tilde{c}_t$

$c_{t_2} = 0.785(0.679) + 0.640(0.092)$

$c_{t_2} = 0.051$

$h_{t_2} = o_{t_2} \tanh(c_{t_2})$

$h_{t_2} = 0.092 \tanh(0.051)$

$h_{t_2} = 0.005$

$h_1 = 0.013 \quad h_2 = 0.005$

Scanned with CamScanner

# Problem 4

c. $y_{1t} = 0.5$ $\qquad$ $y_{2t} = 0.8$

$y_{1g} = 0.013$ $\qquad$ $y_{2g} = 0.005$

$$MSE = \frac{1}{2}\left[(0.013 - 0.5)^2 + (0.005 - 0.9)^2\right]$$

$$MSE = 0.435$$