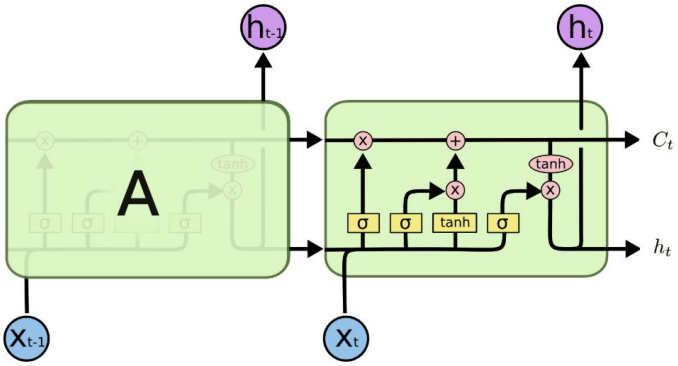
In Class Activity – LSTM (ICA 20) - Solutions

Please solve the following problems and enter your responses at https://tinyurl.com/AIF19-ICA20

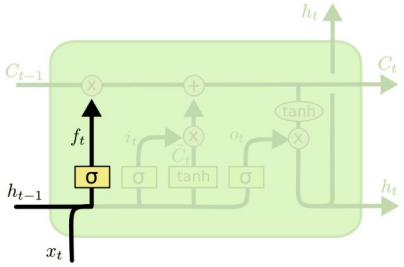
$$W = \begin{bmatrix} w_{i1} & w_{i2} & b_i \\ w_{c1} & w_{c2} & b_c \\ w_{f1} & w_{f2} & b_f \\ w_{o1} & w_{o2} & b_o \\ w_y & 0 & b_y \end{bmatrix} = \begin{bmatrix} 0.5 & 0.25 & 0.01 \\ 0.3 & 0.4 & 0.05 \\ 0.03 & 0.06 & 0.002 \\ 0.02 & 0.04 & 0.001 \\ 0.6 & 0 & 0.025 \end{bmatrix}$$



Consider the LSTM cell above. The first column of weight matrix represents weights from input X_t to input gate (i), memory cell (c), forget gate (f), output gate (o), and output node (y). The second column represents weight from hidden state h_{t-1} to these same entities.

The resulting hidden state after processing cell A is $h_{t-1} = 0.08$, and $X_t = 0.2$, and previous cell state $C_{t-1} = 0.0798$. Using the above information, complete the following computations of the LSTM cell (Sigmoid and tanh values are given below):

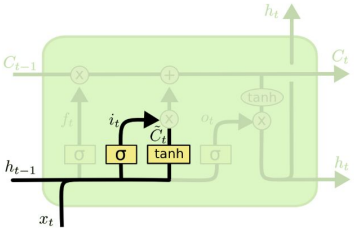
1) Forget Gate:



$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \; + \; b_f\right)$$

$f_t = \sigma(W_{f1} * X_t \; + W_{f2} * h_{t-1} + b_f) = \sigma(0.03 * 0.2 + 0.06 * 0.08 + 0.002) = \; \sigma(0.0128) = 0.503$

2) Input Gate:

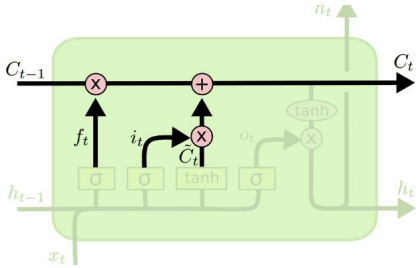


$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \; + \; b_i\right)$$
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \; + \; b_C)$$

$i_t = \sigma(W_{i1} * X_t \; + W_{i2} * h_{t-1} + b_i) = \; \sigma(0.5 * 0.2 + 0.25 * 0.08 + 0.01) = \sigma(0.13) = 0.532$

$\tilde{C}_t = \tanh(W_{C1} * X_t \; + W_{C2} * h_{t-1} + b_C) = \tanh(0.3 * 0.2 + 0.4*0.08 + 0.05) = \tanh(0.142) = 0.141$

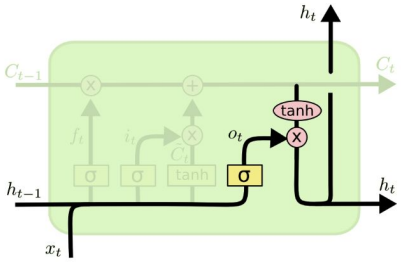
3) Memory Update:



$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t = 0.503 * 0.0798 + 0.532 * 0.141 = 0.115$

4) Output Gate:



$$o_t = \sigma\left(W_o\left[h_{t-1}, x_t\right] + b_o\right)$$
$$h_t = o_t * \tanh\left(C_t\right)$$

$o_t = \sigma(W_{O1} * X_t + W_{O2} * h_{t-1} + b_O) = \sigma(0.02 * 0.2 + 0.04 * 0.08 + 0.001) = \sigma(0.0082) = 0.502$

$h_t = o_t * \tanh(C_t) = 0.502 * \tanh(0.115) = 0.502 * 0.114 = 0.057$

5) Final Output and Error

Now if the final hidden state h_t is connected to an output node (y) with no activation using the weight W_y and bias b_y, what is the final output y of the network? If the target output is 0.08, what is the error?

$y = W_y * h_t + b_y = 0.6*0.057 + 0.025 = 0.0592$

Error = target - y = 0.08 - 0.0592 = 0.02

Use the closest sigmoid and tanh values for your computations from this table:

x:	0	0.008	0.01	0.02	0.13	0.25	0.4
σ(x):	0.5	0.502	0.503	0.505	0.532	0.562	0.598
y:	0.02	0.115	0.142	0.21	0.301	0.45	0.6
tanh(y):	0.01t	0.114	0.141	0.206	0.292	0.421	0.537