ECE 417 Multimedia Signal Processing
Solutions to Homework 6

UNIVERSITY OF ILLINOIS
Department of Electrical and Computer Engineering

Assigned: Tuesday, 11/7/2023; Due: Tuesday, 11/28/2023

Problem 6.1

Suppose you have a recurrent neural network with input $x[n]$, target $y[n] \in \{0, 1\}$, output $h[n]$, and loss function:

$$\mathcal{L} = -\frac{1}{N}\sum_{n=0}^{N-1}\left(y[n]\ln h[n] + (1-y[n])\ln(1-h[n])\right)$$

where

$$\begin{aligned} h[n] &= \sigma(\xi[n]), \\ \xi[n] &= x[n] + \sum_{m=1}^{M-1} w[m]h[n-m], \end{aligned}$$

and where $\sigma(\cdot)$ is the logistic sigmoid. Write $d\mathcal{L}/dw[3]$ in terms of the signals $y[n]$ and $h[m]$. You may assume that $h[n] = 0$ for $n < 0$.

Solution: First step: forward-prop. Assume that $h[n]$ has been calculated. Second step: partial derivatives:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial h[n]} &= \frac{1}{N}\left(\frac{y[n]}{h[n]} - \frac{1-y[n]}{1-h[n]}\right) \\ &= \begin{cases} \frac{1}{N}\frac{1}{h[n]} & y[n] = 1 \\ \frac{1}{N}\frac{1}{h[n]-1} & y[n] = 0 \end{cases} \\ &= \frac{1}{N}\frac{h[n]-y[n]}{h[n](h[n]-1)} \end{aligned}$$

Third step: total derivatives:

$$\begin{aligned} \frac{d\mathcal{L}}{dh[n]} &= \frac{\partial \mathcal{L}}{\partial h[n]} + \sum_{m=1}^{M-1}\frac{d\mathcal{L}}{dh[n+m]}\frac{\partial h[n+m]}{\partial h[n]} \\ &= \frac{\partial \mathcal{L}}{\partial h[n]} + \sum_{m=1}^{M-1}\frac{d\mathcal{L}}{dh[n+m]}\dot{\sigma}\left(\xi[n+m]\right)w[m] \\ &= \frac{1}{N}\frac{h[n]-y[n]}{h[n](h[n]-1)} + \sum_{m=1}^{M-1}\frac{d\mathcal{L}}{dh[n+m]}h[n+m](1-h[n+m])w[m] \end{aligned}$$

Fourth step: weight gradient:

$$\begin{aligned}
\frac{d\mathcal{L}}{dw[3]} &= \sum_{n=0}^{N-1} \frac{d\mathcal{L}}{dh[n]} \frac{\partial h[n]}{\partial w[3]} \\
&= \frac{1}{N} \frac{h[n]-y[n]}{h[n](h[n]-1)} \dot{\sigma}(\xi[n]) h[n-3] + \sum_{m=1}^{M-1} \frac{d\mathcal{L}}{dh[n+m]} h[n+m](1-h[n+m]) w[m] \dot{\sigma}(\xi[n]) h[n-3] \\
&= \frac{1}{N}(h[n]-y[n]) h[n-3] + \sum_{m=1}^{M-1} \frac{d\mathcal{L}}{dh[n+m]} h[n+m](1-h[n+m]) w[m] h[n](1-h[n]) h[n-3]
\end{aligned}$$

Problem 6.2

Suppose that

$$\begin{aligned}
h_0 &= x^3 \\
h_1 &= \cos(x) + \sin(h_0) \\
\hat{y} &= \frac{1}{2}\left(h_1^2 + h_0^2\right)
\end{aligned}$$

What is $d\hat{y}/dx$? Express your answer as a function of x only, without the variables h_0 or h_1 in your answer.

Solution:

$$\begin{aligned}
\frac{dh_0}{dx} &= 3x^2 \\
\frac{dh_1}{dx} &= \frac{\partial h_1}{dx} + \frac{dh_0}{dx} \frac{\partial h_1}{\partial h_0} \\
&= -\sin(x) + 3x^2 \cos(x^3) \\
\frac{d\hat{y}}{dx} &= \frac{dh_0}{dx} \frac{\partial \hat{y}}{\partial h_0} + \frac{dh_1}{dx} \frac{\partial \hat{y}}{\partial h_1} \\
&= h_0\left(3x^2\right) + h_1\left(-\sin(x) + 3x^2 \cos(x^3)\right) \\
&= 3x^5 + \left(\cos(x) + \sin(x^3)\right)\left(-\sin(x) + 3x^2 \cos(x^3)\right)
\end{aligned}$$

Problem 6.3

Consider a one-gate recurrent neural net, defined as follows:

$$\begin{aligned}
c[n] &= c[n-1] + w_c x[n] + u_c h[n-1] + b_c \\
h[n] &= o[n] c[n] \\
o[n] &= \sigma(w_o x[n] + u_o h[n-1] + b_o)
\end{aligned}$$

where $\sigma(\cdot)$ is the logistic sigmoid, $x[n]$ is the network input, $c[n]$ is the cell, $o[n]$ is the output gate, and and $h[n]$ is the output. Suppose that you've already completed synchronous back-prop, which has given you the following quantity:

$$\epsilon[n] = \frac{\partial \mathcal{L}}{\partial h[n]}$$

Find asynchronous back-prop formulas, i.e., find formulas for the following quantities, in terms of one another, and/or in terms of the other quantities defined above:

$$\delta_h[n] = \frac{d\mathcal{L}}{dh[n]}$$
$$\delta_o[n] = \frac{d\mathcal{L}}{do[n]}$$
$$\delta_c[n] = \frac{d\mathcal{L}}{dc[n]}$$

Solution:

$$\begin{aligned}
\delta_h[n] &= \epsilon[n] + \delta_c[n+1]u_c + \delta_o[n+1]\dot{\sigma}(w_o x[n+1] + u_o h[n] + b_1)u_o \\
&= \epsilon[n] + \delta_c[n+1]u_c + \delta_o[n+1]o[n+1](1 - o[n+1])u_o \\
\delta_o[n] &= \delta_h[n]c[n] \\
\delta_c[n] &= \delta_c[n+1] + \delta_h[n+1]o[n]
\end{aligned}$$

Problem 6.4

Remember that in lecture we defined a super-simplified nonlinearity called the clipped ReLU or CReLU, which is:

$$\text{CReLU}(x) = \max(0, \min(1, x))$$

Using the CReLU nonlinearity for both σ_h and σ_g in an LSTM, choose weights and biases,

$$\{b_c, u_c, w_c, b_f, u_f, w_f, b_i, u_i, w_i, b_o, u_o, w_o\},$$

that will cause an LSTM to count the number of nonzero inputs, and output the tally only when the input is zero:

$$h[n] = \begin{cases} \sum_{m=0}^{n} \mathbf{1}[x[m] \geq 1] & x[n] = 0 \\ 0 & x[n] \geq 1 \end{cases}$$

where $\mathbf{1}[\cdot]$ is the unit indicator function, and you may assume that $x[n]$ is always a non-negative integer.

Solution: First, we only want to generate output when $x[n] = 0$, so

$$o[n] = \text{CReLU}(x[n]), \quad \Rightarrow \quad b_o = 1, w_o = -1, u_o = 0$$

Second, we want $c[n]$ to increase by exactly 1, every time $x[n] \geq 1$, and otherwise to keep its previous value, i.e.,

$$c[n] = c[n-1] + \text{CReLU}(x[n]),$$

but we know that $c[n]$ is defined to be

$$c[n] = f[n]c[n-1] + i[n]\text{CReLU}\left(w_c x[n] + u_c h[n-1] + b_c\right)$$

so we want $f[n] = 1$ always, and $i[n] = 1$ whenever $x[n] \geq 0$, so

$$b_f = 1, w_f \geq 0, u_f \geq 0, (w_i \geq 1 \text{ or } b_i \geq 1), u_i \geq 0, w_c = 1, u_c = 0, b_c = 0$$

Putting it all together, we have

$$\{b_c = 0, u_c = 0, w_c = 1, b_f = 1, u_f \geq 0, w_f \geq 0, (b_i \geq 1 \text{ or } w_i \geq 1), u_i \geq 0, b_o = 1, u_o = 0, w_o = 1\}$$