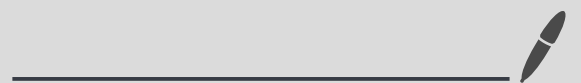


Assignment - 1

(Bonus)



(2)(a)

$$\begin{aligned}i^{(t)} &= \sigma(\omega_{i_x} x^{(t)} + \omega_{i_h} h^{(t-1)}) \\f^{(t)} &= \sigma(\omega_{f_x} x^{(t)} + \omega_{f_h} h^{(t-1)}) \\o^{(t)} &= \sigma(\omega_{o_x} x^{(t)} + \omega_{o_h} h^{(t-1)}) \\g^{(t)} &= \tanh(\omega_{g_x} x^{(t)} + \omega_{g_h} h^{(t-1)}) \\c^{(t)} &= f^{(t)} c^{(t-1)} + i^{(t)} g^{(t)} \\h^{(t)} &= o^{(t)} \tanh(c^{(t)})\end{aligned}$$

let the gradient pass down by the LSTM cell be $\bar{h}^{(t)}$

where $\bar{h}^{(t)} = \frac{\partial L}{\partial h^t}$

if we are using MSE, then $\bar{h}^{(t)} = y - h^{(t)}$
where $h^{(t)}$ is the predicted value & y is the original value.

\Rightarrow gradient w.r.t output gate

$$\begin{aligned}\bar{o}^{(t)} &= \frac{\partial L}{\partial o^{(t)}} = \frac{\partial L}{\partial h^{(t)}} \times \frac{\partial h^{(t)}}{\partial o^{(t)}} \\&= \bar{h}^{(t)} \frac{\partial}{\partial o^{(t)}} (o^{(t)} \tanh(c^{(t)})) \\&= \bar{h}^{(t)} (\tanh c^{(t)})\end{aligned}$$

\Rightarrow gradient w.r.t $c^{(t)}$

$$\begin{aligned}\bar{c}^{(t)} &= \frac{\partial L}{\partial c^{(t)}} = \frac{\partial L}{\partial h^{(t)}} \times \frac{\partial h^{(t)}}{\partial c^{(t)}} \\&= \bar{h}^{(t)} \frac{\partial}{\partial c^{(t)}} (o^{(t)} \tanh(c^{(t)})) \\&= \bar{h}^{(t)} o^{(t)} (1 - \tanh^2 c^{(t)})\end{aligned}$$

\Rightarrow gradient w.r.t input gate $i^{(t)}$ & $g^{(t)}$

$$\begin{aligned}\bar{i}^{(t)} &= \frac{\partial L}{\partial i^{(t)}} = \frac{\partial L}{\partial c^{(t)}} \times \frac{\partial c^{(t)}}{\partial i^{(t)}} \\&= \bar{h}^{(t)} o^{(t)} (1 - \tanh^2 c^{(t)}) \times \frac{\partial}{\partial i^{(t)}} (f^{(t)} c^{(t-1)} + i^{(t)} g^{(t)}) \\&= \bar{h}^{(t)} o^{(t)} (1 - \tanh^2 c^{(t)}) g^{(t)}\end{aligned}$$

$$\begin{aligned}
 \bar{g}^{(t)} &= \frac{\partial L}{\partial g^{(t)}} = \frac{\partial L}{\partial c^{(t)}} \times \frac{\partial c^{(t)}}{\partial g^{(t)}} \\
 &= \bar{h}^{(t)} o^{(t)} (1 - \tanh^2 c^{(t)}) \times \frac{\partial}{\partial g^{(t)}} (f^{(t)} c^{(t-1)} + i^{(t)} g^{(t)}) \\
 &= \bar{h}^{(t)} o^{(t)} (1 - \tanh^2 c^{(t)}) i^{(t)}
 \end{aligned}$$

gradient w.r.t forget gate

$$\begin{aligned}
 \bar{f}^{(t)} &= \frac{\partial L}{\partial f^{(t)}} = \frac{\partial L}{\partial c^{(t)}} \times \frac{\partial c^{(t)}}{\partial f^{(t)}} \\
 &= \bar{h}^{(t)} o^{(t)} (1 - \tanh^2 c^{(t)}) \times \frac{\partial}{\partial f^{(t)}} (f^{(t)} c^{(t-1)} + i^{(t)} g^{(t)}) \\
 &= \bar{h}^{(t)} o^{(t)} (1 - \tanh^2 c^{(t)}) c^{(t-1)}
 \end{aligned}$$

(b)

$$\begin{aligned}
 \bar{w}_{ix} &= \frac{\partial L}{\partial w_{ix}} = \frac{\partial L}{\partial i^{(t)}} \times \frac{\partial i^{(t)}}{\partial w_{ix}} \\
 &= \bar{h}^{(t)} o^{(t)} (1 - \tanh^2 c^{(t)}) g^{(t)} \times \frac{\partial}{\partial w_{ix}} (\sigma(w_{ix} x^{(t)} + w_{ih} h^{(t-1)})) \\
 &= \bar{h}^{(t)} o^{(t)} (1 - \tanh^2 c^{(t)}) g^{(t)} \sigma'(w_{ix} x^{(t)} + w_{ih} h^{(t-1)}) x^{(t)}
 \end{aligned}$$