

Bogaziçi University  
Guided Research  
LSTM

Alptekin Orbah  
Student ID: 2017700090

10 November 2017

# 1 BPPT

## 1.1 Forward

$$z^t = g(W_z x^t + R_z h^{t-1}) \quad (1)$$

$$i^t = \sigma(W_i x^t + R_i h^{t-1}) \quad (2)$$

$$f^t = \sigma(W_f x^t + R_f h^{t-1}) \quad (3)$$

$$o^t = \sigma(W_o x^t + R_o h^{t-1}) \quad (4)$$

$$c^t = (z^t \odot i^t) + (c^{t-1} \odot f^t) \quad (5)$$

$$h^t = h(c^t) \odot o^t \quad (6)$$

Note that

$$\xi = \sigma(\bar{\xi}) \quad (7)$$

$$\xi = g(\bar{\xi}) \quad (8)$$

## 1.2 Backward for $W_f$

$$\frac{\partial E}{\partial W_f} = \frac{\partial E}{\partial h} \frac{\partial h}{\partial W_f} \quad (1)$$

$$\frac{\partial h}{\partial W_f} = \frac{\partial h}{\partial c^t} \frac{\partial c^t}{\partial W_f} + \frac{\partial h}{\partial o^t} \frac{\partial o^t}{\partial W_f} \quad (2)$$

$$\frac{\partial c^t}{\partial W_f} = \frac{\partial f^t}{\partial z^t} \frac{\partial z^t}{\partial W_f} + \frac{\partial c^t}{\partial i^t} \frac{\partial i^t}{\partial W_f} + \frac{\partial c^t}{\partial c^{t-1}} \frac{\partial c^{t-1}}{\partial W_f} + \frac{\partial c^t}{\partial f^t} \frac{\partial f^t}{\partial W_f} \quad (3)$$

$$\frac{\partial f^t}{\partial W_f} = \frac{\partial f^t}{\partial \phi} \frac{\partial \phi}{\partial W_f} + \frac{\partial f^t}{\partial h^{t-1}} \quad \text{Note that} \quad \phi = W_f x^t \quad (4)$$

$$\frac{\partial i^t}{\partial W_f} = \frac{\partial i^t}{\partial h^{t-1}} \frac{\partial h^{t-1}}{\partial W_f} \quad (5)$$

$$\frac{\partial z^t}{\partial W_f} = \frac{\partial z^t}{\partial h^{t-1}} \frac{\partial h^{t-1}}{\partial W_f} \quad (6)$$

$$\frac{\partial o^t}{\partial W_f} = \frac{\partial o^t}{\partial h^{t-1}} \frac{\partial h^{t-1}}{\partial W_f} \quad (7)$$

### 1.3 Generic Formula

$$\frac{\partial E}{\partial h^t} = \begin{cases} R_z^\top \frac{\partial E}{\partial \bar{z}^{t+1}} + R_i^\top \frac{\partial E}{\partial i^{t+1}} + R_o^\top \frac{\partial E}{\partial \bar{o}^{t+1}} + R_f^\top \frac{\partial E}{\partial f^{t+1}} & t \text{ is not last time step.} \\ \frac{\partial E}{\partial y} \frac{\partial y}{\partial h^t} & t \text{ is last time step and } y \text{ is output.} \end{cases}$$

$$\frac{\partial E}{\partial \bar{o}^t} = \frac{\partial E}{\partial h^t} \odot h(c^t) \odot \sigma'(\bar{o}^t) \quad (1)$$

$$\frac{\partial E}{\partial c^t} = \frac{\partial E}{\partial h^t} \odot o^t \odot h'(c^t) + \frac{\partial E}{\partial c^{t+1}} \odot f^{t+1} \quad (2)$$

$$\frac{\partial E}{\partial \bar{f}^t} = \frac{\partial E}{\partial c^t} \odot c^{t-1} \odot \sigma'(\bar{f}^t) \quad (3)$$

$$\frac{\partial E}{\partial \bar{i}^t} = \frac{\partial E}{\partial c^t} \odot z^t \odot \sigma'(\bar{i}^t) \quad (4)$$

$$\frac{\partial E}{\partial \bar{z}^t} = \frac{\partial E}{\partial c^t} \odot i^t \odot \sigma'(\bar{z}^t) \quad (5)$$

We know that all  $W$ s and  $R$ s are shared parameter. But, we can write  $\frac{\partial E}{\partial W} = \sum W^t$  where  $W^t$  is  $W$  at time step  $t$ . And the same manner is applied for  $R$ .

$$\frac{\partial E}{\partial W_f^t} = \frac{\partial E}{\partial \bar{f}^t} \otimes x^t \quad (6)$$

$$\frac{\partial E}{\partial W_f} = \sum_0^T W_f^t = \frac{\partial E}{\partial \bar{f}^t} \otimes x^t \quad (7)$$

$$\frac{\partial E}{\partial R_f^t} = \frac{\partial E}{\partial \bar{f}^t} \otimes h^{t-1} \quad (8)$$

$$\frac{\partial E}{\partial R_f} = \sum_1^T R_f^t = \frac{\partial E}{\partial \bar{f}^t} \otimes h^{t-1} \quad (9)$$