forwarding:

$$a^{(t)} = b + Wh^{(t-1)} + Ux^{(t)} \quad \text{——— ①}$$

$$h^{(t)} = \tanh(a^{(t)}) \quad \text{——— ②}$$

$$O^{(t)} = c + Vh^{(t)} \quad \text{——— ③}$$

$$\hat{y}^{(t)} = \mathrm{softmax}(O^{(t)}) \quad \text{——— ④}$$

L 為所有時間步 t 損失之總和，求 $\nabla_{h^{(t)}} L$？

當 $t = \tau =$ 最後一步，$h^{(\tau)} \longrightarrow O^{(\tau)} \longrightarrow L$

$$\Rightarrow \frac{\partial L}{\partial h^{(t)}} = \frac{\partial O^{(t)}}{\partial h^{(t)}} \frac{\partial L}{\partial O^{(t)}}$$

$$\Rightarrow \nabla_{h^{(t)}} L = V^T \nabla_{O^{(t)}} L$$

當 t 不為最後一步，

$$\begin{array}{ccc} & O^{(t)} & \\ h^{(t)} & & L \\ & h^{(t+1)} & \end{array}$$

$$\Rightarrow \frac{\partial L}{\partial h^{(t)}} = \frac{\partial h^{(t+1)}}{\partial h^{(t)}} \cdot \frac{\partial L}{\partial h^{(t+1)}} + \frac{\partial O^{(t)}}{\partial h^{(t)}} \cdot \frac{\partial L}{\partial O^{(t)}}$$

$$= \boxed{\frac{\partial h^{(t+1)}}{\partial h^{(t)}}} \cdot \nabla_{h^{(t+1)}} L + V^T \nabla_{O^{(t)}} L = W^T \overbrace{\left(1 - (h^{(t+1)})^2\right)}^{H^{(t+1)}} \left(\nabla_{h^{(t+1)}} L\right) + V^T \left(\nabla_{O^{(t)}} L\right)$$

$$\Rightarrow \frac{\partial h^{(t+1)}}{\partial h^{(t)}} = \frac{\partial a^{(t+1)}}{\partial h^{(t)}} \cdot \frac{\partial h^{(t+1)}}{\partial a^{(t+1)}}$$

① $\Rightarrow a^{(t+1)} = b + Wh^{(t)} + Ux^{(t+1)} \quad \Rightarrow \dfrac{\partial a^{(t+1)}}{\partial h^{(t)}} = W^T$

② $\Rightarrow h^{(t+1)} = \tanh(a^{(t+1)})$

$$= \tanh(b + Wh^{(t)} + Ux^{(t+1)}) \Rightarrow \frac{\partial h^{(t+1)}}{\partial a^{(t+1)}} = 1 - \tanh^2(a^{(t+1)})$$

②左右同時平方

$$= 1 - (h^{(t+1)})^2 \quad \text{——— ⑤}$$

$$\boxed{\frac{d}{dx} \tanh x = 1 - \tanh^2 x}$$

承上:

$$\nabla_W L = \sum_t \frac{\partial a^{(t)}}{\partial W} \cdot \frac{\partial h^{(t)}}{\partial a^{(t)}} \cdot \frac{\partial L}{\partial h^{(t)}}$$

$\downarrow$ by①' $\quad \downarrow$ by⑤ $\quad \downarrow$

$$= \sum_t h^{(t-1)T} \cdot \left(1-\left(h^{(t)}\right)^2\right) \cdot \nabla_{h^{(t)}} L$$

$$= \sum_t H^{(t)} \cdot \left(\nabla_{h^{(t)}} L\right) \cdot h^{(t-1)T} \quad \#$$

$$\nabla_U L = \sum_t \frac{\partial a^{(t)}}{\partial U} \cdot \frac{\partial h^{(t)}}{\partial a^{(t)}} \cdot \frac{\partial L}{\partial h^{(t)}}$$

by①'

$$= \sum_t H^{(t)} \left(\nabla_{h^{(t)}} L\right) x^{(t)T} \quad \#$$

$$\nabla_V L = \sum_t \frac{\partial o^{(t)}}{\partial V} \cdot \frac{\partial L}{\partial o^{(t)}}$$

by③'

$$= \sum_t \left(\nabla_{o^{(t)}} L\right) \cdot h^{(t)T} \quad \#$$

$$\nabla_b L = \sum_t \frac{\partial a^{(t)}}{\partial b} \cdot \frac{\partial h^{(t)}}{\partial a^{(t)}} \cdot \frac{\partial L}{\partial h^{(t)}}$$

$\downarrow$ by①' $\quad \downarrow \quad \downarrow$

$$= \sum_t 1 \cdot H^{(t)} \cdot \nabla_{h^{(t)}} L$$

$$= \sum_t H^{(t)} \left(\nabla_{h^{(t)}} L\right) \quad \#$$

$$\nabla_c L = \sum_t \frac{\partial o^{(t)}}{\partial c} \frac{\partial L}{\partial o^{(t)}}$$

$\downarrow$ by③' $\quad \downarrow$

$$= \sum_t 1 \cdot \nabla_{o^{(t)}} L$$

$$= \sum_t \nabla_{o^{(t)}} L \quad \#$$