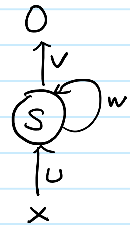
RNN(BPTT)

2017년 7월 1일 토요일 오후 1:17

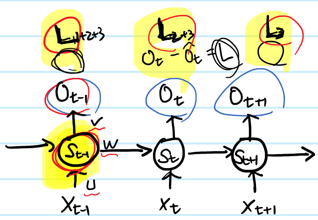
$$\begin{cases} a_t = b + W\cdot S_{t-1} + U\cdot x_t \\ S_t = \tanh(a_t) \\ O_t = V\cdot S_t + C \end{cases}$$

$L_n = \sum_{i=n}^{t} L_i$ (t = the end of sequence)

ex) $L_3 = L_3 + L_4 + \cdots + L_t$



$L_n = \sum_{i=n}^{t} L_i$ (t = the end of sequence)



※. $\Sigma L_t = L_t + \cdots + L_T$ (T = the end of sequence)

$$\frac{\partial \Sigma L_1}{\partial O_1} = \frac{\partial(L_1+L_2+L_3)}{\partial O_1} = \frac{\partial(L_1+\cancel{L_2}+\cancel{L_3})}{\partial L_1} \times \frac{\partial L_1}{\partial O_1} = \frac{\partial L_1}{\partial O_1}$$

$$\frac{\partial \Sigma L_t}{\partial O_t} = \frac{\partial(L_t + L_{t+1} + \cdots + L_T)}{\partial O_t} = \frac{\partial L_t}{\partial O_t} \times \frac{\partial(L_t + L_{t+1} + \cdots + L_T)}{\partial L_t} = \frac{\partial L_t}{O_t} = L'$$

Dependent on only present time.

Chain rule

$$\begin{cases} \frac{\partial \Sigma L_1}{\partial O_1} = \frac{\partial(L_1+L_2+L_3)}{\partial O_1} = \frac{\partial L_1}{\partial O_1} \times \frac{\partial(L_1+\cancel{L_2}+\cancel{L_3})}{\partial L_1} = \frac{\partial L_1}{\partial O_1} \\ \frac{\partial \Sigma L_t}{\partial O_t} = \frac{\partial(L_t+L_{t+1}+\cdots+L_T)}{\partial O_t} = \frac{\partial L_t}{\partial O_t} \times \frac{\partial(L_t+\cancel{L_{t+1}}+\cancel{L_{t+2}+\cdots+L_T})}{\partial L_t} = \frac{\partial L_t}{\partial O_t} = \text{Can be known} = L' \end{cases}$$

Constant

function for O_t : $O_t - \hat{O}_t$

$$\left(\frac{\partial \Sigma L_t}{\partial S_t} = \frac{\partial L_t}{\partial O_t}\cdot\frac{\partial O_t}{\partial S_t}\cdot\frac{\partial \Sigma L_t}{\partial L_t}\right)$$

$$\frac{\partial \Sigma L_t}{\partial S_t} = \underbrace{\frac{\partial L_t}{\partial S_t} + \frac{\partial L_{t+1}}{\partial S_t}}_{\text{Pair wise}} + \cdots + \frac{\partial L_T}{\partial S_t}$$

— known due to finite sequence

$$\frac{\partial \Sigma L_t}{\partial S_t} = \frac{\partial L_t}{\partial S_t} + \frac{\partial L_{t+1}}{\partial S_t}$$

$$= \frac{\partial L_t}{\partial O_t}\cdot\frac{\partial O_t}{\partial S_t} + \frac{\partial L_{t+1}}{\partial S_t}$$ next step (unknown, but...)

$$= \underbrace{L' \times V}_{\text{current (known)}} + \frac{\partial L_{t+1}}{\partial O_{t+1}} \times \frac{\partial O_{t+1}}{\partial S_{t+1}} \times \frac{\partial S_{t+1}}{\partial S_t}$$

$$\frac{\partial V\cdot S_t + C}{\partial S_t}$$

$$\frac{\partial \Sigma L_t}{\partial L_t} = 1$$

$$= L'\cdot V + \frac{\partial \Sigma L_t}{\partial S_{t+1}}\cdot\frac{\partial S_{t+1}}{\partial S_t}\cdot 1 = \delta \text{ (current layer's Loss)}$$

δ^t

$$\frac{\partial S_{t+1}}{\partial S_t} = W\cdot(1 - S_{t+1}^2)$$

$\cdot\ S_{t+1} = \tanh(b + W\cdot S_t + U\cdot x_{t+1})$

$\cdot\ S_t = \tanh(a_t)$

※ $\tanh(x) \xrightarrow{\text{diff}} 1 - \tanh^2(x)$

$$= L'\cdot V + \delta^t \cdot W(1 - S_{t+1}^2)$$

Gradient Vanishing

$0 < ? < 1,\ 0 < ? < 1$

$$\frac{\partial S_3}{\partial S_1} = \frac{\partial S_3}{\partial S_2}\cdot\frac{\partial S_2}{\partial S_1}$$

$$\frac{\partial L_3}{\partial W} = \sum_{i=1}^{3} \frac{\partial L_3}{\partial O_3}\cdot\frac{\partial O_3}{\partial S_3}\cdot\frac{S_3}{S_i}\cdot\frac{S_i}{\partial W}$$