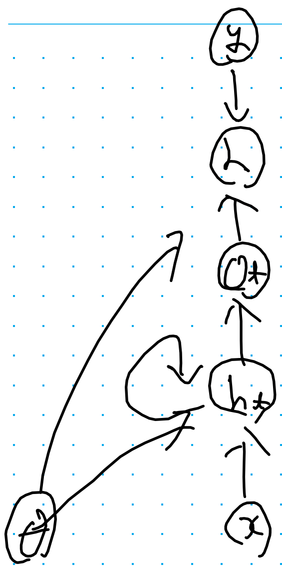
Subject: Answer 9-(c)

Date: 10

Attendants



$$\frac{\partial L_t}{\partial \theta} = \frac{\partial L_t}{\partial o_t}\frac{\partial o_t}{\partial \theta}$$

$$= \frac{\partial L_t}{\partial o_t}\left\{\frac{\partial o_t}{\partial \theta} + \frac{\partial o_t}{\partial h_t}\frac{\partial h_t}{\partial \theta}\right\}$$

$$= \frac{\partial L_t}{\partial o_t}\left\{\frac{\partial o_t}{\partial \theta} + \frac{\partial o_t}{\partial h_t}\left\{\frac{\partial h_t}{\partial \theta} + \frac{\partial h_t}{\partial h_{t+1}}\frac{\partial h_{t+1}}{\partial \theta}\right\}\right\}$$

$$= \frac{\partial L_t}{\partial o_t}\left\{\frac{\partial o_t}{\partial \theta} + \frac{\partial o_t}{\partial h_t}\left\{\frac{\partial h_t}{\partial \theta} + \underline{\frac{\partial h_t}{\partial h_{t+1}}}\left\{\frac{\partial h_{t+1}}{\partial \theta} + \underline{\frac{\partial h_{t+1}}{\partial h_{t+2}}}\frac{\partial h_{t+2}}{\partial \theta}\right\}\right\}\right\}$$

this recurrence has

$$T = c \prod_{i=1}^{n}\left(\frac{\partial h_{n-1}}{\partial h_n}\right) \quad \text{where } c \text{ is constant}$$

when $n \to \infty$

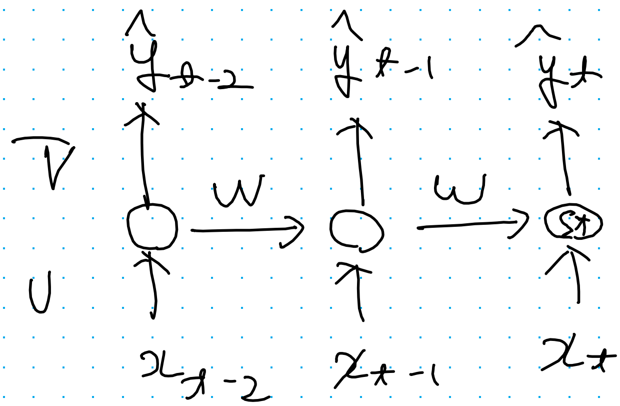
if $\left(\frac{\partial h_{n-1}}{\partial h_n}\right) > 1 \quad \frac{\partial L_t}{\partial \theta} \to \infty$

if $\left(\frac{\partial h_{n-1}}{\partial h_n}\right) < 1 \quad \frac{\partial L_t}{\partial \theta} \to 0$

so, prevent gradient explosion and vanishing gradient, $\frac{\partial h_{n-1}}{\partial h_n}$ should be 1 //

Subject: Extra.　　　　Date: / /

Attendants: 京大医の資料



$$s_t = \sigma(Ux_t + Ws_{t-1})$$

$$y_t = Vs_t$$

$$\frac{\partial s_t}{\partial s_{t-1}} = \frac{\partial}{\partial s_{t-1}} \sigma(Ux_t + Ws_{t-1})$$

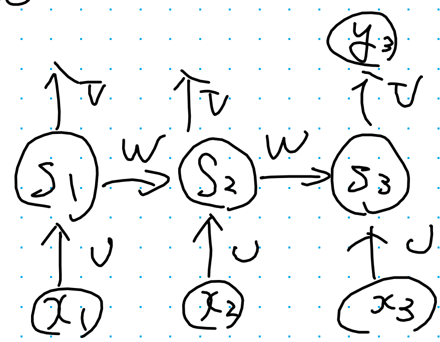
$$= \sigma' \times (Ux_t + Ws_{t-1})$$

$$= \sigma' \times W$$

$$\frac{\partial E}{\partial W} = \sum_{t=1}^{T} \frac{\partial E_t}{\partial W}$$

$$\frac{\partial E_t}{\partial W} = \sum_{k=1}^{t} \frac{\partial E_t}{\partial s_t} \boxed{\frac{\partial s_t}{\partial s_k}} \frac{\partial s_k}{\partial W}$$

$$\frac{\partial s_t}{\partial s_k} = \prod_{i=k+1}^{t} \frac{\partial s_i}{\partial s_{i-1}}$$



e.g. $\frac{\partial E_3}{\partial W} = \frac{\partial E_3}{\partial s_3} \frac{\partial s_3}{\partial s_2} \frac{\partial s_2}{\partial s_1} \frac{\partial s_1}{\partial W}$

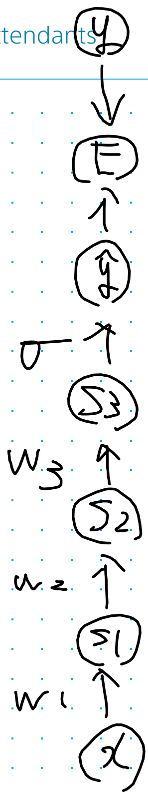
深いDNNと同じ状況になっているね

Wを深い部分と浅い部分で共有するので $\frac{\partial s_t}{\partial s_{t-1}} = \sigma'(Ux_t + Ws_{t-1})W$ が1より大か小で発散 or 消失

消失だけじゃなく発散もする

Subject 深いDNNの場合　　Date 12

Attendants



$$\frac{\partial E}{\partial w_3} = \frac{\partial E}{\partial s_3}\frac{\partial s_3}{\partial w_3}$$

$$s_t = \sigma(W_t s_{t-1} + b)$$

$$\frac{\partial E}{\partial w_2} = \frac{\partial E}{\partial s_3}\frac{\partial s_3}{\partial s_2}\frac{\partial s_2}{\partial w_2}$$

$$\frac{\partial E}{\partial w_1} = \frac{\partial E}{\partial s_3}\frac{\partial s_3}{\partial s_2}\frac{\partial s_2}{\partial s_1}\frac{\partial s_1}{\partial w_1}$$

$\sigma' w_2$　　$\sigma' w_1$

$$\boxed{\frac{\partial}{\partial s_2}\sigma(W_3 s_2 + b) = \sigma' w_3}$$

$$\frac{\partial E}{\partial w_1} = \frac{\partial E}{\partial s_3} \times (\sigma')^3 \times \underline{w_3 w_2 w_1}$$

同じwじゃない