

Vanilla RNN

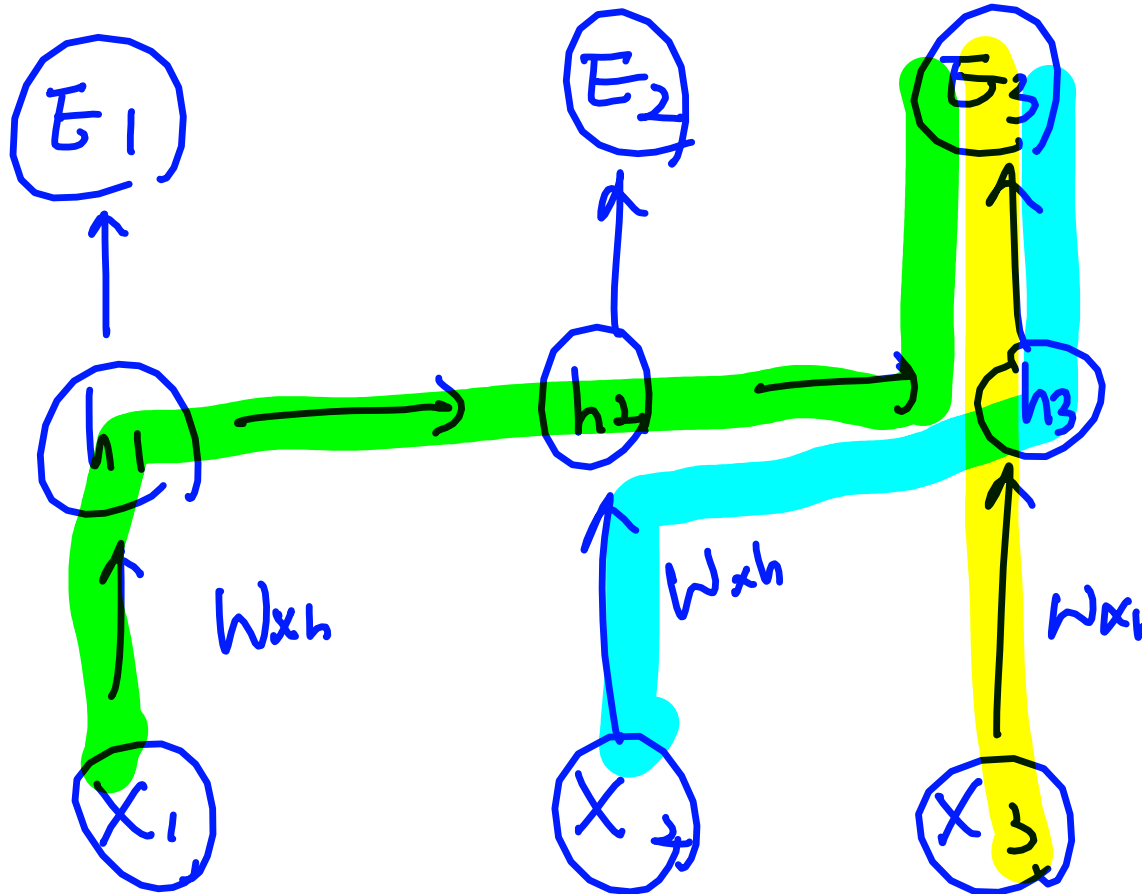
seq가 길어지면...
(ex. 문장 번역기)

LSTM이 좋음.
gradient exploding
gradient vanishing

$$① W = W - lr \cdot \frac{\partial E}{\partial W}$$

BPTT ... → gradient vanishing

$$② \frac{\partial E}{\partial W} = \frac{\partial E_1}{\partial W} + \frac{\partial E_2}{\partial W} + \frac{\partial E_3}{\partial W}$$

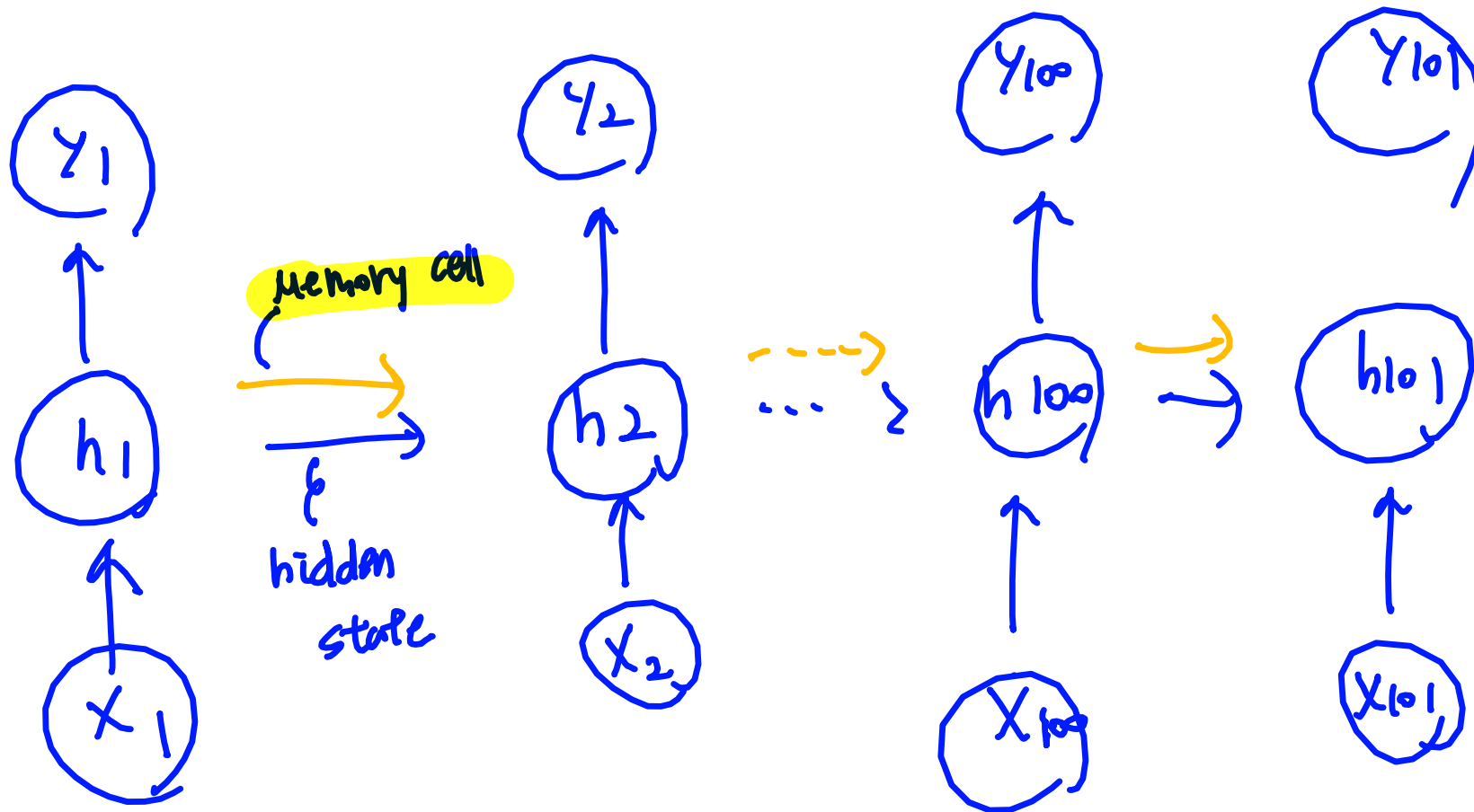


$$\frac{\partial E_3}{\partial W} = \frac{\partial E_3}{\partial h_3} \times \frac{\partial h_3}{\partial W_{xh}}$$

$$+ \frac{\partial E_3}{\partial h_2} \times \frac{\partial h_3}{\partial h_2} \times \frac{\partial h_2}{\partial W_{xh}}$$

$$+ \frac{\partial E_3}{\partial h_3} \times \frac{\partial h_3}{\partial h_1} \times \frac{\partial h_2}{\partial b_1} \times \frac{\partial h_1}{\partial W_{xh}}$$

LSTM \rightarrow Memory cell



• 잊는 과정

