

Long-Short Term Memory

”

NLP 1조 황익린

LSTM



목차

”

01

Introduction

02

LSTM

03

Model architecture

04

Forward & back
propagation

05

LSTM 모델의 변형



”

Introduction

01

● ● ○ Introduction

RNN(Recurrent neural network)

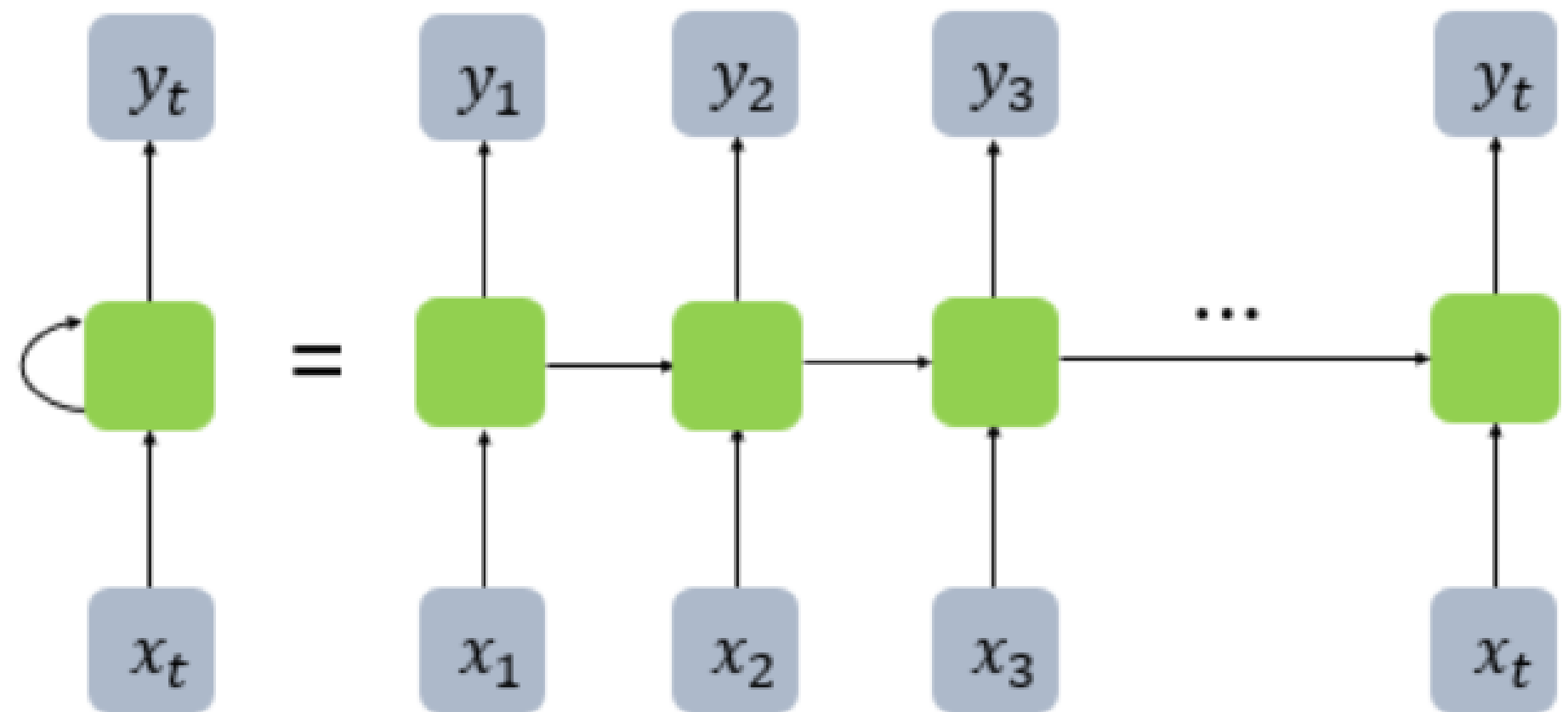
‘기억’을 가지는 모델

01 Hidden state

맥락을 이해하기 위한 기억력을 가지는 상태

02 Vanilla RNN -> LSTM, GRU

가장 기본적인 RNN 모델에서 다른 모델들이 파생됨

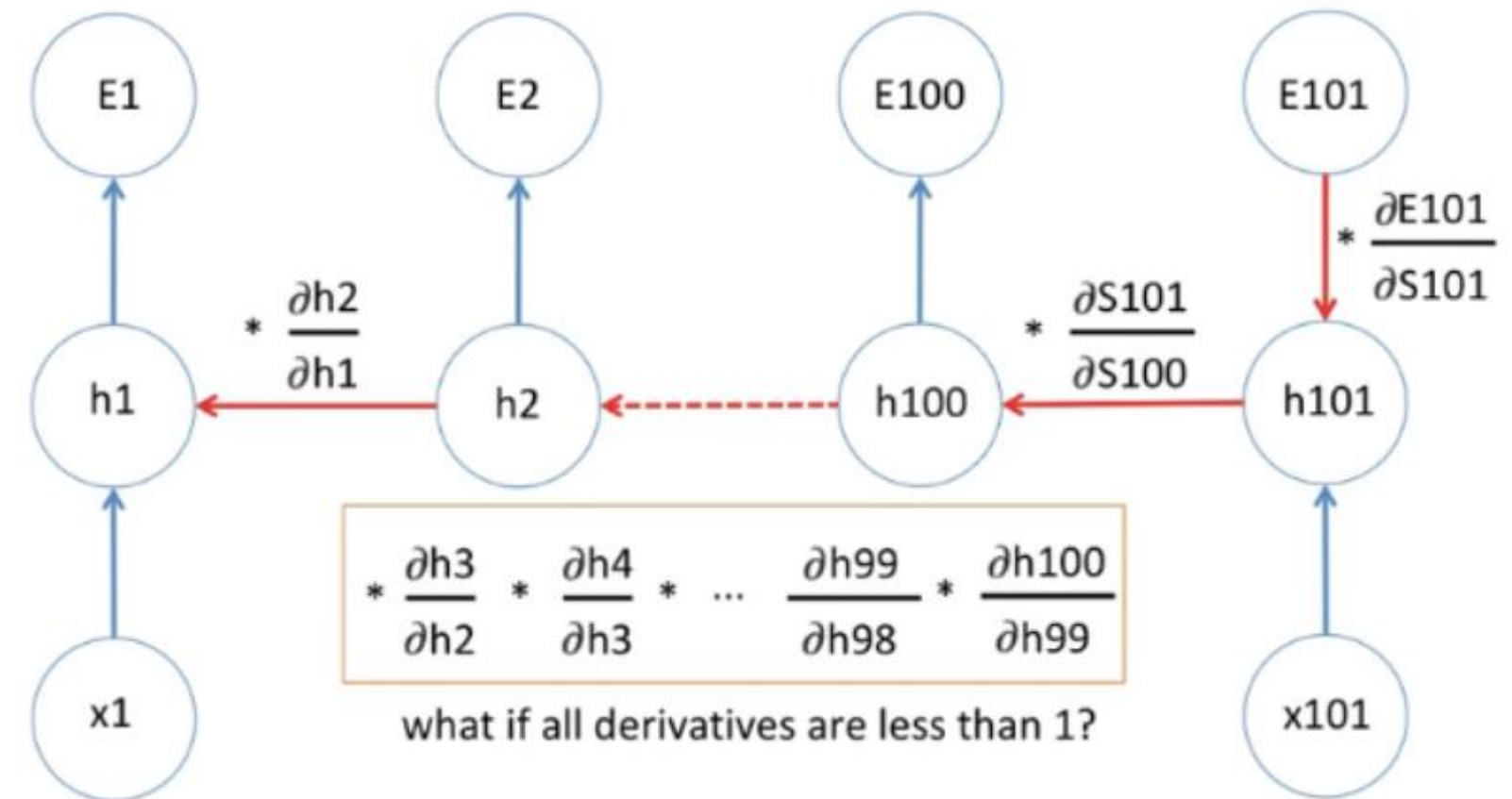


● ● ○ Introduction

BPTT(**backpropagation through time**)

:RNN 모델 학습시 back propagation 이
시점(time step)마다 이루어짐

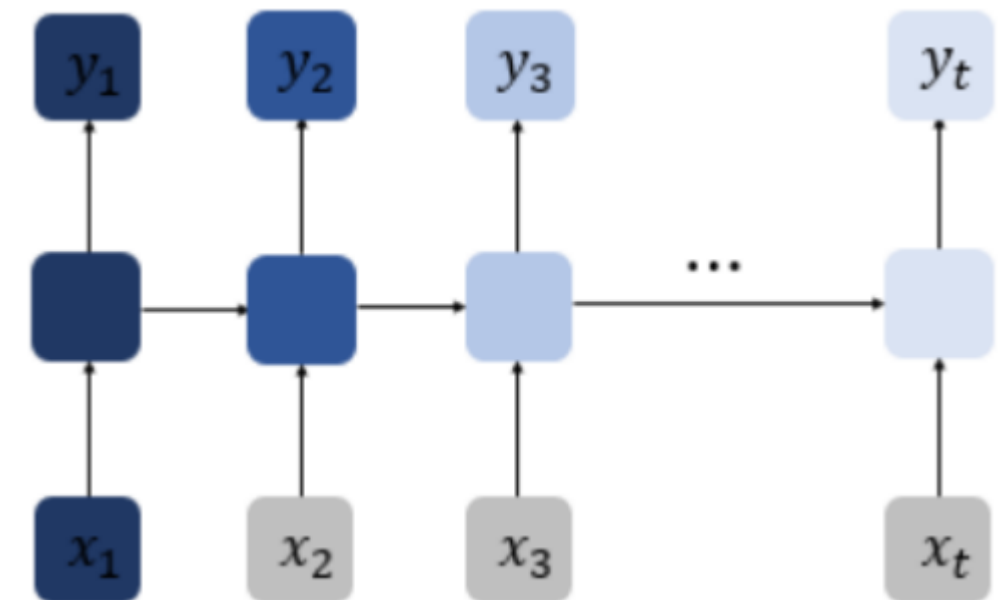
→ gradient vanishing / exploding



● ● ○ Introduction

장기 의존성 문제 (the problem of Long-Term Dependencies)

: 문장의 길이가 길어질 수록 앞의 정보가 뒤로 충분히 전달되지 못함



나는 한국에 살고 내가 쓰는 언어는 ‘_____’



내가 어제 서점에 갔는데 볼만한 책이 많았어.
그리고 오랜만에 밖에 나오니까 좋더라구. 그런데
학교 선배한테 전화가 와서 어디냐고 묻더라?
그래서 나는 말했지. 나는 지금 ‘_____’



”

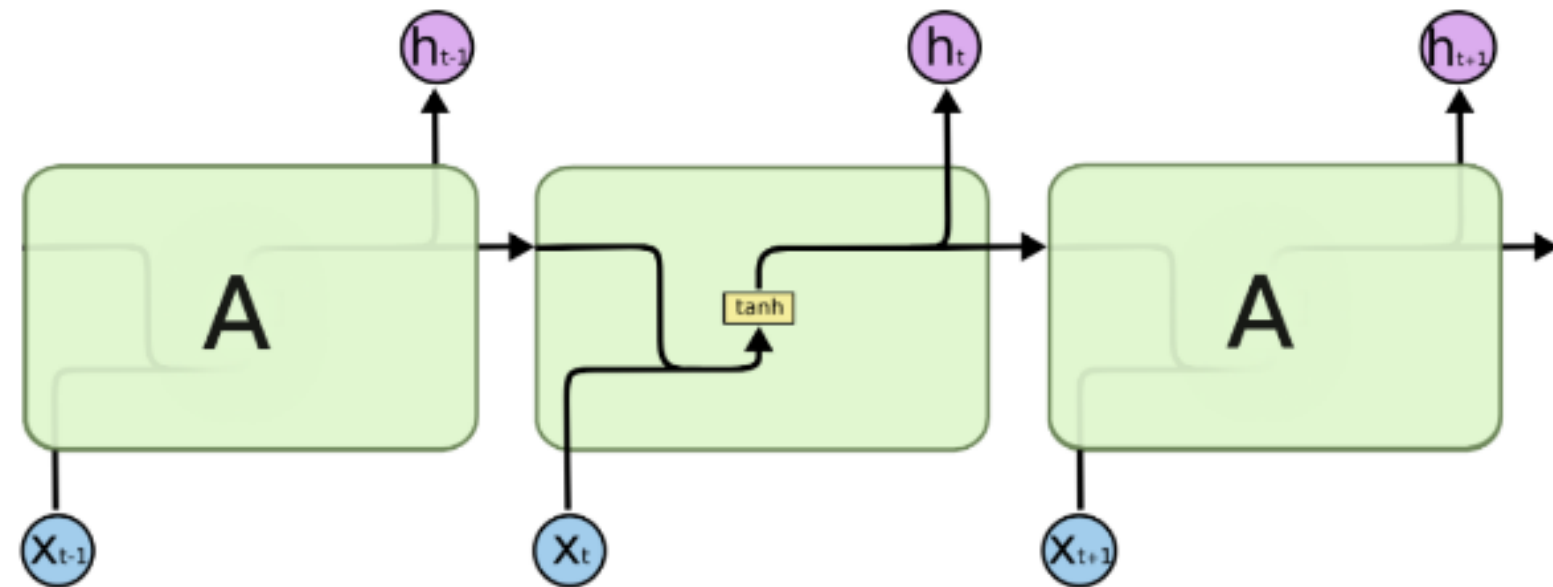
LSTM



02

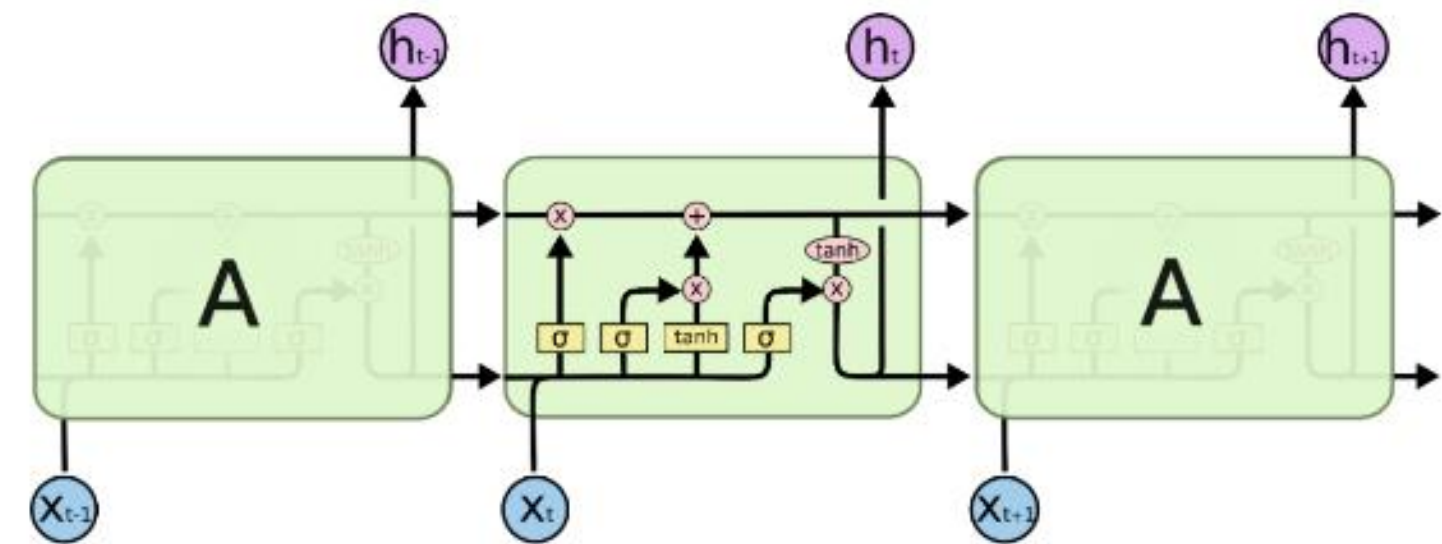
● ● ○ LSTM

RNN



- Input으로 x_t 와 h_{t-1} 을 받음
- 단일 tanh layer
- 단일 layer가 chain 형태로 반복

LSTM



- Cell state가 추가됨
- Tanh layer와 sigmoid layer

● ● ○ LSTM

State

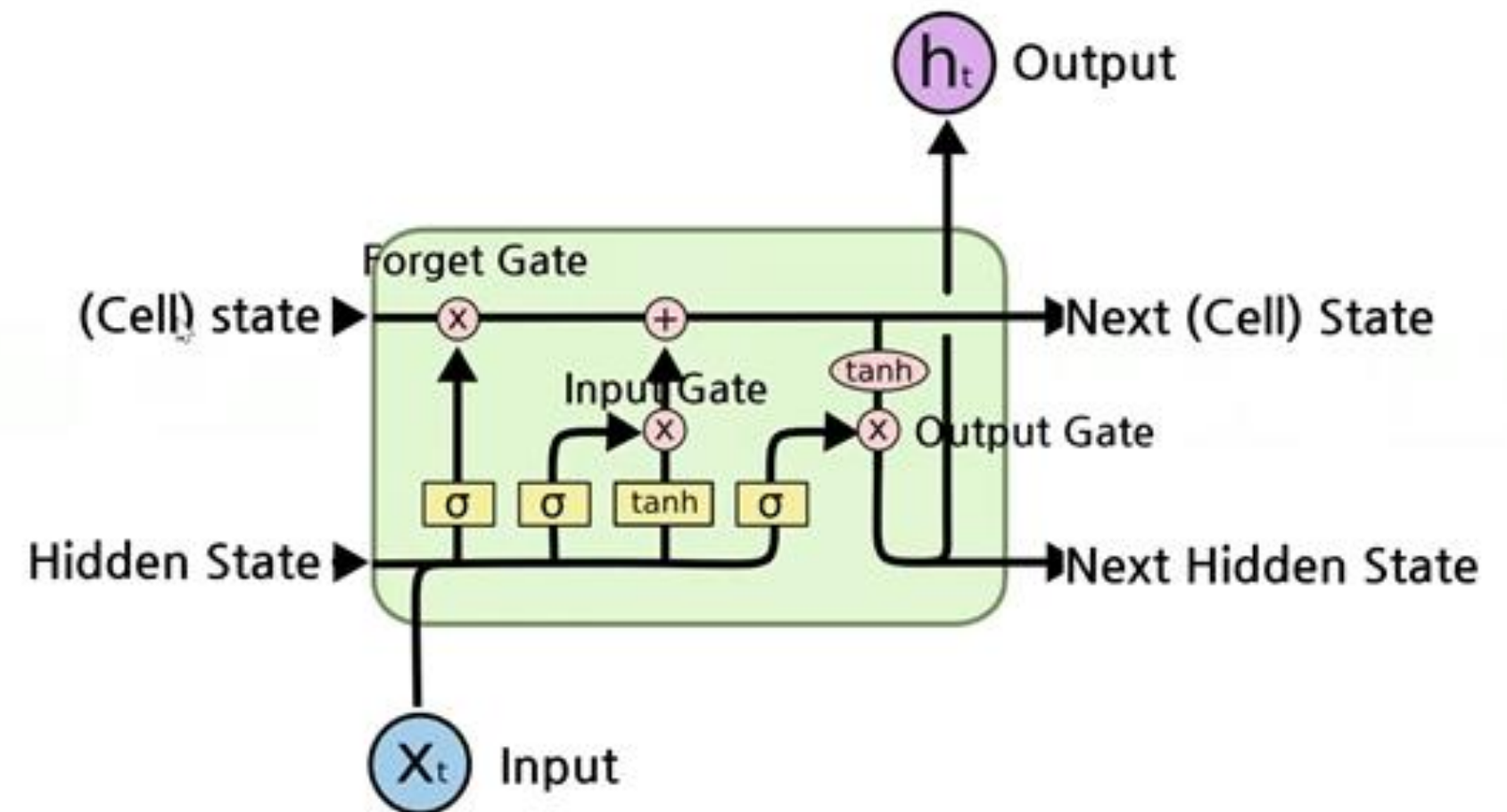
- cell state: 장기 상태
- Hidden state: 단기 상태

Cell state를 조절하는 gate

- 삭제 게이트(forget gate), f
- 입력 게이트(input gate), i , g
- 출력 게이트(output gate), o

Neural network layer

Sigmoid layer을 거쳐 0~1 사이 값을 얻음



“

LSTM은 불필요한 기억은 지우고
기억해야 할 것은 기억한다

—

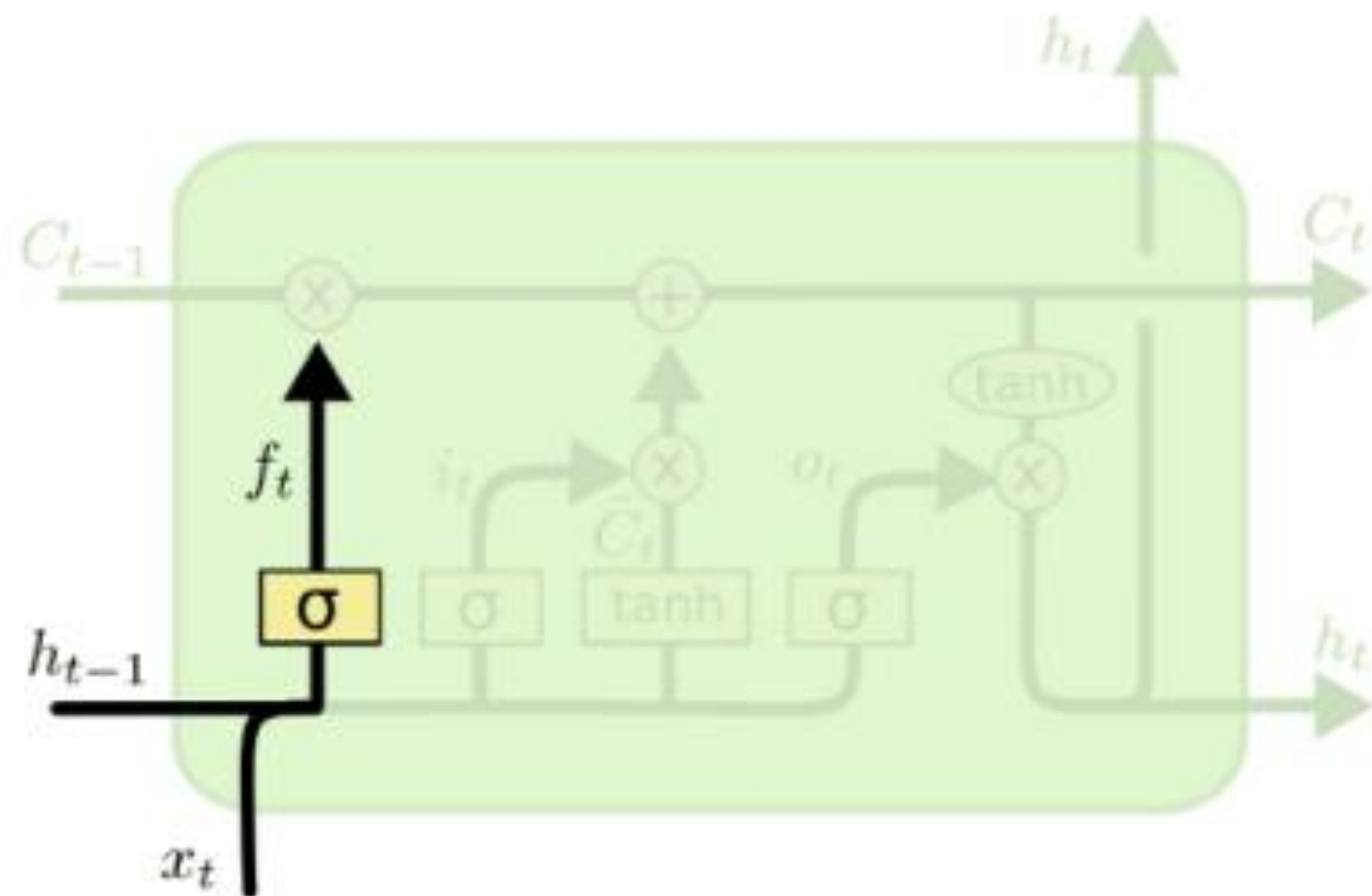


”

Model architecture 03

- Model architecture

Forget gate

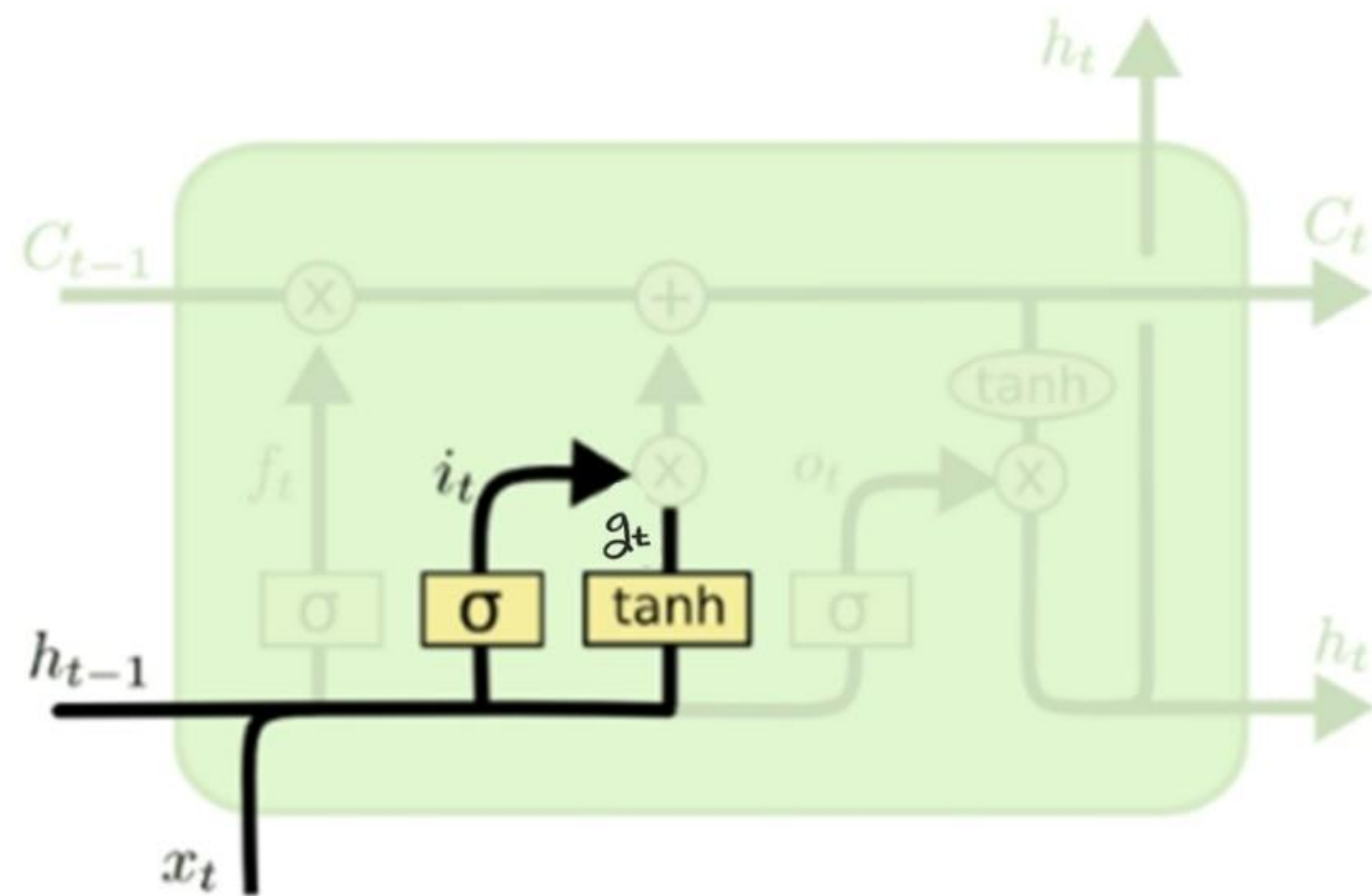


$$f_t = \sigma (W_f \cdot [h_{t-1}, x_t] + b_f)$$

잊을 만한 정보를 잊을 수 있다.

● ● ○ Model architecture

Input gate



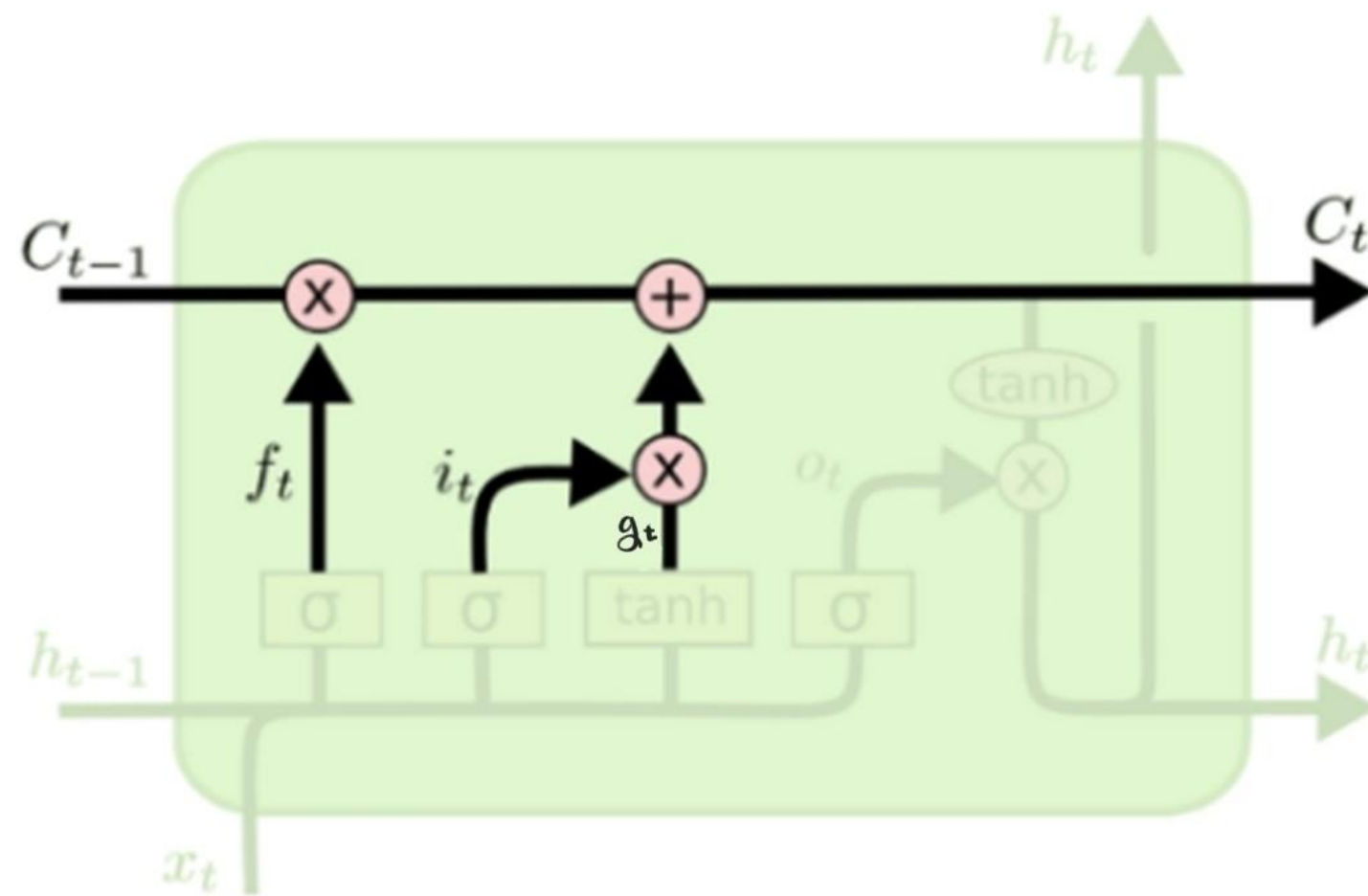
$$i_t = \sigma (W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$g_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_t)$$

새로운 주어가 나왔을 때 이전 주어를 잊어버리고,
새로운 주어를 기억할 수 있다.

• • ○ Model architecture

Cell state update

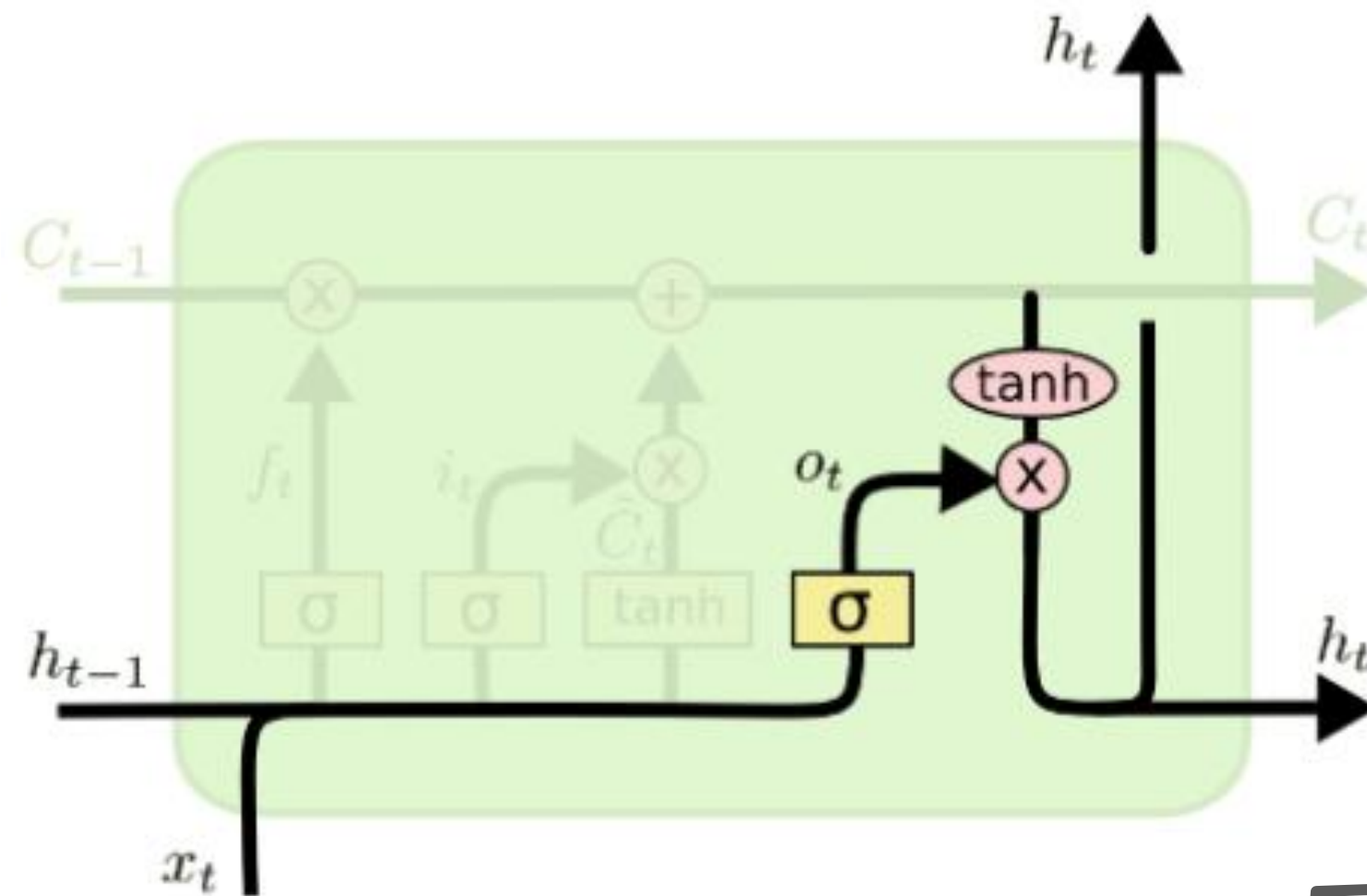


$$C_t = f_t * C_{t-1} + i_t * g_t$$

버릴 것은 버리고, 기억할 것은 기억하는 C_t 가 새롭게 계산됨.

• • ○ Model architecture

output gate



$$o_t = \sigma (W_o [h_{t-1}, x_t] + b_o)$$

$$h_t = o_t * \tanh (C_t)$$

각 스텝에서 다음 hidden state를 계산할 때 cell state가 얼마나 노출될 지 결정된다.



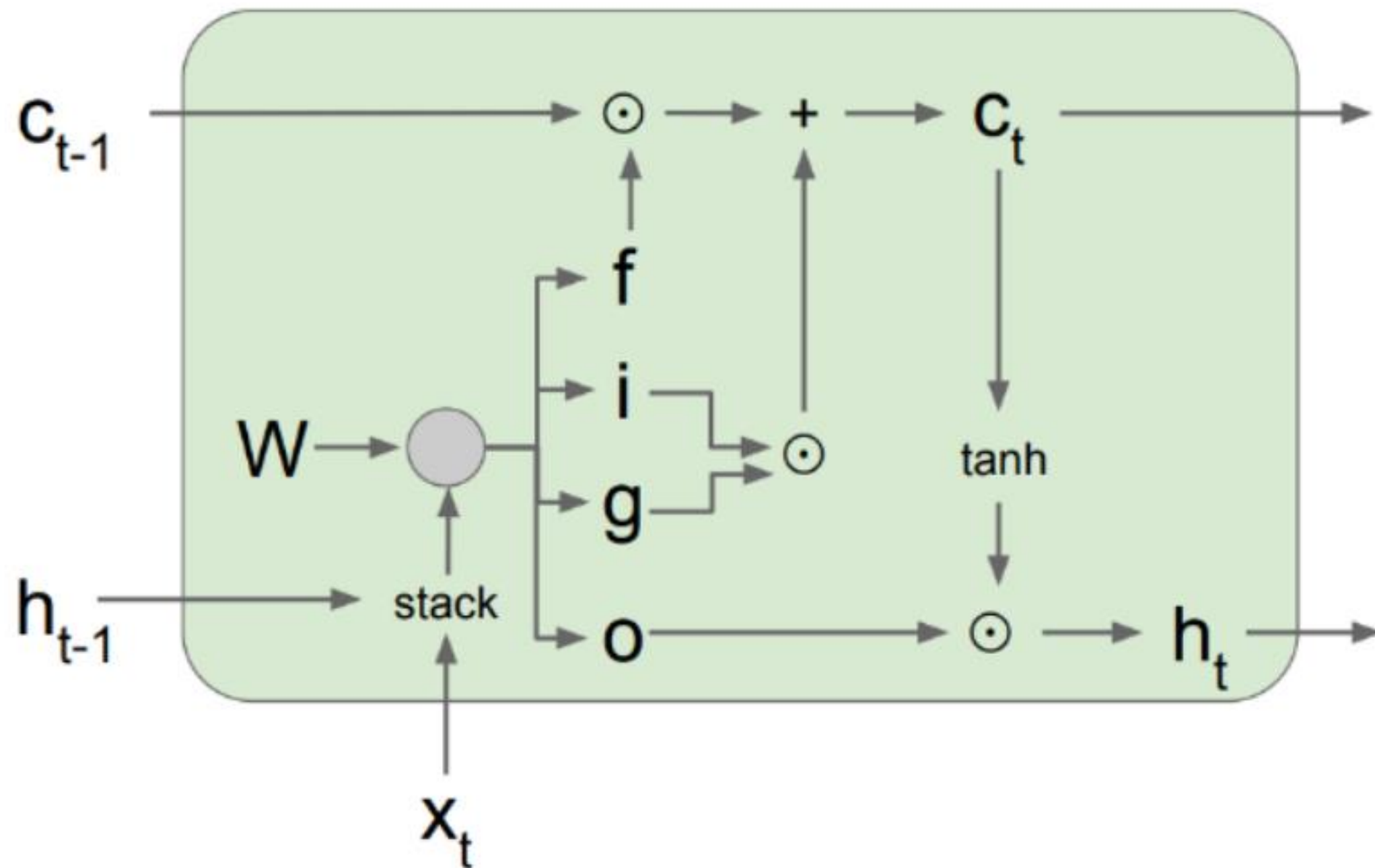
”

forward and back
propagation

04

- • ◦ LSTM forward and backward propagation

Forward propagation



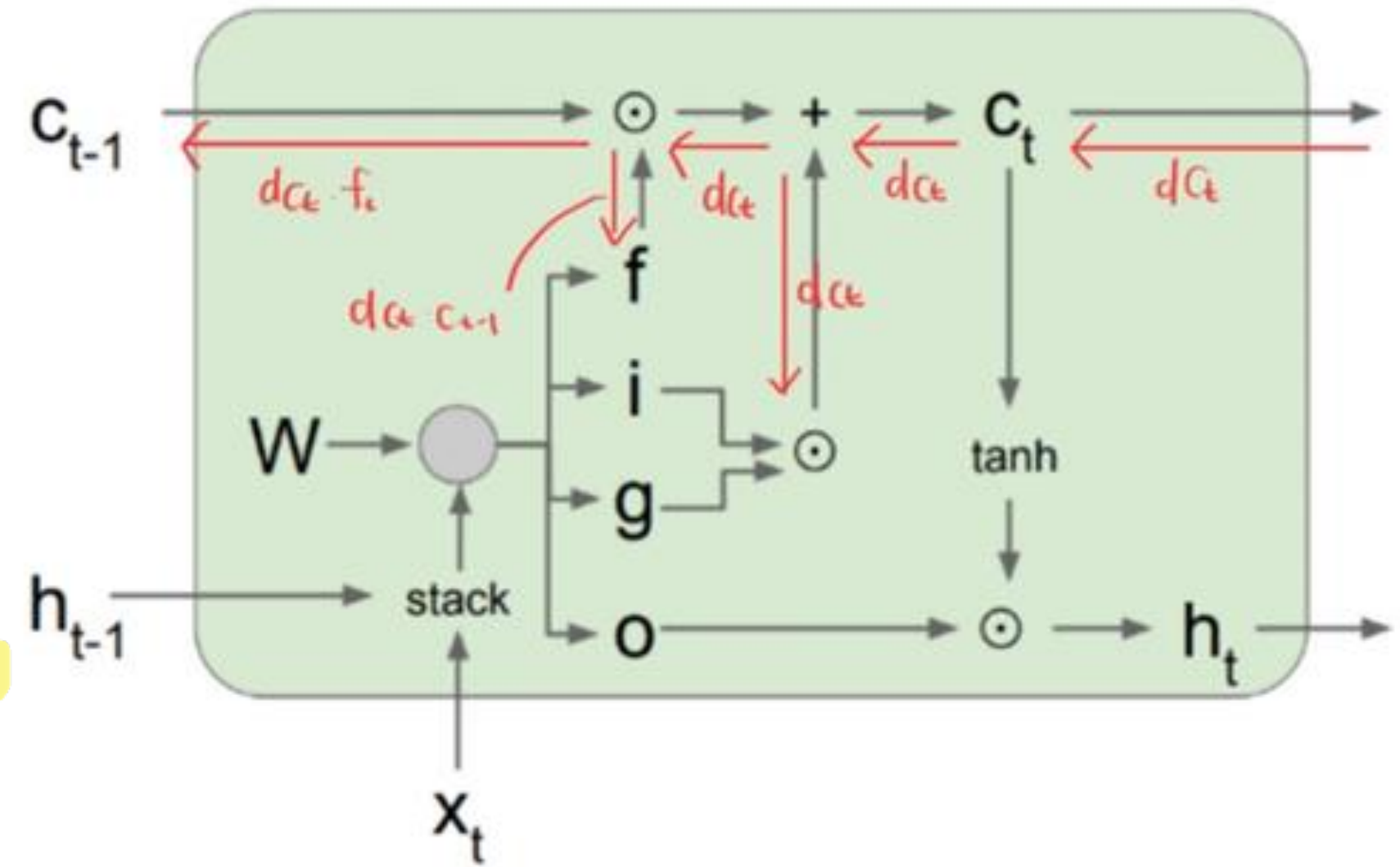
• • ◦ LSTM forward and backward propagation

backward propagation

$$\frac{\partial C_T}{\partial C_{T-1}} = f_T, \frac{\partial C_{T-1}}{\partial C_{T-2}} = f_{T-1}, \dots, \frac{\partial C_{t+1}}{\partial C_t} = f_{t+1}$$

$$\frac{\partial C_T}{\partial C_t} = \prod_{i=t+1}^T f_i$$

- Gradient가 forget gate에만 영향을 받음
- Gradient vanishing / exploding 문제 완호





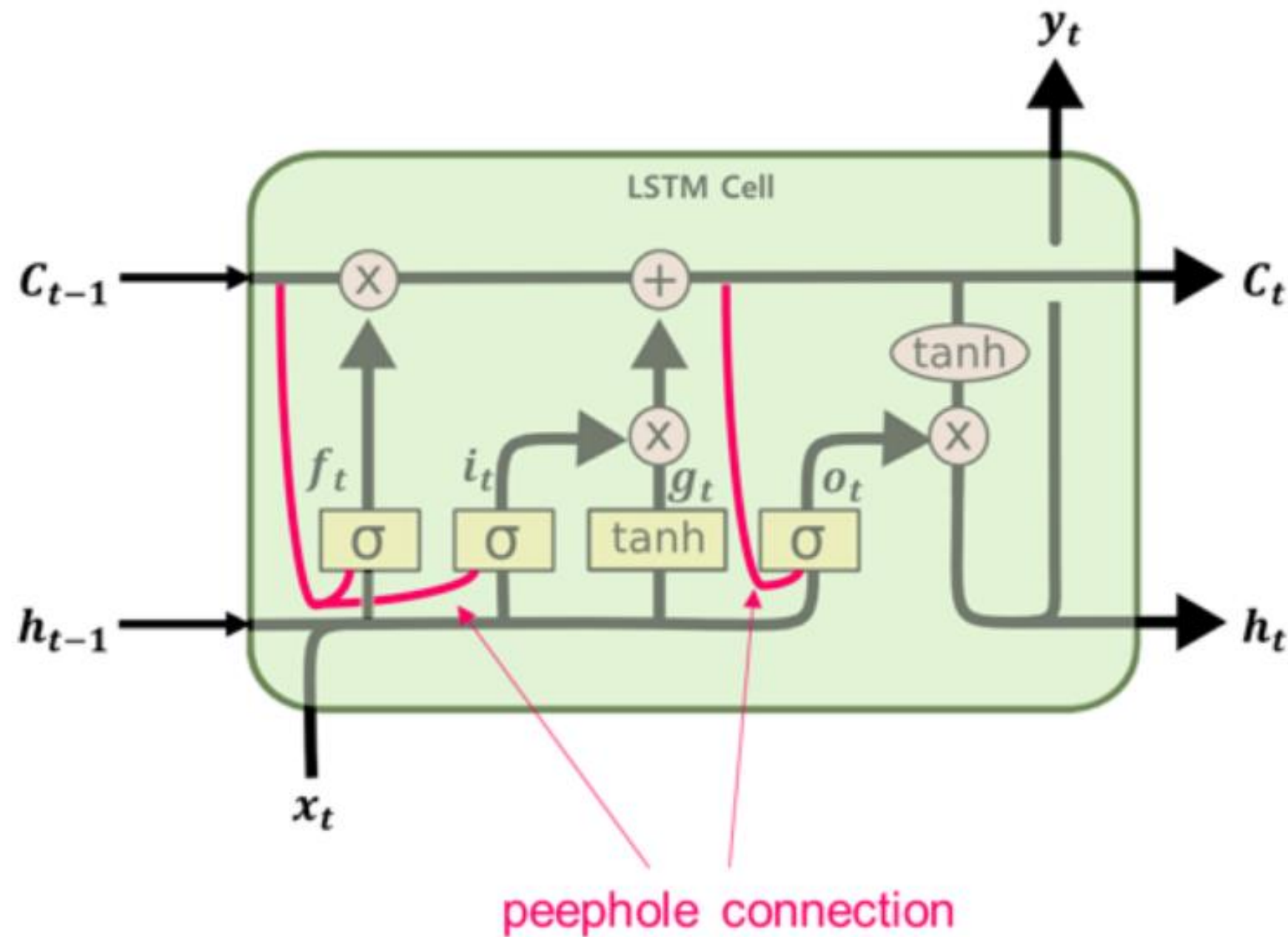
”

LSTM의 변형 모델

05

• • ◦ LSTM의 변형 모델

Peephole connection



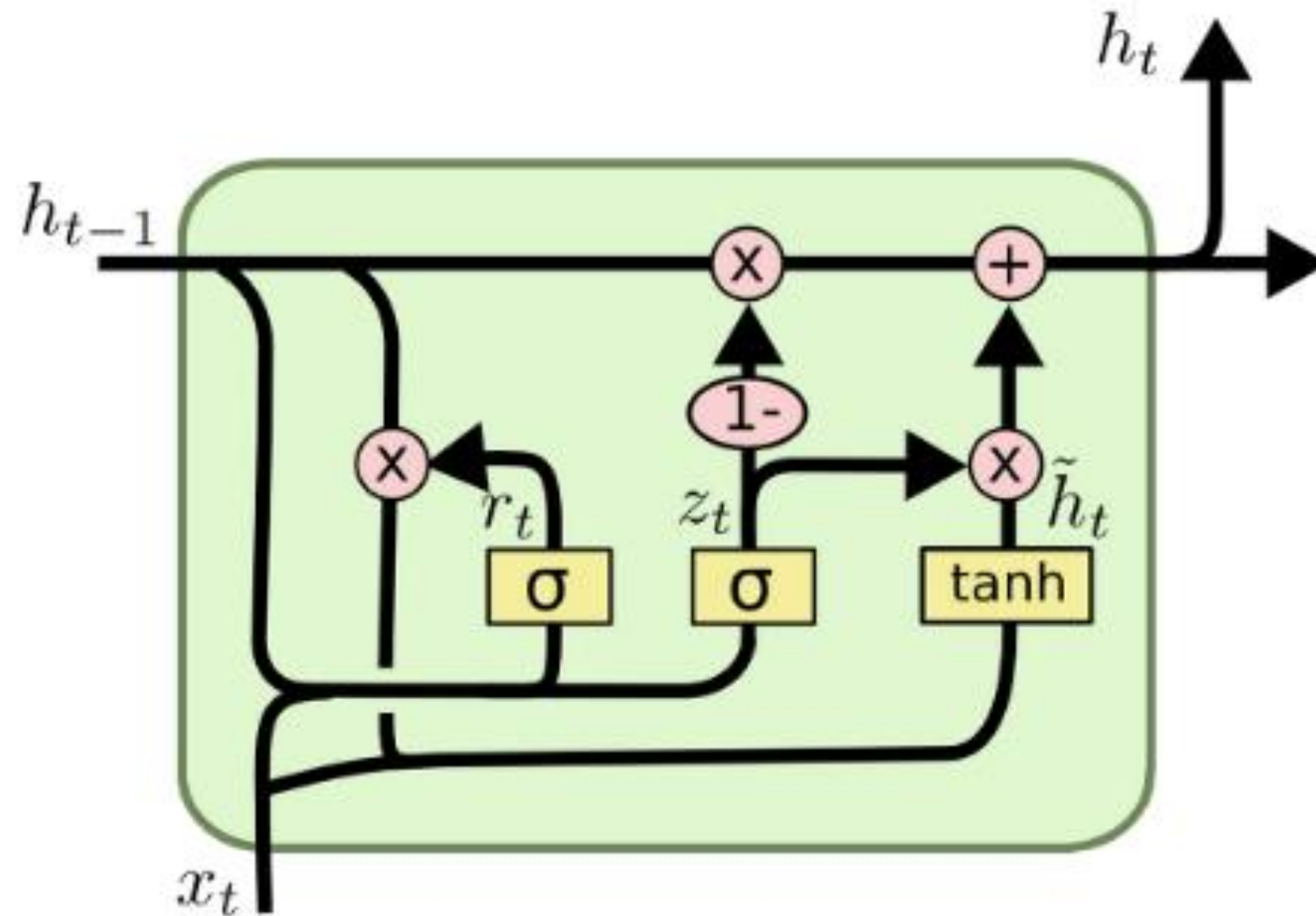
$$f_t = \sigma(W_f \cdot [C_{t-1}, h_{t-1}, x_t] + b_f)$$

$$i_t = \sigma(W_i \cdot [C_{t-1}, h_{t-1}, x_t] + b_i)$$

$$o_t = \sigma(W_o \cdot [C_t, h_{t-1}, x_t] + b_o)$$

• • ◦ LSTM의 변형 모델

GRU(Gated Recurrent Unit)



- Reset gate (r_t)
- Update gate (z_t)

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t])$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t])$$

$$\tilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t])$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$



감사합니다

”

발표 끝 !