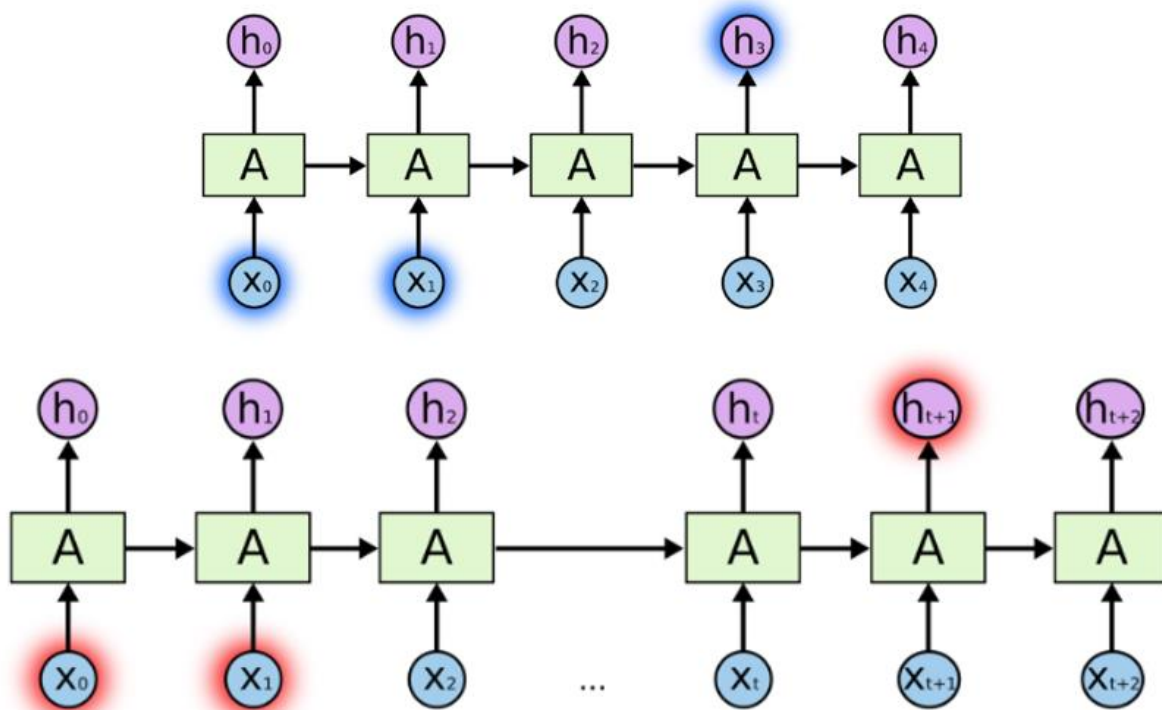


LSTM

RNN의 Vanishing Gradient 문제

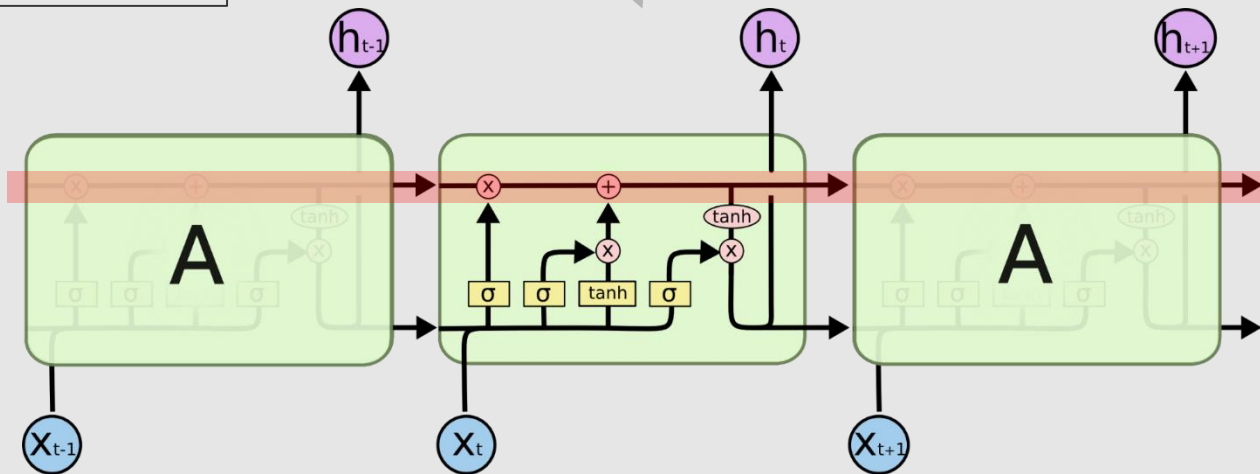
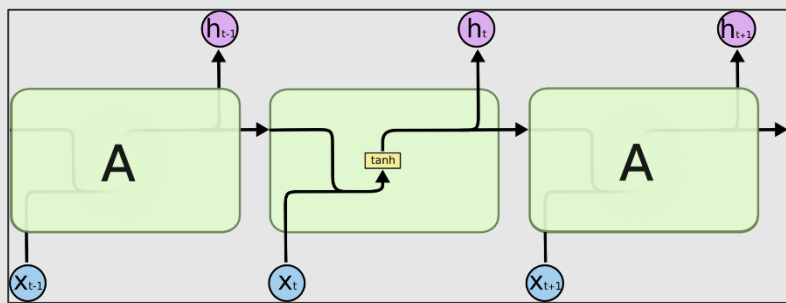
- ✓ RNN은 관련 정보와 그 정보를 사용하는 지점 사이 거리가 멀 경우 역전파시 그래디언트가 점차 줄어 학습능력이 크게 저하



<관련 정보와 그 정보를 사용하는 지점 사이 거리가 멀 경우 RNN 학습능력 저하>

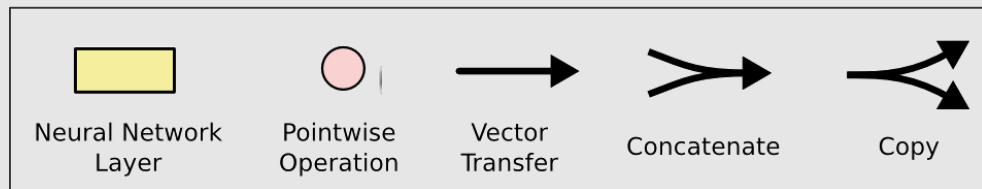
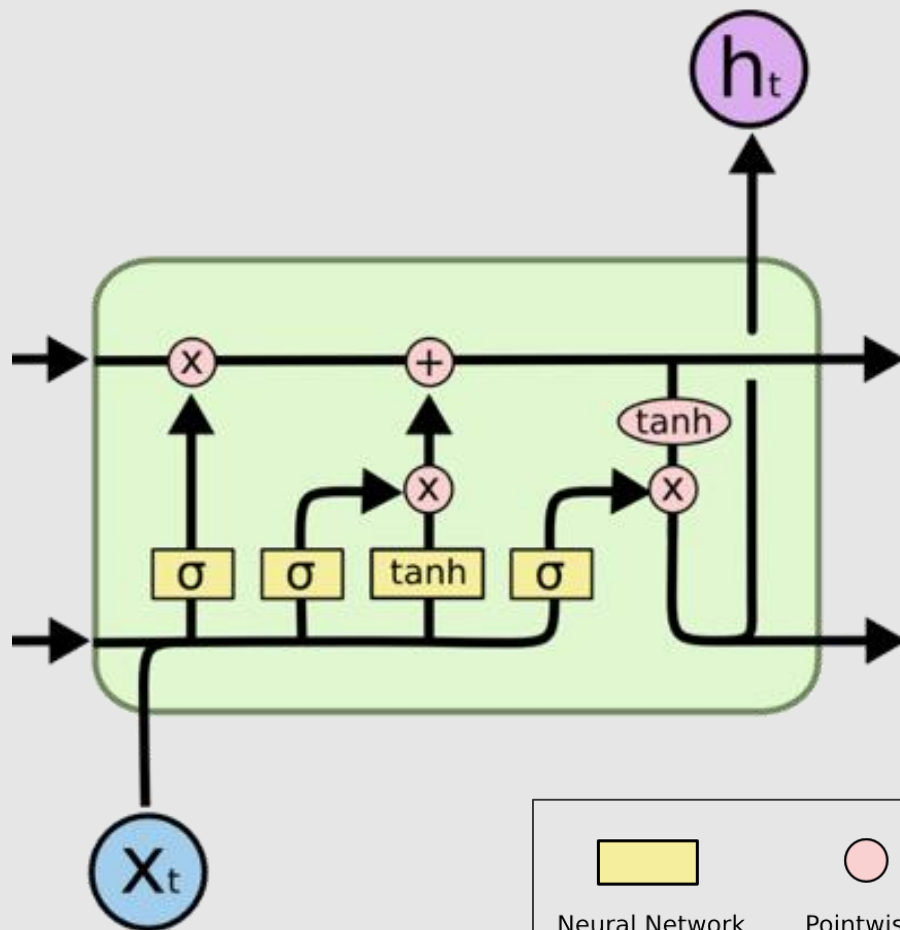
RNN의 Vanishing Gradient 문제

- ✓ 이 문제를 극복하기 위해서 고안된 것이 바로 LSTM
- ✓ RNN의 hidden state에 cell state(장기 기억 메모리)를 추가한 구조



LSTM(Long Short-Term Memory Model)

✓그림 설명



LSTM(Long Short-Term Memory Model)

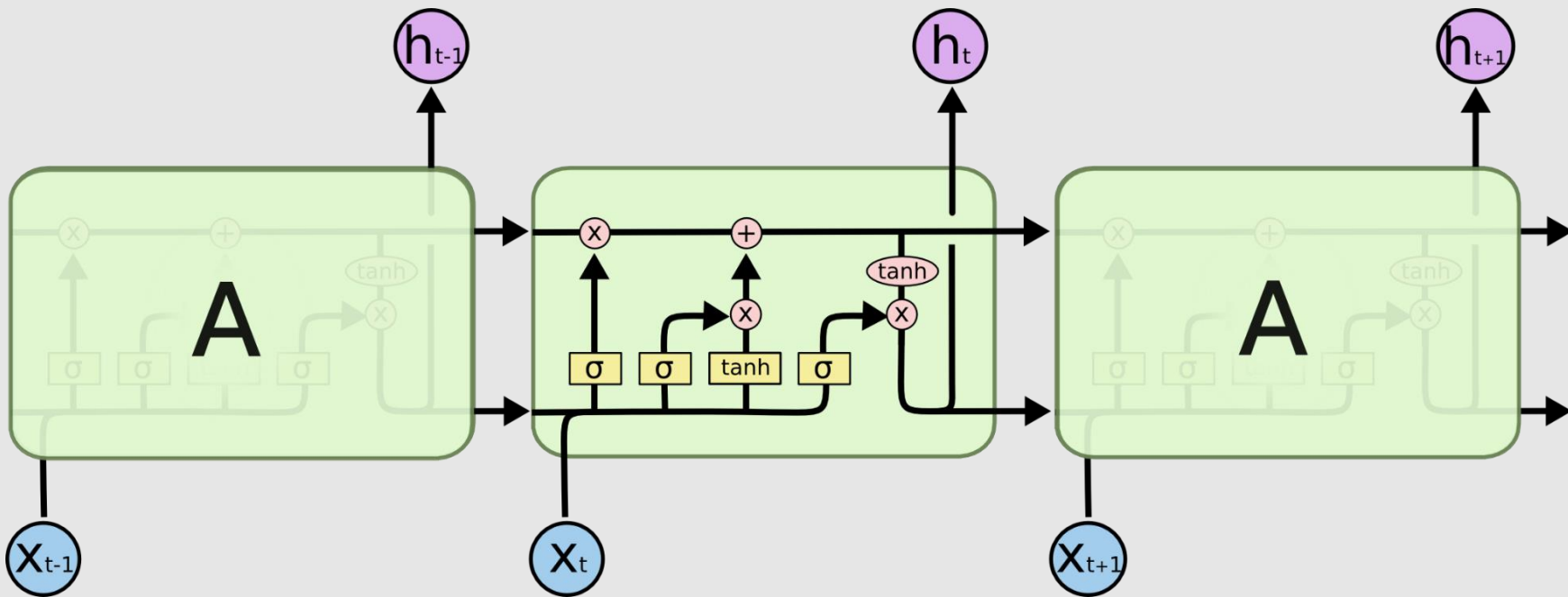
✓ 구조적인 특징

- Cell State
- 3개의 Gate

① Forget Gate

② Input Gate

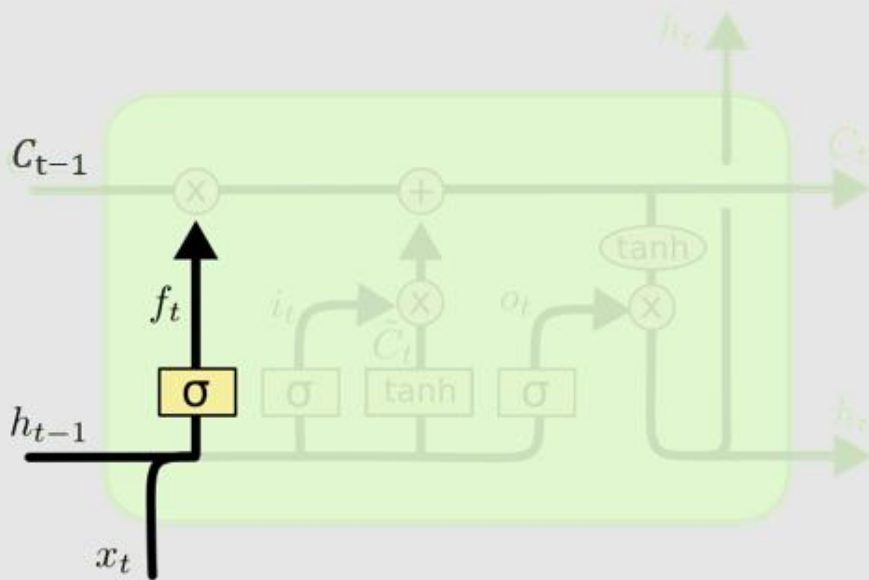
③ Output Gate



LSTM : ① forget gate

✓ C_{t-1} 는 이전 time step에서 넘어온 Cell State 값입니다.

C_{t-1}		f_t		?
2.5	\otimes	0.01	=	
0.45		0.9		
1.2		0.5		
-0.5		0.99		



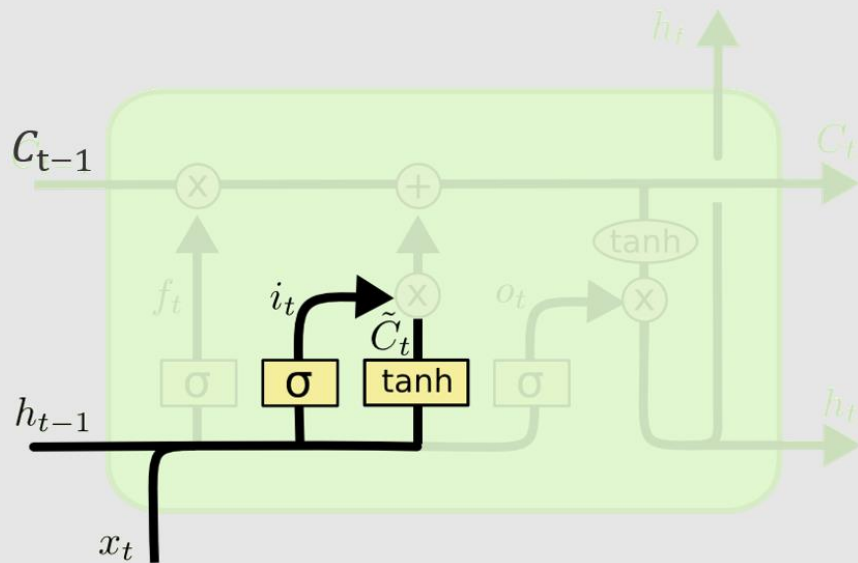
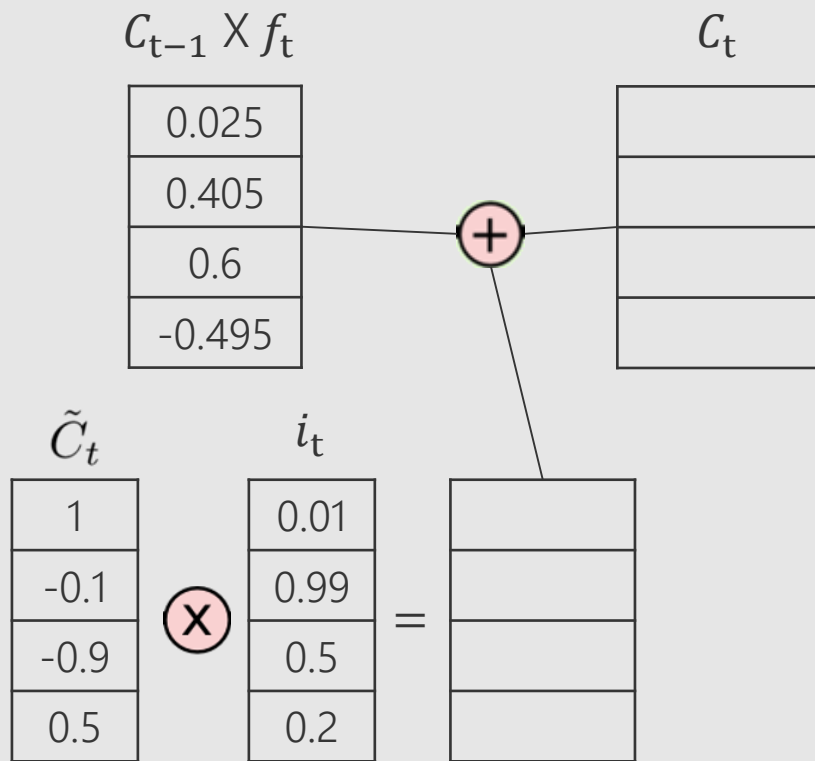
$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

✓ f_t 는 어떤 값을 어떤 비율로 잊을지 정해주는 역할(정보 소실 비율 결정)

LSTM : ② input gate

✓ i_t : 어떤 정보를 어느정도 비율을 반영할지 결정

✓ \tilde{C}_t : Cell state 후보

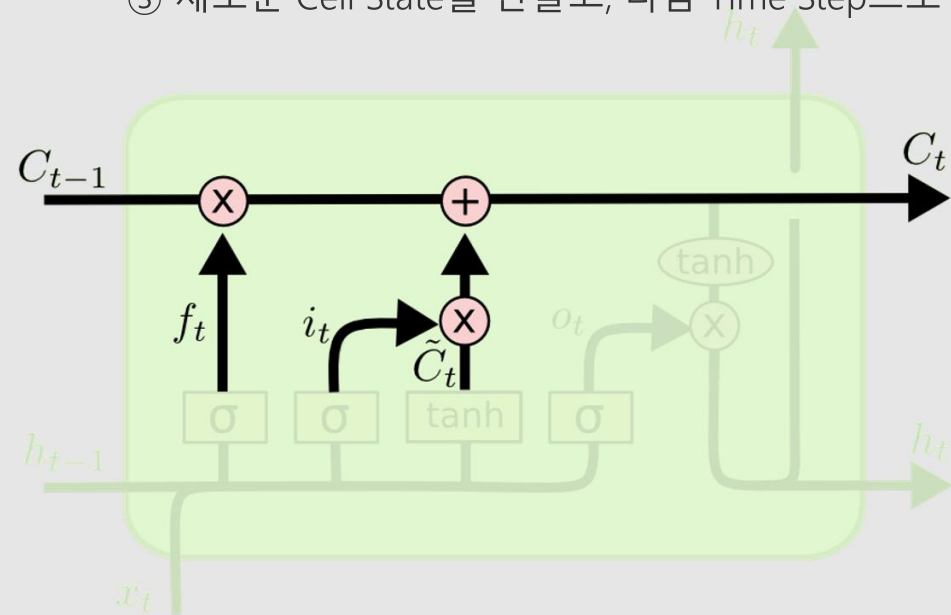


$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

LSTM : Cell State 업데이트 정리

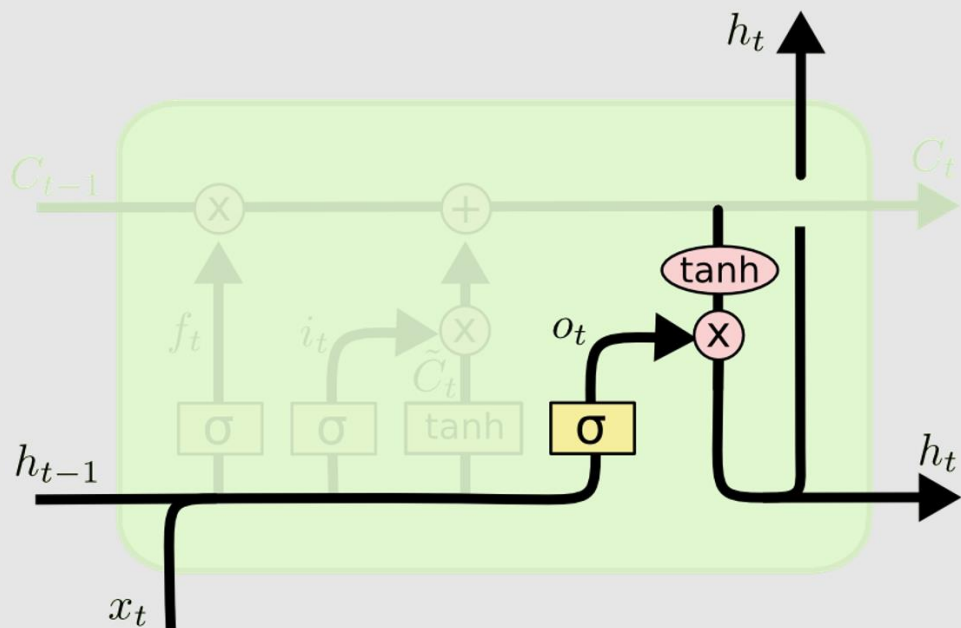
- ✓ Hidden State는 과거의 hidden state를 받고 현시점의 입력데이터를 고려하여 특징을 추출하는 역할
- ✓ Cell State는 메모리 역할
 - ① 과거 정보 중 중요한 것 덜 중요한 것 필터링하고,
 - ② 현시점 Cell State 후보 중 중요한 정보를 반영하여
 - ③ 새로운 Cell State를 만들고, 다음 Time Step으로 전달



$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

LSTM : ③ output gate

- ✓ 새로운 Cell State(C_t)에 의존하여 새로운 Hidden State(h_t) 만드는 부분
- ✓ 역시, C_t 에서 중요한 정보와 그렇지 않은 정보 비율을 적용하여
- ✓ 다음 Hidden State(h_t)로 넘김



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

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

LSTM의 문제점

✓ 최적화 해야 할 파라미터가 많다. → Training Cost가 높다

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

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

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

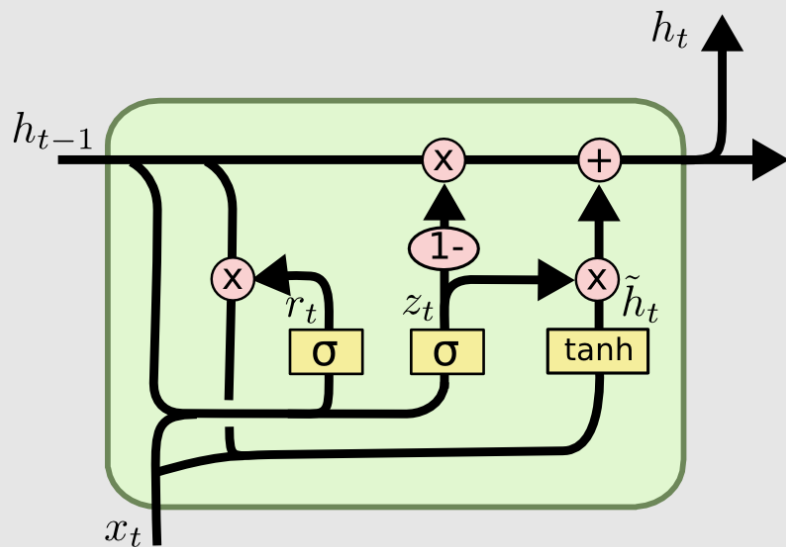
✓ Output gate : $o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$

✓ RNN보다 4배 많다. 그런데, RNN보다 성능이 4배 좋을까?

- 가중치가 많으면 성능을 높일 수 있는 여지가 있기도 하고,
- 최적화가 더 어려울 수도 있다.(가끔 RNN보다 성능이 떨어지기도 한다.)
- 당연히 학습시간은 오래 걸린다.

GRU(Gated Recurrent Units)

- ✓ 2014년에 K. Cho(조경현) 박사에게 의해 제안됨
- ✓ 근본 원리는 LSTM과 유사.
- ✓ LSTM 셀의 경량화 된 버전(Cell State를 Hidden State에 합침)
- ✓ LSTM보다 가벼우면서 성능은 비슷.



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

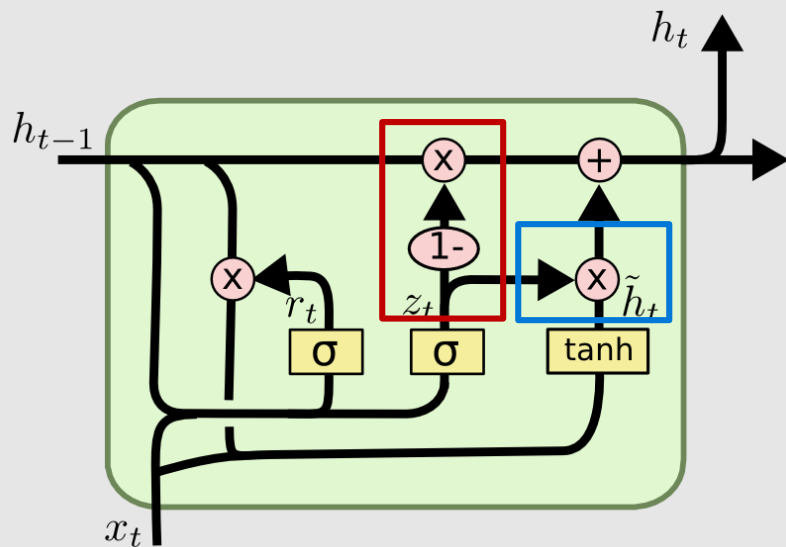
$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t])$$

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

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

GRU(Gated Recurrent Units)

- ✓ z_t : 반영 비율
- ✓ 이전 hidden state(h_{t-1})와 hidden state 후보(\tilde{h}_t)에 반영 비율을 적용하여
- ✓ 새로운 hidden state 값(h_t) 결정



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

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

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

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

더 공부해야 할 부분!

- ✓ Time Distributed : <https://yjjo.tistory.com/32>
- ✓ Bidirectional layer : <https://engineer-mole.tistory.com/26>
- ✓ Conv1D : <https://jiho-ml.com/weekly-nlp-9/>