

# PyTorch로 딥러닝 제대로 배우기

- 중급편 -

Part7. 순환 신경망(RNN) 이론

강사: 김 동 희

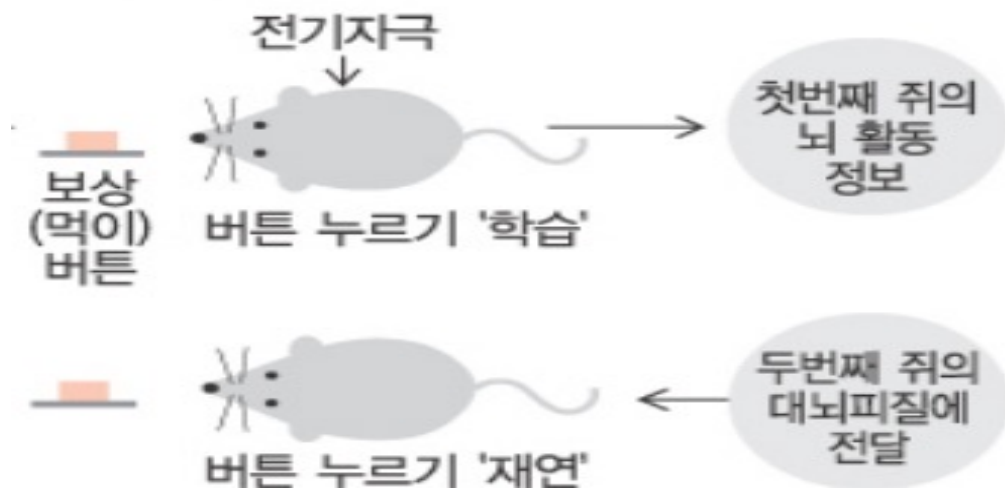


## II. LSTM(Long Short Term Memory, 1997)

## 1. LSTM

### □ 개요 - 쥐 전기자극 실험

- 짧은 시간 내의 기억을 떠올릴 때는 해마 쪽 신경세포 활발 → Hidden state
- 오래된 기억을 떠올릴 때는 대뇌피질 신경세포 활발 → Cell state

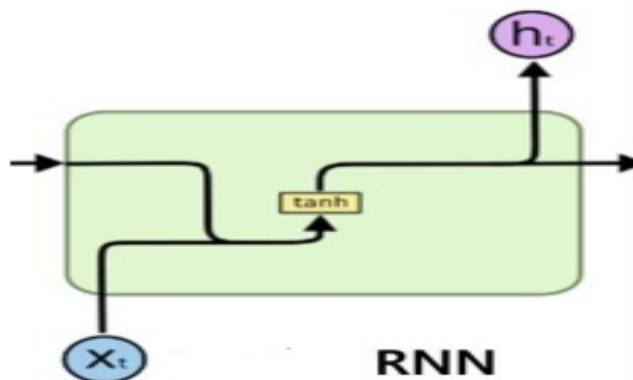


[그림] 쥐 전기자극 실험

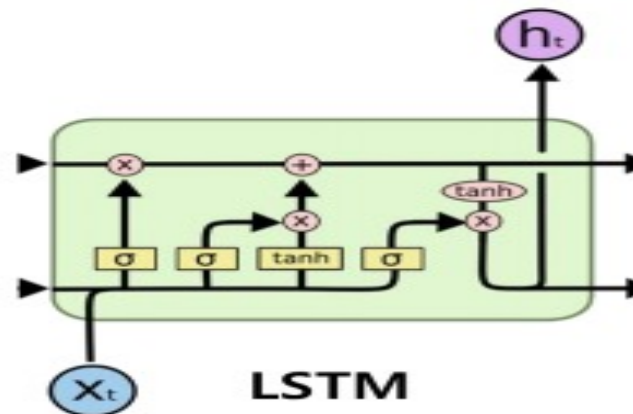
## II. LSTM (Long Short Term Memory)

### 1. LSTM

#### □ RNN vs. LSTM



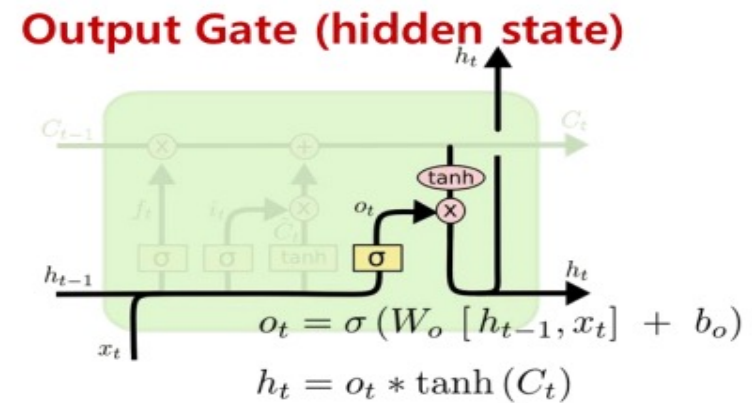
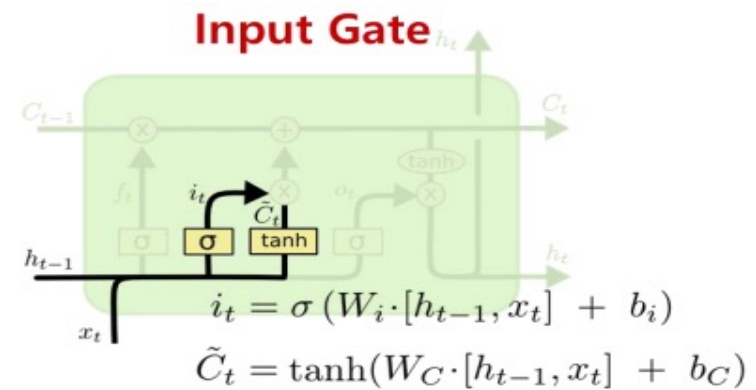
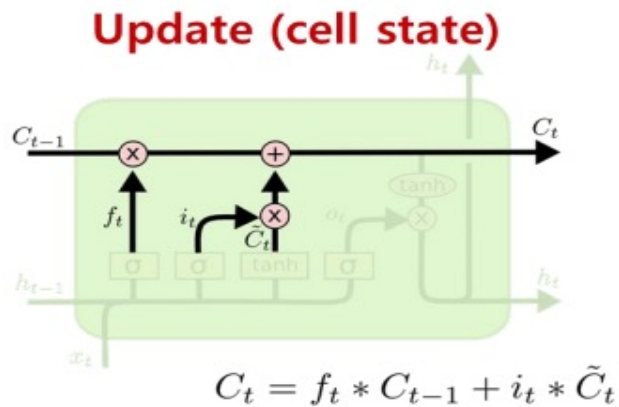
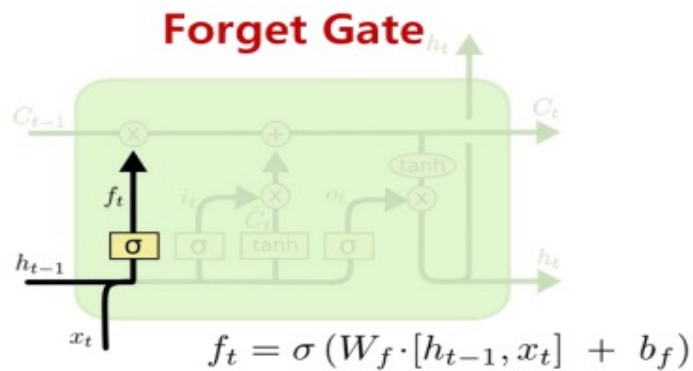
- Hidden Stage만으로 구성
- 장기 의존성 문제로 인해 긴 시퀀스에서 성능 저하



- 단기 기억을 위한 Hidden Stage와 장기 기억을 위한 Cell Stage로 구성
- Hidden Stage 내에서도 Forget, Input, Output Gate 각각 역할 수행  
⇒ 긴 시퀀스에서 활용 가능

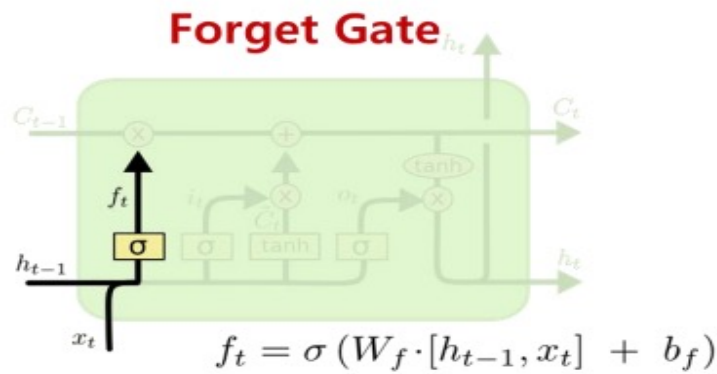
## 2. LSTM 구조

### □ Overview



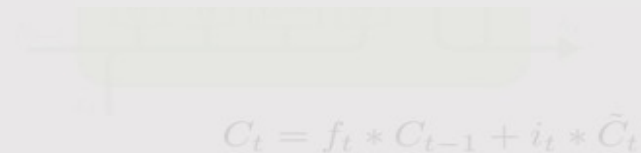
## 2. LSTM 구조

### ❑ Forget Gate



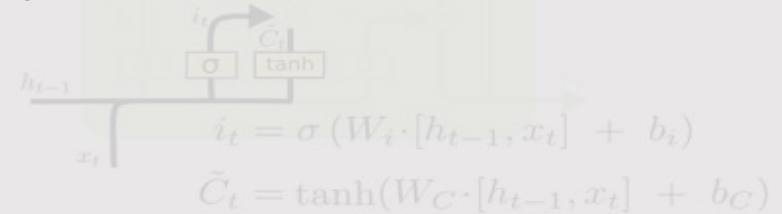
### Update (cell state)

이전 단기 기억( $h_{t-1}$ ) 과 현재 입력( $x_t$ ) 으로 부터  
이전 장기 기억( $C_{t-1}$ ) 중 **지워줄** 비율( $f_t$ ) 결정

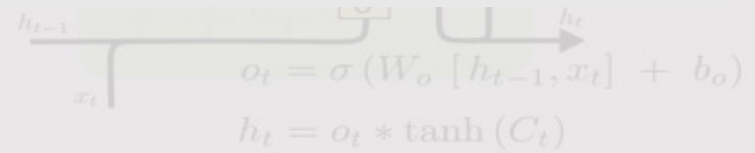


Cell Stage에 이를 곱해주어 **얼만큼 잊을지** 결정

### Input Gate



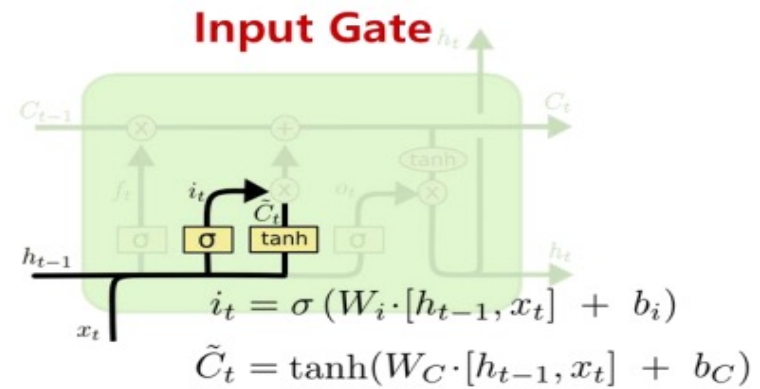
### Output Gate (hidden state)



## 2. LSTM 구조

### □ Input Gate

새롭게 추출한 특징을 얼마 만큼 사용할지 결정



이전 단기 기억( $h_{t-1}$ ) 과 현재 입력( $x_t$ ) 으로 부터  
현재 장기 기억후보 비율( $i_t$ ) 결정 및  
현재 장기 기억후보 ( $\tilde{C}_t$ ) 계산

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

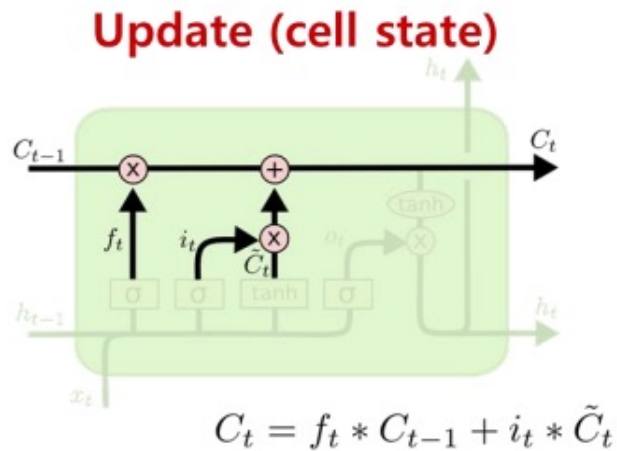
$$o_t = \sigma(W_o [h_{t-1}, x_t] + b_o)$$
$$h_t = o_t * \tanh(C_t)$$

## 2. LSTM 구조

### ❑ Cell state

- (1) 이전 장기 기억( $C_{t-1}$ )에  $f_t$ (forget gate) 곱하여 지워줄 정보 삭제
- (2) 현재 장기 기억후보( $\tilde{C}_t$ )에  $i_t$ (input gate) 곱하여 중요한 정보 추가

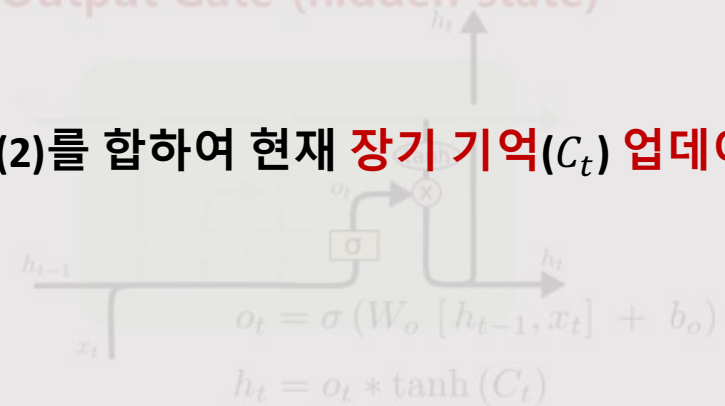
$$x_t \mid f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$



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

**Output Gate (hidden state)**

(1)과 (2)를 합하여 현재 **장기 기억( $C_t$ ) 업데이트**

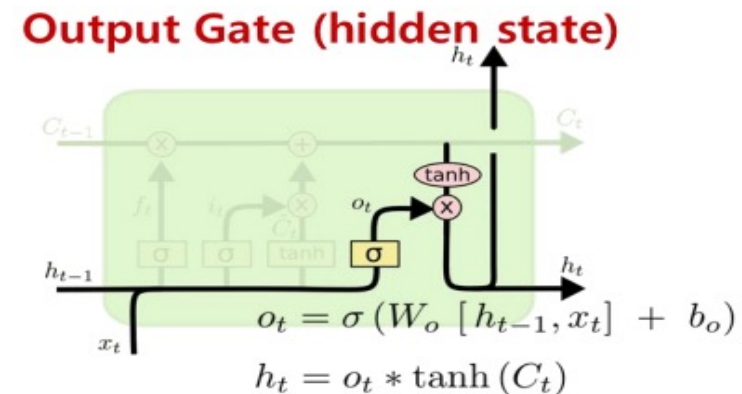
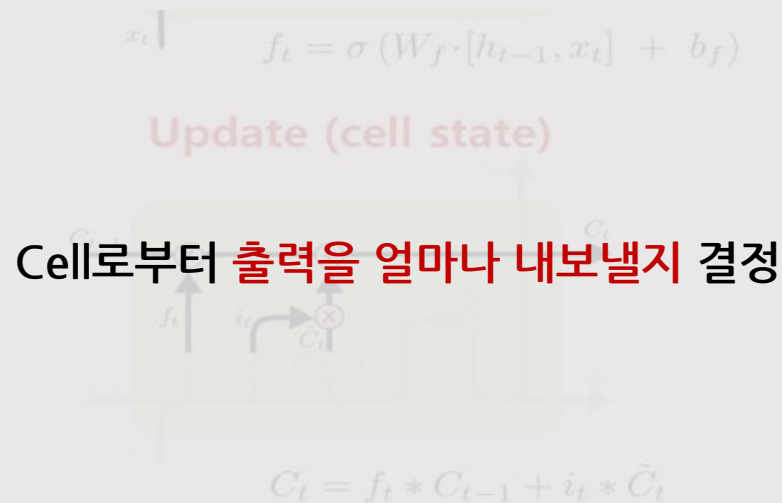




## 2. LSTM 구조

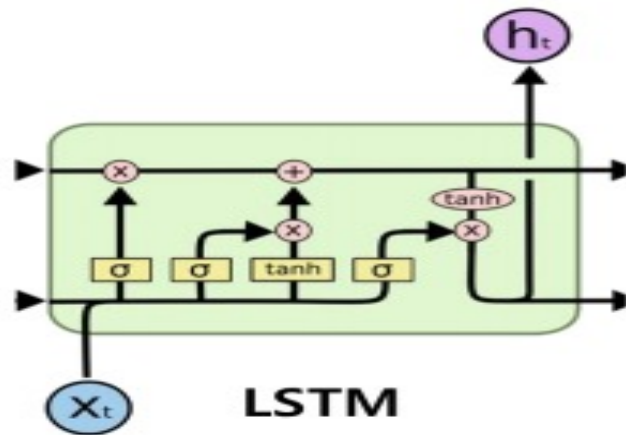
### □ Output Gate

이전 단기 기억( $h_{t-1}$ ) 과 현재 입력( $x_t$ ) 으로 부터  
출력을 통해 내보낼 비율( $o_t$ ) 결정 및  
현재 장기 기억( $C_t$ ) 에  $o_t$  를 곱하여 **현재 단기 기억( $h_t$ ) 업데이트**



### 3. LSTM 단점

#### □ LSTM 단점



- 단기기억을 위한 Hidden Stage와 장기기억을 위한 Cell Stage로 나누어짐
- Hidden Stage 내에서도 Forget, Input, Output Gate 각각 역할 수행
- 복잡한 구조로 학습에 시간  $\uparrow$

감사합니다.