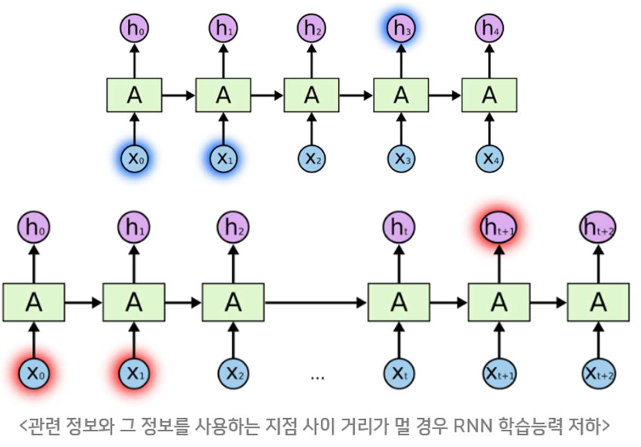
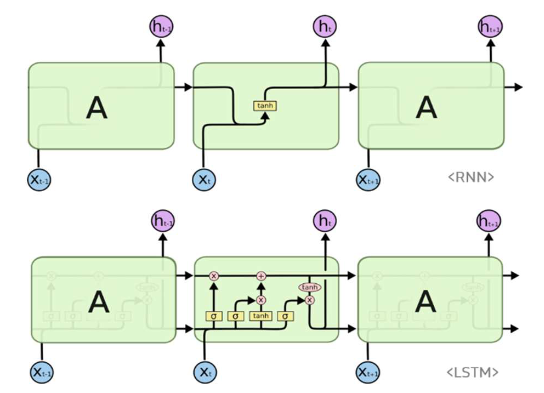
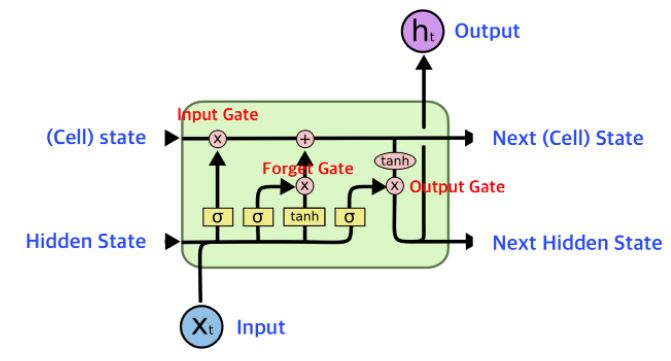
LSTM모델

RNN은 출력 결과가 이전의 계산 결과에 의존하여 시퀀스(Sequence)길이가 길어질 경우 앞의 정보가 뒤로 충분히 전달되지 못하는 Vanashing gradient problem이 발생한다.  
RNN의 단점인 장기 의존성 문제(The problem of Long-Term Dependencies)를 보완하여 장/단기 기억을 가능하게 설계한 신경망의 구조가 LSTM(Long Short Term Memory)이다.

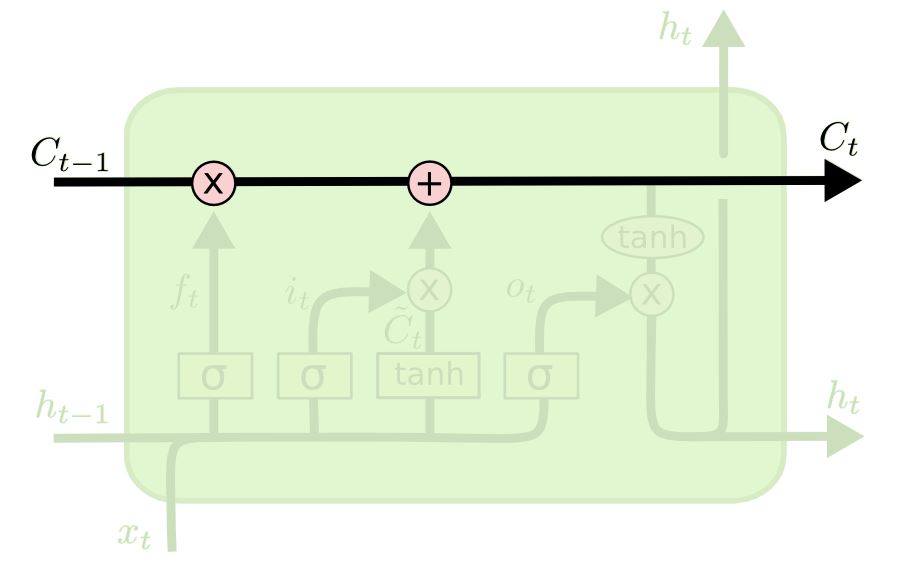
# 구조

LSTM은 RNN과 똑같은 체인구조를 가지고 있지만 4개의 Layer가 추가되어 정보를 주고 받도록 되어있다. ht = 단기상태, ct = 장기상태

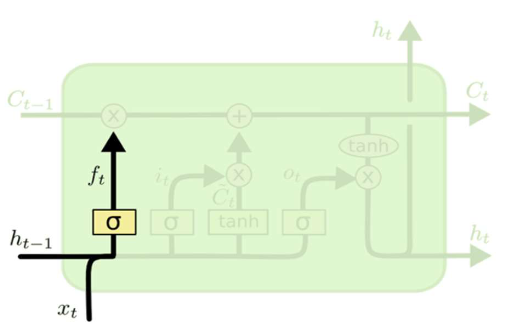


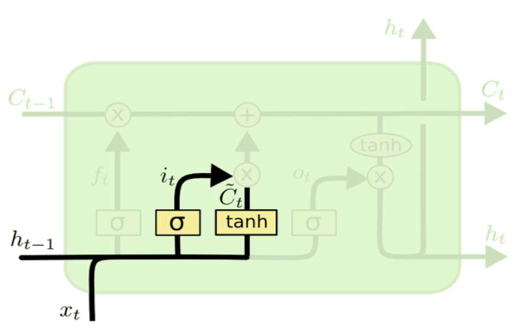
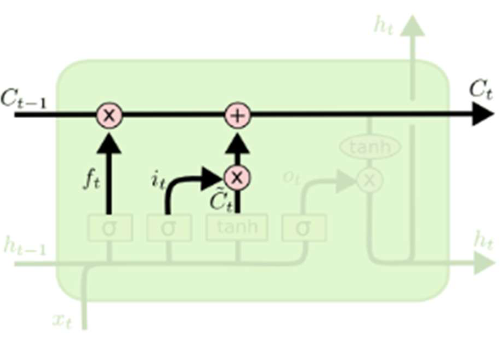
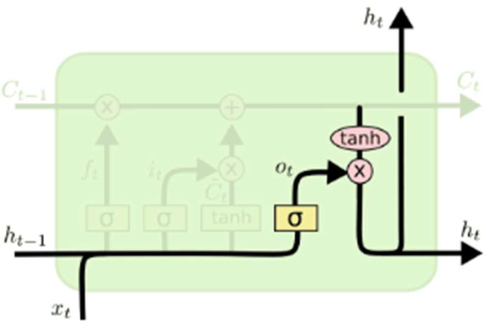


* **Cell state**정보가 바뀌지 않고 그대로 흐르도록 하는 역할이다. Gate를 통해서 정보가 추가되거나 제거되며,   
  Gate들은 Training을 통해 어떤 정보를 유지하고 버릴지 학습된다.



* **Forget gate**  
  업데이트 할 과거 정보량을 정하는 역할이다. Sigmoid layer를 거쳐 어떤 정보를 버릴지 정한다.  
  값이 0이라면 이전 상태의 정보는 잊고, 1이라면 이전 상태의 정보를 모두 기억한다.



* **Input gate**  
  현재정보 중 어떤 것을 cell state에 저장할지 정한다. Sigmoid layer를 거쳐 업데이트할 값을 정한 후   
  than layer에서 새로운 Vector를 만든다.
* **Cell state update**  
  이전Gate(Forget, Input)에서 정한 버릴 정보와 업데이트할 정보를 업데이트한다.
* **Output gate**  
  어떤 정보를 output으로 보낼 지 정한다. Sigmoid layer에서 output정보를 정한 후 tanh layer에서  
  sigmoid layer의 output과 곱하여 output으로 출력한다.