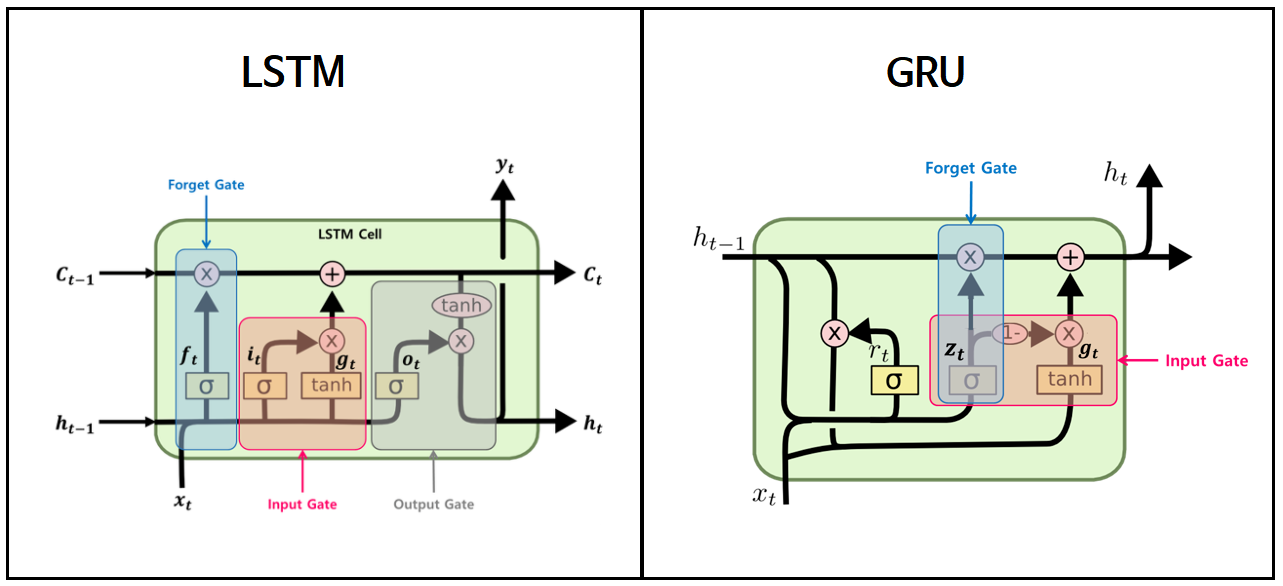
GRU모델

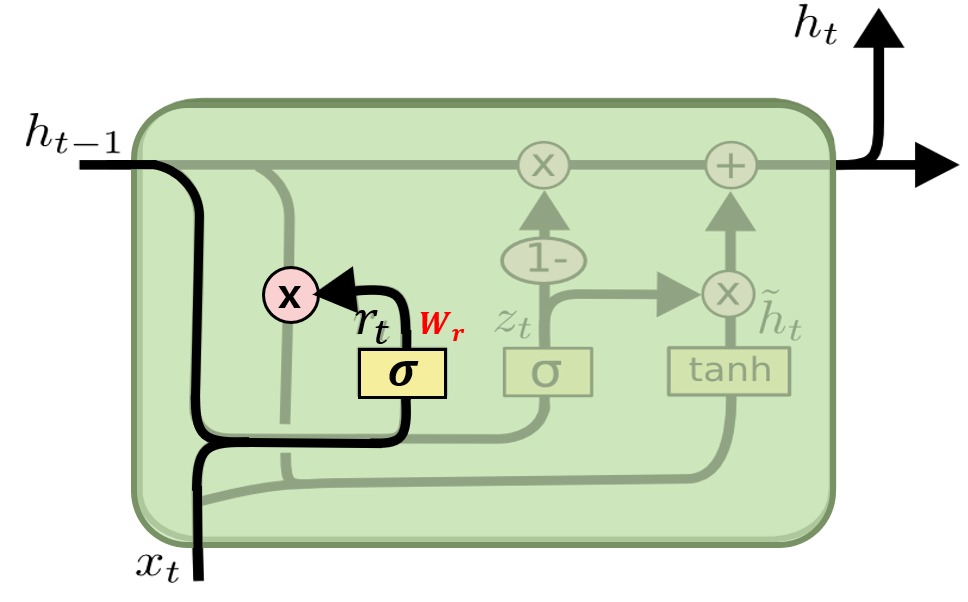
LSTM은 RNN의 치명적인 한계점이었던 장기의존성 문제를 해결하면서 긴 sequence를 가진  
데이터에서도 좋은 성능을 내는 모델이 되었지만, 복잡한 구조 때문에 RNN에 비하여 많은 파라미터가   
필요하게 되었다. 파라미터가 많아지는데 데이터가 충분하지 않은 경우 오버피팅이 발생하는데  
이를 방지하고자 LSTM의 변형인 GRU가 등장하게 되었다.

# 구조

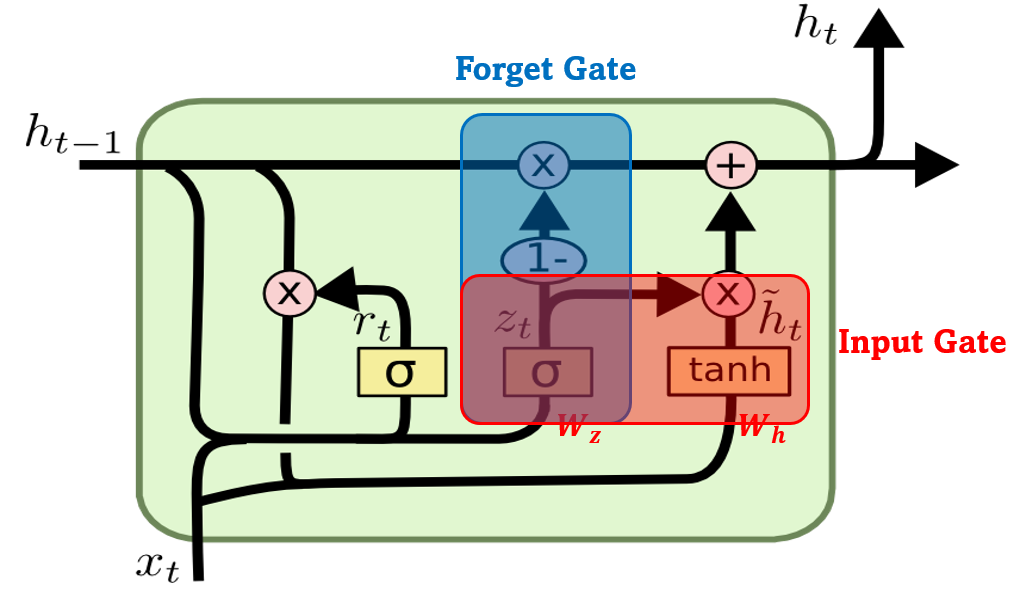
LSTM의 cell state와 hidden state가 GRU에서는 하나의 벡터로 합쳐졌다  
또한 forget gate, input gate는 update gate로 통합되었다.  
LSTM에서는 forget과 input이 서로 독립적이었으나, GRU에서는 전체 양이 정해져 있어   
forget한만큼 input하는 방식으로 조절된다.



* **Reset gate**  
  새로운 입력 정보를 이전 메모리와 어떻게 합칠지 정하는 역할이다. 이전 hidden state에서 얼만큼 값을  
  반영할지를 정한다. LSTM과 달리 output gate가 없어서 전체 상태 벡터인 hidden state가 매 timestamps  
  마다 출력되며, 직전 hidden state의 어느 부분을 출력될지를 정하게 된다.



* **Update gate**  
  과거, 현재 정보를 각각 얼마나 반영할지 비율을 구하는 단계이다.  
  LSTM의 forge, input gate와 동일한 역할을 수행한다.



GRU는 LSTM에 비해 상당히 간단한 구조를 가지고 있으며 성능면에서는 LSTM보더 더 우월하다고 할 수는 없지만 학습할 파라미터가 훨씬 적다는 특징을 가지고 있다.