CHAPTER 6
순환 신경망

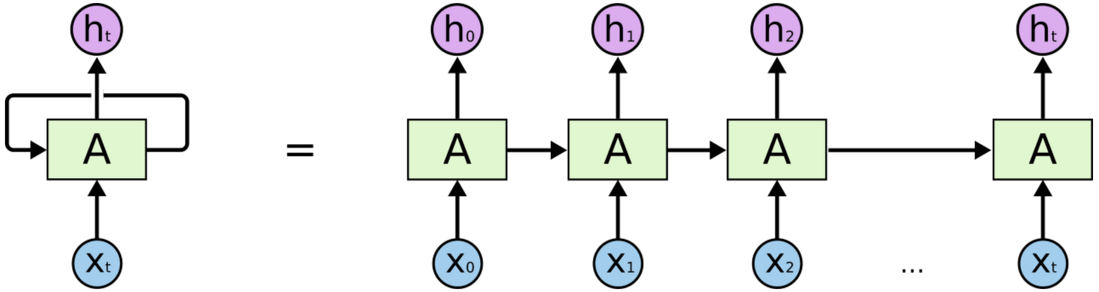
6.1 순환 신경망의 발달 과정

- Sequence data : 순서가 존재하는 데이터
- Time-series data : 시간에 따른 의미가 존재하는 데이터

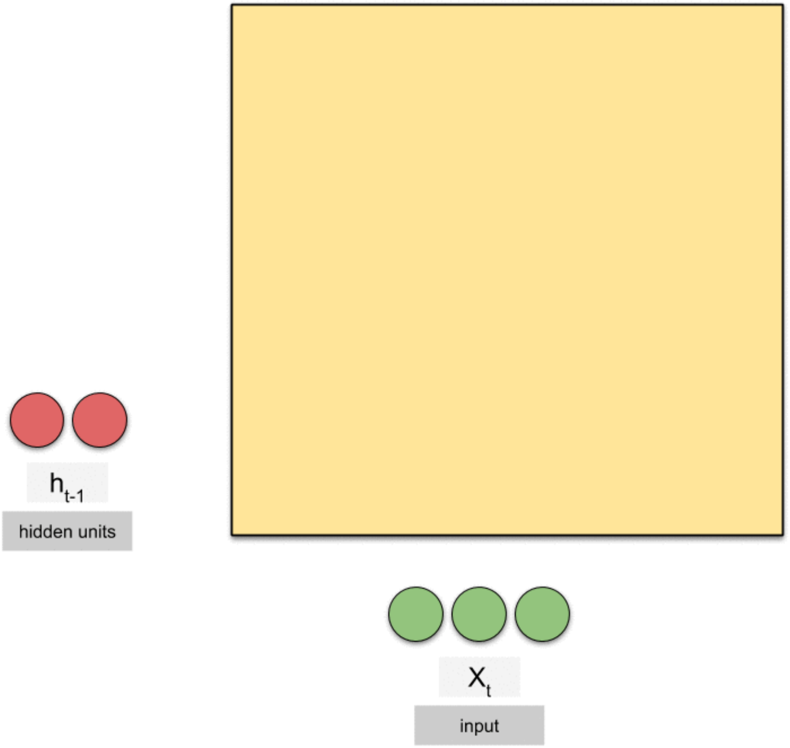
- 순환 신경망
  = 순서가 있는 데이터에서 의미를 찾아내기 위해 고안된 모델

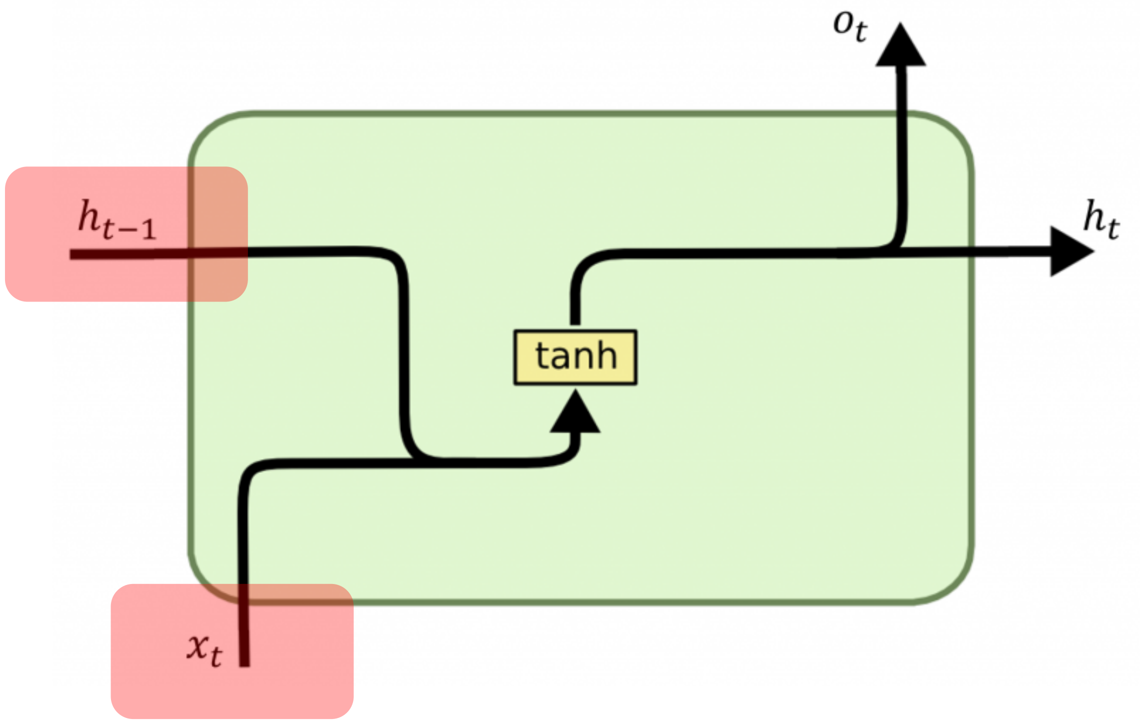
6.2 순환 신경망의 동작 원리

- 각 노드에 순환성을 추가
- 이전 시간(h0)의 출력값을 다음 노드의 입력값(x1)을 concat하여 출력(h1)

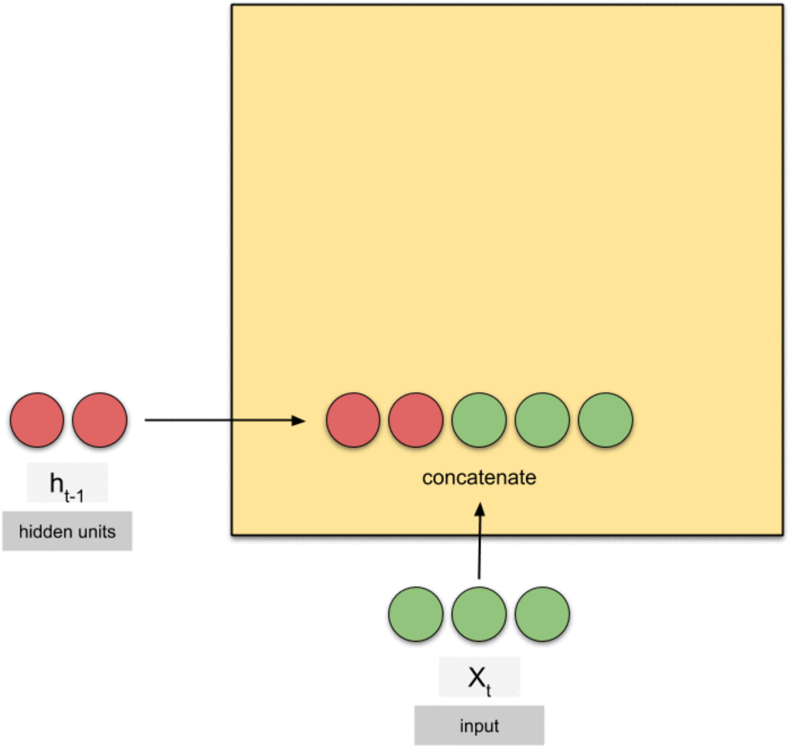


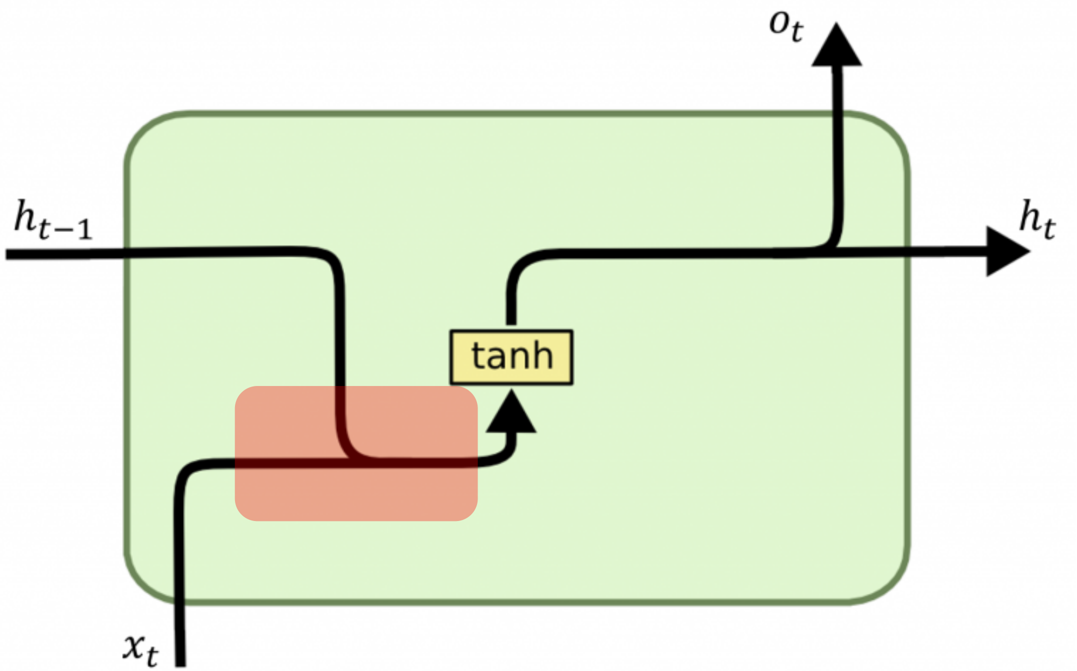
6.2 순환 신경망의 동작 원리





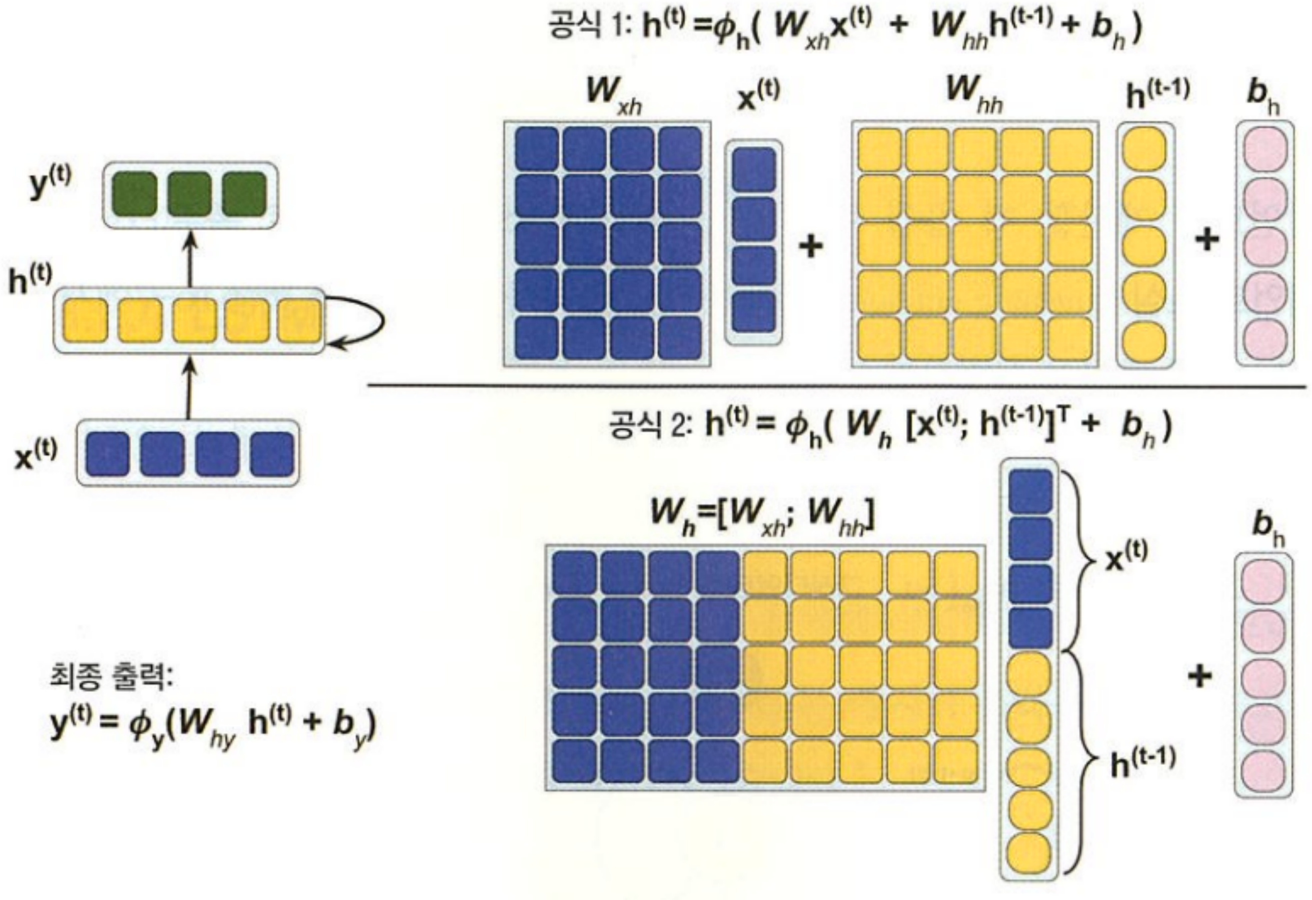
6.2 순환 신경망의 발달 과정



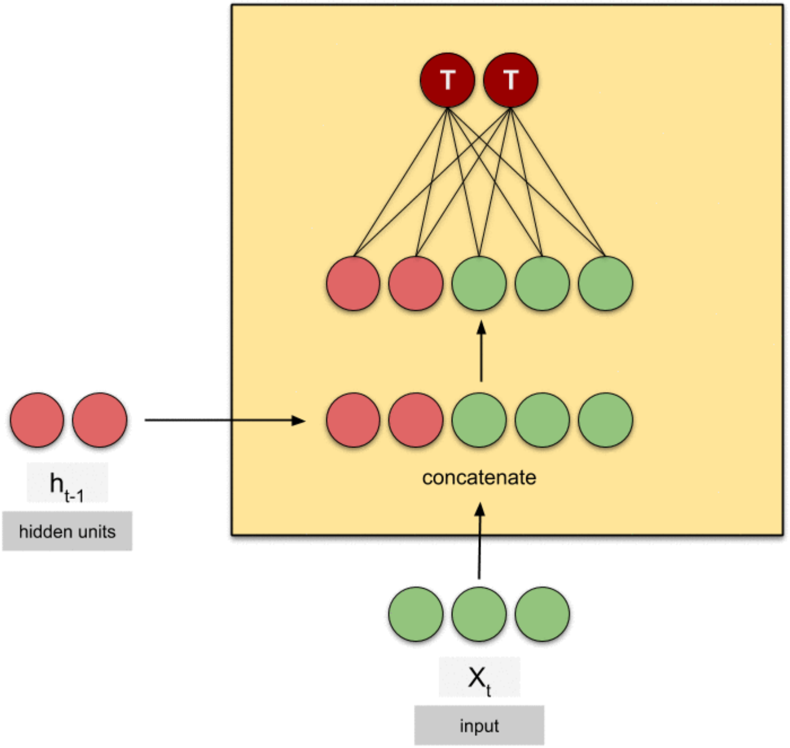


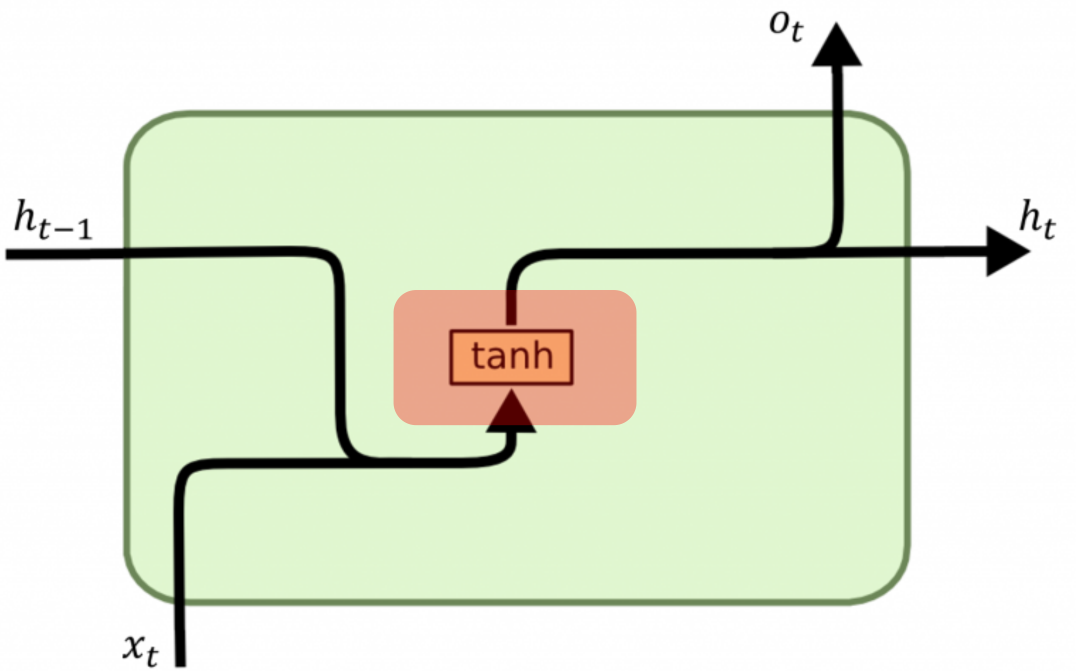
6.2 순환 신경망의 발달 과정

concat의 의미 =>

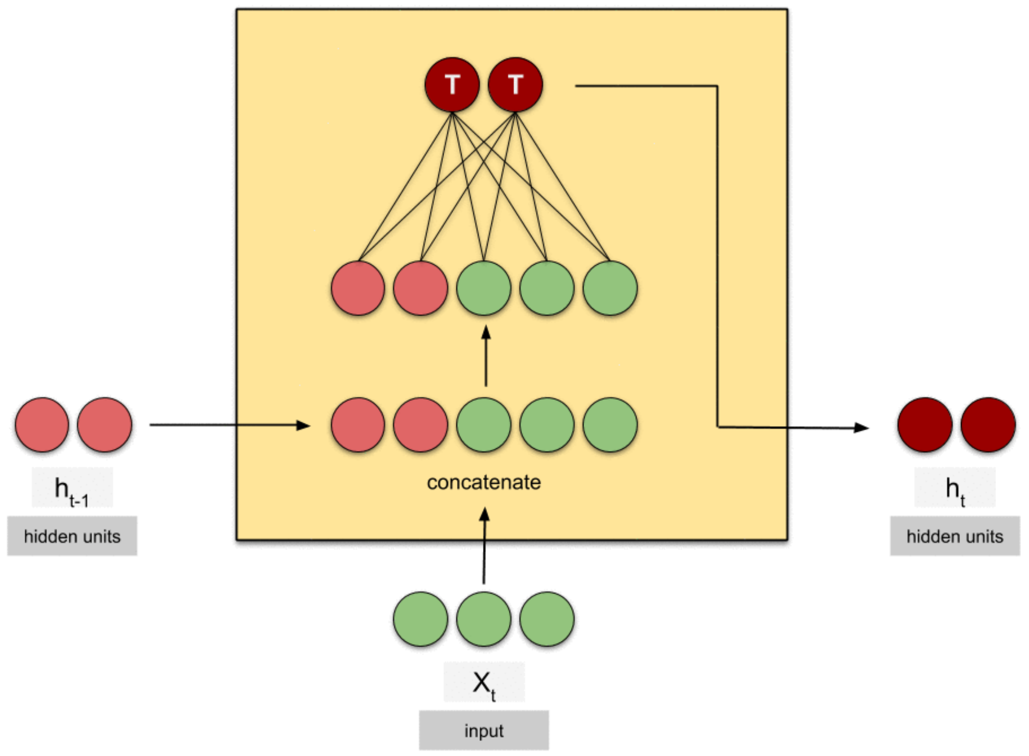


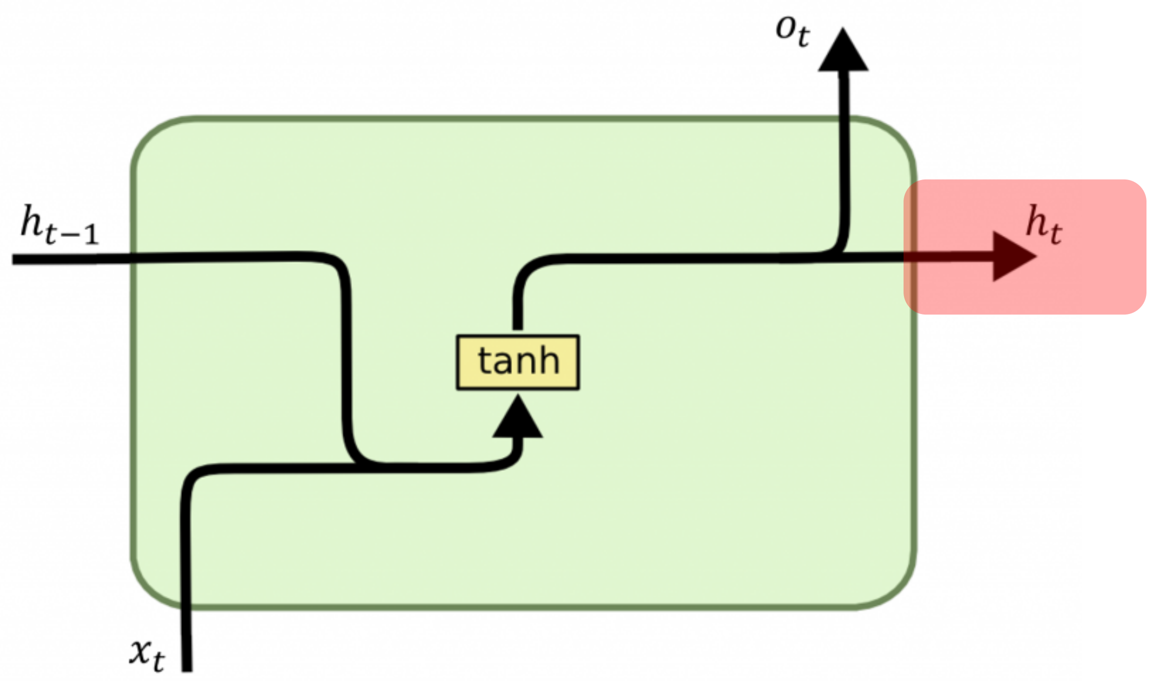
6.2 순환 신경망의 발달 과정



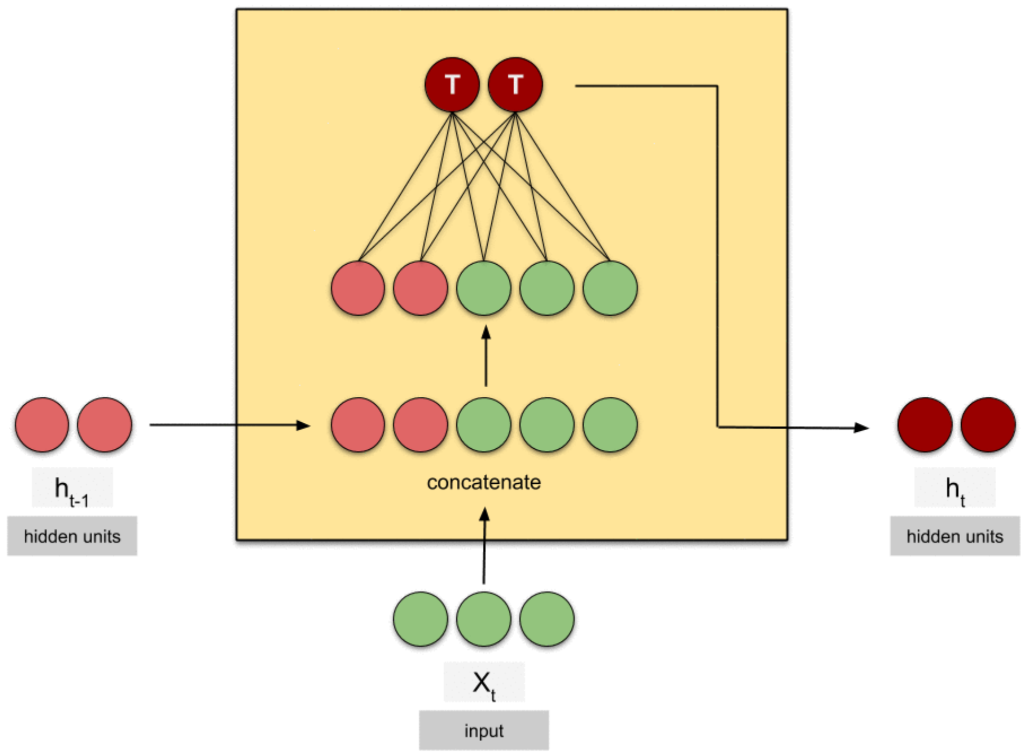


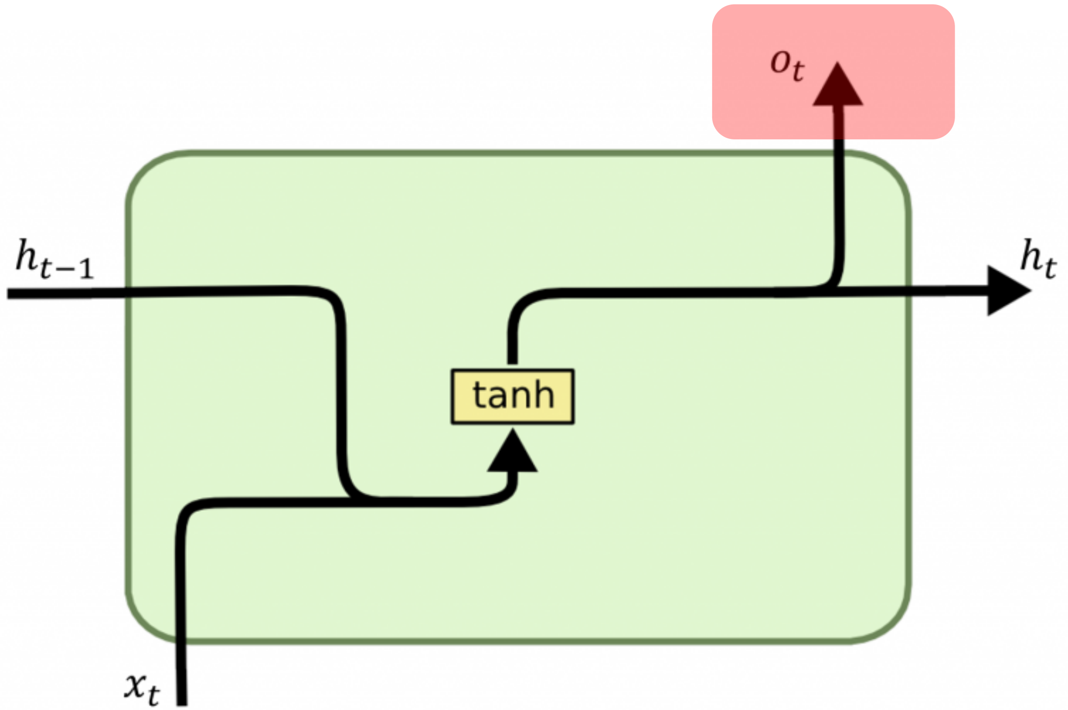
6.2 순환 신경망의 발달 과정





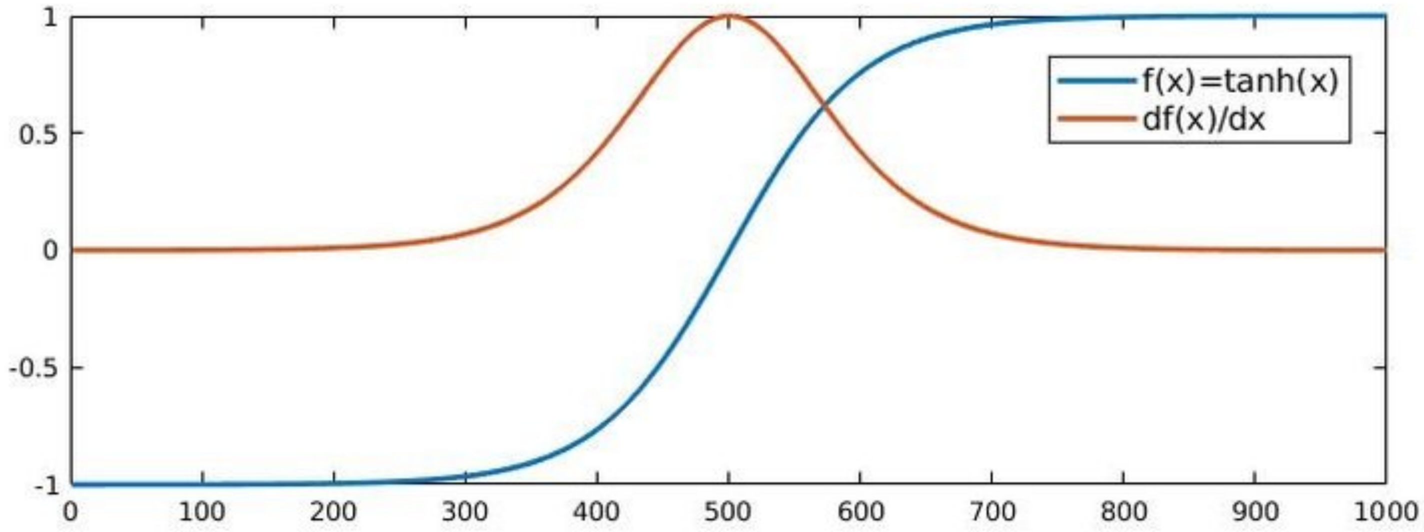
6.2 순환 신경망의 발달 과정





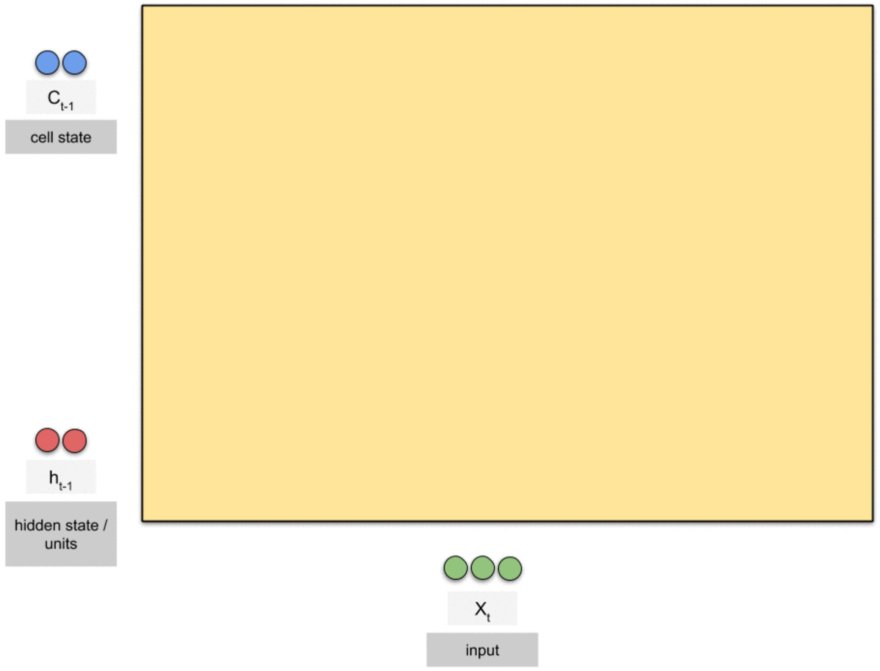
6.4 순환 신경망의 한계 및 개선 방안

- 역전파 시 하이퍼볼릭 탄젠트 함수의 미분값이 -1에서 1사이의 값이 나와서, 타임 시퀀스가 길어질 수록 제대로 학습이 이루어지지 않는 기울기 소실
- 기울기 소실 문제를 해결하기 위해 LSTM이 등장



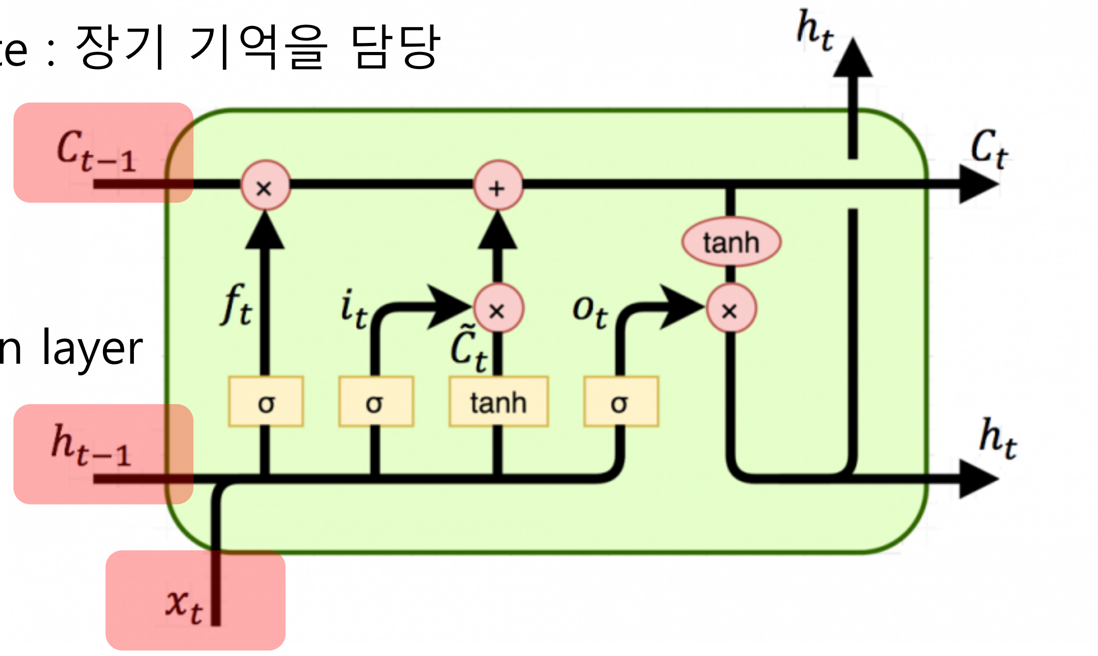
6.4 순환 신경망의 한계 및 개선 방안 : LSTM

LSTM : 기존의 순환 신경망 모델에 장기기억을 담당하는 cell state를 추가한 모델
Input gate, forget gate, output gate로 출력값을 조절



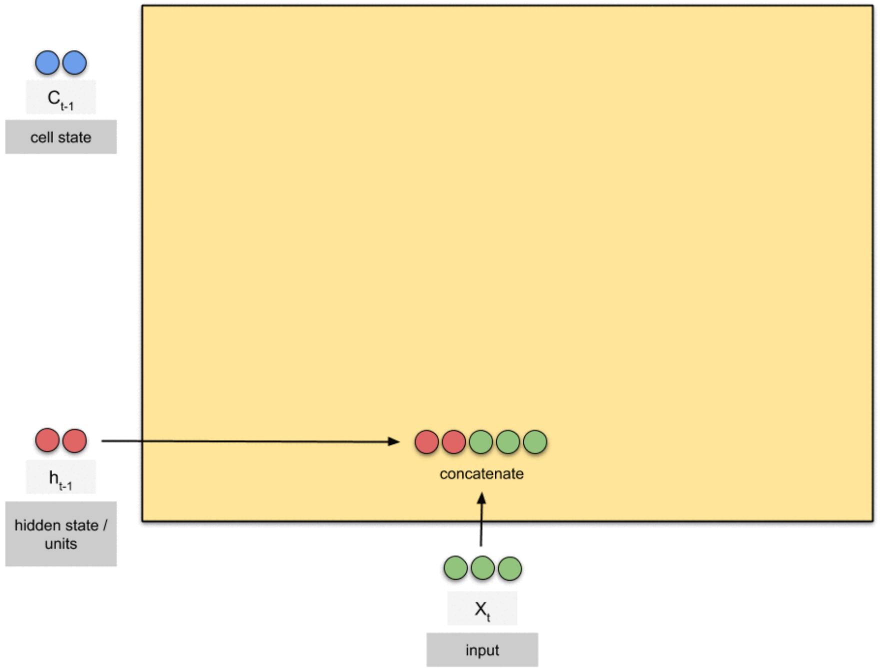
cell state : 장기 기억을 담당

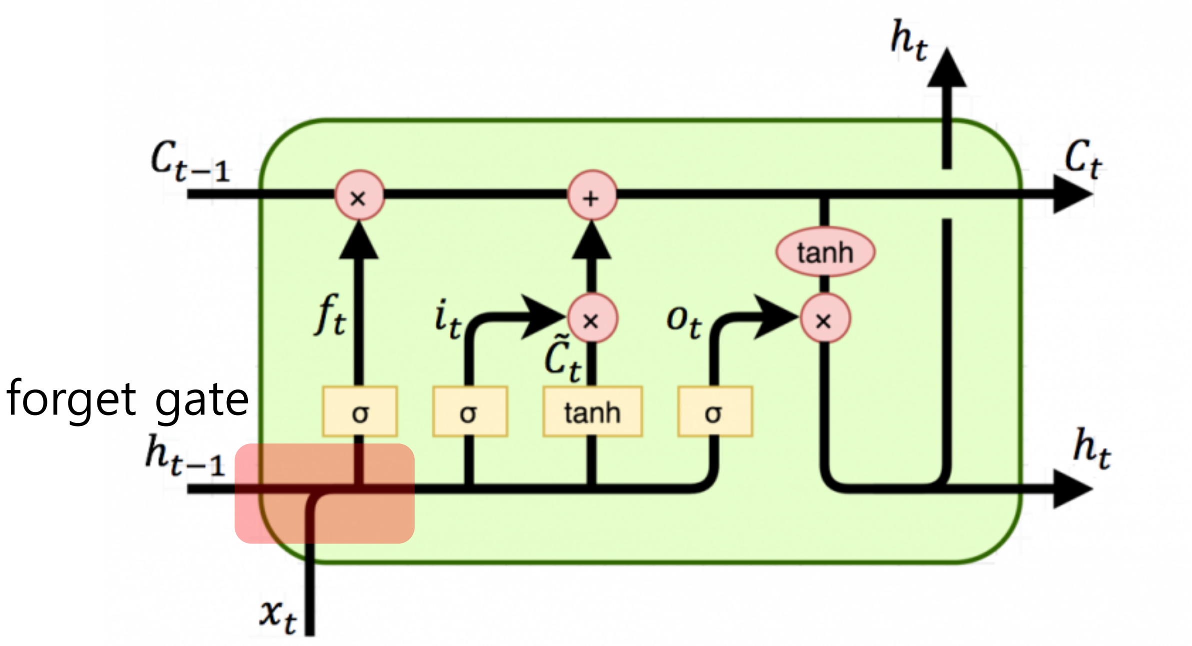
이전
hidden layer



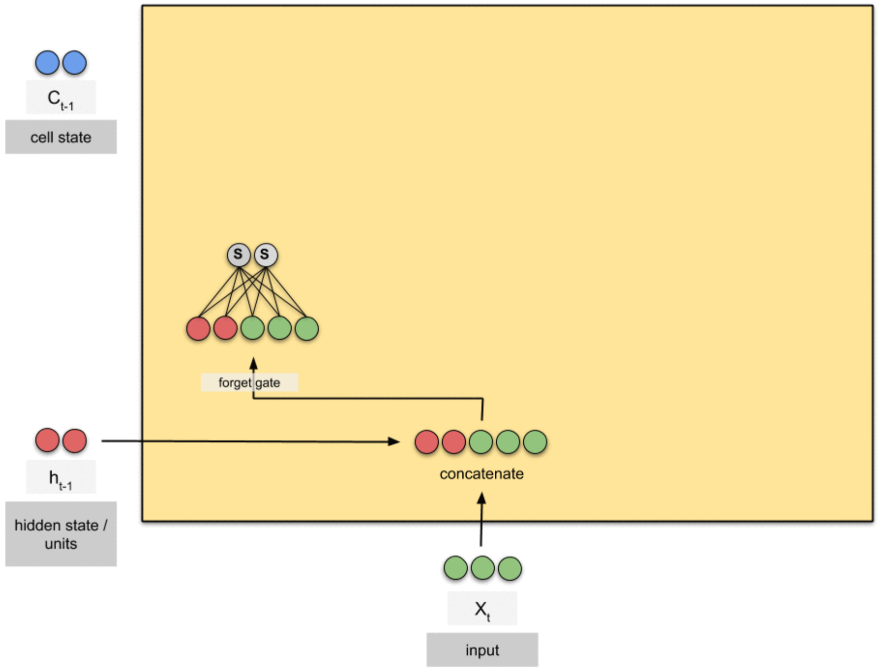
입력값

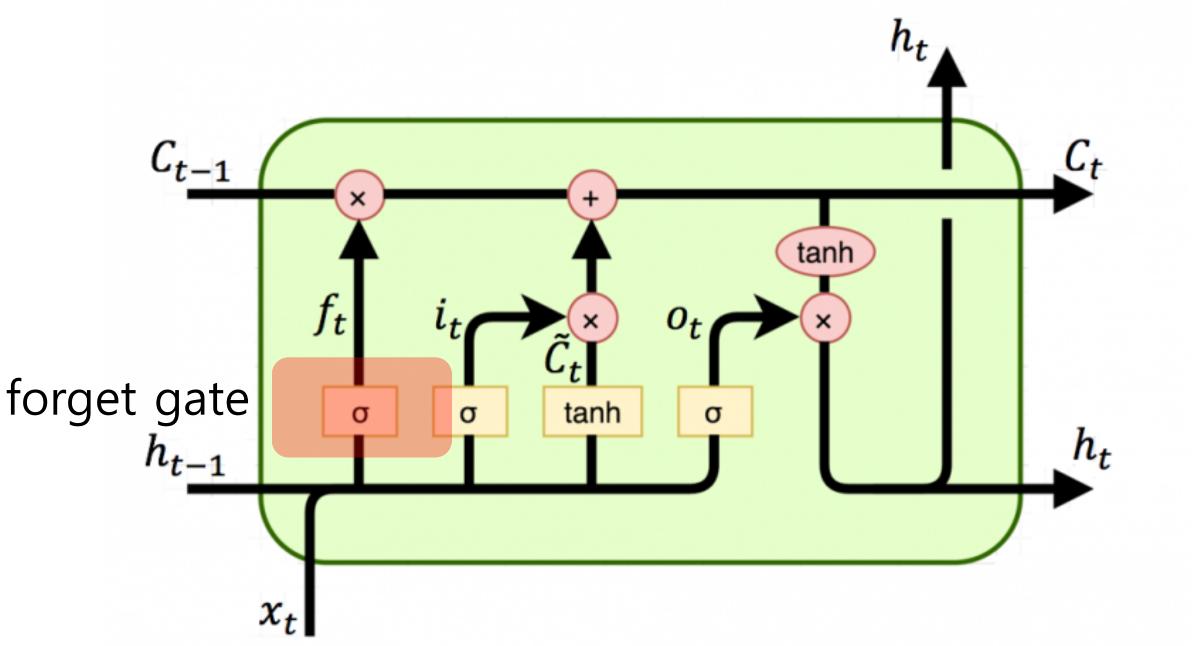
6.4 순환 신경망의 한계 및 개선 방안 : LSTM





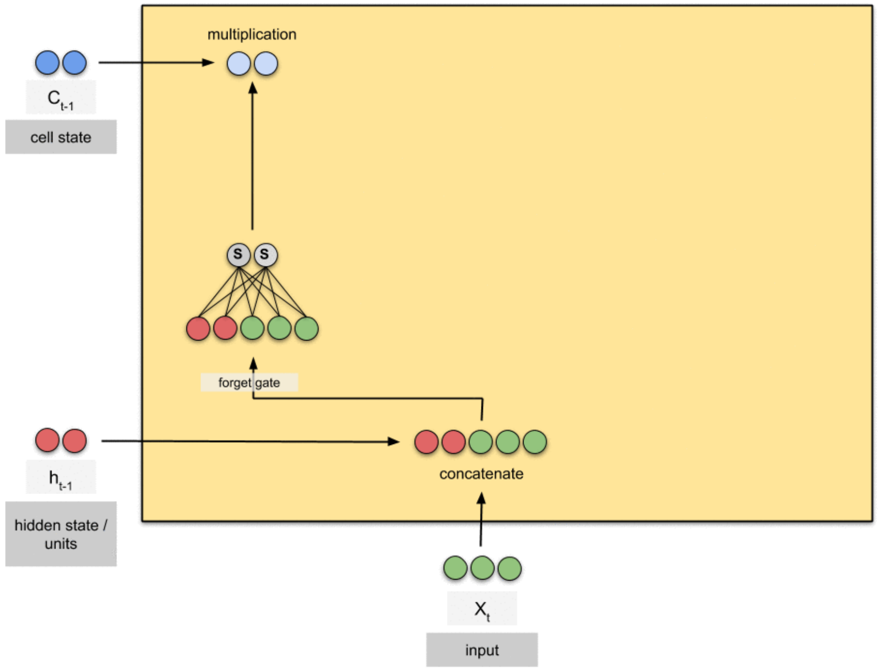
6.4 순환 신경망의 한계 및 개선 방안 : LSTM



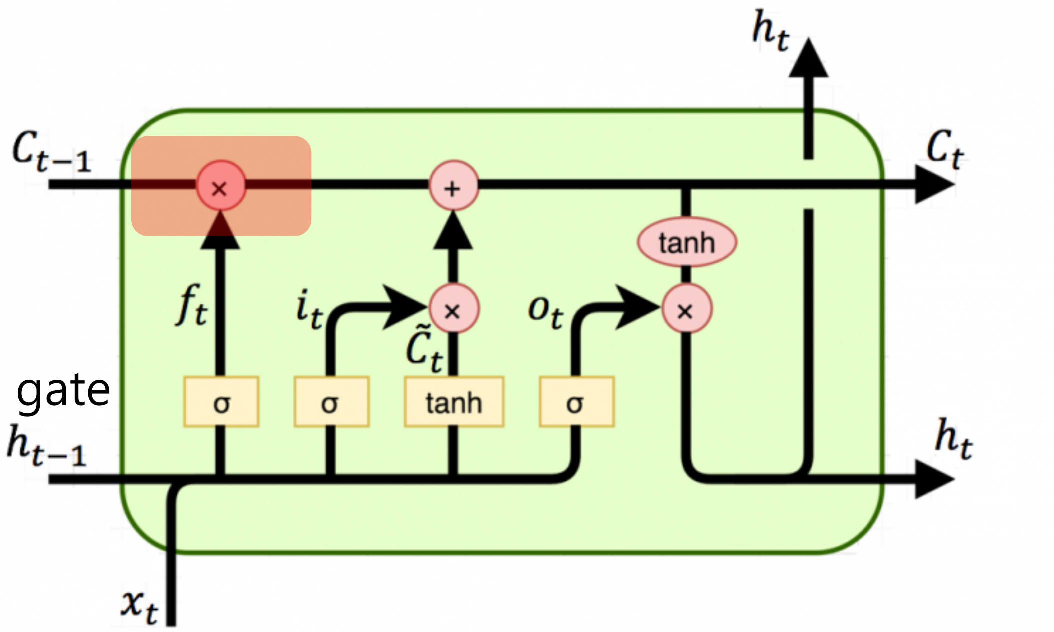


6.4 순환 신경망의 한계 및 개선 방안 : LSTM

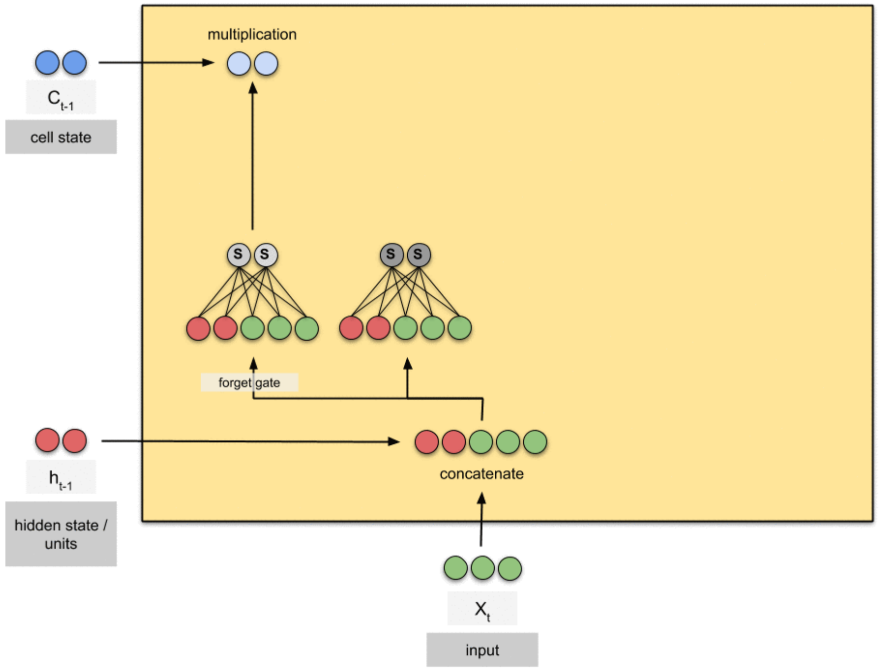
forget gate : sigmoid 함수의 0과 1 사이의 값으로 기존의 정보를 얼마나 전달할 지 결정

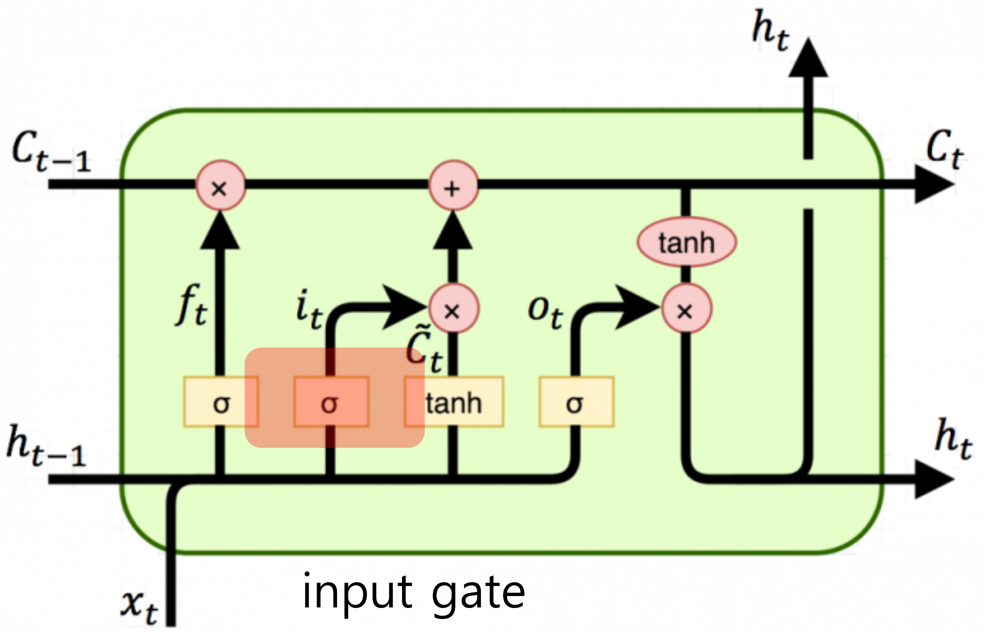


forget gate



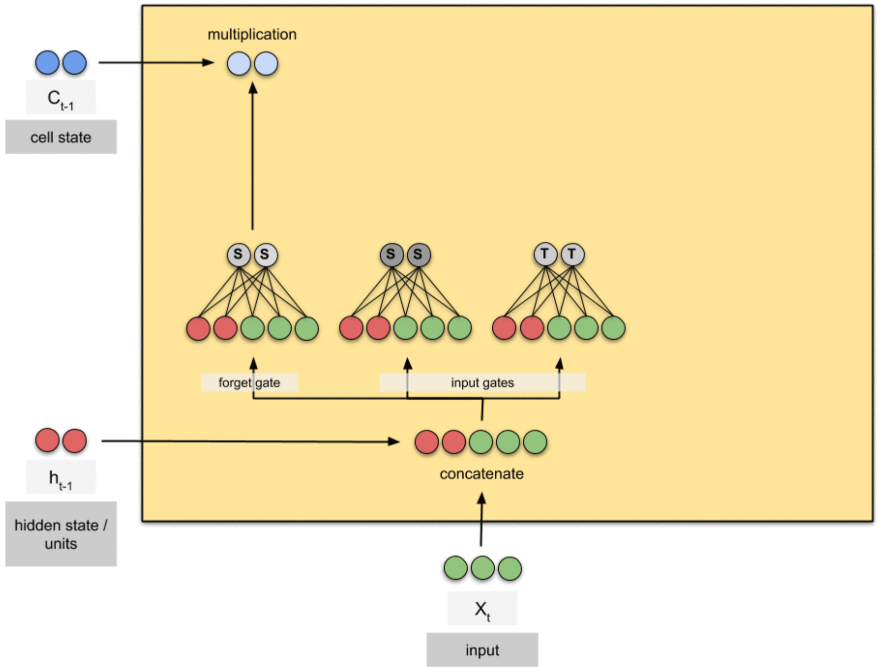
6.4 순환 신경망의 한계 및 개선 방안 : LSTM

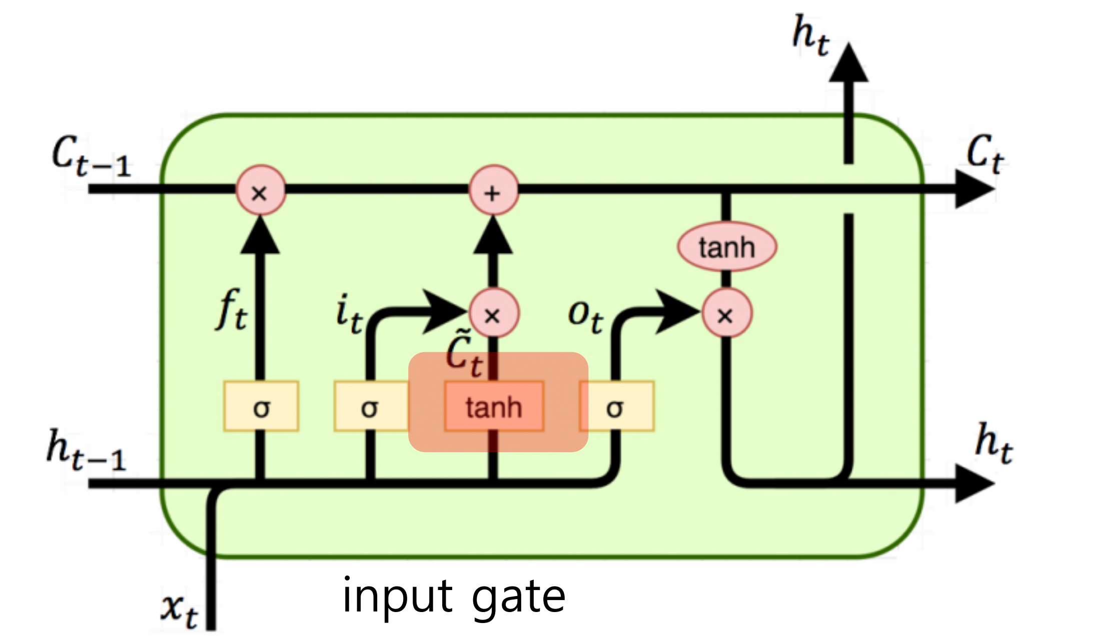




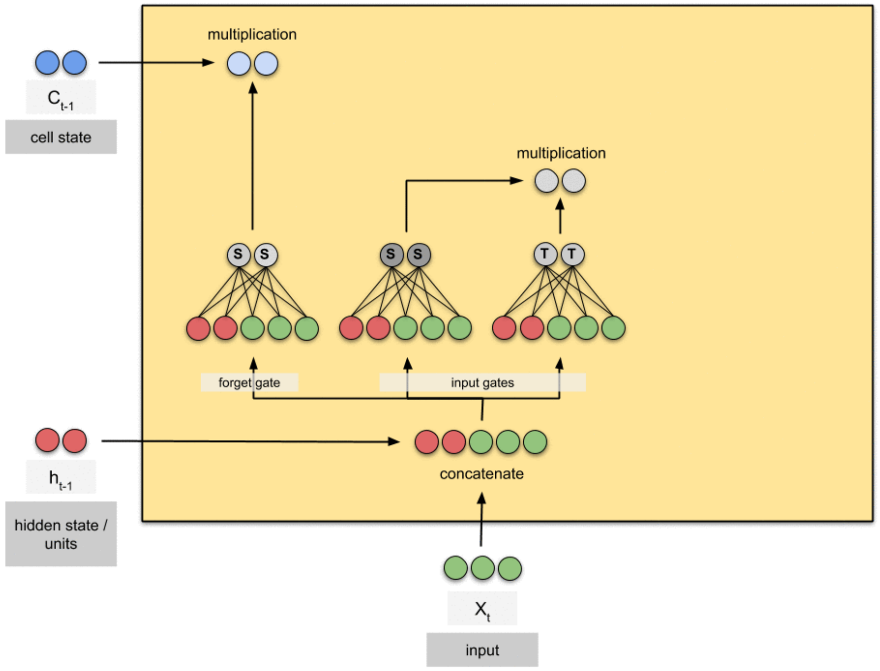
6.4 순환 신경망의 한계 및 개선 방안 : LSTM

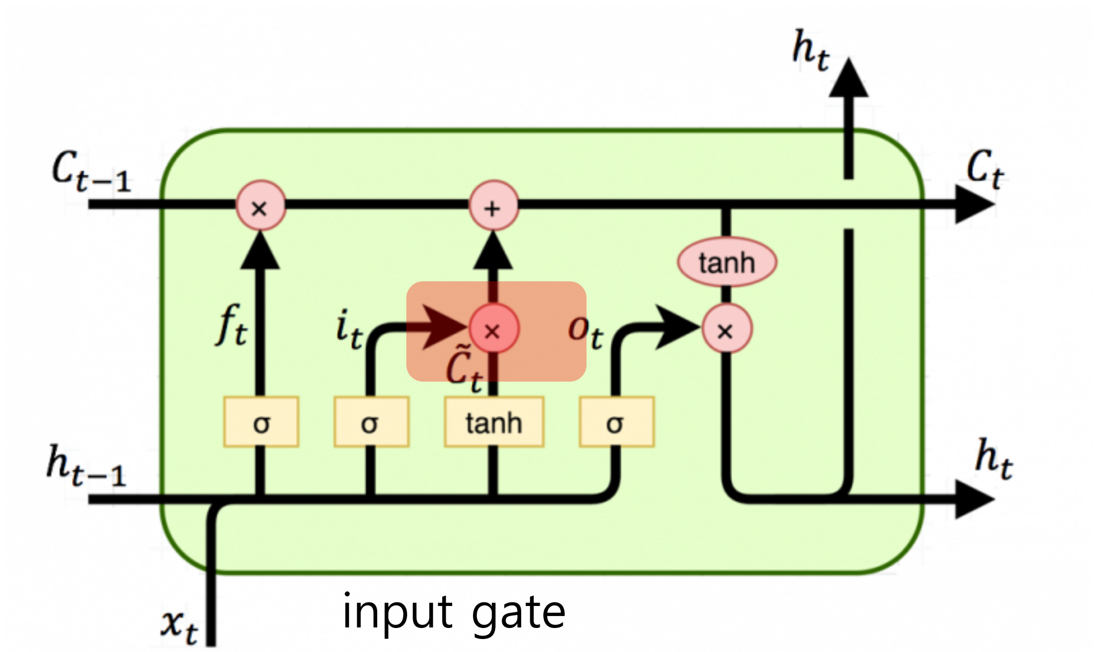
tanh : -1과 1 사이의 값을 가지고 새롭게 셀 상태에 추가할 정보
sigmoid : 0과 1 사이의 비중으로 새롭게 추가할 정보를 얼마만큼 셀 상태에 더할지 결정



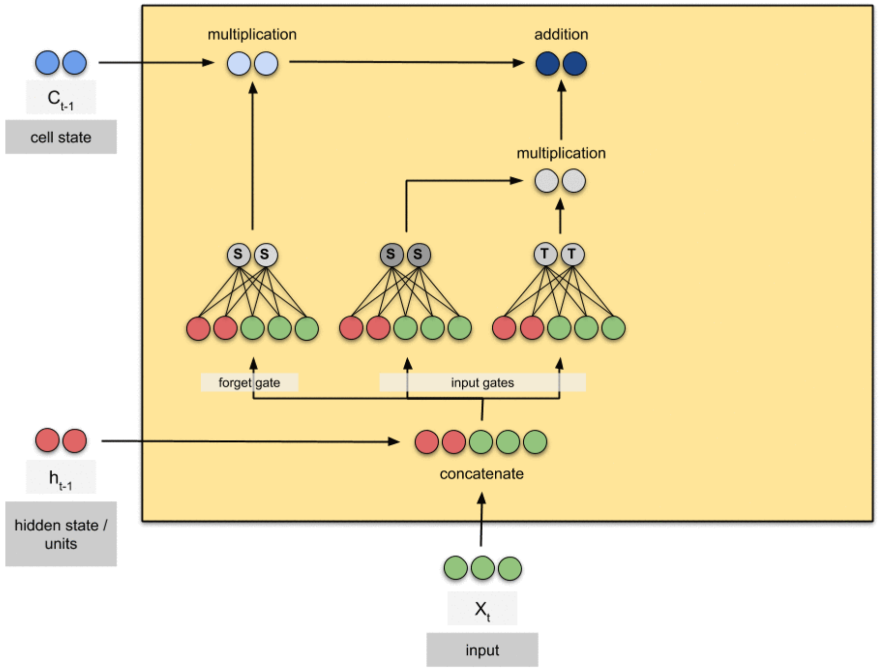


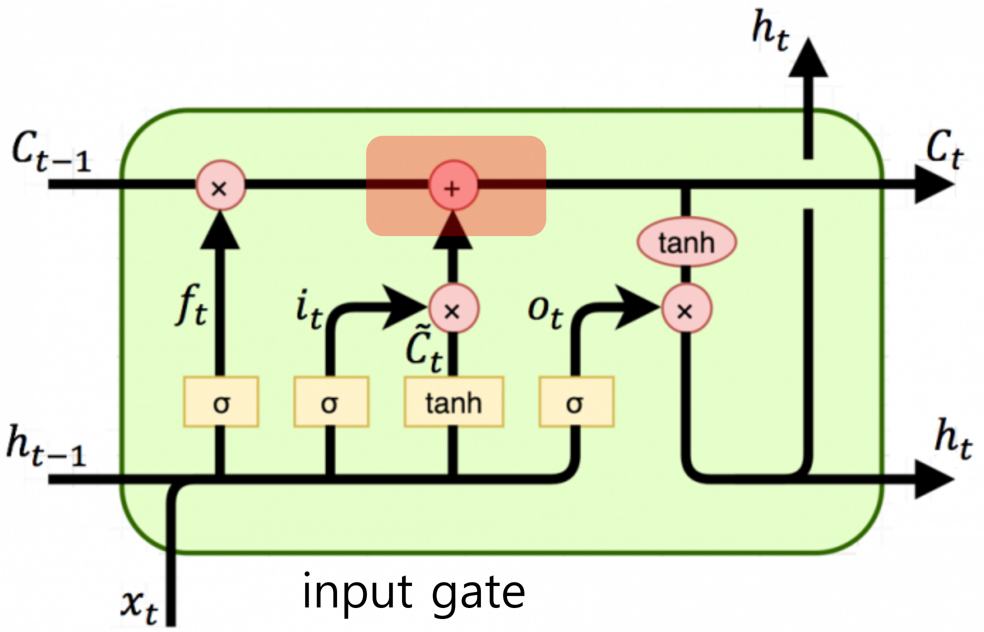
6.4 순환 신경망의 한계 및 개선 방안 : LSTM





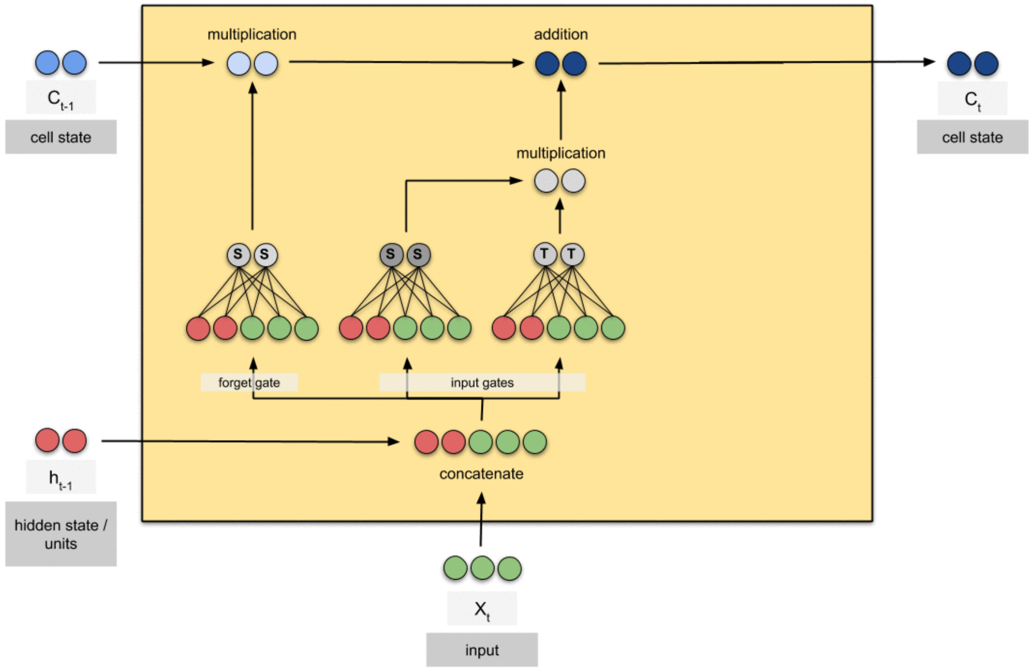
6.4 순환 신경망의 한계 및 개선 방안 : LSTM

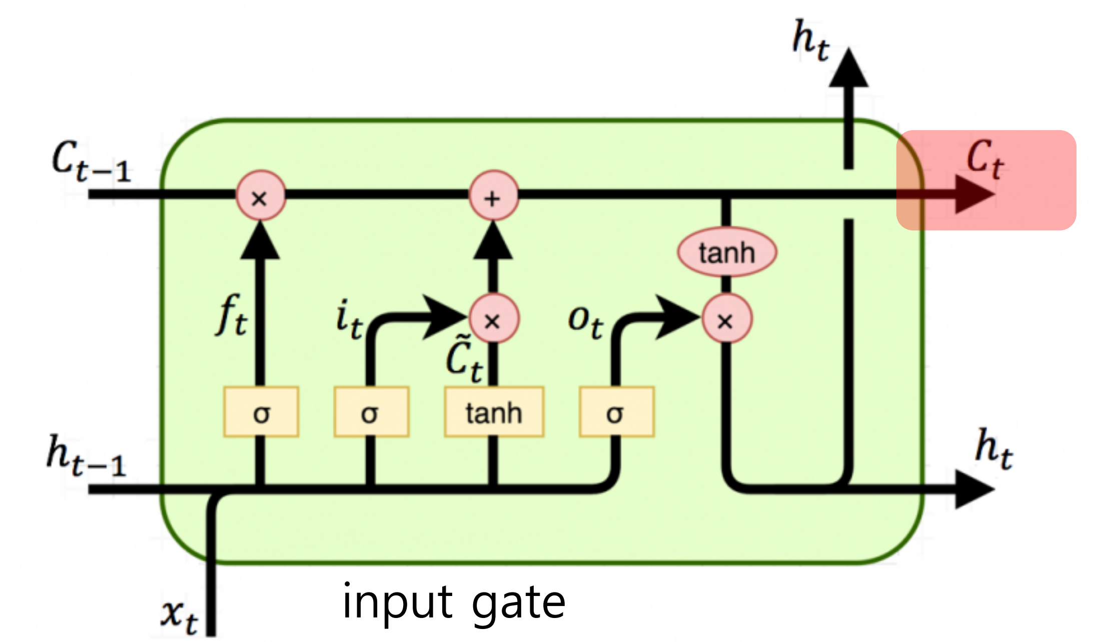




input gate

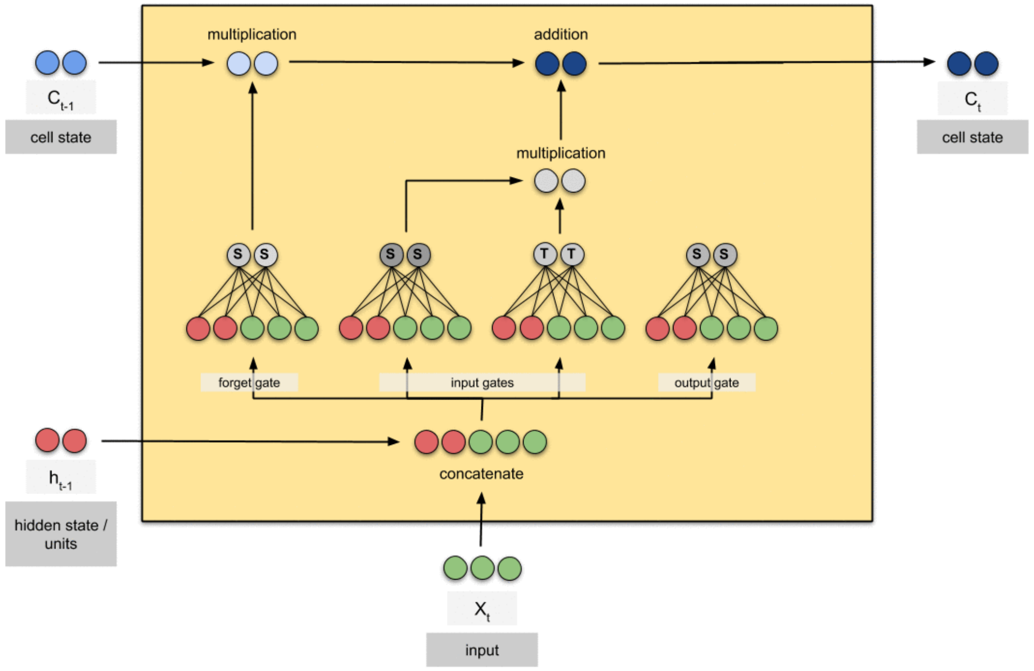
6.4 순환 신경망의 한계 및 개선 방안 : LSTM

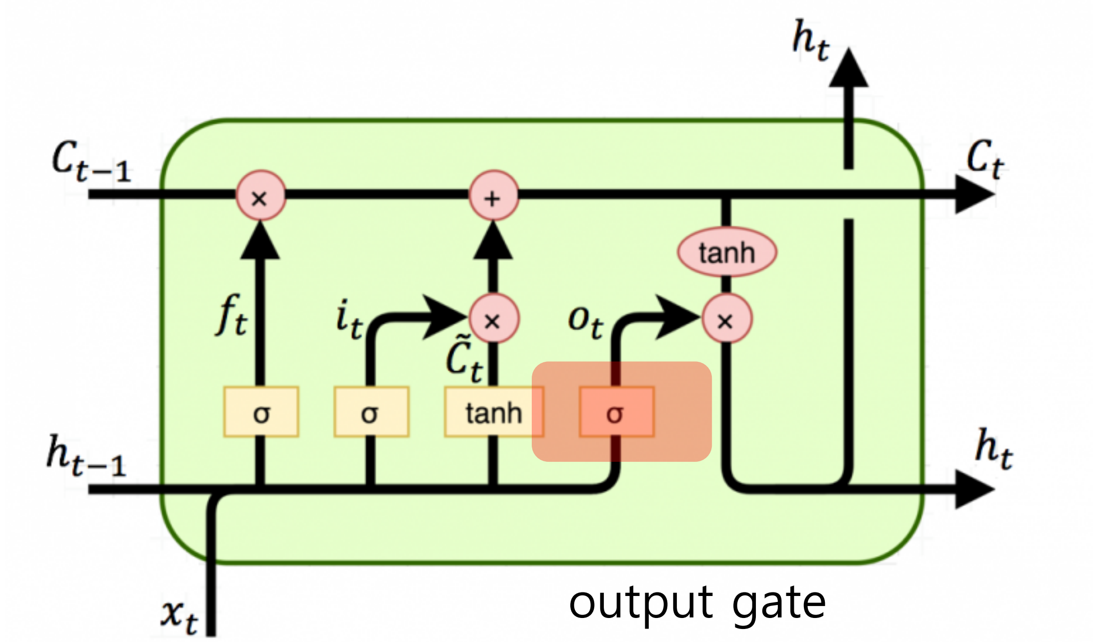




input gate

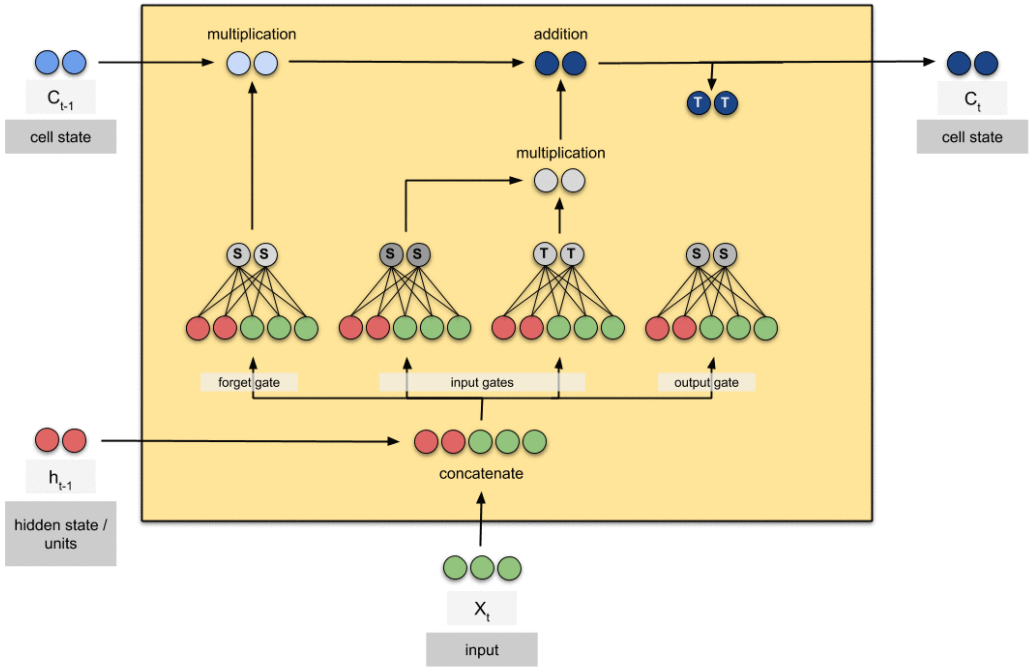
6.4 순환 신경망의 한계 및 개선 방안 : LSTM

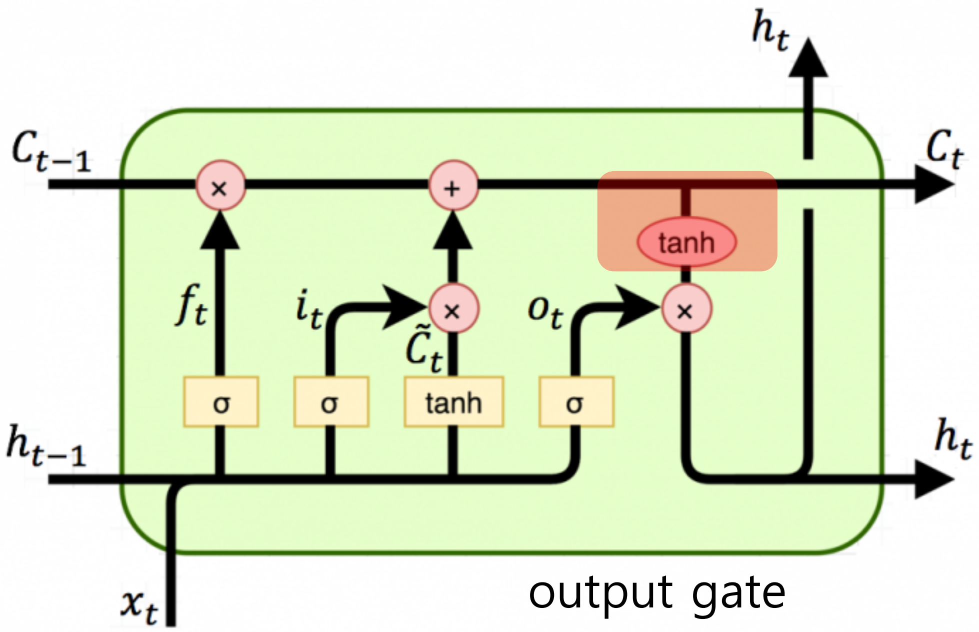




output gate

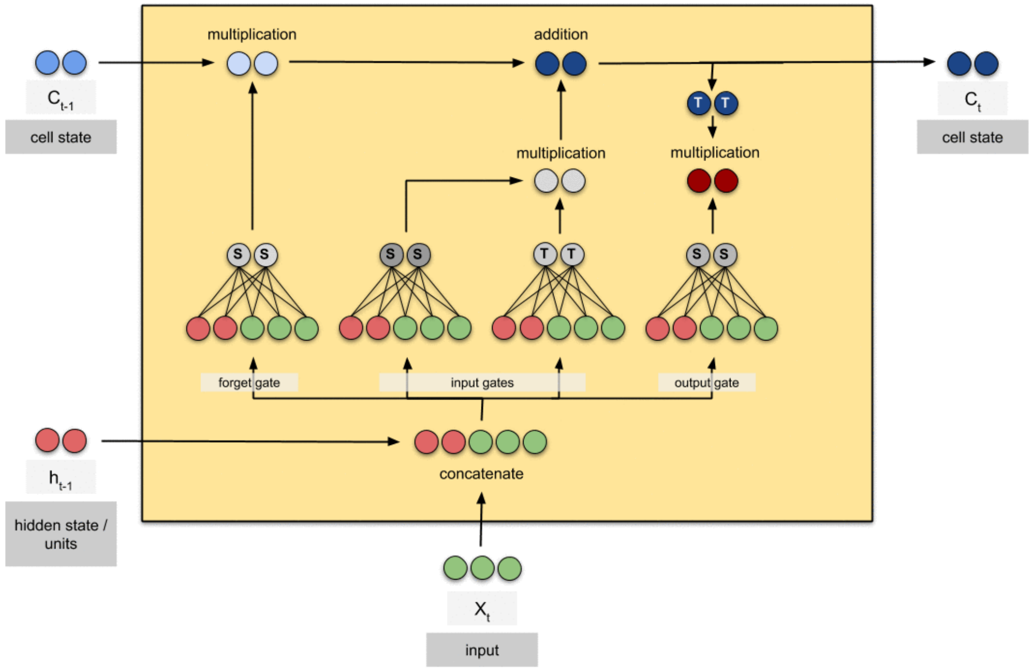
6.4 순환 신경망의 한계 및 개선 방안 : LSTM

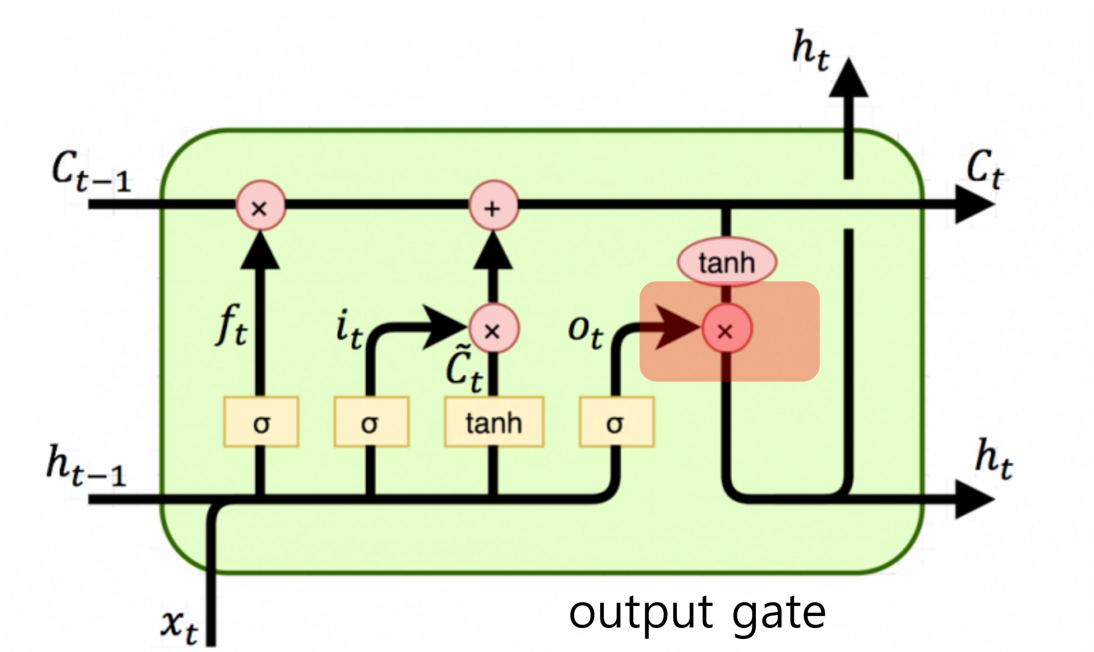




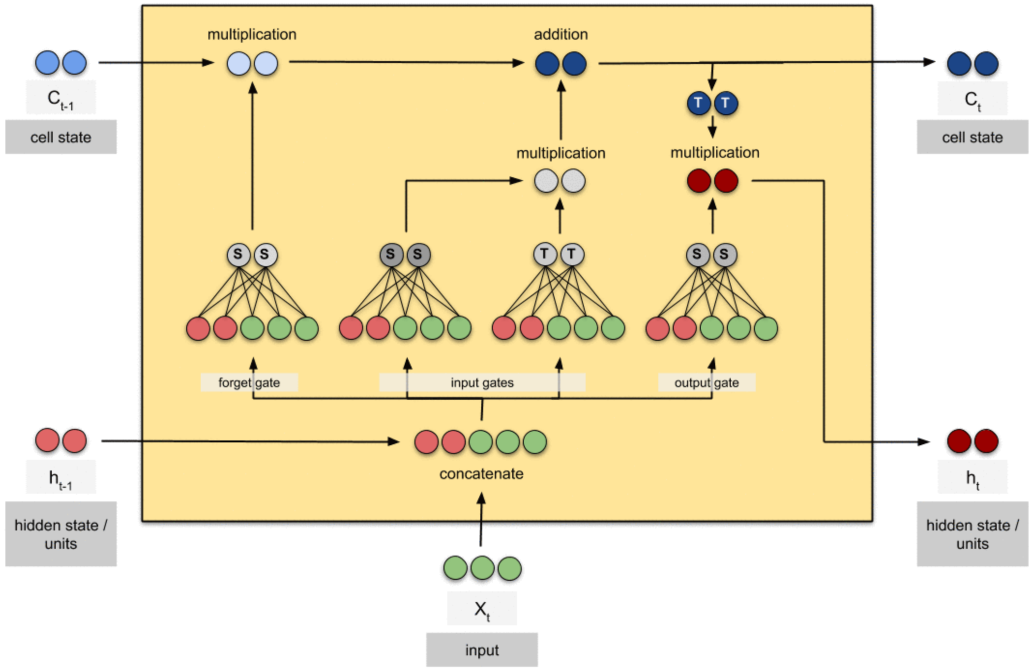
output gate

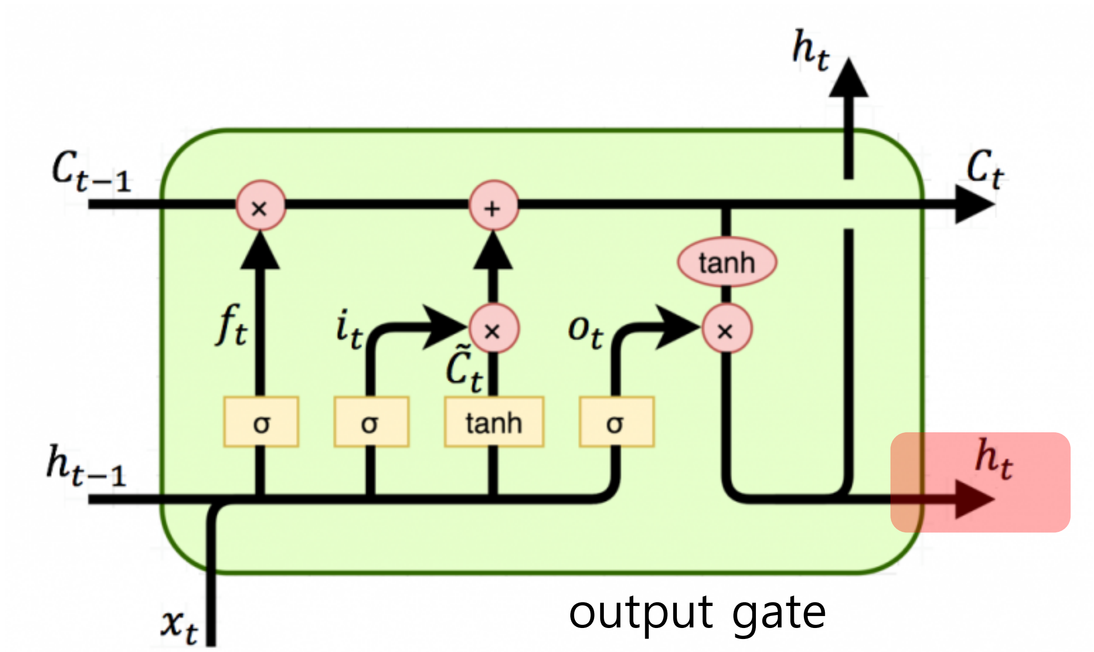
6.4 순환 신경망의 한계 및 개선 방안 : LSTM





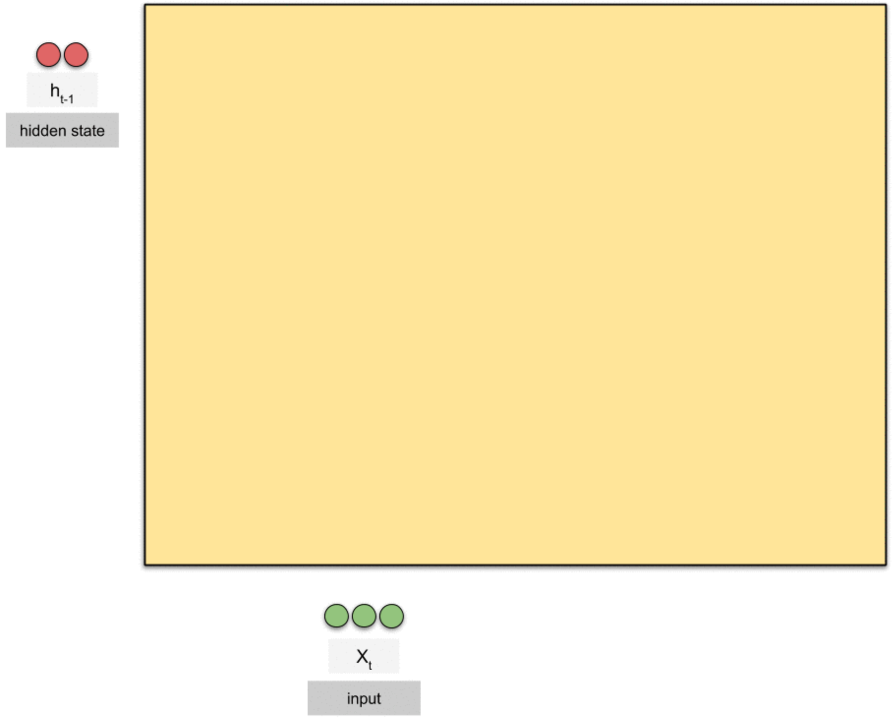
6.4 순환 신경망의 한계 및 개선 방안 : LSTM

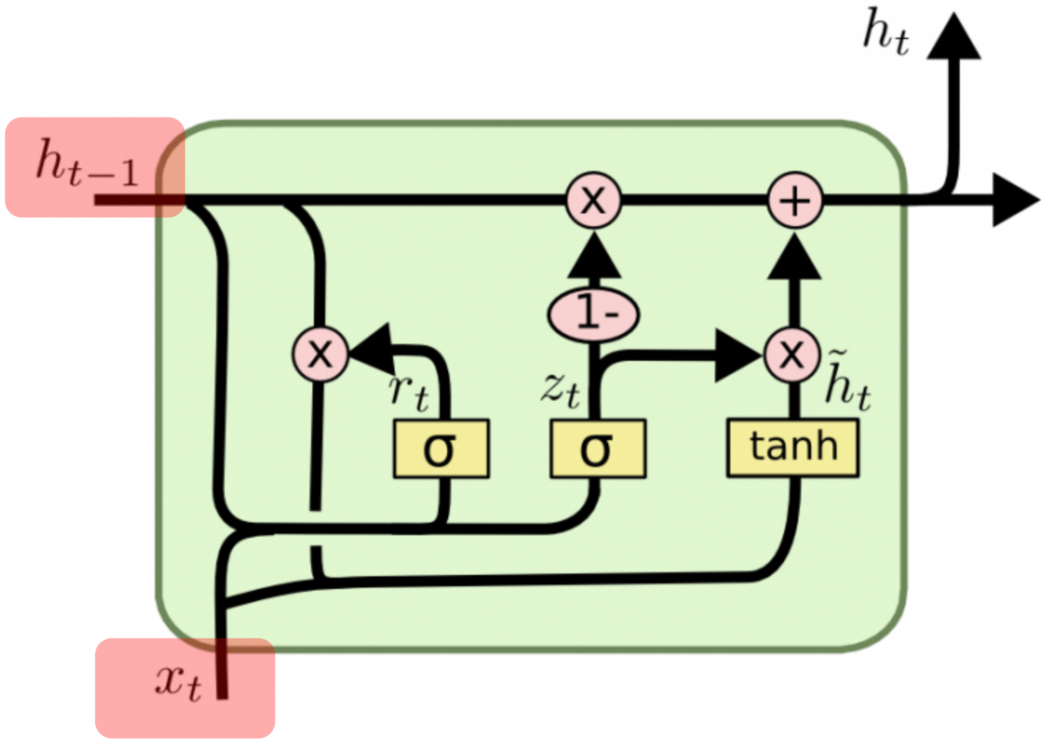




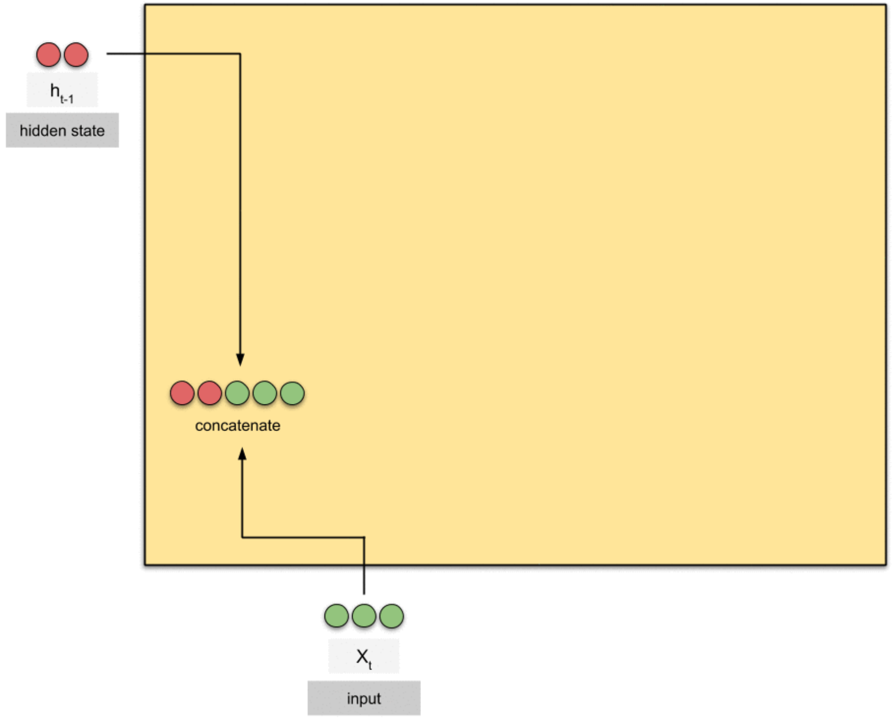
6.4 순환 신경망의 한계 및 개선 방안 : GRU

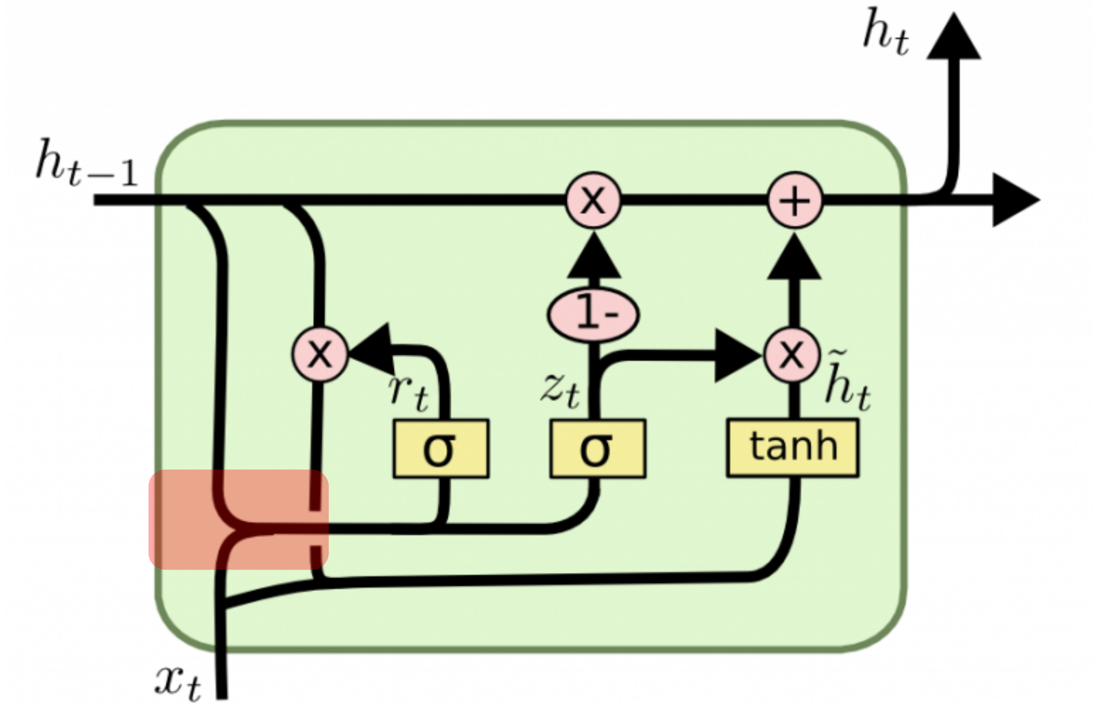
GRU : 셀 상태와 은닉 상태를 분리하지 않고, 은닉 상태 하나로 합침
LSTM의 forget gate, input gate를 하나의 update gate로 합침



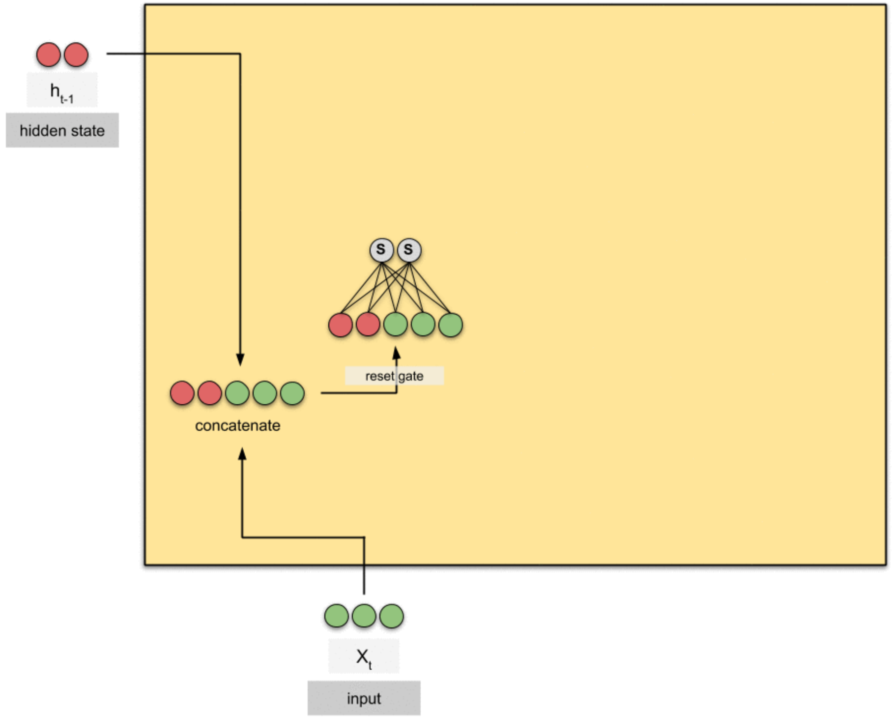


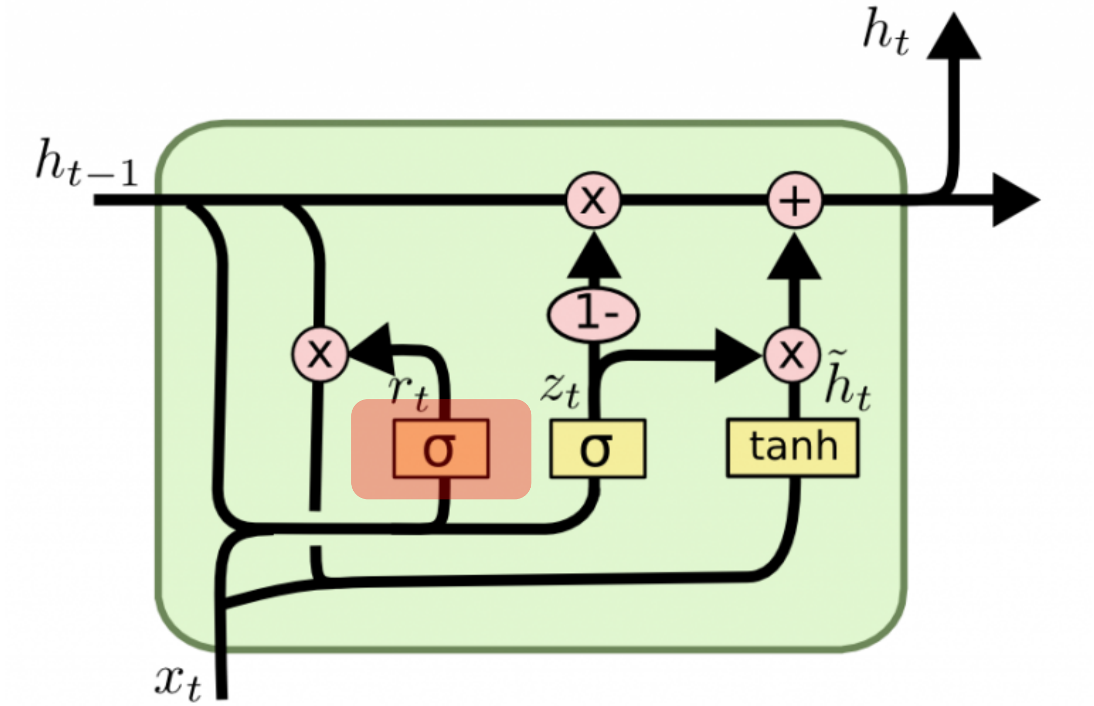
6.4 순환 신경망의 한계 및 개선 방안 : GRU



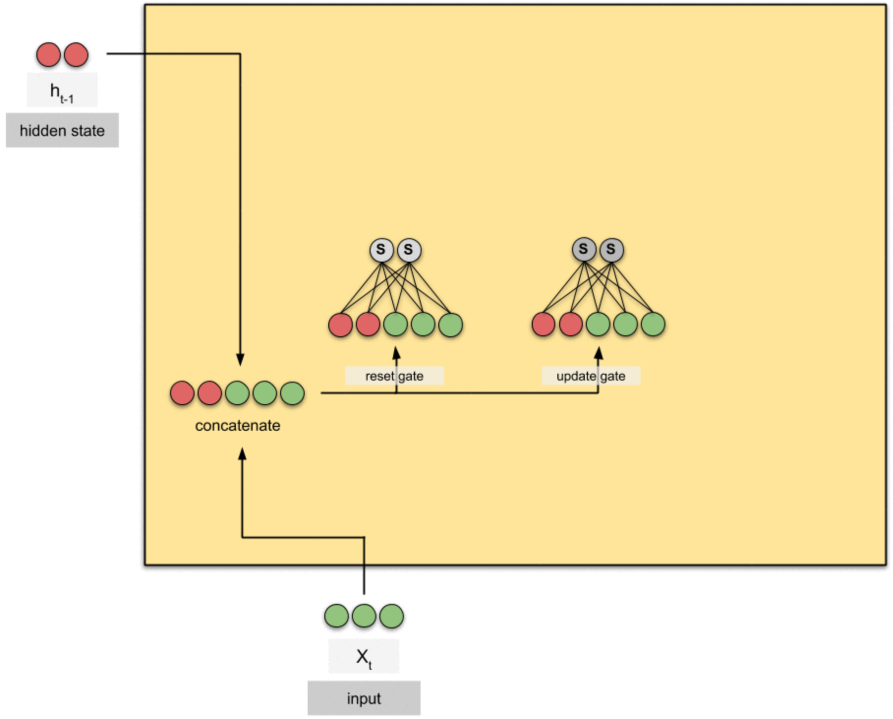


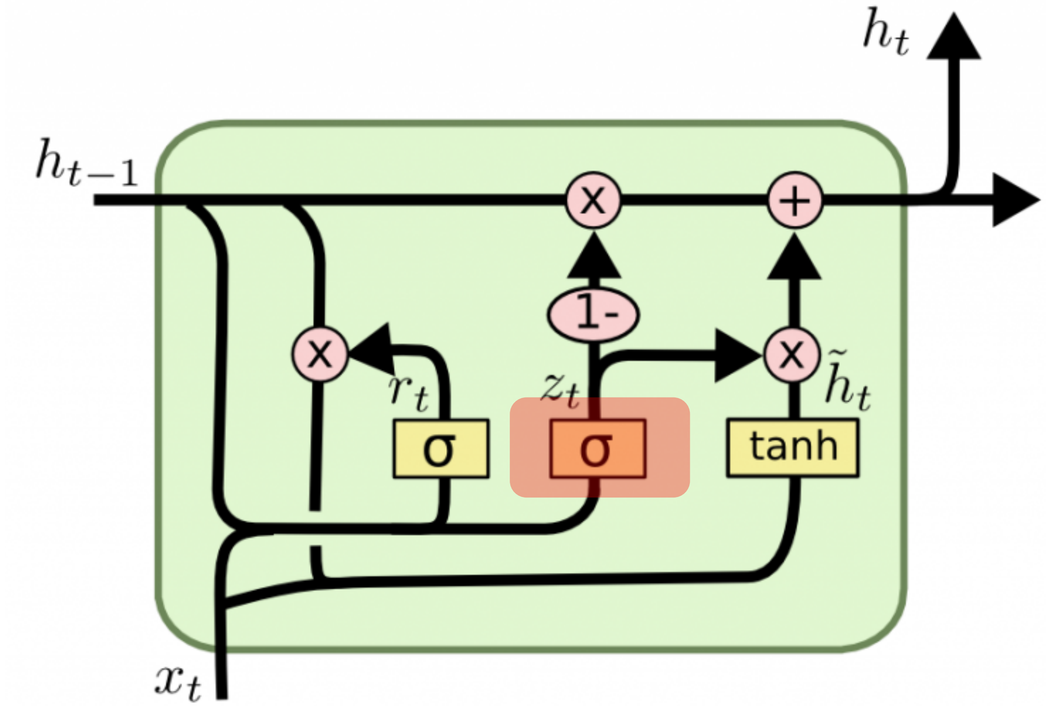
6.4 순환 신경망의 한계 및 개선 방안 : GRU



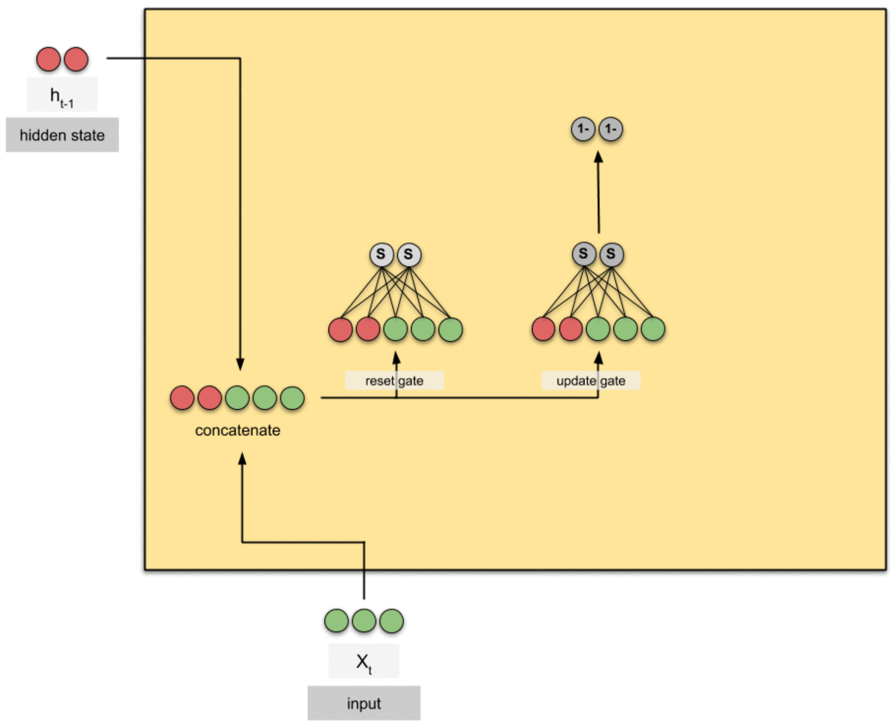


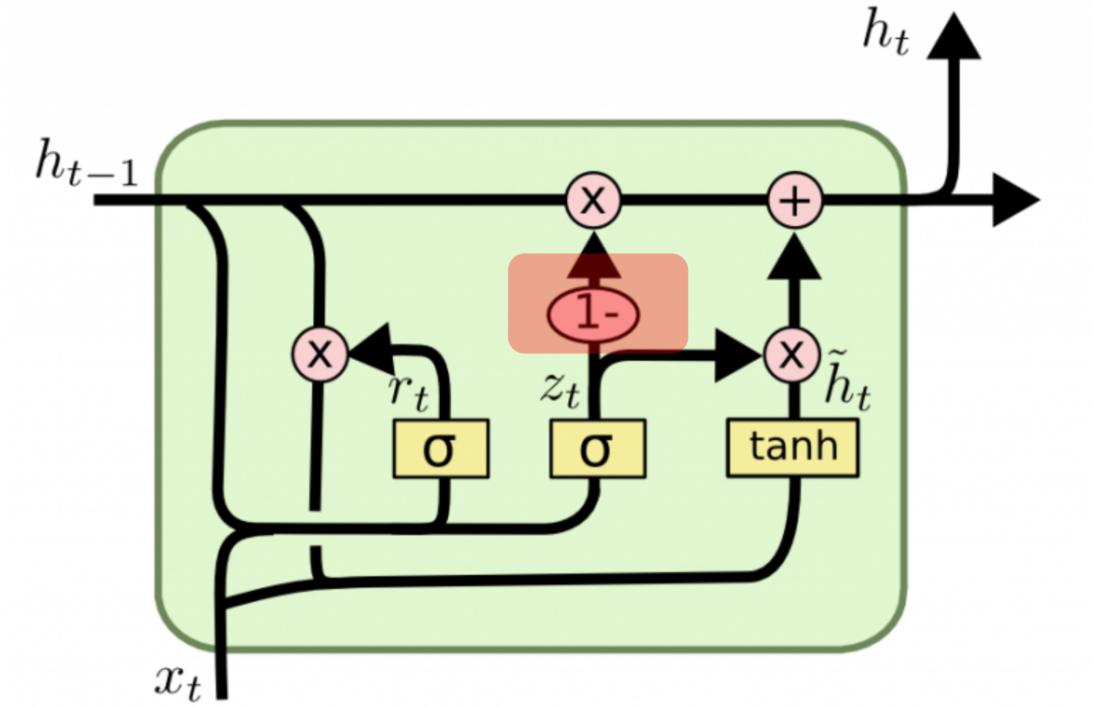
6.4 순환 신경망의 한계 및 개선 방안 : GRU



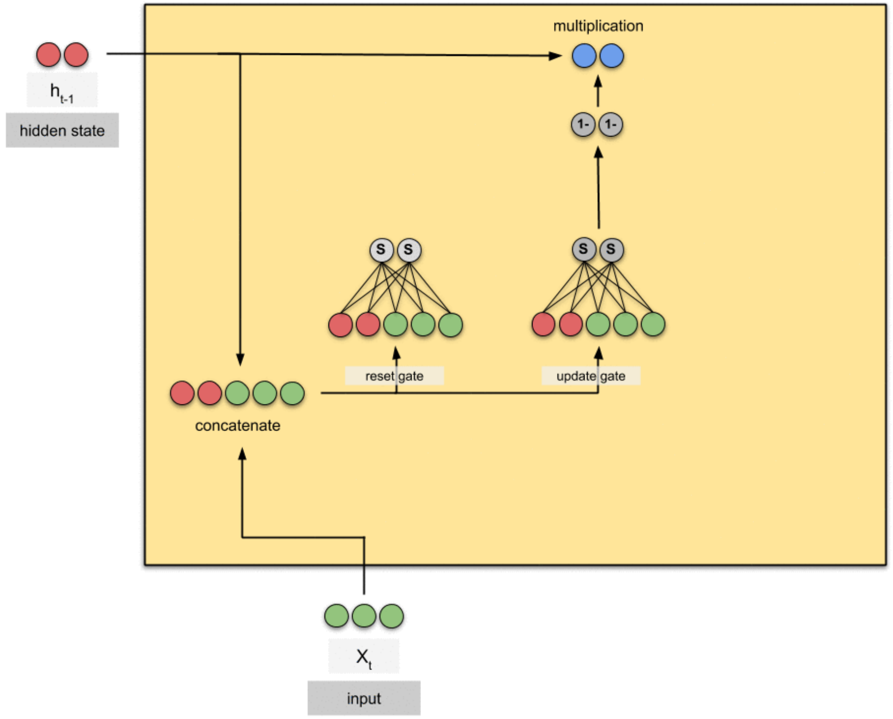


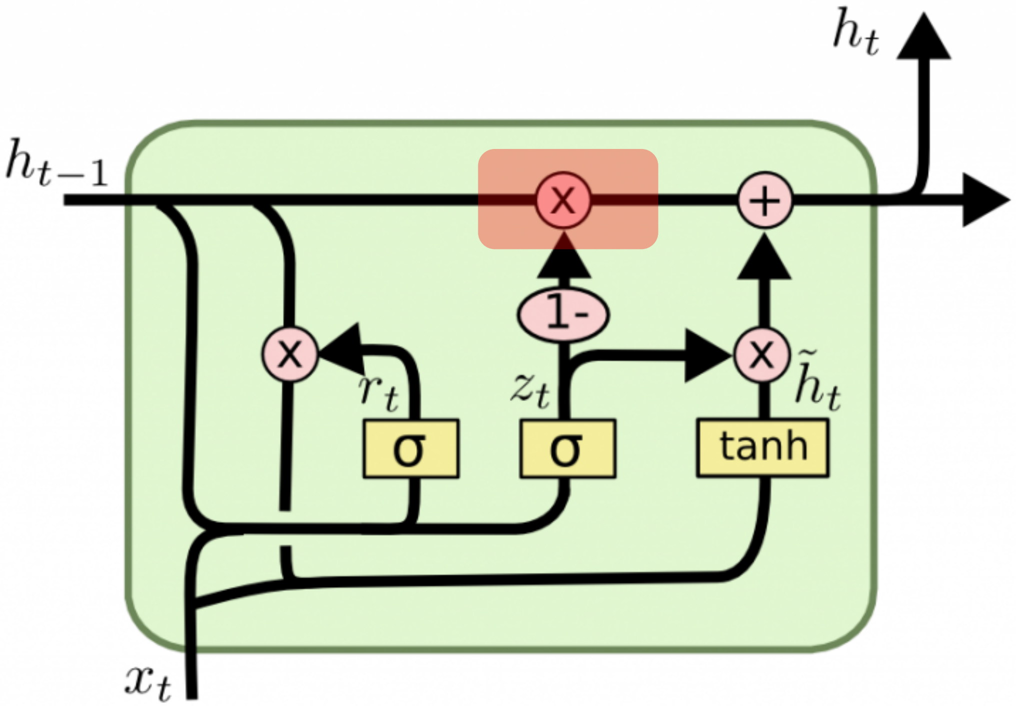
6.4 순환 신경망의 한계 및 개선 방안 : GRU



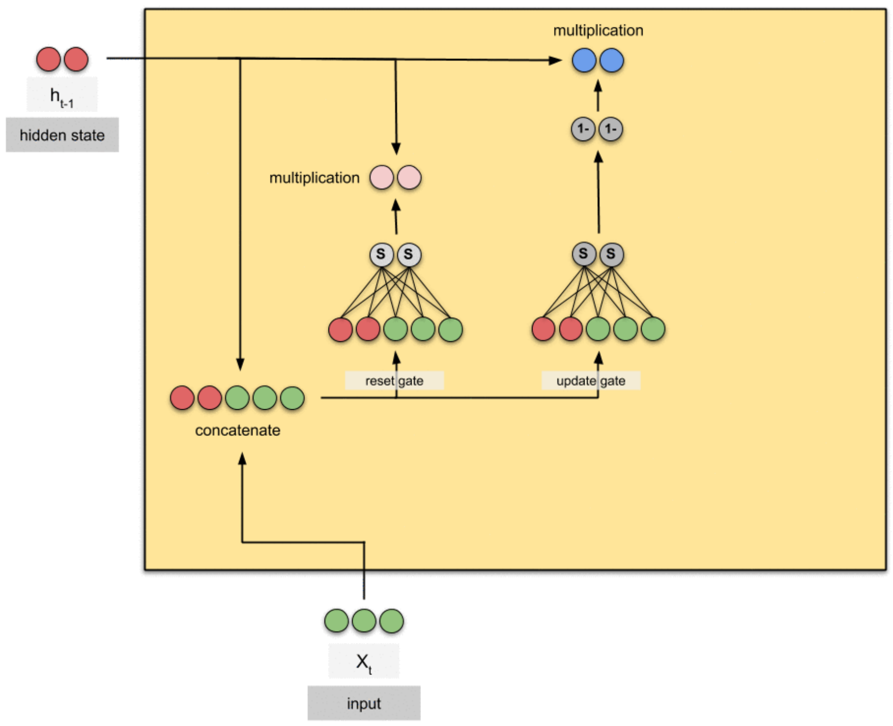


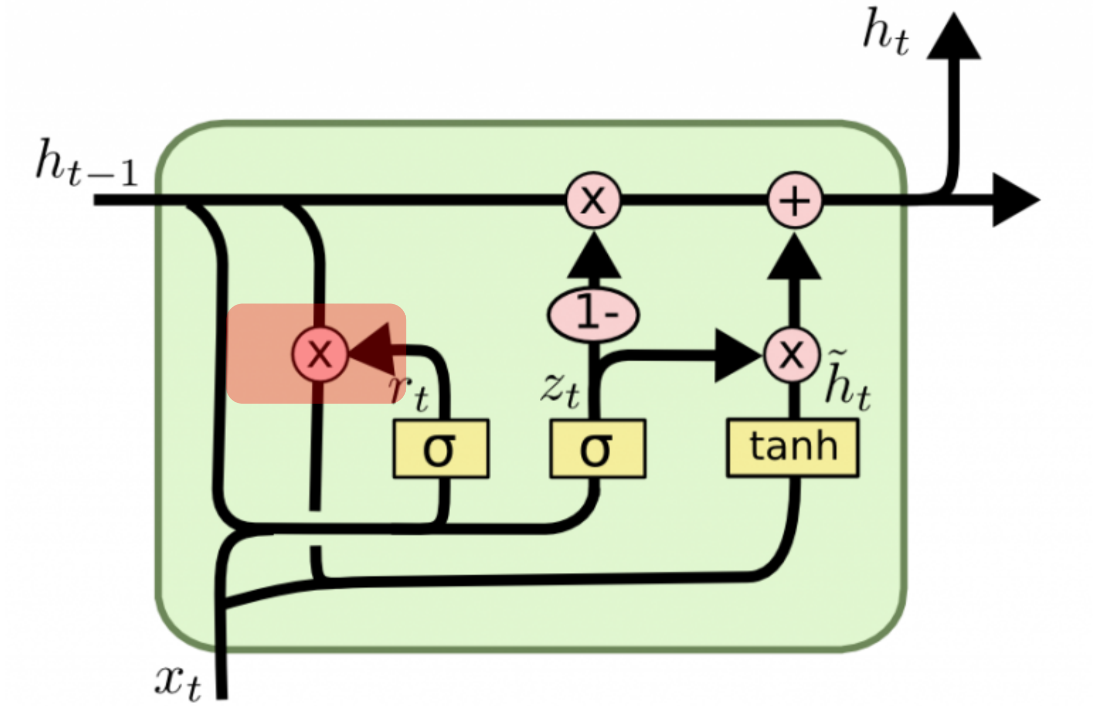
6.4 순환 신경망의 한계 및 개선 방안 : GRU



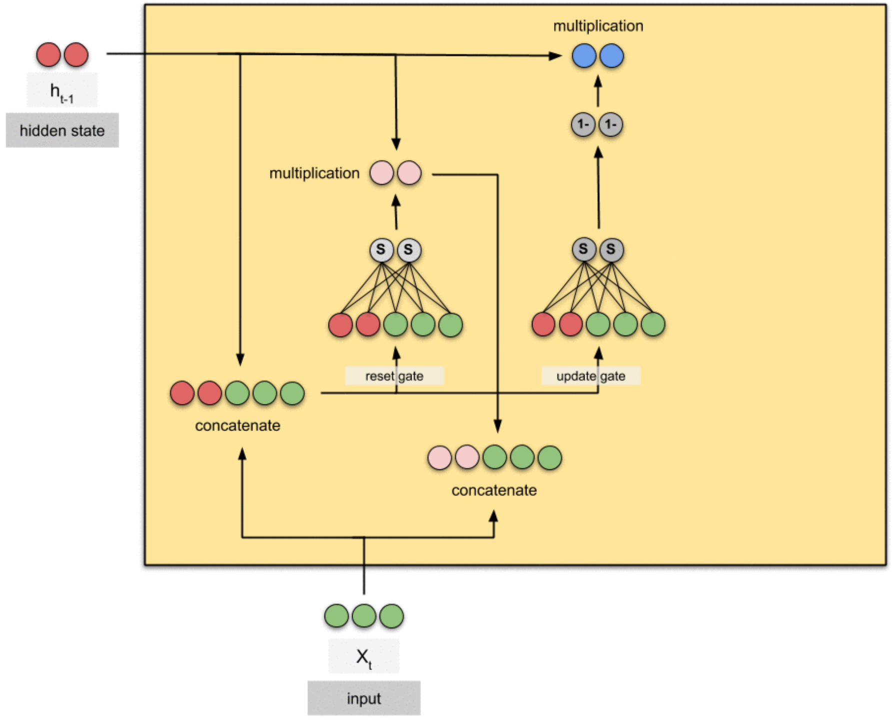


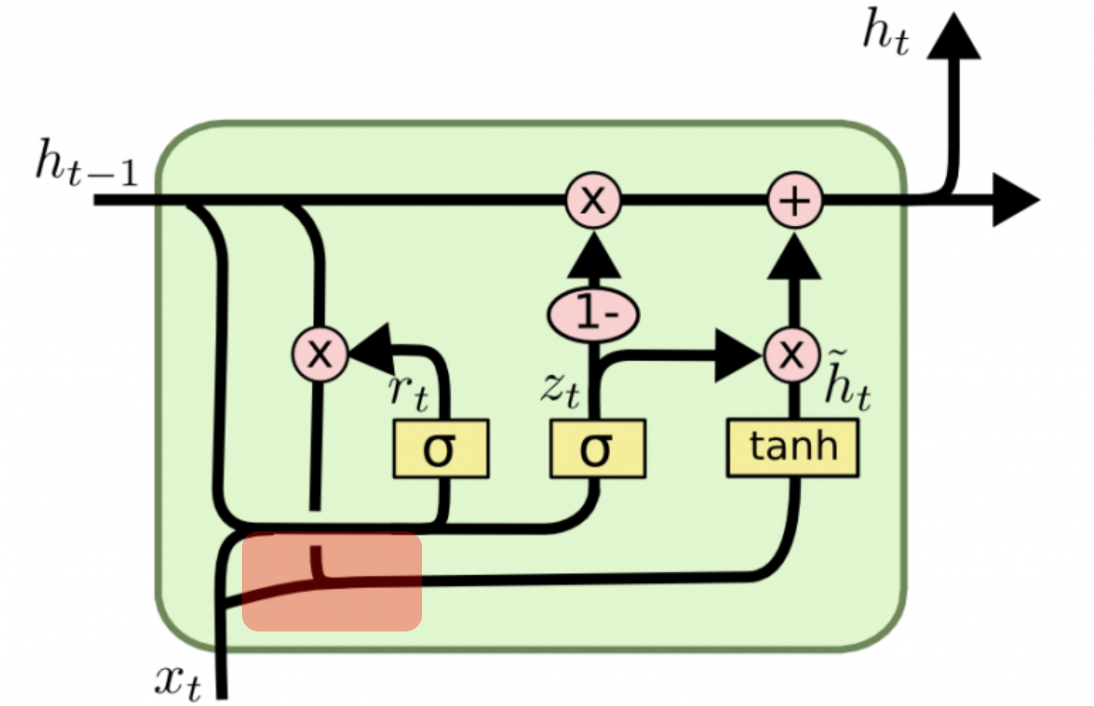
6.4 순환 신경망의 한계 및 개선 방안 : GRU



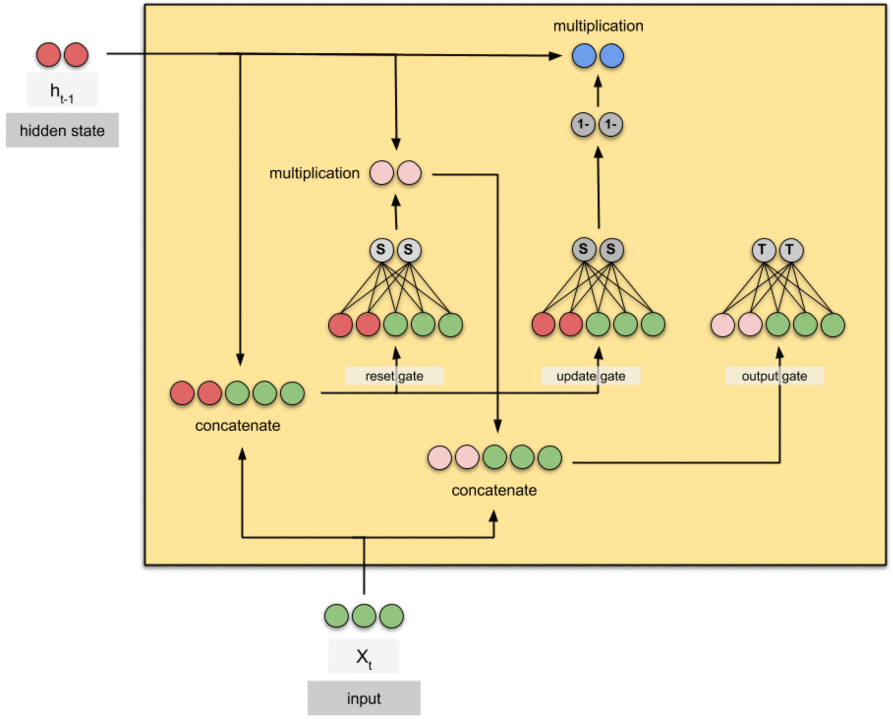


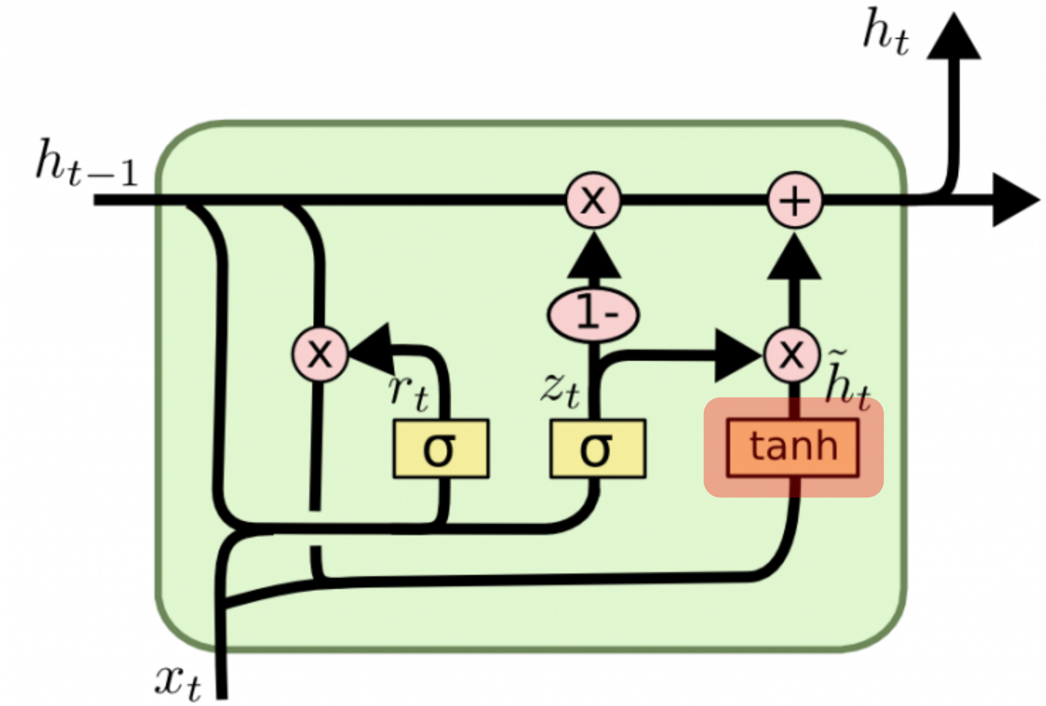
6.4 순환 신경망의 한계 및 개선 방안 : GRU



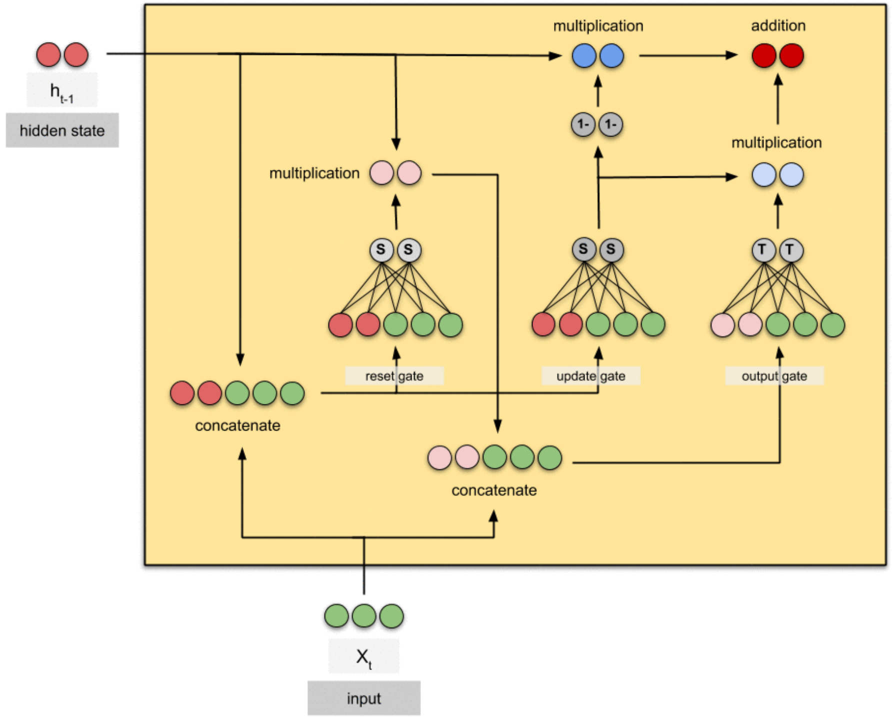


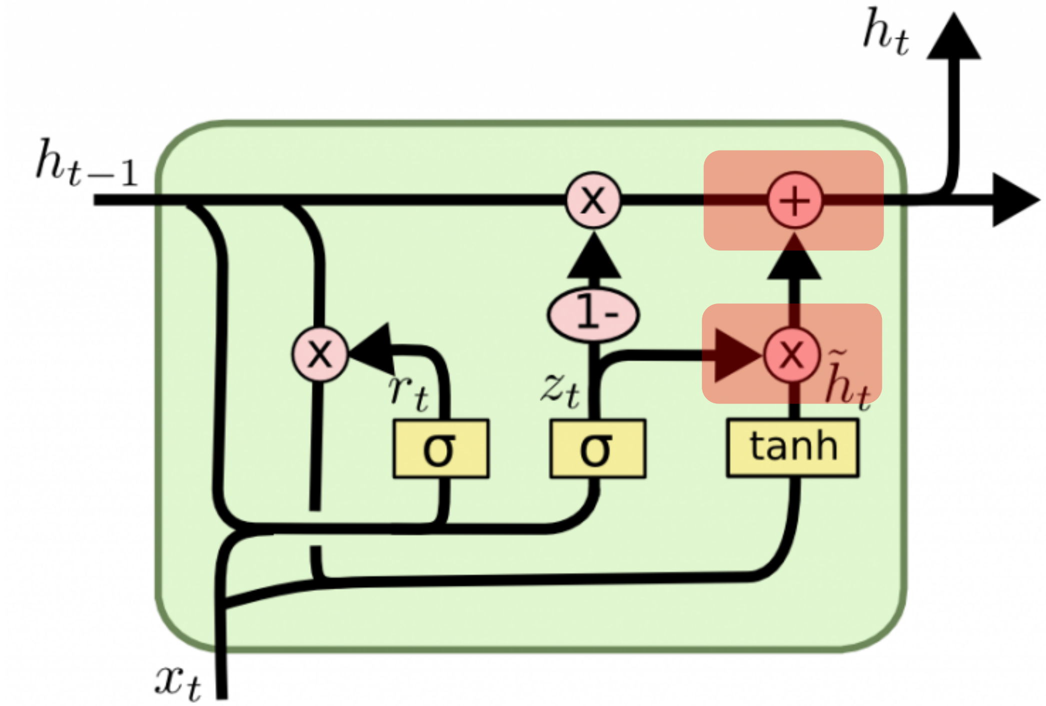
6.4 순환 신경망의 한계 및 개선 방안 : GRU



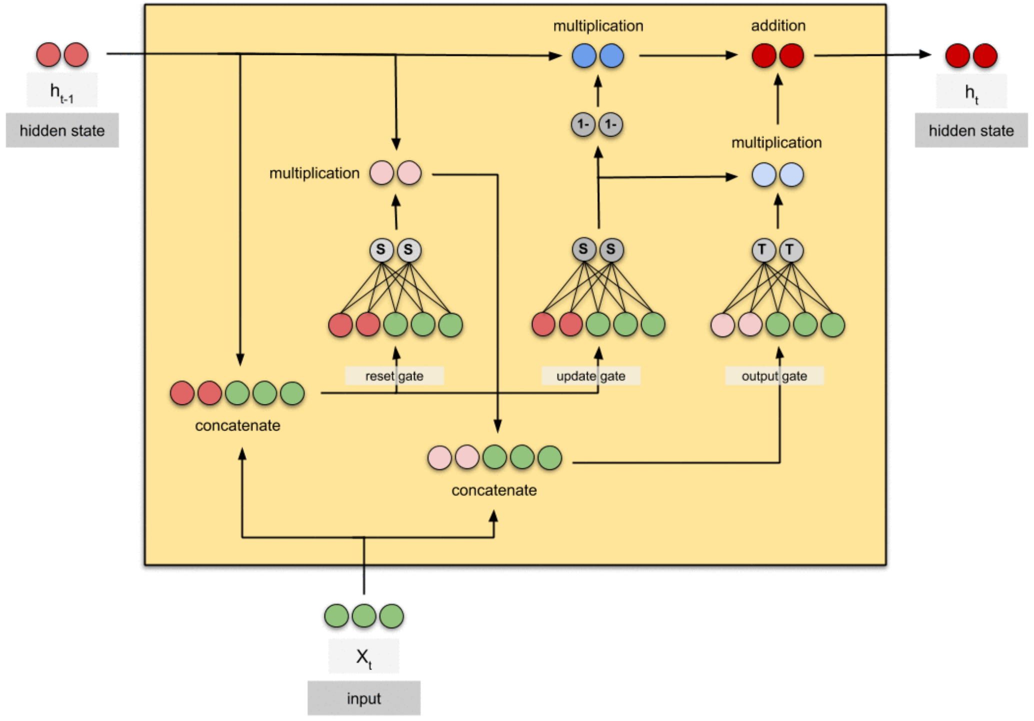


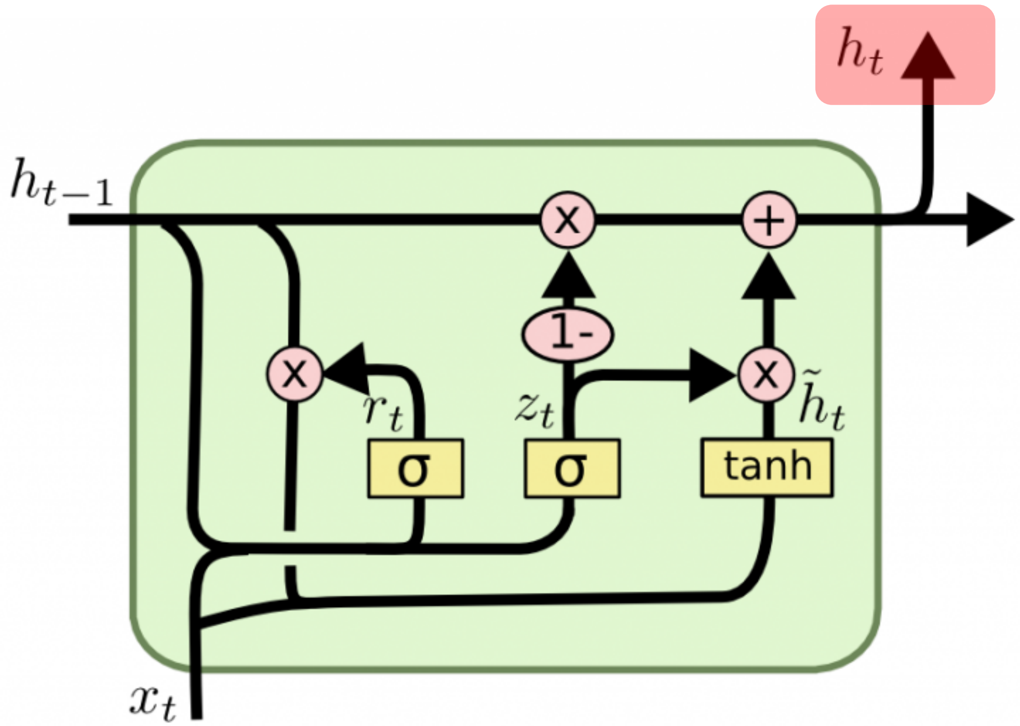
6.4 순환 신경망의 한계 및 개선 방안 : GRU





6.4 순환 신경망의 한계 및 개선 방안 : GRU





임베딩 개념 설명

- 이전까지는 알파벳이나 문자를 하나하나의 one-hot 벡터로 연산
- 두 벡터의 분포를 비교하기 위해 코사인 유사도(내적과 같은 의미)
- 두 벡터가 일치하지 않는 이상 one-hot 벡터끼리의 내적은 0
- 두 벡터의 유사도를 구할 수 없음
- 이를 극복하기 위해 임베딩 방법을 사용

word2vec

- 임베딩의 대표적인 기법: CBOW, skip-gram
- CBOW : 주변 단어로부터 가운데 들어갈 단어가 나오도록 임베딩
- skip-gram : 중심 단어로부터 주변 단어들이 나오도록 임베딩, 같은 말뭉치로도 더 많은 학습 데이터를 확보할 수 있어 임베딩 품질이 CBOW보다 좋은 경향이 있음

출처

http://colah.github.io/posts/2015-08-Understanding-LSTMs/

<한국어 임베딩>, 이기창, 에이콘, 2019