



5. LSTM

5-0. Intro

- RNN은 가중치가 업데이트되는 과정에서 기울기가 1보다 작은 값이 계속 곱해짐
 - 기울기 소멸 문제 발생
- 이를 해결하기 위해 LSTM이나 GRU 같은 확장된 RNN 방식들을 활용하고 있음

5-1. LSTM 구조

LSTM 순전파

- 기울기 소멸 문제를 해결하기 위해 망각 게이트, 입력 게이트, 출력 게이트라는 새로운 요소를 은닉층에 추가함

망각 게이트(forget gate)

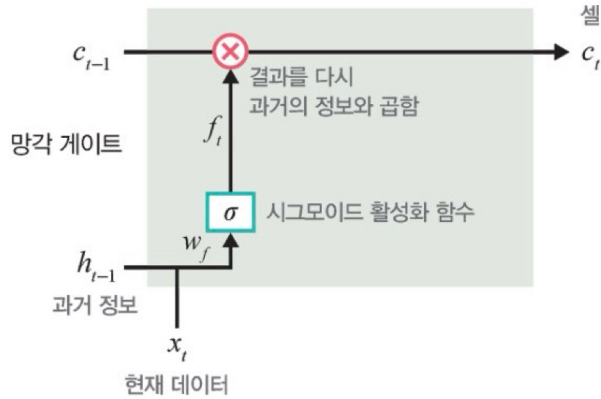
- 과거의 정보를 어느 정도로 기억할 지를 결정
 - 과거 정보와 현재 데이터를 입력받아 `sigmoid()`를 취한 후 해당 값을 과거 정보에 곱해줌
 - 출력이 0이면 과거 정보를 아예 버리고, 1이면 과거 정보를 온전히 보존함
- 입력: x_t (새로운 입력 값), h_{t-1} (이전 은닉층에서 입력되는 값)
 - 계산 값이 1이면 바로 직전의 정보를 메모리에 유지
 - 계산 값이 0이면 초기화

망각 게이트에 대한 수식은 다음과 같습니다.

$$f_t = \sigma(w_f[h_{t-1}, x_t])$$

$$c_t = f_t \cdot c_{t-1}$$

그림으로는 다음과 같이 표현할 수 있습니다.



▲ 7-16 망각 게이트

입력 게이트(input gate)

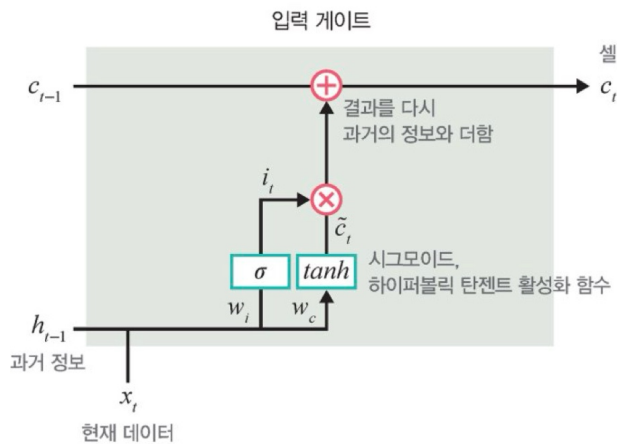
- 현재 정보를 기억하기 위한 부분
 - 과거 정보와 현재 데이터를 입력받아 `sigmoid()` 와 `tanh()` 를 기반으로 현재 정보에 대한 보존량 결정
- 현재 메모리에 새로운 정보를 반영할 지 결정
 - 계산값이 1이면 입력(x_t)가 들어올 수 있도록 허용(→ open)
 - 계산값이 0이면 차단

$$i_t = \sigma(w_i[h_{t-1}, x_t])$$

$$\tilde{c}_t = \tanh(w_c[h_{t-1}, x_t])$$

$$c_t = c_{t-1} + i_t \cdot \tilde{c}_t$$

그림으로는 다음과 같이 표현할 수 있습니다.



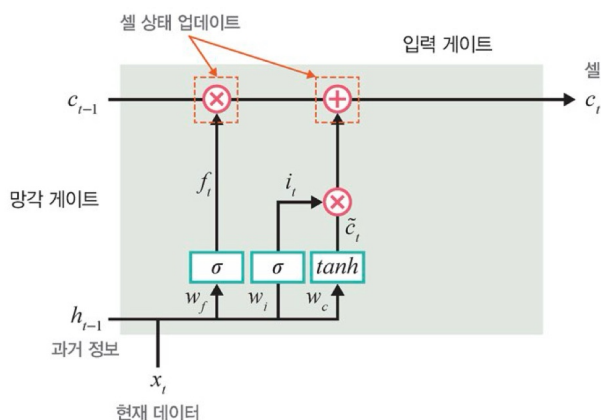
셀

- 각 단계에 대한 은닉 노드(hidden node)
- 총합을 사용하여 셀값을 반영 → 기울기 소멸 문제 해결 가능
- 망각 게이트와 입력 게이트의 이전 단계 셀 정보를 계산하여 현재 단계의 셀 상태를 업데이트

$$f_t = \sigma(w_f[h_{t-1}, x_t])$$

$$c_t = c_{t-1} + i_t \cdot \tilde{c}_t$$

그림으로는 다음과 같이 표현할 수 있습니다.



▲ 7-18 셀

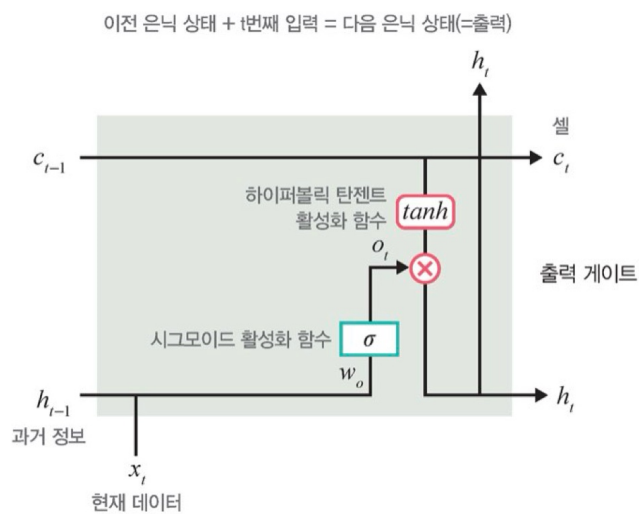
출력 게이트(output gate)

- 과거 정보와 현재 데이터를 사용하여 뉴런의 출력을 결정함
 - 이전 은닉 상태와 t번째 입력을 고려하여 다음 은닉 상태를 계산
 - LSTM에서는 해당 은닉 상태가 그 시점에서의 출력이 됨
- 갱신된 메모리의 출력 값을 제어하는 역할
 - 계산 값이 1이면 의미 있는 결과로 최종 출력
 - 계산 값이 0이면 해당 연산 출력 x

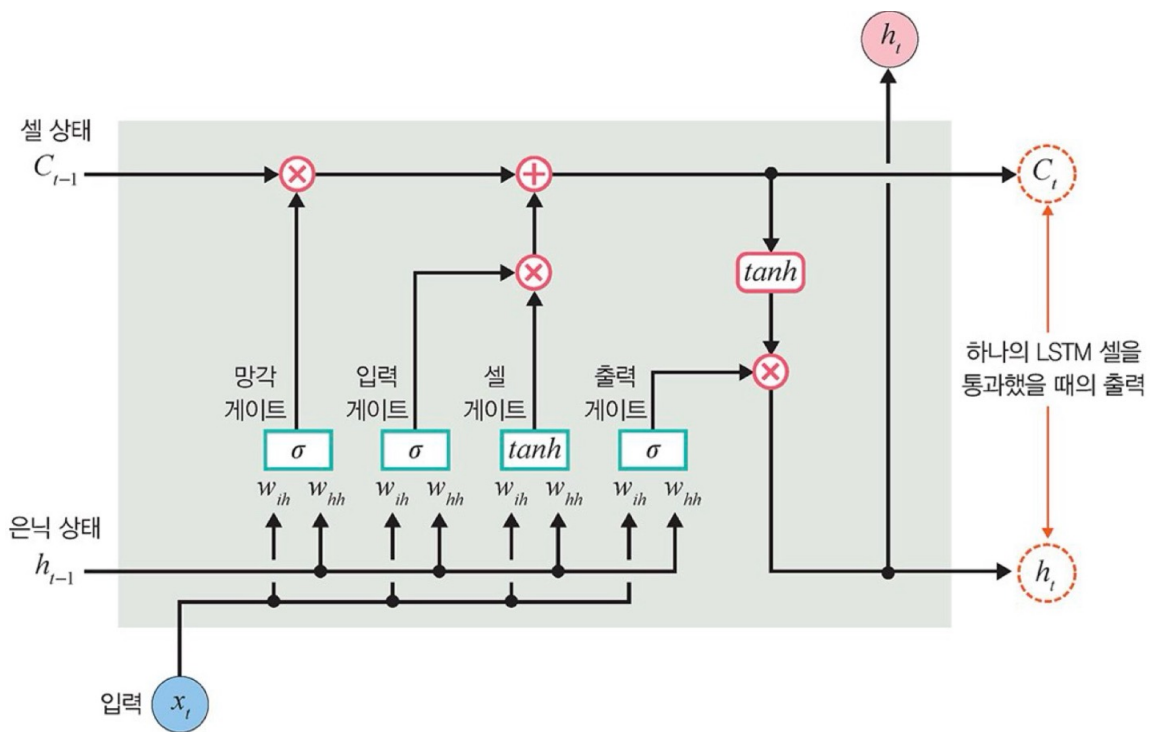
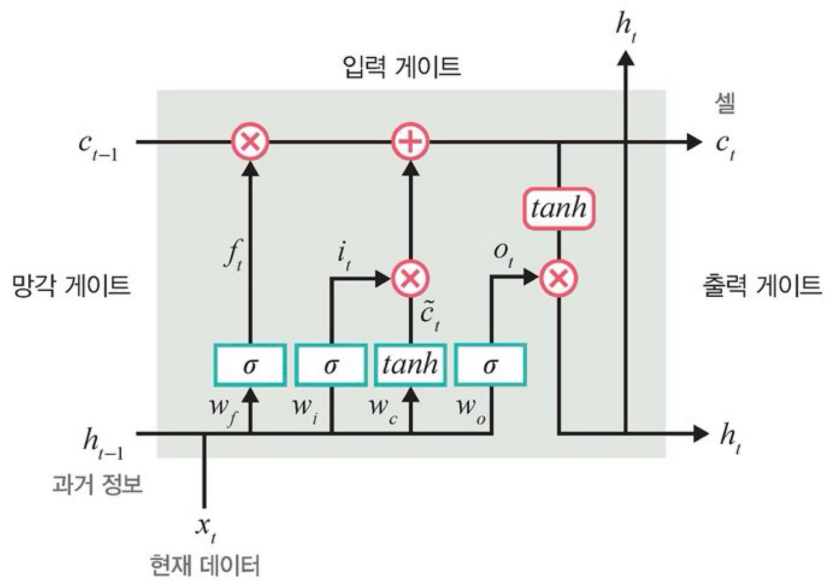
$$o_t = \sigma(w_o [h_{t-1}, x_t])$$

$$h_t = o_t \cdot \tanh(c_{t-1})$$

그림으로는 다음과 같이 표현할 수 있습니다.

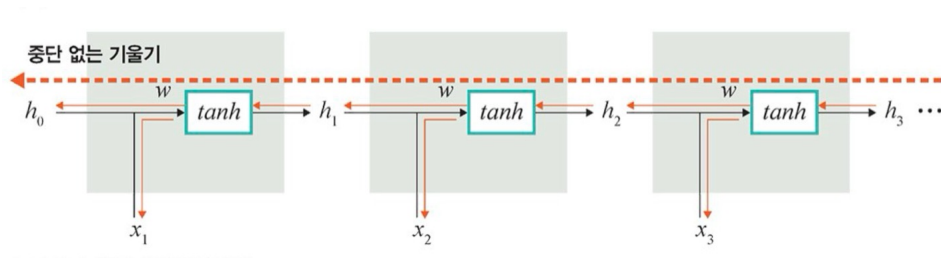


전체 게이트



LSTM 역전파

- LSTM은 셀을 통해 역전파를 수행
 - 중단 없는 기울기(uninterrupted gradient flow)라고도 함
 - 최종 오차 자체는 모든 노드에 전파됨



- 수식

$$\begin{aligned}
 t_t &= \tanh(w_{hh}h_{t-1} + w_{xh}x_t) \\
 &= \tanh\left((w_{hh} \ w_{xh}) \begin{pmatrix} h_{t-1} \\ x_t \end{pmatrix}\right) \\
 &= \tanh\left(w \begin{pmatrix} h_{t-1} \\ x_t \end{pmatrix}\right)
 \end{aligned}$$

! 입력 방향으로도 오차가 전파됨

5-2. LSTM 셀 구현

- 교재 & 예제 code 참고
 - p. 532 ~ 537

5-3. LSTM 계층 구현

- 교재 & 예제 code 참고