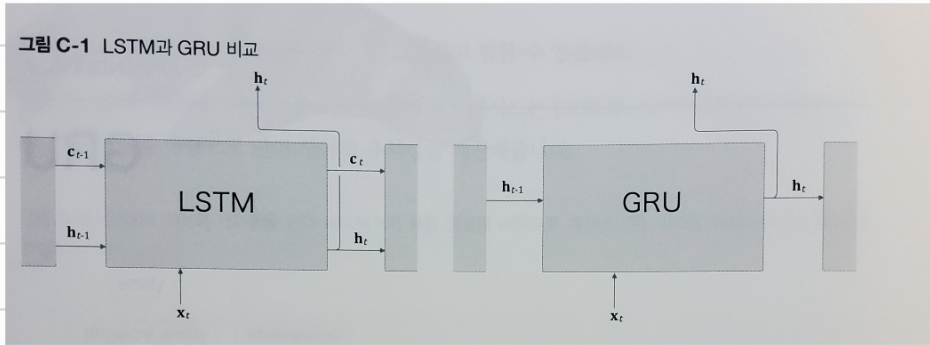


GRU

- 게이트를 사용하는 개념은 유지한 채, 매개변수 수를 줄여 계산시간을 줄여줌
- GRU는 은닉상태만을 사용



$$z = \sigma(x_t W_x^{(z)} + h_{t-1} W_h^{(z)} + b^{(z)})$$

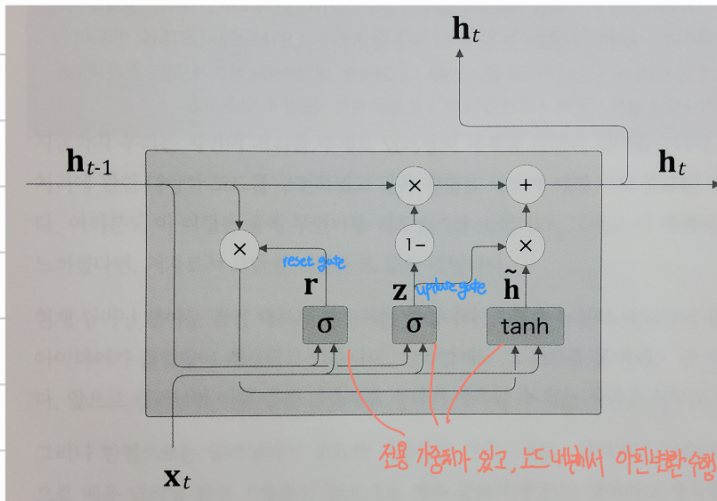
$$r = \sigma(x_t W_x^{(r)} + h_{t-1} W_h^{(r)} + b^{(r)})$$

$$\tilde{h} = \tanh(x_t W_x + (r \odot h_{t-1}) W_h + b)$$

$$h_t = (1 - z) \odot h_{t-1} + z \odot \tilde{h}$$

은닉상태 갱신. LSTM의 forget + input

과거 은닉 상태를 얼마나 무시할지 정함, $r=0$ 이면 z 는 x_t 만으로 결정



LSTM vs GRU

- 최근에는 LSTM (이나 변종)이 많이 사용됨
- GRU는 데이터셋이 작거나 모델 설계시 반복시도를 많이 해야 하는 경우 특히 적합.