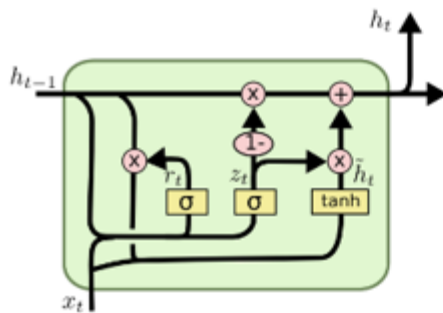


GRU: 기존 LSTM 의 구조를 조금 더 간단하게 개선한 모델

- GRU 는 기존 LSTM 에 비해 더 간단한 구조를 가지고 있음.
- 마지막 출력값에 활성화함수를 적용하지 않음.
- 성능 면에서는 LSTM 과 비교해서 우월하다고 할 수 없지만, 학습할 파라미터가 더 적은 것이 장점임.
- GRU 에서는 reset gate, update gate 2 개의 gate 만을 사용함.
- cell state, hidden state 가 합쳐져 하나의 hidden state 로 표현함.

[모델의 구조 및 수식]



$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \quad \text{---(1)}$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \quad \text{---(2)}$$

$$\tilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t]) \quad \text{---(3)}$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \quad \text{---(4)}$$

- Reset Gate

: 이전 시점의 hidden state 와 현 시점의 x 를 시그모이드 활성화함수를 적용하여 구할 수 있음.
결과값은 0~1 사이의 값을 가지며, 이전 hidden state 의 값을 얼마나 활용할 것인지에 대한
정보로 해석할 수 있음. 식 (3)에서는 전 시점의 hidden state 에 reset gate 를 곱하여 계산함.

- Update Gate

: LSTM 의 input, forget gate 와 비슷한 역할을 하며 과거와 현재의 정보를 각각 얼마나
반영할지에 대한 비율을 구하는 것이 핵심. 식 (4)을 통해 현 시점의 출력값 hidden state 를
구할 수 있음.