



Long Short-Term Memory (LSTM)

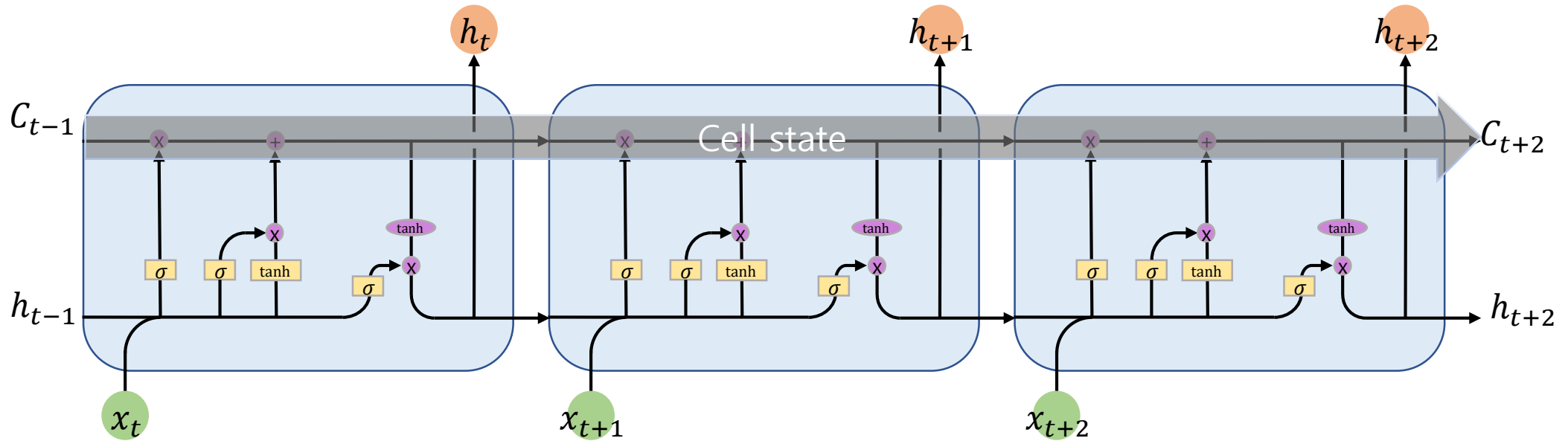
Tae-Won Lee,
leetee95@gmail.com

2020. 12. 22.

Introduction

- Long Short-Term memory (LSTM)은 기존 Recurrent Neural Network (RNN)의 Gradient vanishing and exploding problem을 위해 개발됨.
- 기존의 RNN에 Cell state를 추가한 형태.
- Gradient vanishing and exploding problem을 어느정도 해소.

Structure

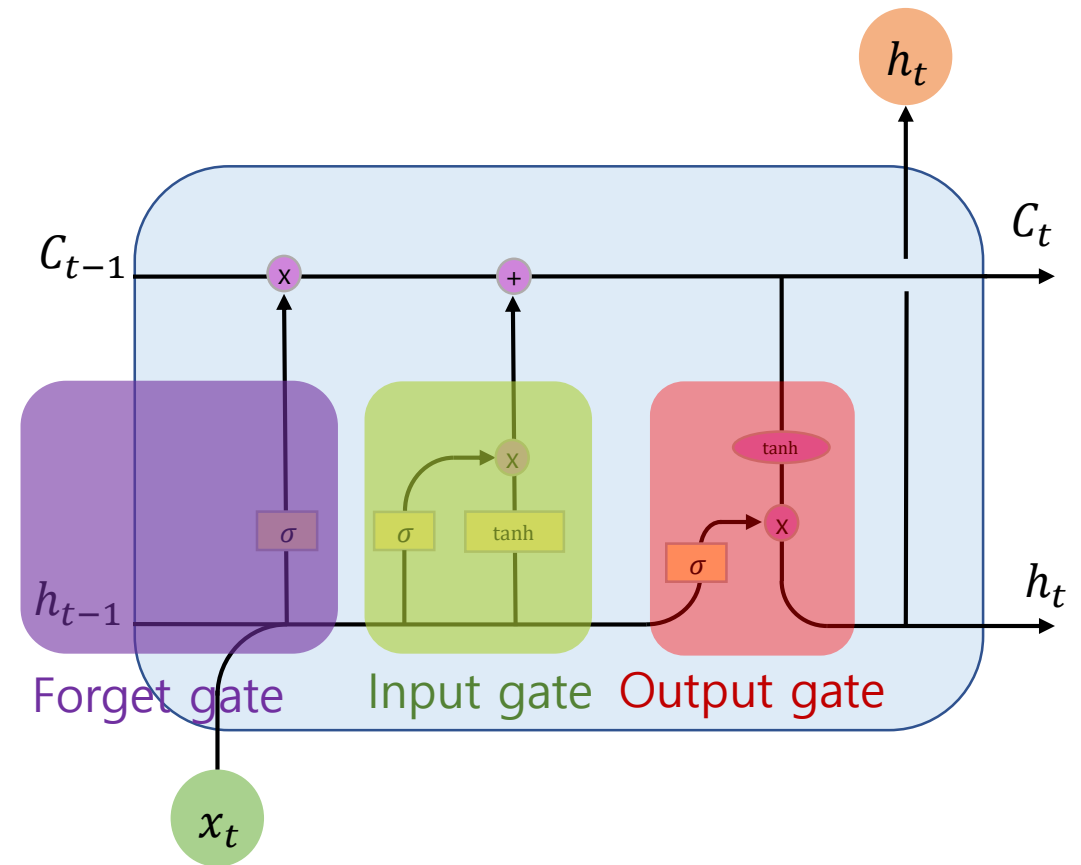


- non-linear activation function을 거치지 않고 전체 흐름을 관통하는 Cell state를 이용.

Structure

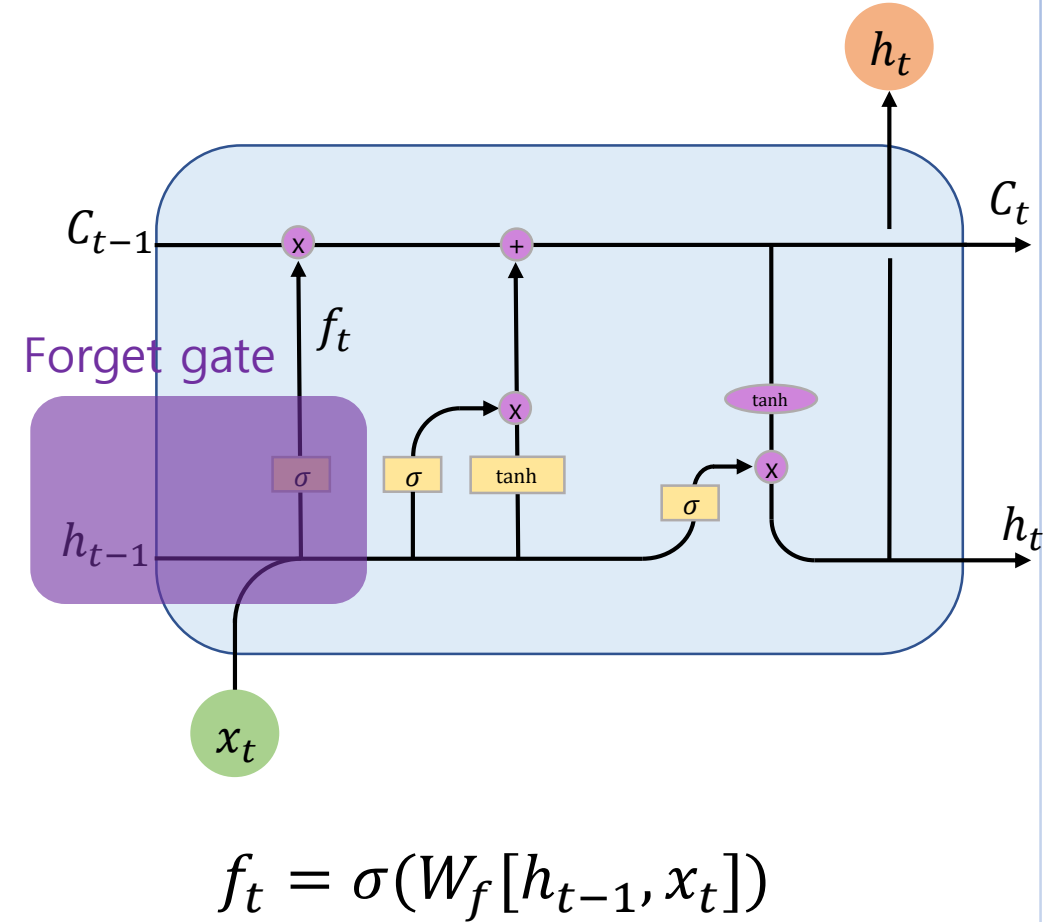
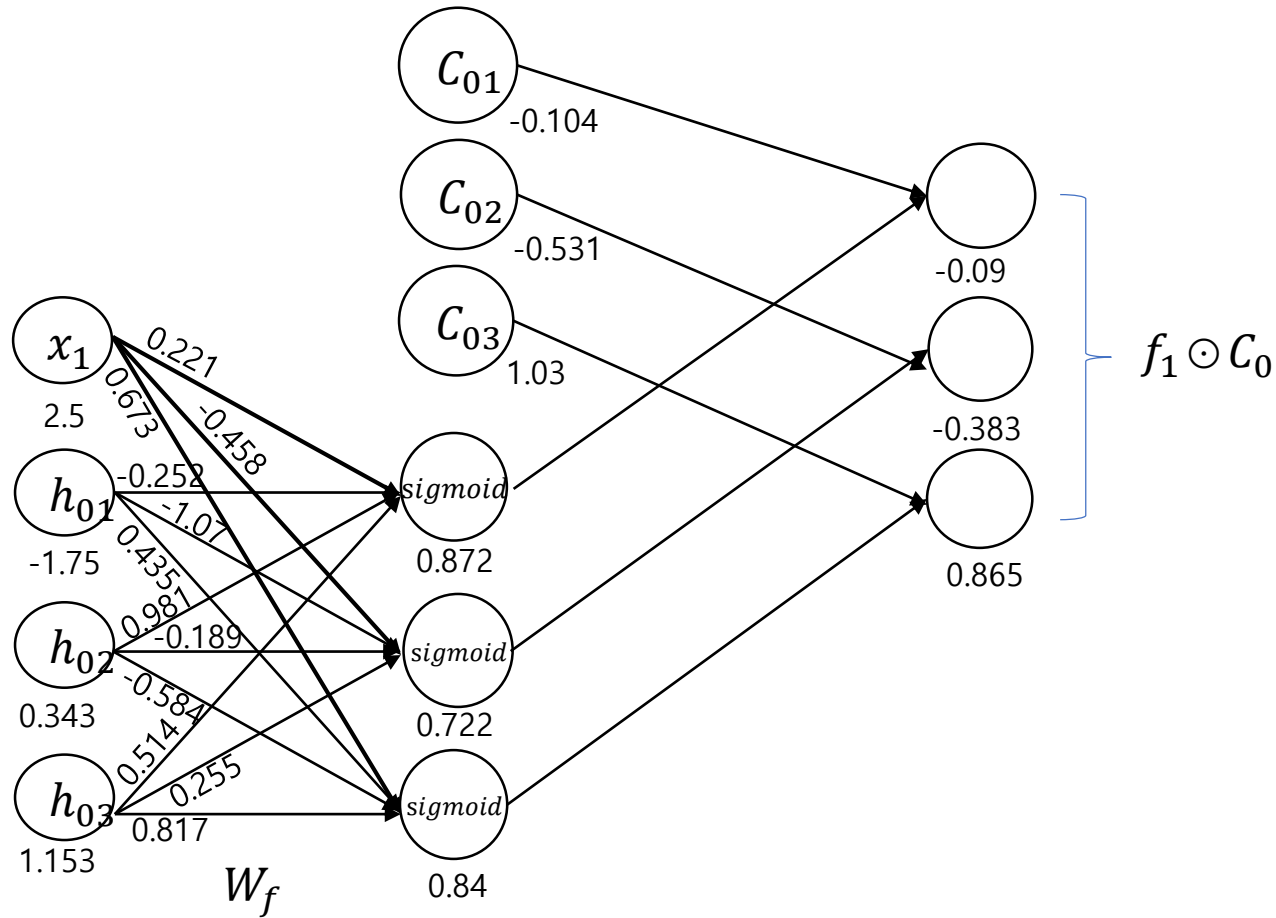
- Gate

- Forget gate : 과거의 cell state의 어떤 부분을 얼마나 잊을지 결정
- Input gate : cell state에 현재 input값 중 어떤 부분을 얼마나 더할지 결정
- Output gate : cell state에 현재 input값 중 어떤 부분을 얼마나 다음 hidden state로 전달할지 결정



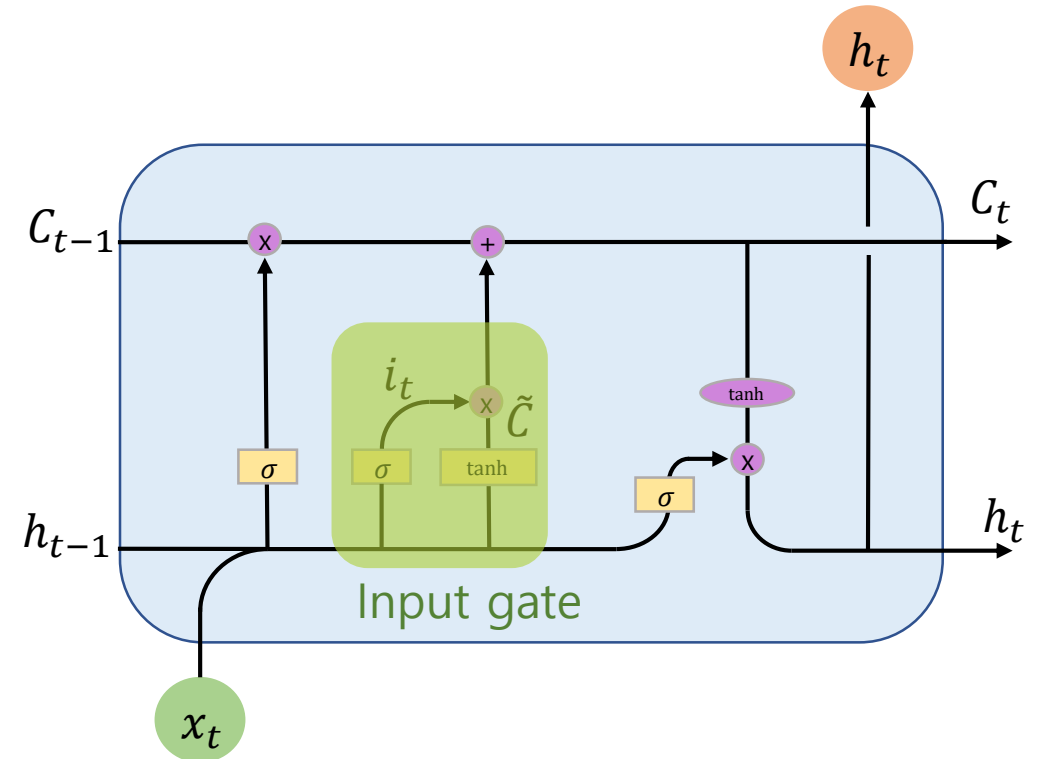
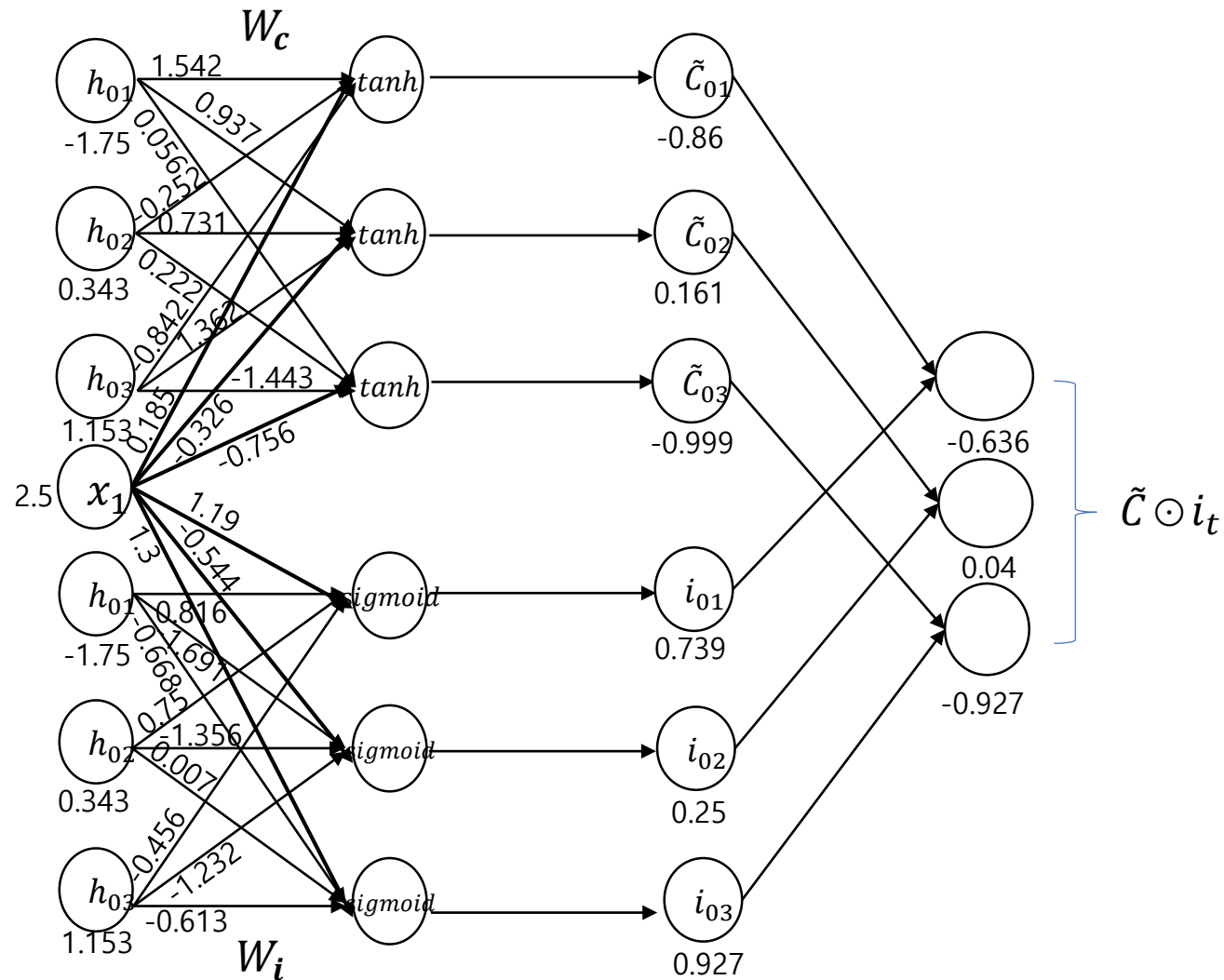
Structure

- Forget gate



Structure

- Input gate

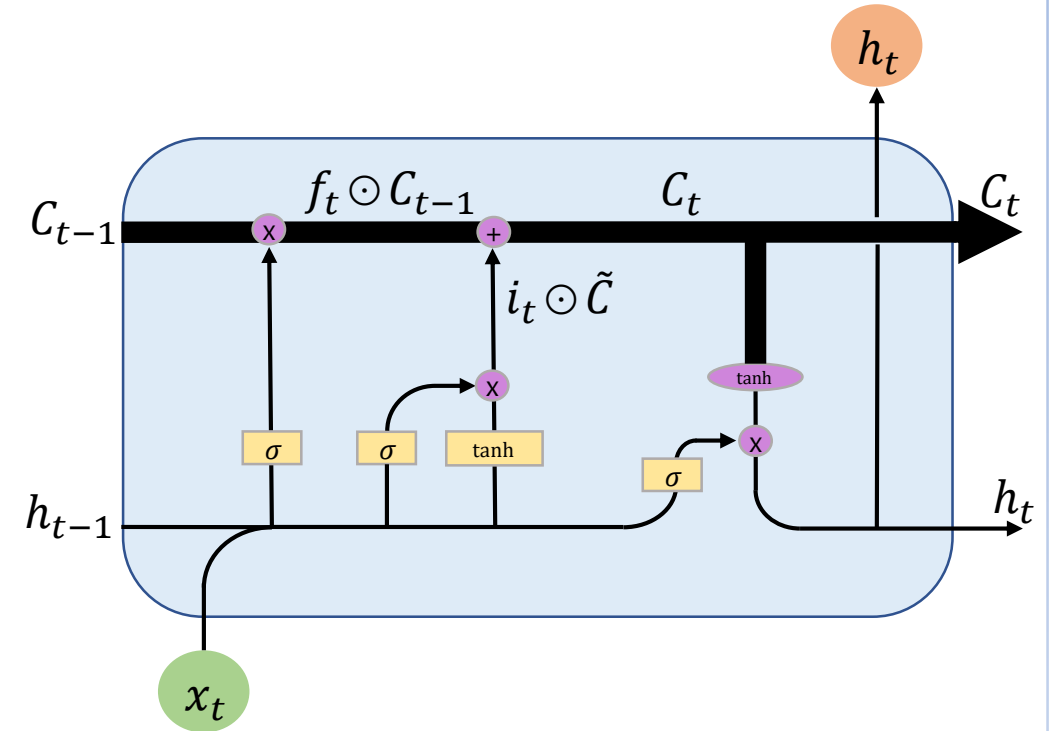
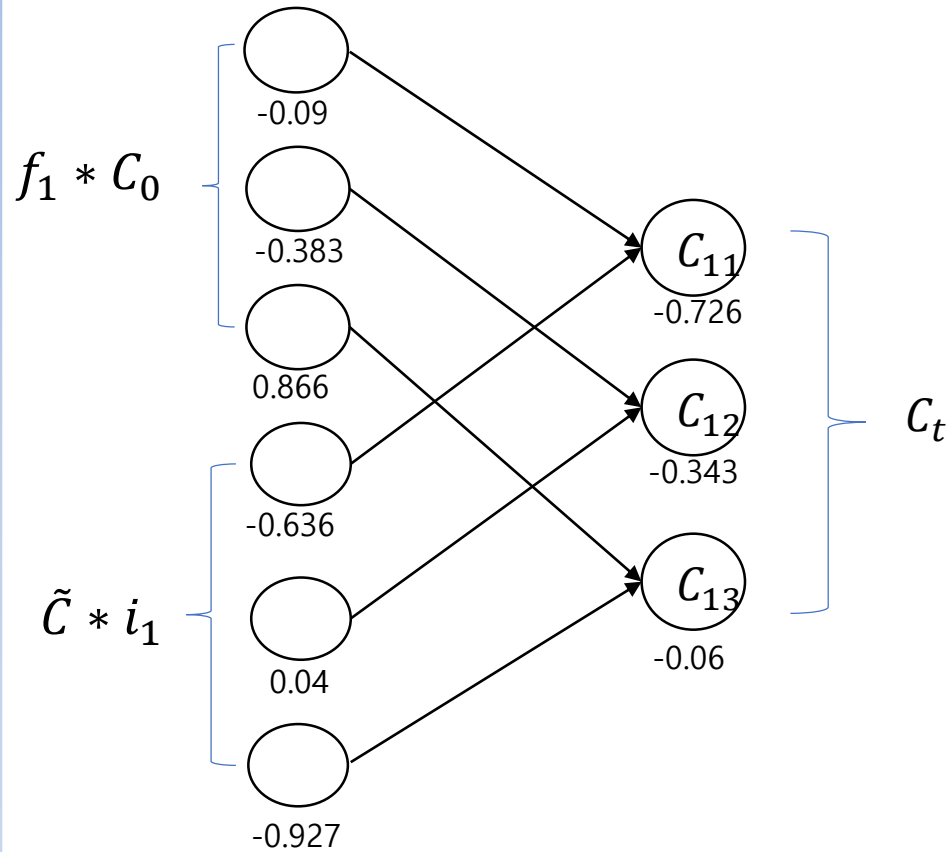


$$i_t = \sigma(W_i[h_{t-1}, x_t])$$

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

Structure

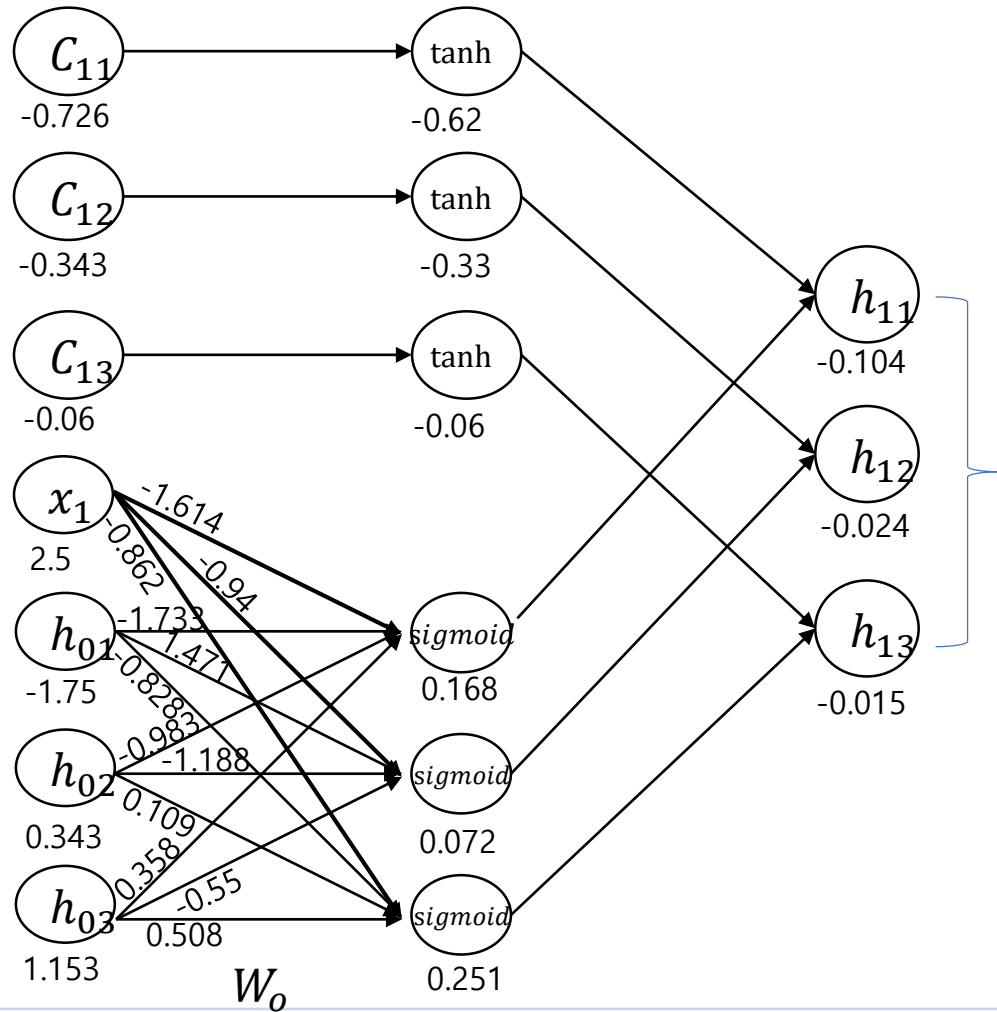
- Cell state update



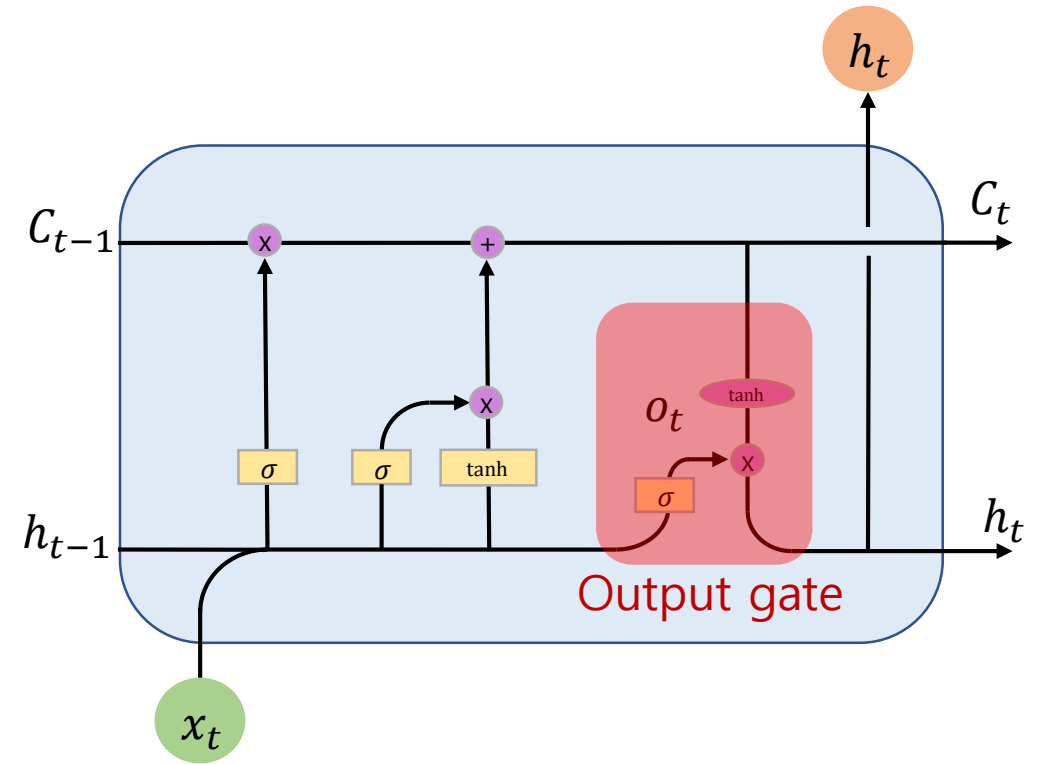
$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}$$

Structure

- Output gate



$$o_t \odot \tanh(C_t)$$



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

$$h_t = o_t \odot \tanh(C_t)$$