Long Short-Term Memory (LSTM)

Tae-Won Lee,
leetee95@gmail.com

2020. 12. 29.

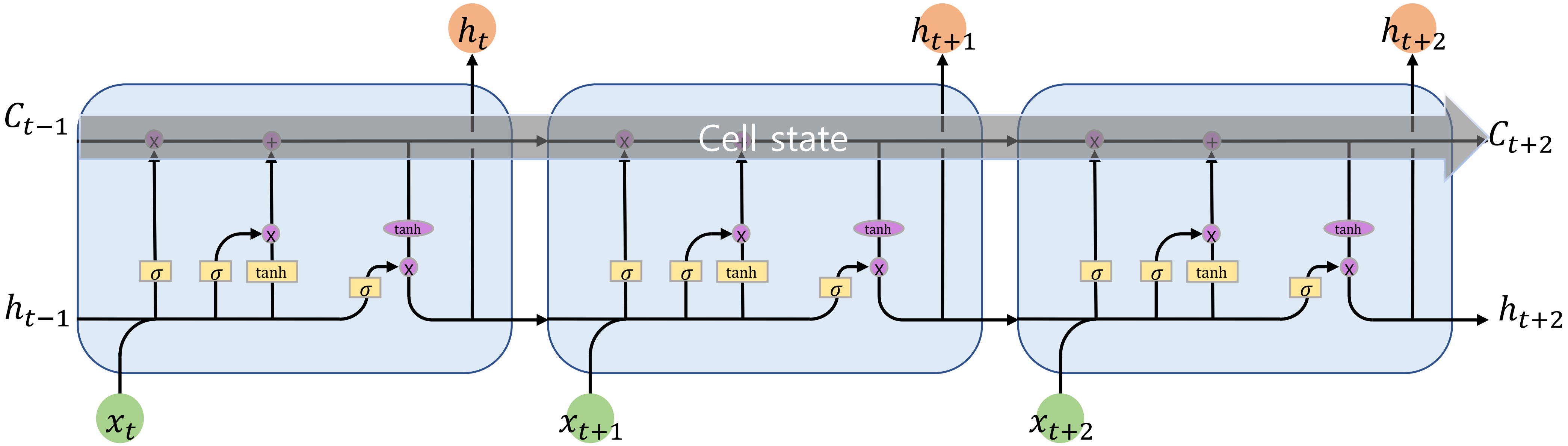
Introduction

- 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.

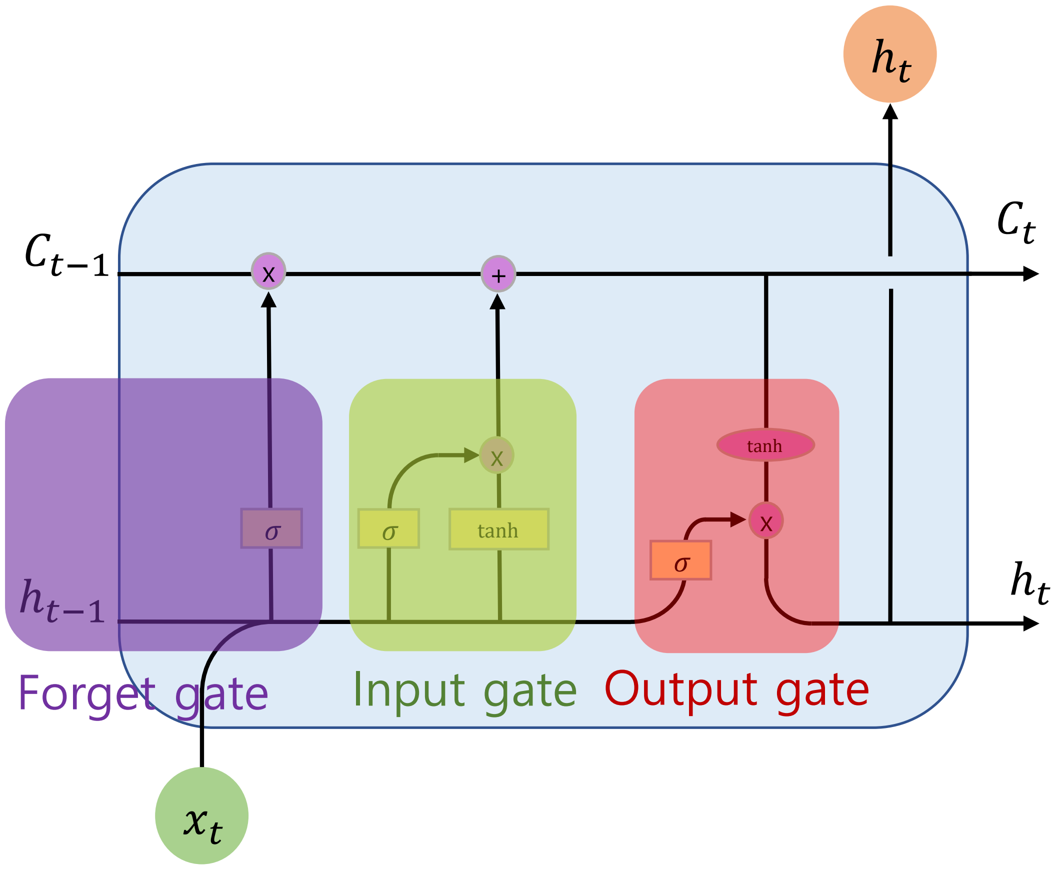
Structure

- **Gate**

  - Forget gate : 과거의 cell state의 어떤 부분을 얼마나 잊을지 결정

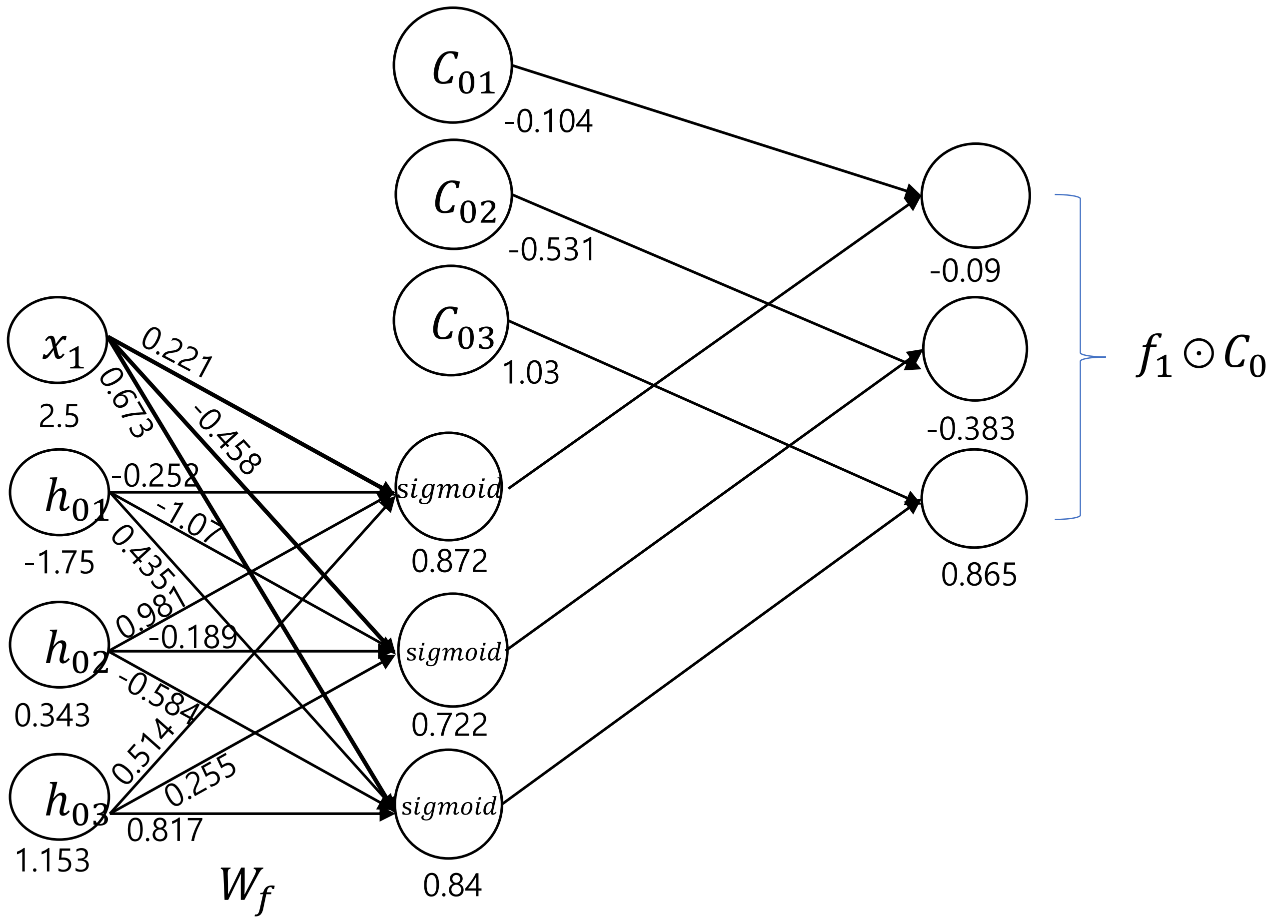
  - Input gate : cell state에 현재 input값 중 어떤 부분을 얼마나 더할지 결정

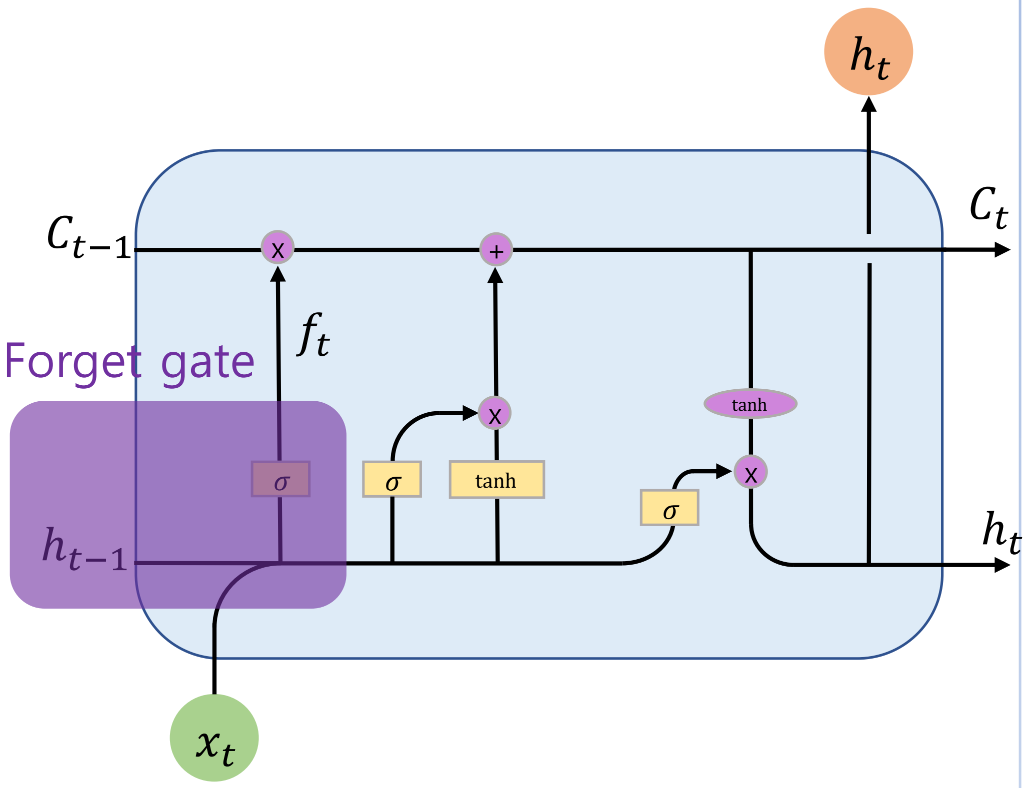
  - Output gate : cell state에 현재 input값 중 어떤 부분을 얼마나 다음 hidden state로 전달할지 결정



Structure

- **Forget gate**

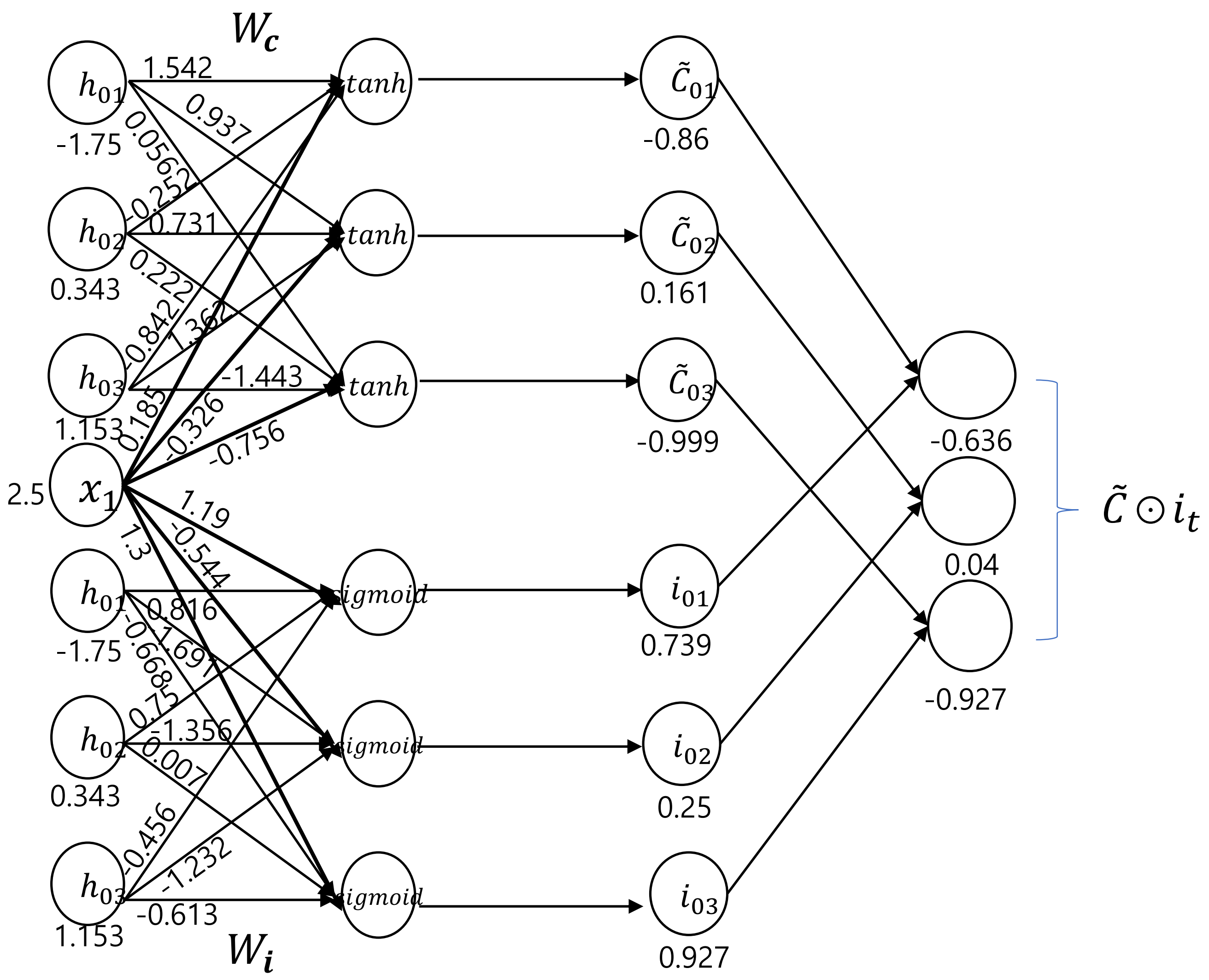


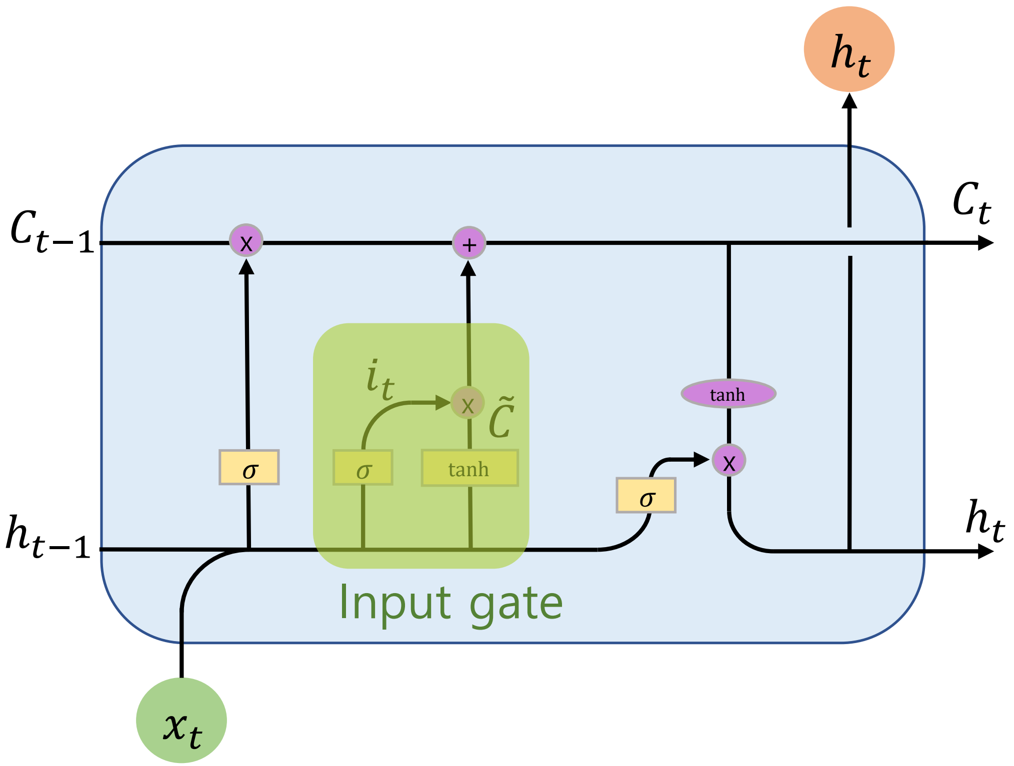


$$f_t = \sigma(W_f[h_{t-1}, x_t])$$

Structure

- **Input gate**



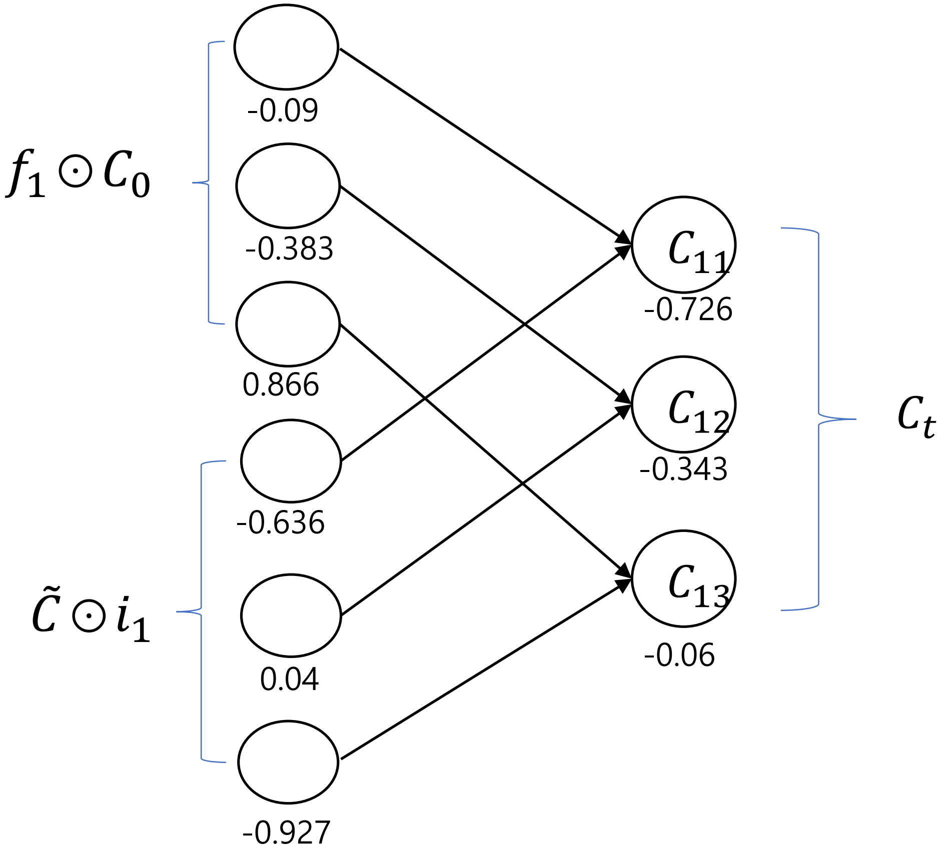


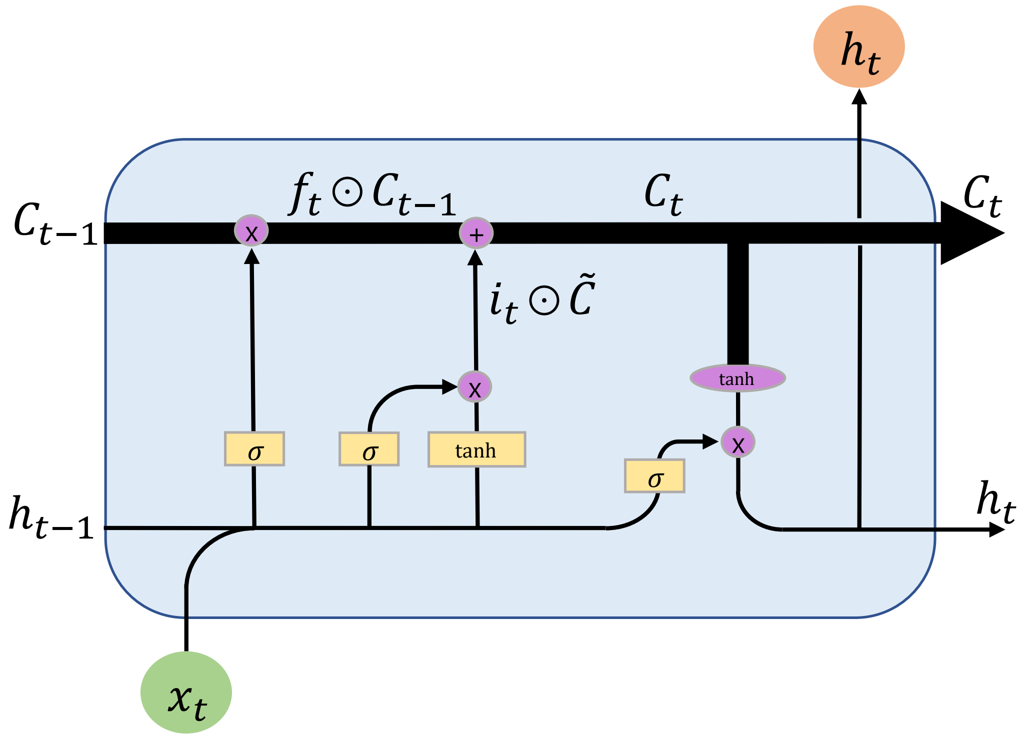
$$i_t = \sigma(W_i[h_{t-1}, x_t])$$

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

Structure

- **Cell state update**

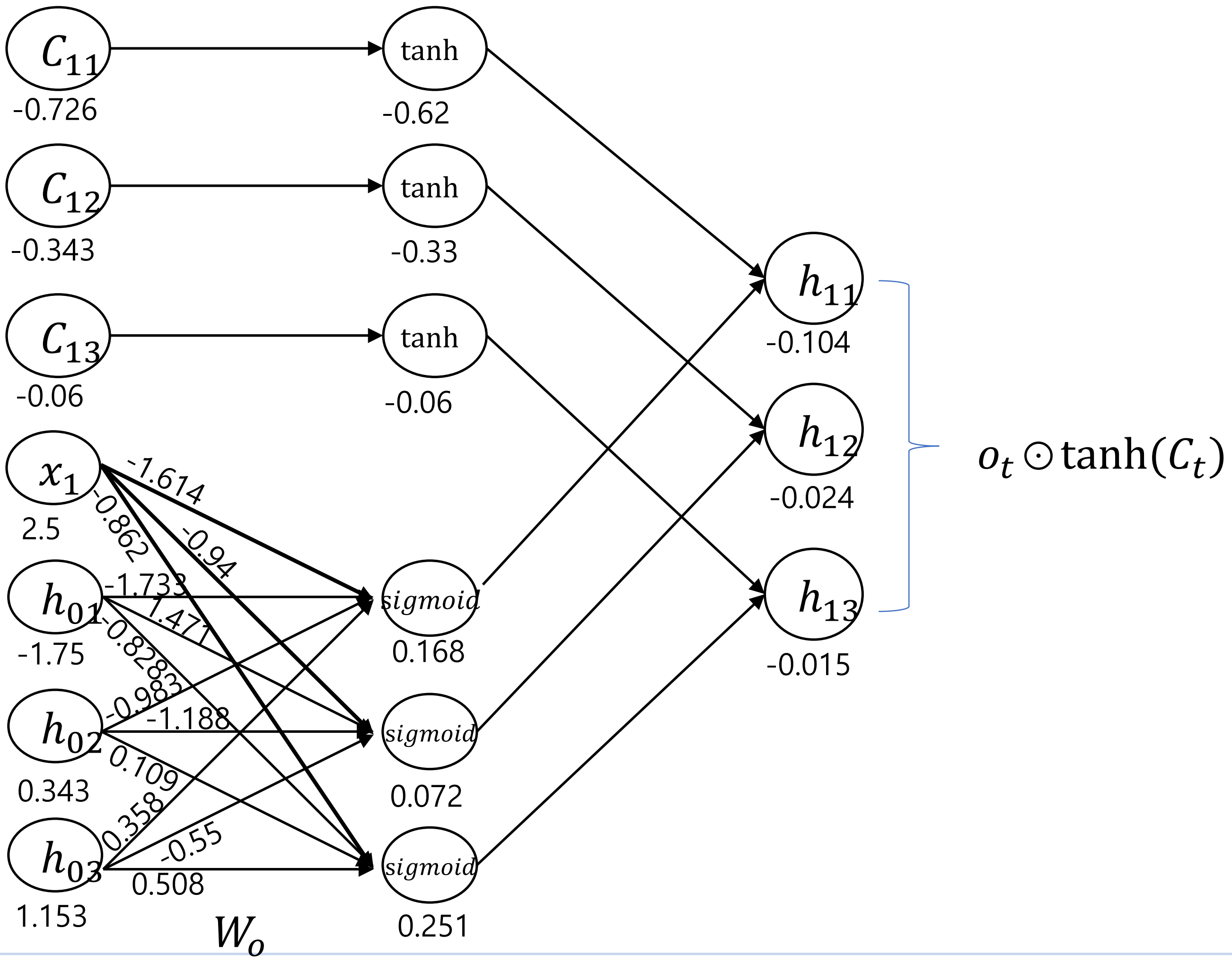


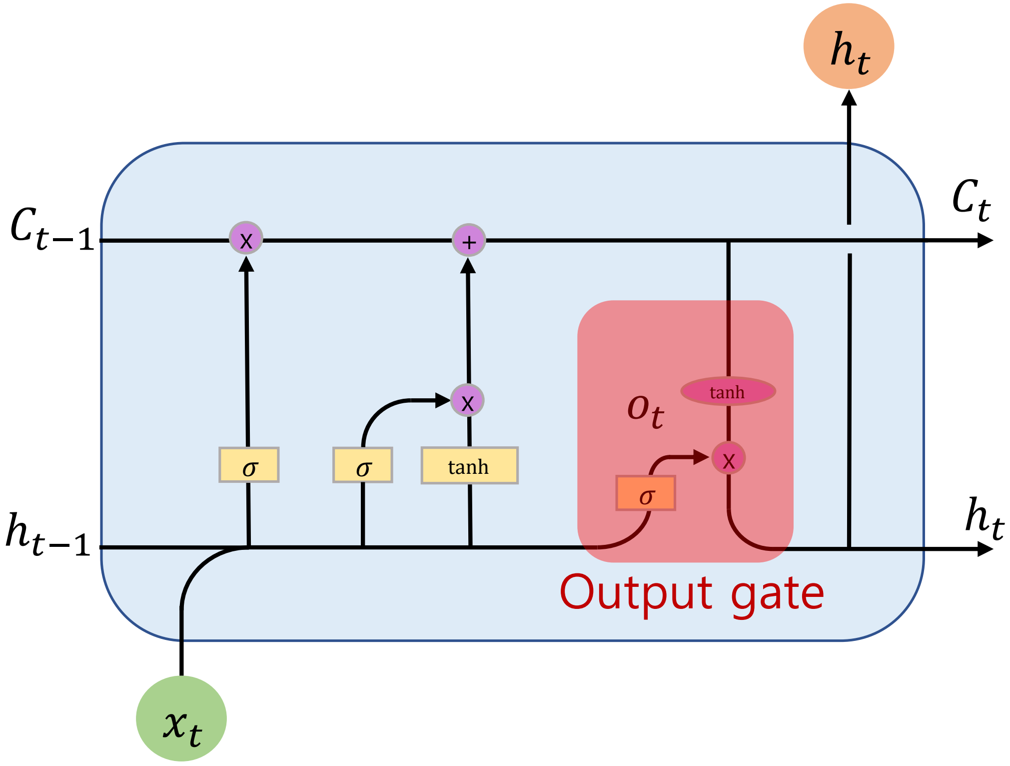


$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}$$

Structure

- **Output gate**





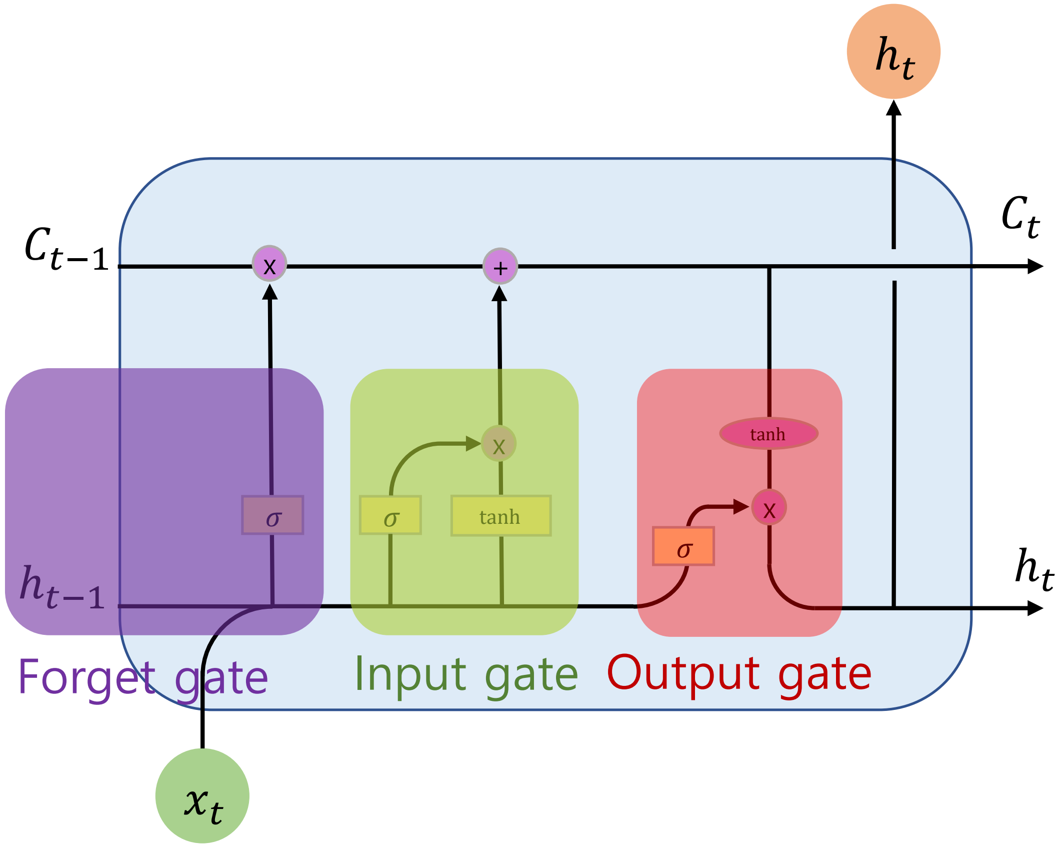
$$o_t = \sigma(W_o[h_{t-1}, x_t])$$
$$h_t = o_t \odot \tanh(C_t)$$

Feedback

- 1)

  - Forget Gate의 "Forget"의 의미.

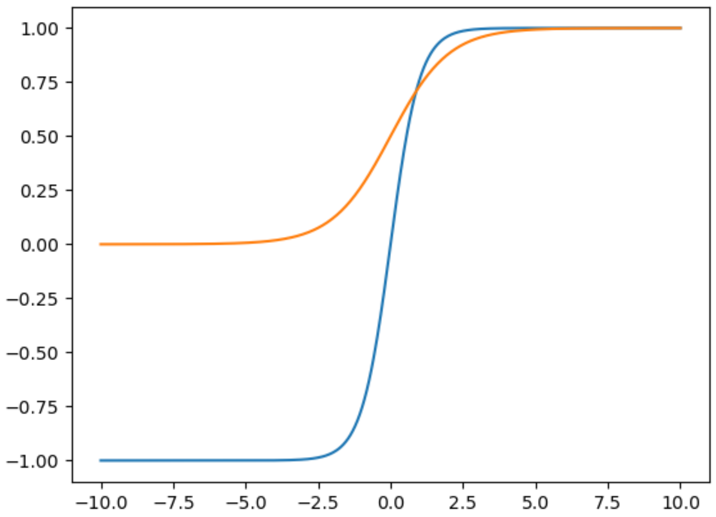
$f_t \odot C_{t-1}$	0.0387	2.0915	-1.0494	1.696	-0.6048	0.099
C_{t-1}	3.87	2.35	-1.06	3.2	-2.88	0.9
f_t	0.01	0.89	0.99	0.53	0.21	0.11

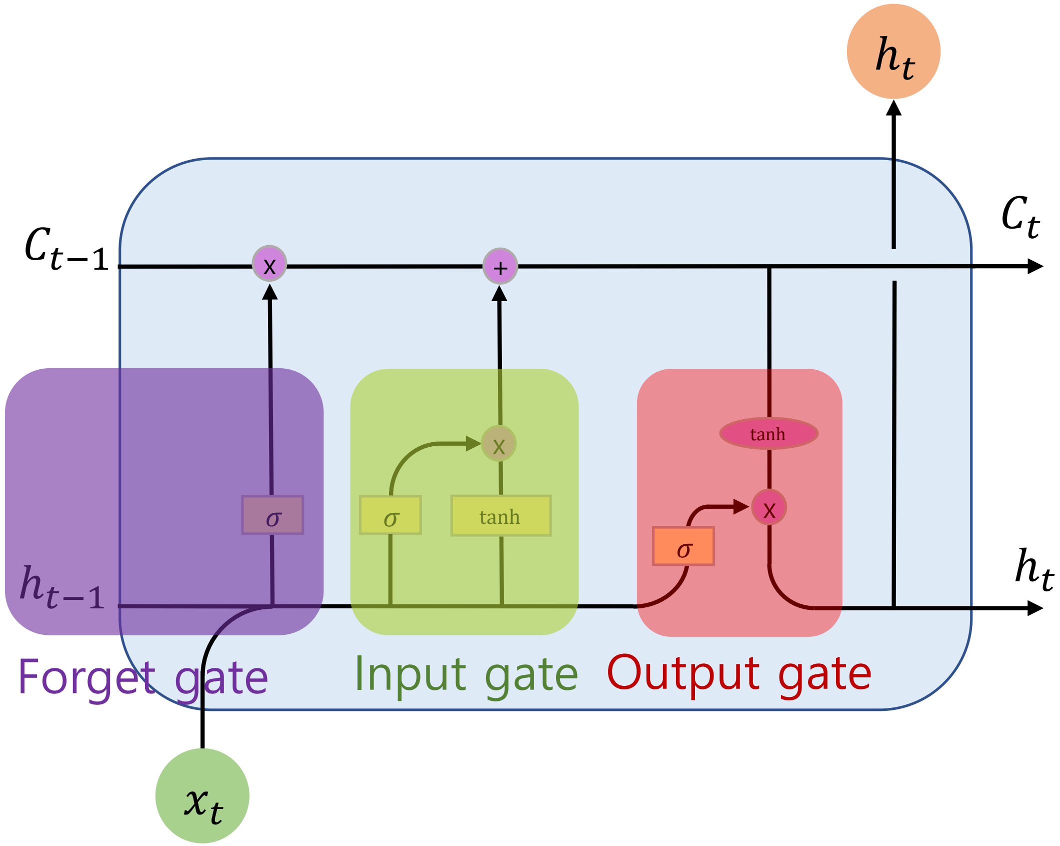


Feedback

- **2), 3)**

  - tanh, sigmoid함수 사용의 차이.





  - +(더하기), *(곱하기)기호의 차이

Experiment-lstm

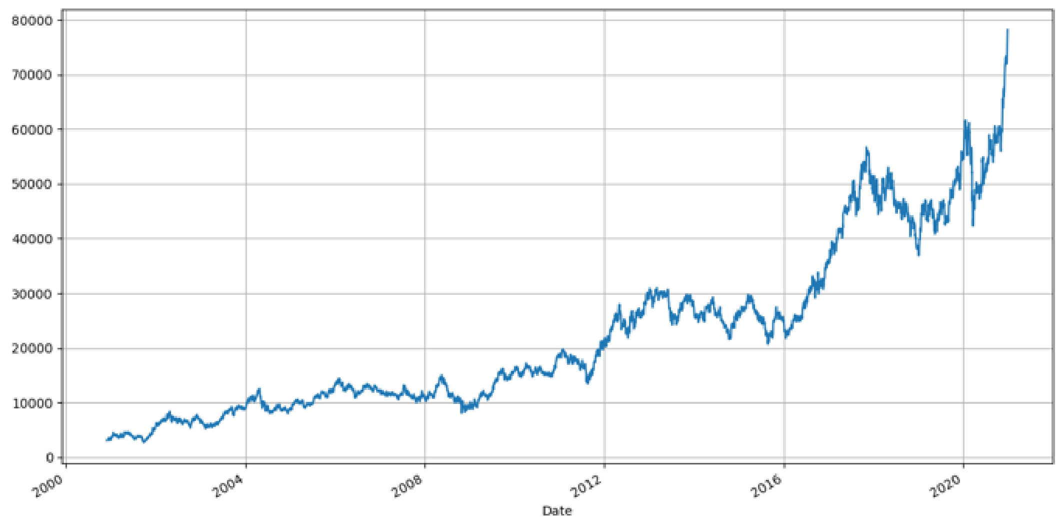
- Dataset

  - 삼성전자 주가 데이터

  - train data : 2000.1.1 ~ 2011.12.31 (3010개)

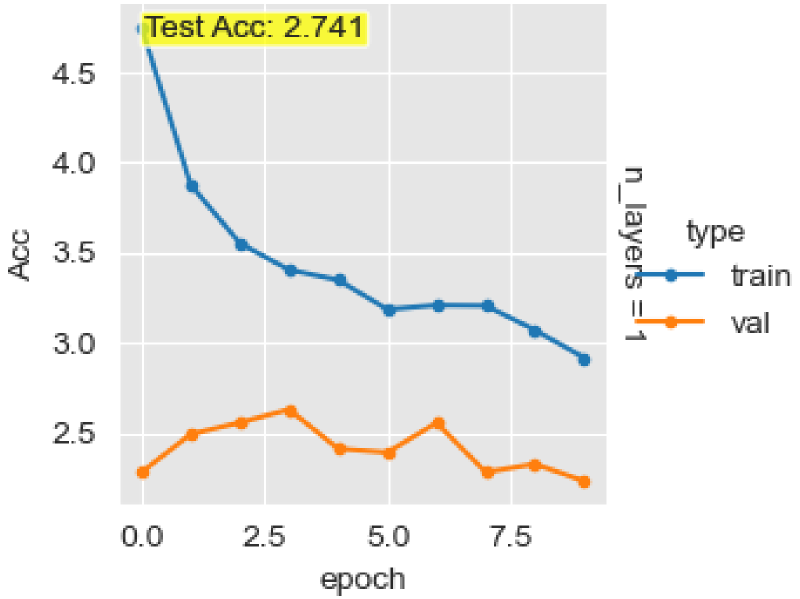
  - validation data : 2012.1.1 ~ 2015.12.31 (975개)

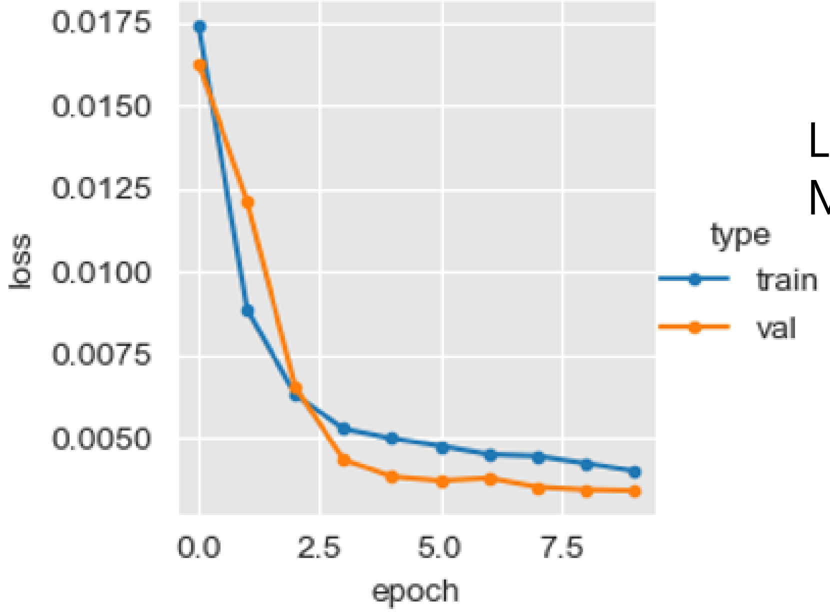
  - test data : 2016.1.1 ~ 2019.2.1(736)



Experiment

Accuracy의 계산에 MAE를 사용





Loss의 계산에 MSELoss를 사용

- **hidden state, cell state dimension : 50**
- **input sequence length = 10, output sequence length = 5**
- **epoch = 10**