

RNN

(Recurrent Neural Network)

기초

임세진

<https://youtu.be/wKtAvV0pBPw>

Contents

01. RNN

02. RNN의 한계

03. LSTM



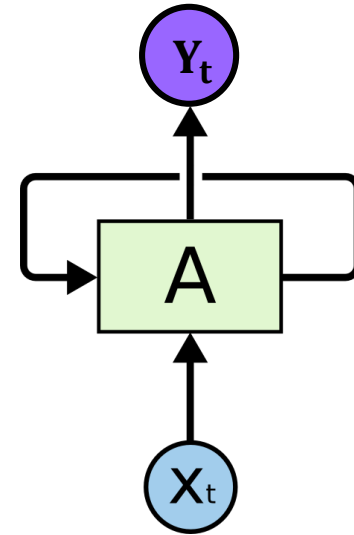
01. RNN



01. RNN

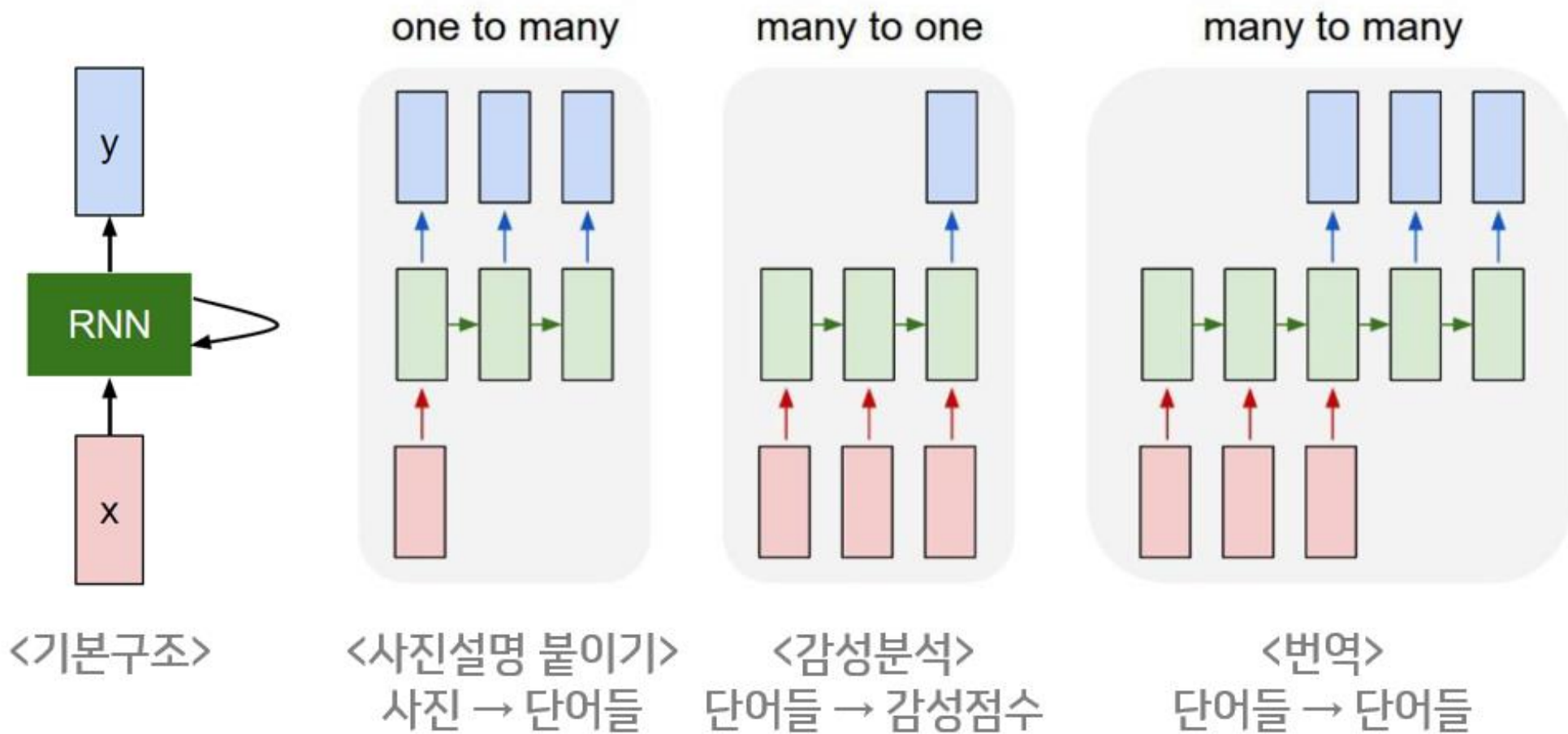
- RNN (^{반복되는}Recurrent Neural Network)

- 순환 신경망
- 앞의 데이터를 잊지 않고 유기적으로 연결시키는 방법
- 전통적인 신경망에서는 모든 입/출력이 서로 독립적이라고 가정
 - ⇒ [문장 > 다음 단어 예측] 시 불리함
- 앞뒤 순서가 존재하는 시계열 데이터에 강력한 성능을 보여줌
 - ⇒ 언어 모델링 및 텍스트 생성, 번역, 음성인식, 이미지 > 설명 생성



RNN 구조

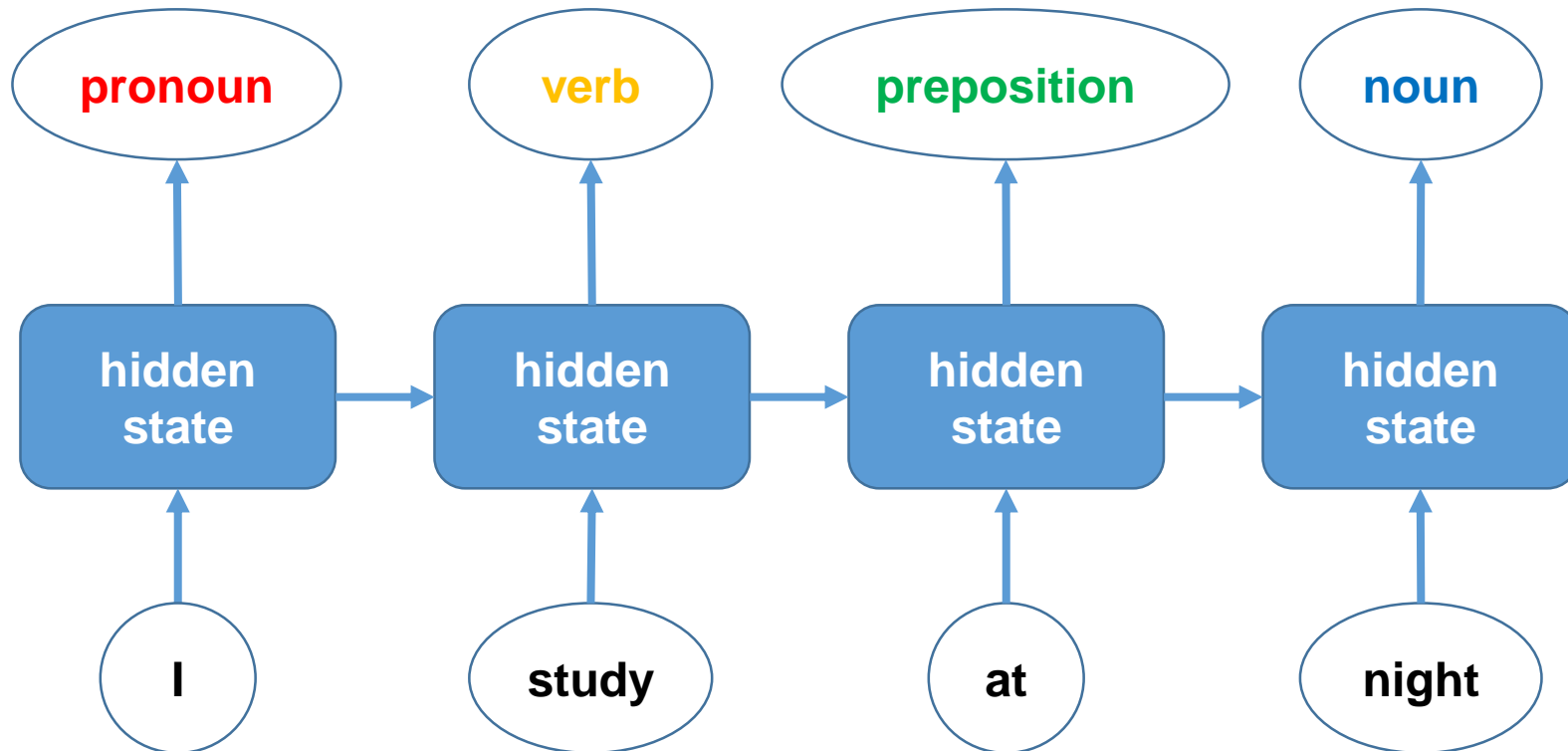
01. RNN



01. RNN

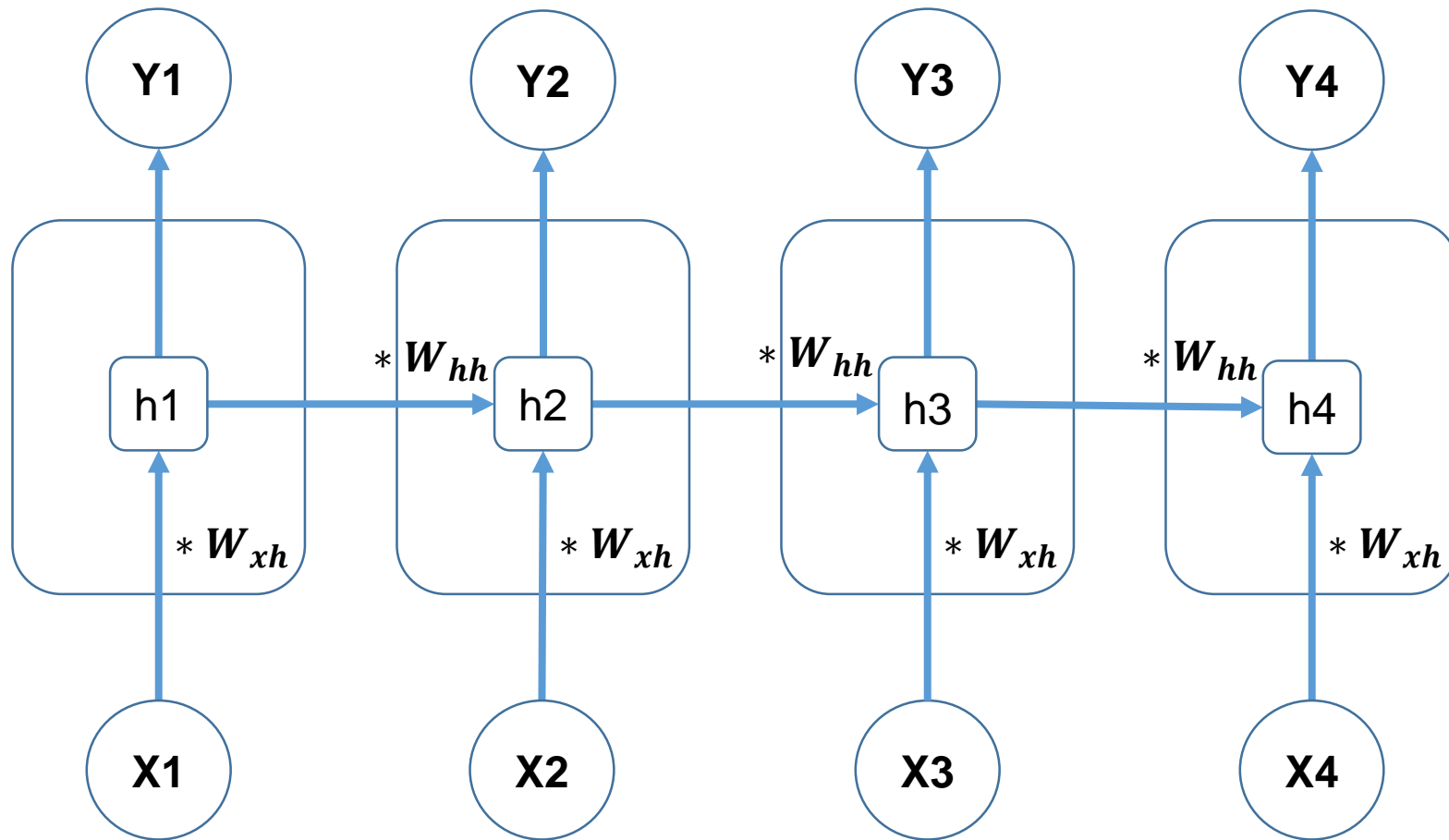
예 1) POS tagging (문법 태깅)

I study at night
pronoun verb preposition noun

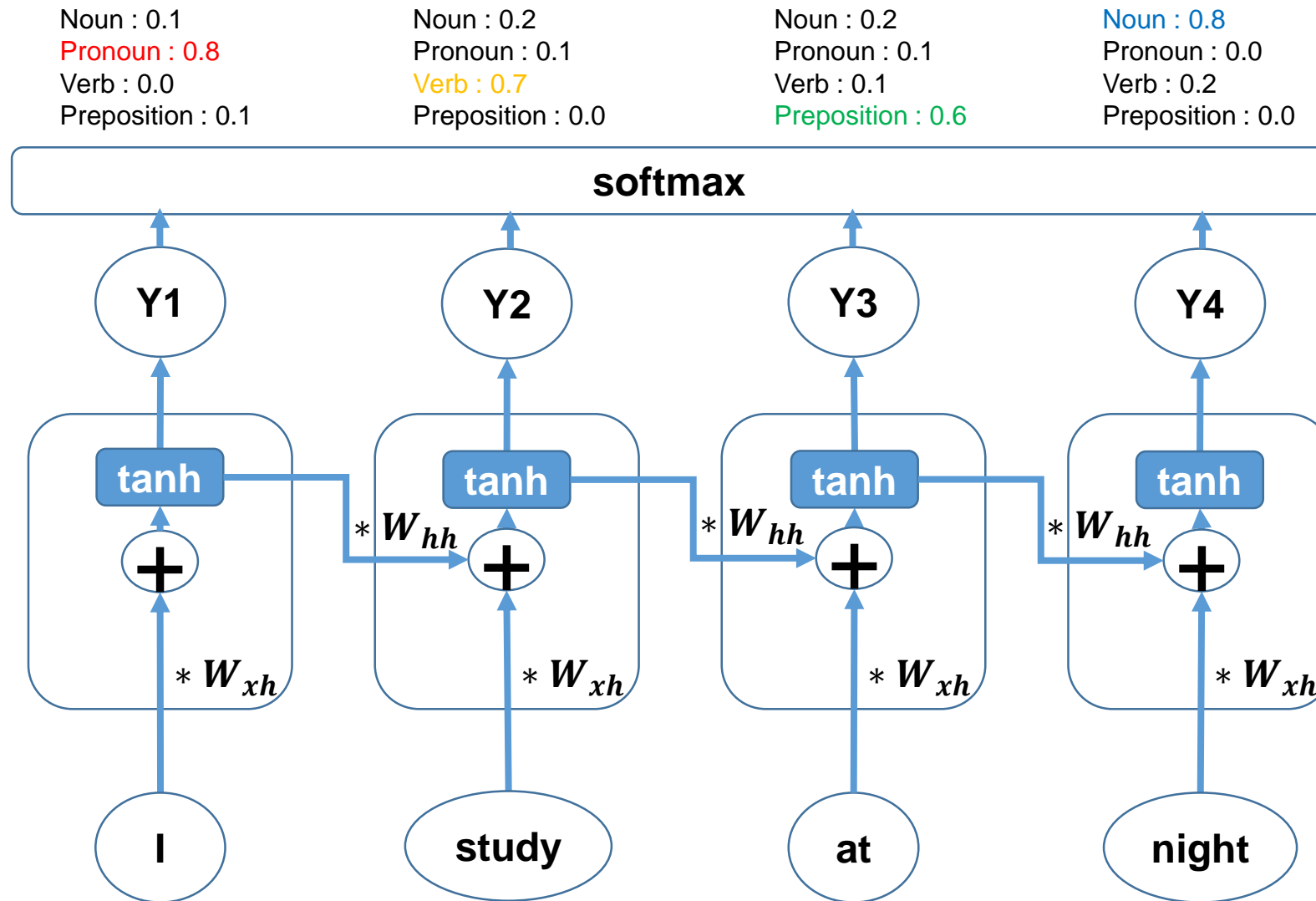


01. RNN

예 1) POS tagging (문법 태깅)

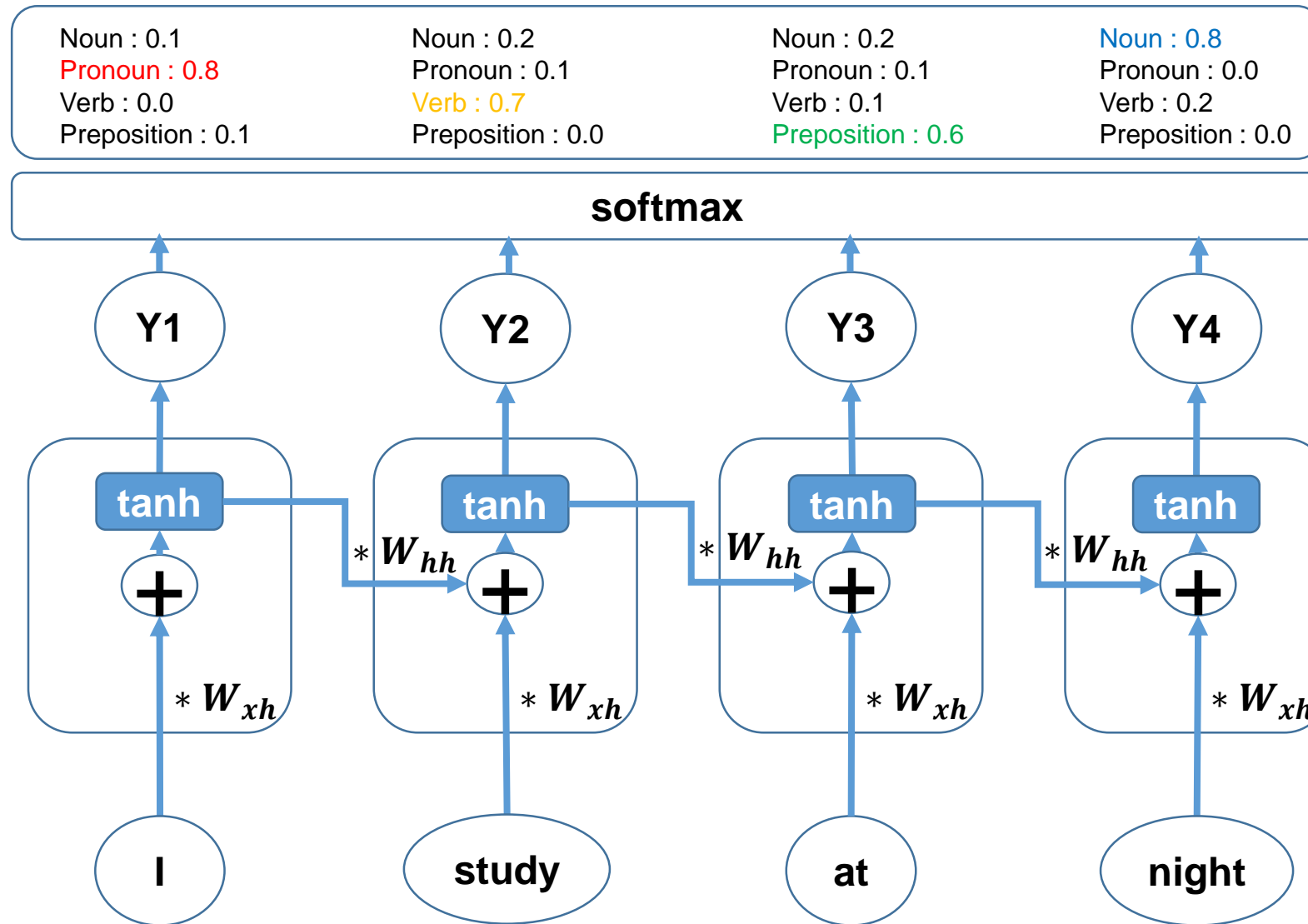


01. RNN



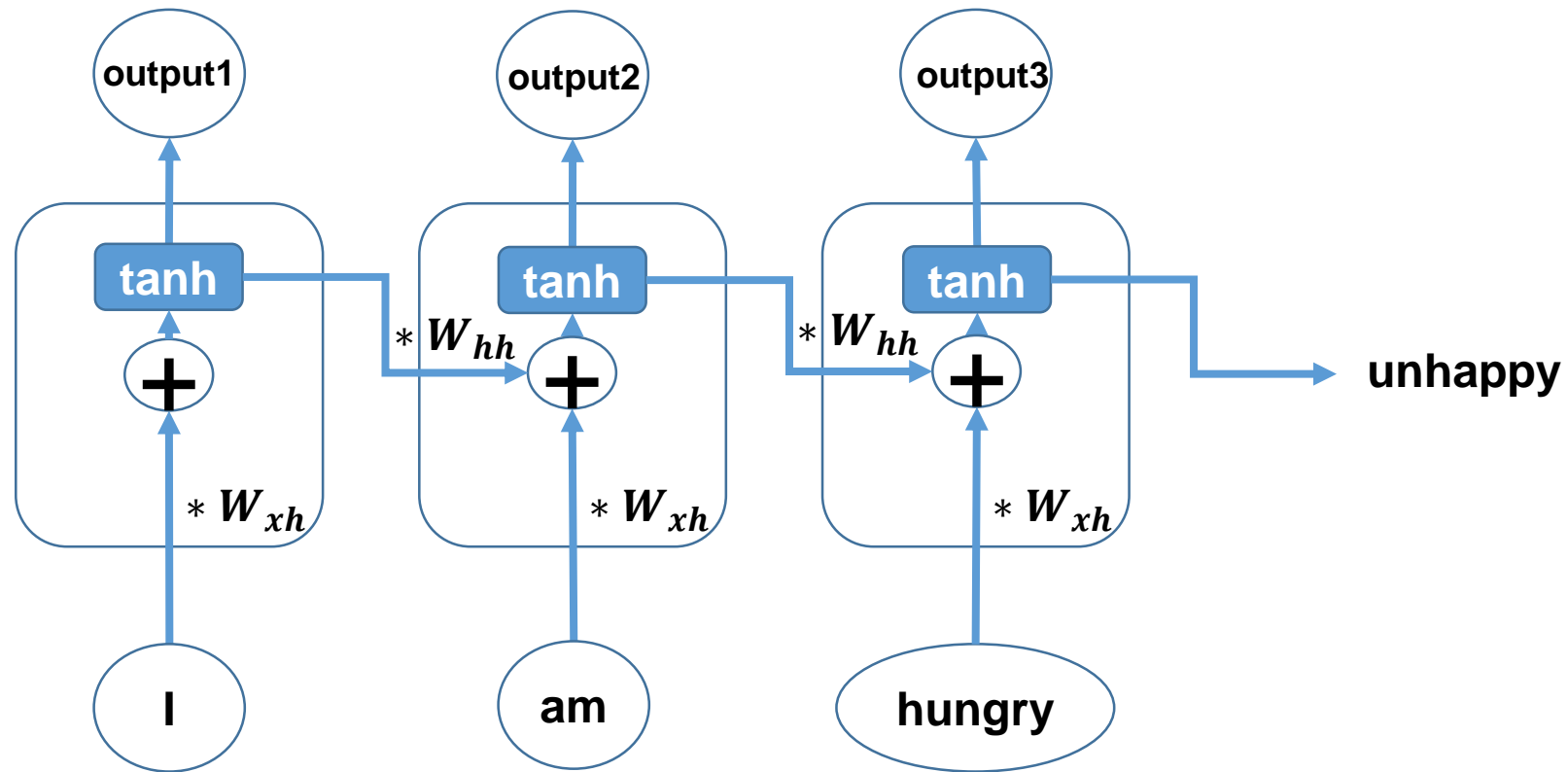
01. RNN

Minimize
difference



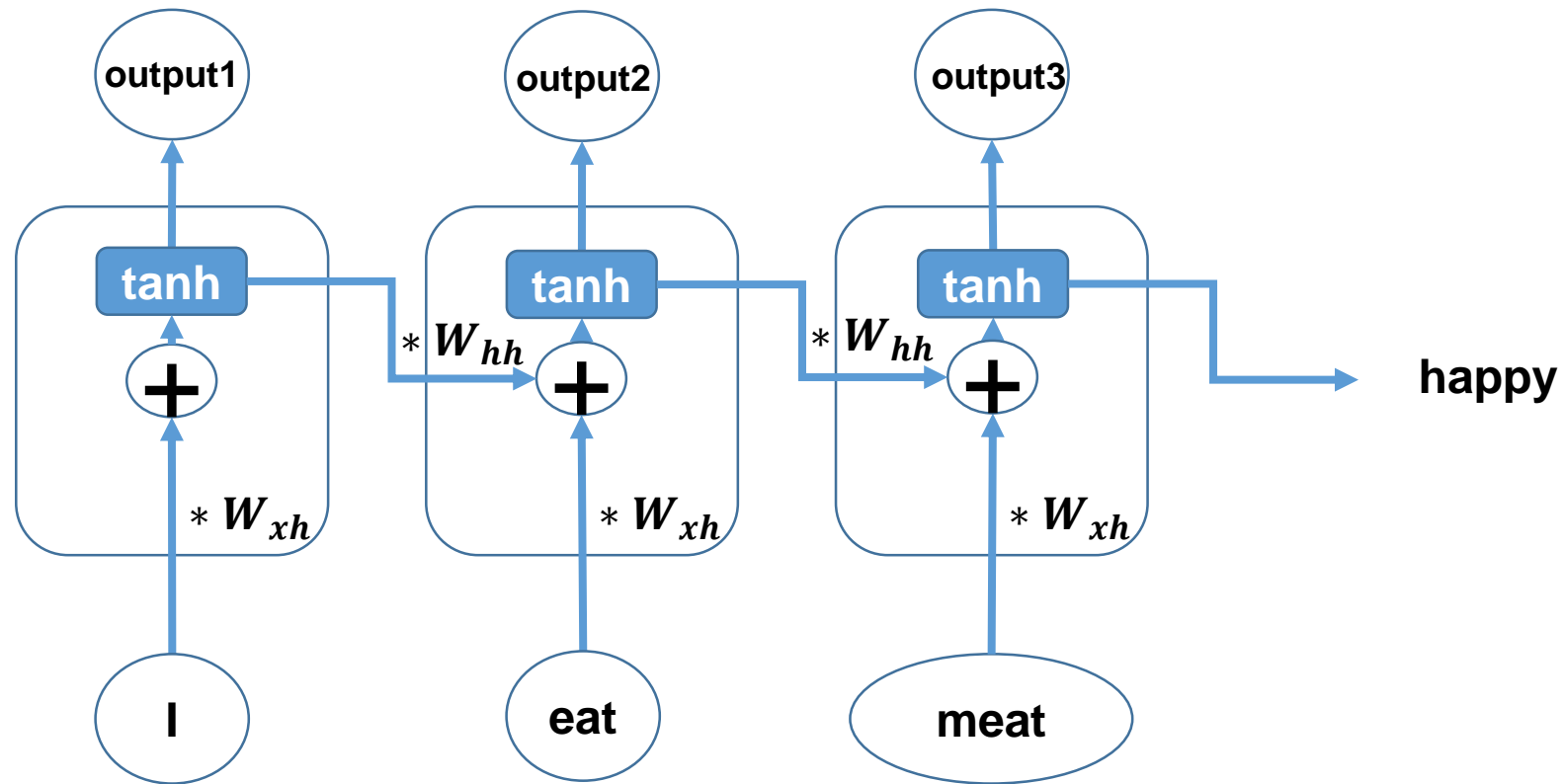
01. RNN

예 2) Sentiment Analysis (감성 분석)



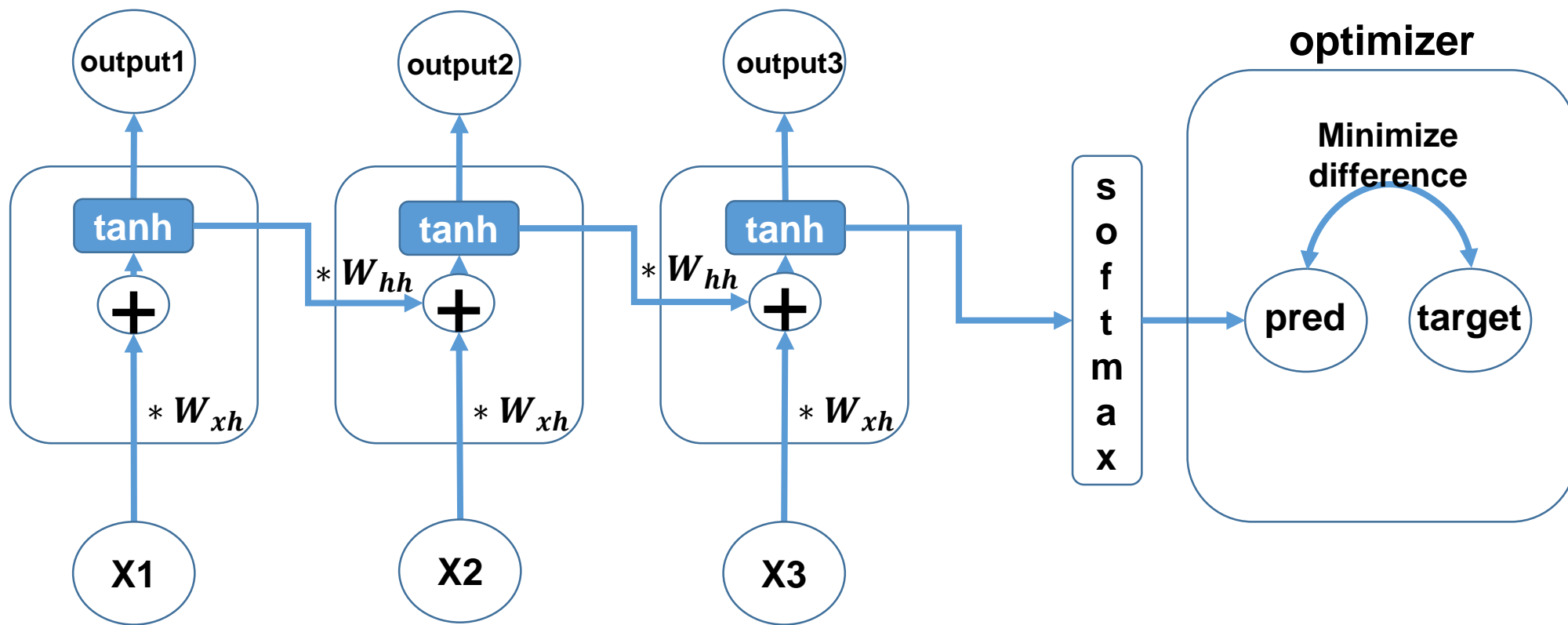
01. RNN

예 2) Sentiment Analysis (감성 분석)



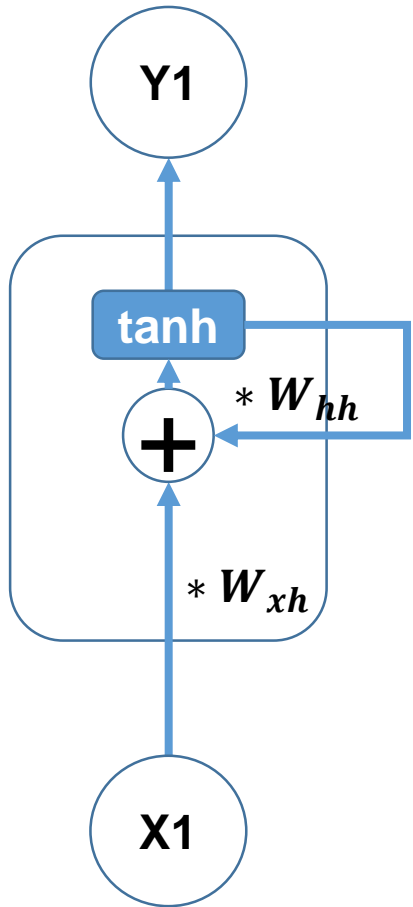
01. RNN

예 2) Sentiment Analysis (감성 분석)



01. RNN

- RNN의 기본 구조

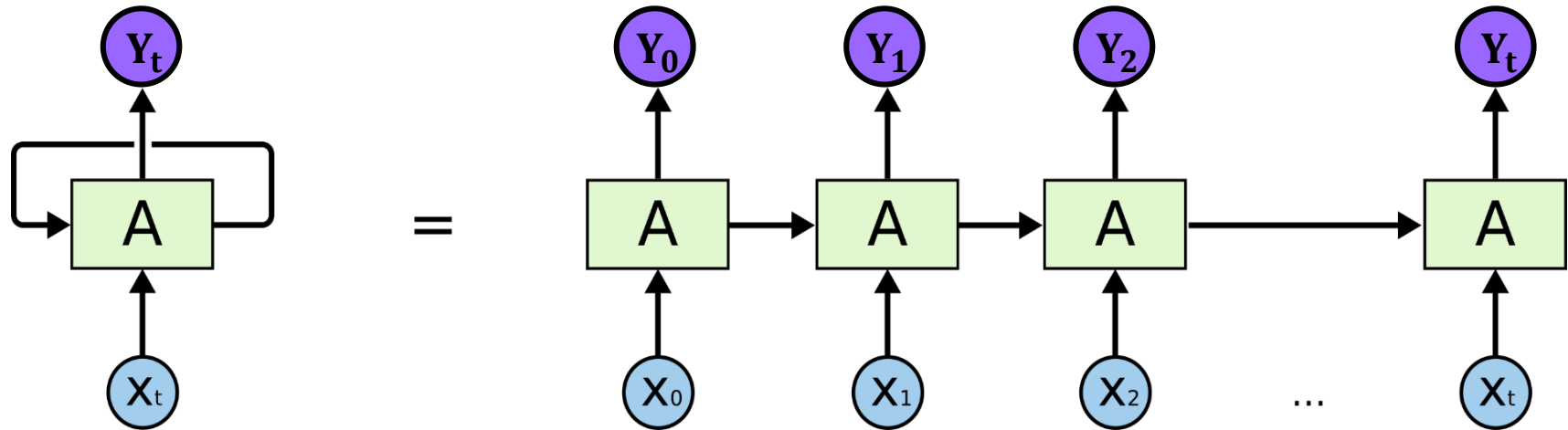


New state Old state

$$h_t = fw(h_{t-1}, x_t)$$

어떤 함수

Input vector at this time step



$$h_t = \tanh(W_{xh} * x_t + W_{hh} * h_{t-1})$$

01. RNN

- Activation function (활성화 함수) – tanh

- Hidden state의 활성화 함수로 비선형 함수인 tanh [하이퍼볼릭탄젠트] 사용

- 활성화 함수로 **비선형 함수**를 사용하는 이유

- ⇒ 선형함수인 $h(x) = cx$ 를 활성화함수로 사용한 3층 네트워크를 가정

- ⇒ 식으로 표현하면 $y(x) = h(h(h(x)))$ 즉 $y(x) = c * c * c * x$ 이다.

- ⇒ 세번의 곱셈을 수행하지만 $y(x) = ax$ 와 같다.

- ⇒ 즉, hidden layer가 없는 레이어와 같아짐

- ⇒ 따라서 층을 쌓는 효과를 내고 싶다면 활성화 함수는 반드시 비선형함수를 사용해야 함

02. RNN의 한계

02. RNN의 한계

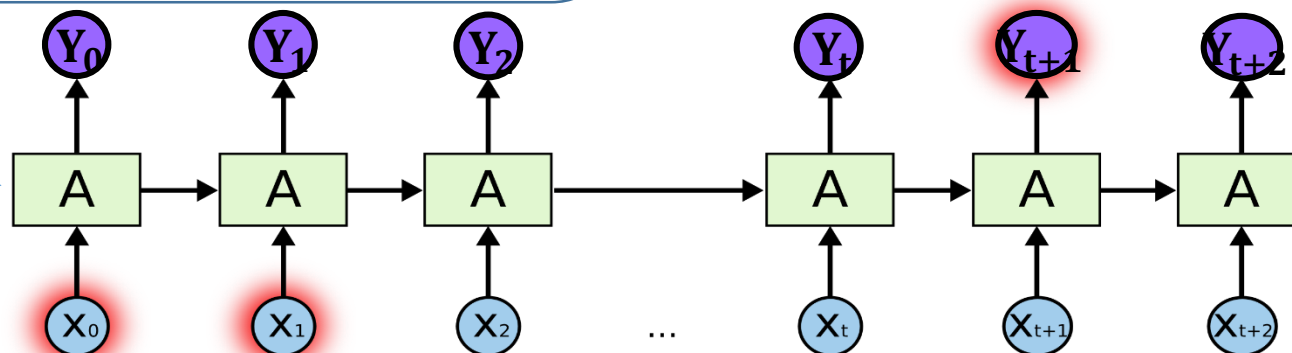
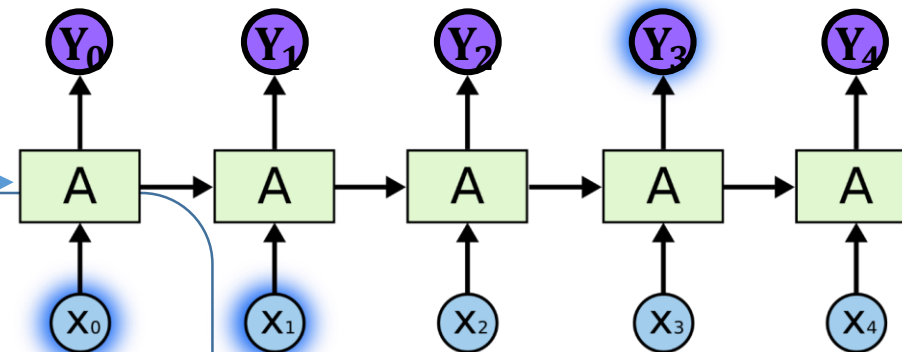
- RNN의 한계

- 시퀀스가 길어졌을 때 문제 발생

나는 한국에서 자랐습니다. 그래서 (한국어)는 너무 편합니다.
요즘은 영어도 잘하고 싶어 공부하고 있지만 쉽지 않습니다.

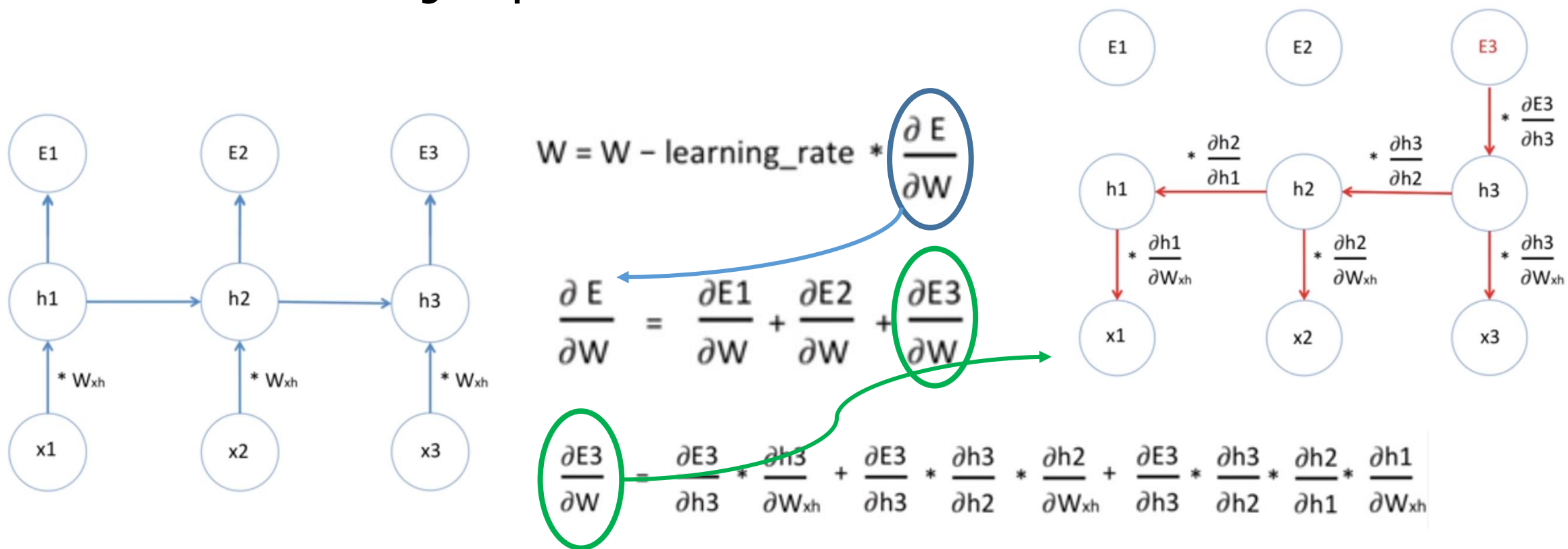
...

영어가 _(???)_만큼 편해지는 그날까지 최선을 다해 공부하겠습니다.



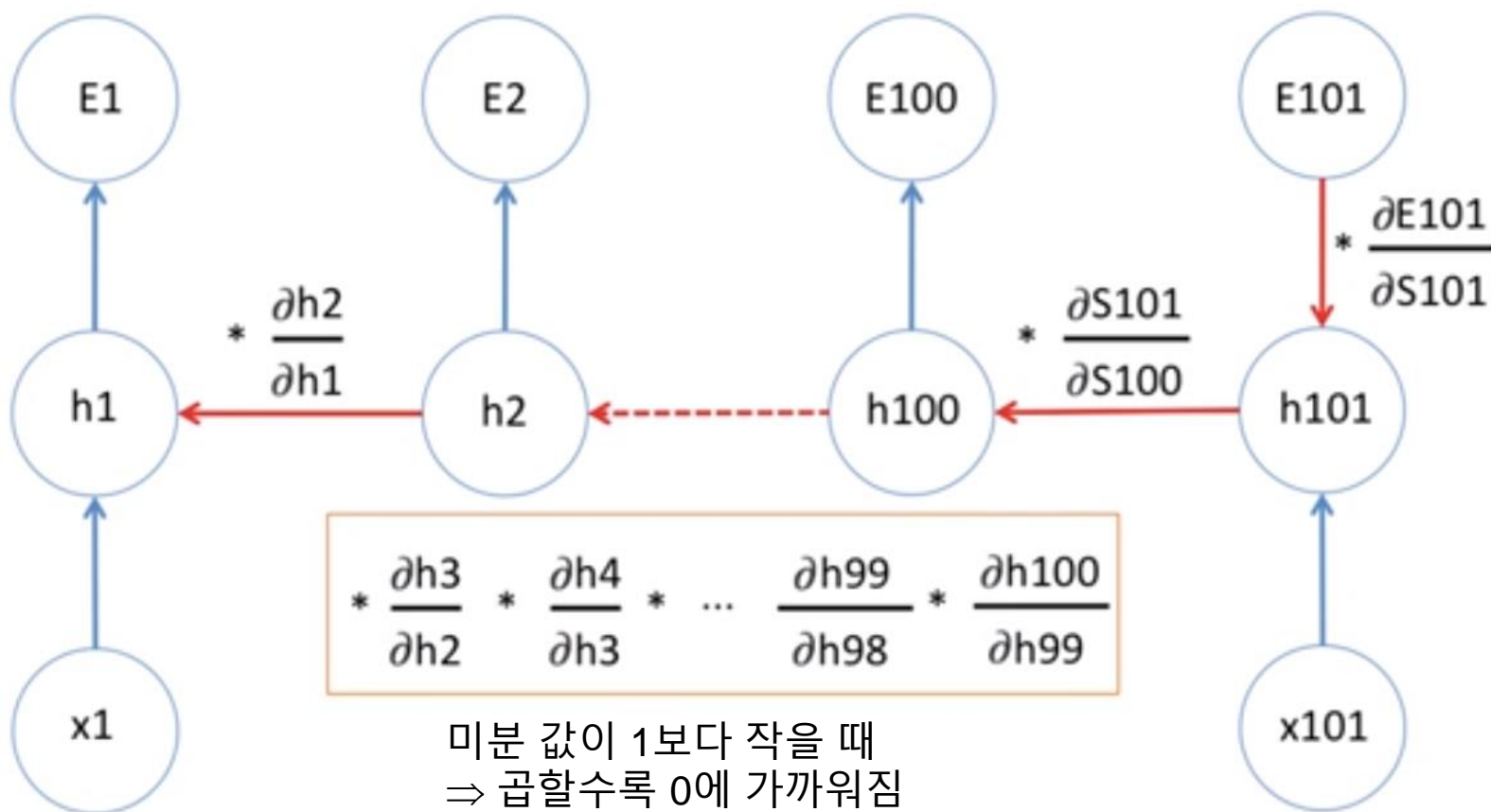
02. RNN의 한계

- Gradient descent weight optimization



02. RNN의 한계

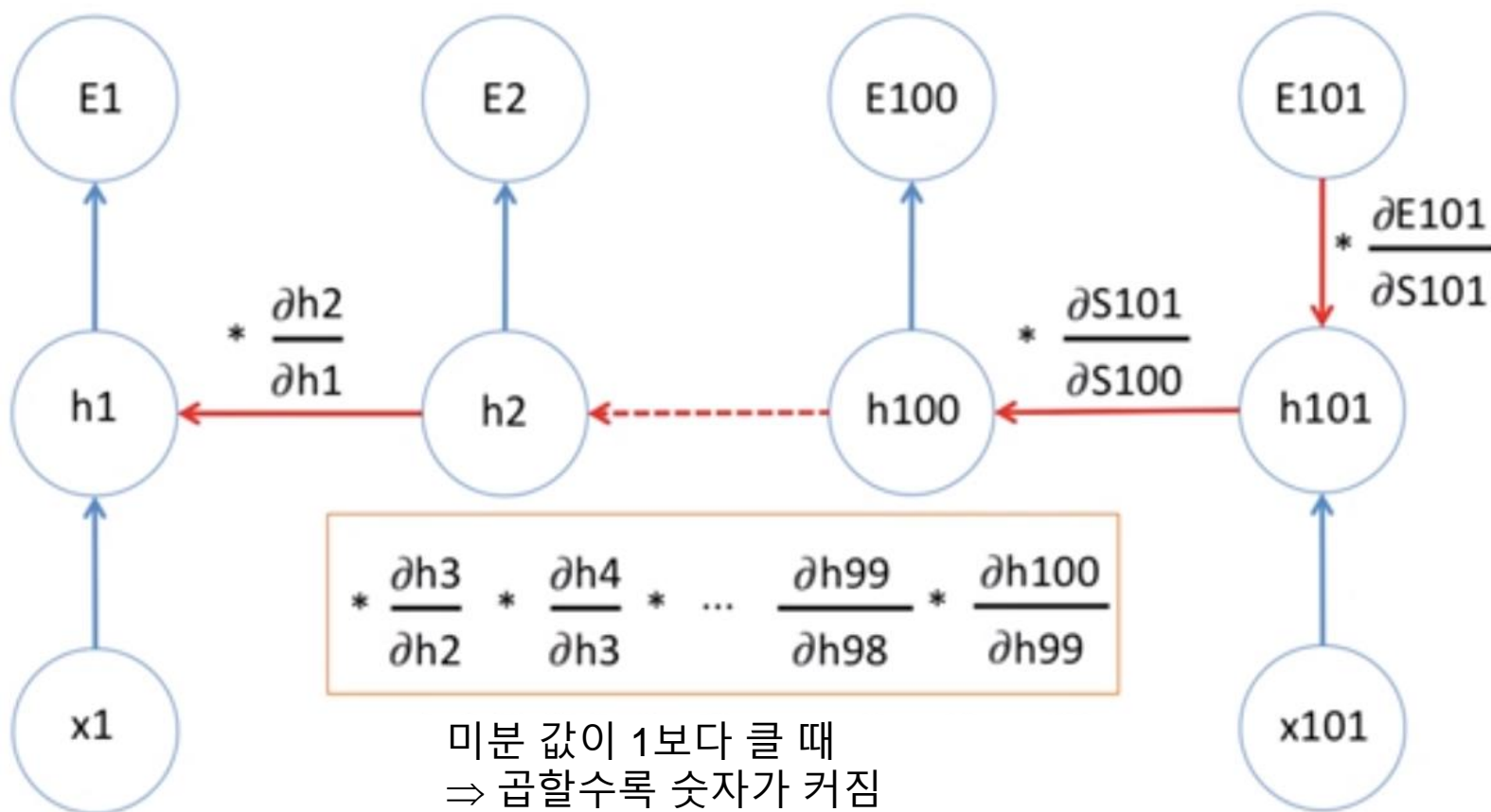
- 시퀀스가 길 때 - Gradient Vanishing 문제



미분 값이 1보다 작을 때
⇒ 곱할수록 0에 가까워짐
⇒ 새로운 Weight value는 기존과 차이 X
⇒ 학습은 길어지고 비효율적

02. RNN의 한계

- 시퀀스가 길 때 - Gradient Exploding 이슈



미분 값이 1보다 클 때
⇒ 곱할수록 숫자가 커짐
⇒ 새로운 Weight value의 변화폭이 큼
⇒ 트레이닝의 방향이 중구난방

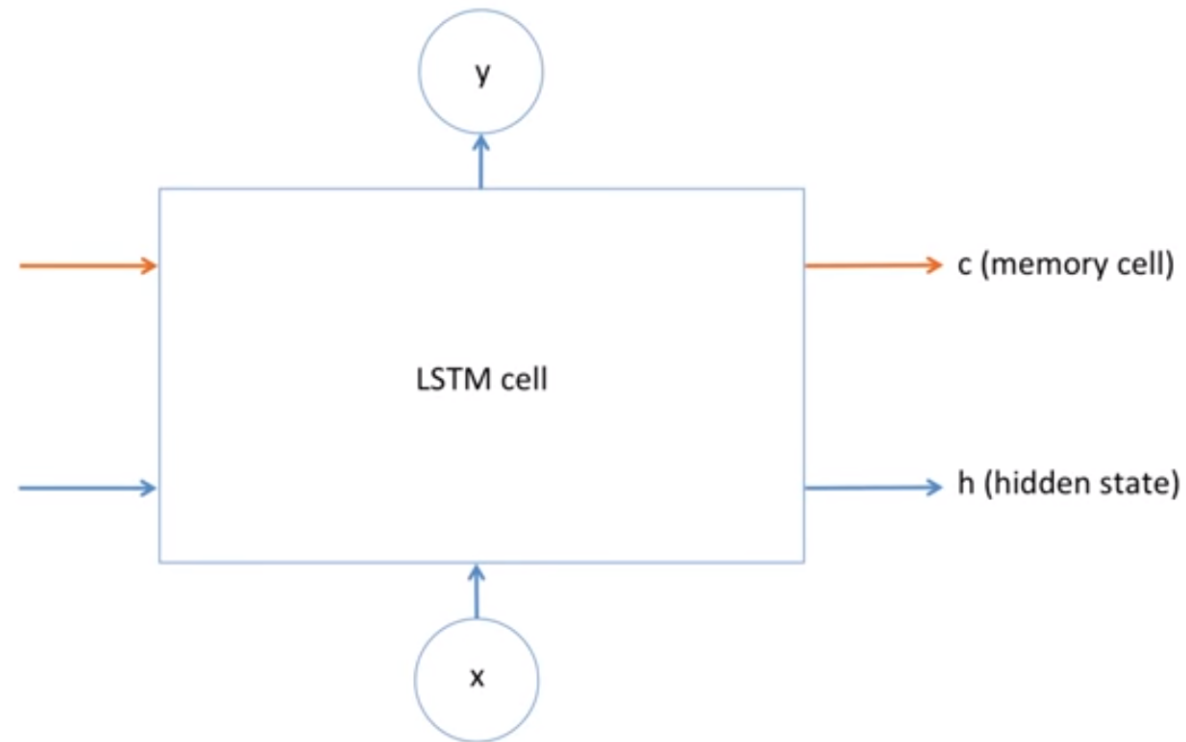
03. LSTM



03. LSTM

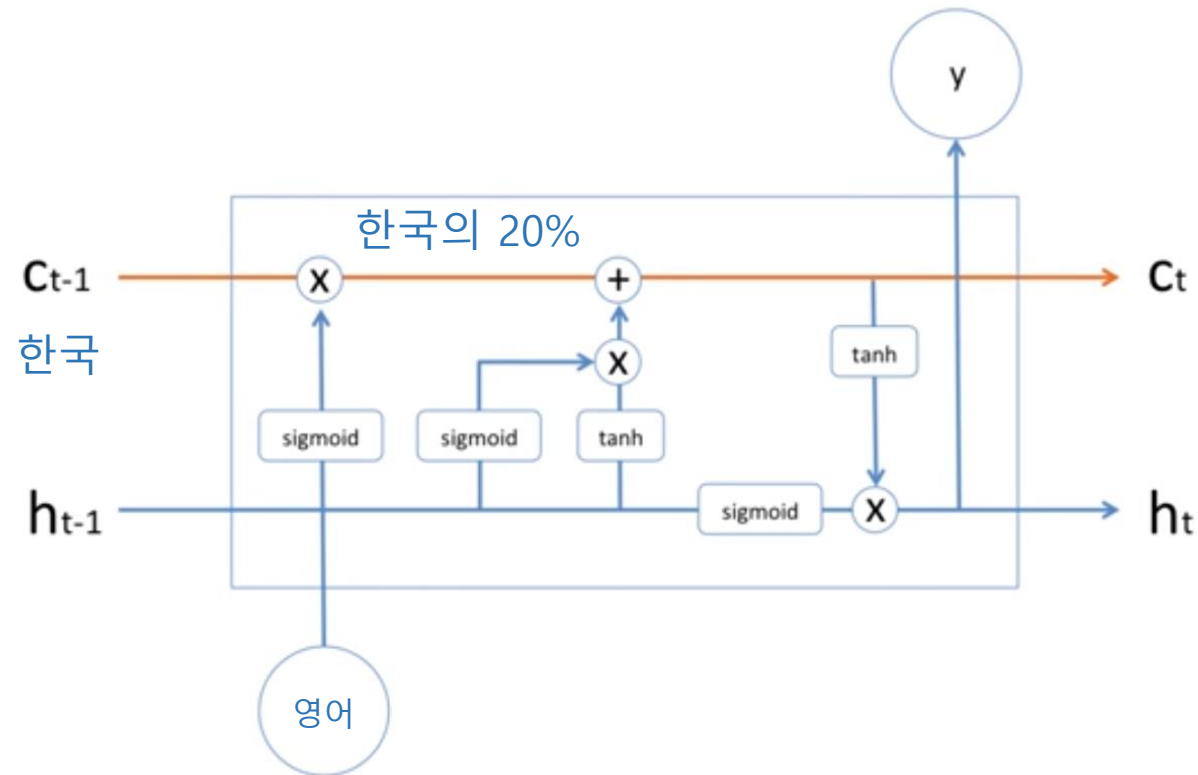
- LSTM (Long Short-Term Memory models)

- 앞의 문제점 해결 >> memory cell 추가 (정보를 기억 / 잊는 메커니즘 추가)



03. LSTM

- Cell state 를 통해 context 유지하여 장기 종속성 확보
- LSTM cell에는 각각 Forget / Input / Output mechanism이 있음



감사합니다

