



BiTamin

순차 데이터와 순환신경망

2조 : 이두진, 이주석, 임홍주

■ CHAPTER 01

Sequence data

Sequential data란?

데이터 집합 내의 객체들이 어떤 **순서**를 가진 데이터

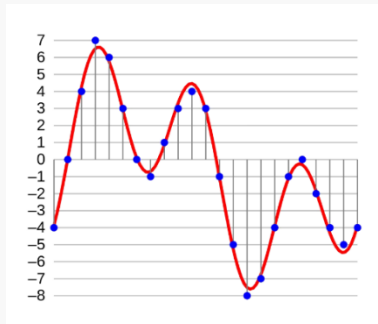
일상에서 벌어지는 대부분의 데이터들은 위치나 순서 혹은 그 사건이 벌어진 시간에 따라 영향을 받고 이에 대한 정보가 데이터의 특징에 많은 기여를 하고 있다.

→ *Sequential data*

01 시계열 데이터

시간	데이터
0	x_0
1	x_1
2	x_2

02 음성 데이터



03 문자 데이터

나는 사과 를 좋아한다

x_0 x_1 x_2 x_3

✓ 이외에도 동영상 데이터, 단백질 염기서열 데이터 등등

머신러닝 모델의 가정(독립항등분포)

일반적인 머신러닝 모델들은 독립항등분포(줄여서 iid: independent and identically distributed)를 가정한다.

T(시간)	X	Y
0	x_0	y_0
1	x_1	y_1
2	x_2	y_2
3	x_3	y_3
4	x_4	y_4
5	x_5	y_5
6	x_6	y_6

- ✓ 하지만, sequential data는 시간이라는 변수에 나머지 변수들이 명확하게 종속
- ✓ 데이터 인스턴스간에도 종속관계이다.

→ 입력변수들이 독립항등분포라는 가정을 위배

독립항등분포란?

각 확률변수들이 상호독립이고 모두 동일한 확률분포를 가지고 있는 상태

왜 sequential model이 필요할까?

일반적인 머신러닝 모델은 기존의 데이터의 집합을 갖고 처음 접하는 입력변수가 주어졌을 때, 기존 데이터의 분포 혹은 동일한 상황을 따른다는 가정하에, 새로운 타겟을 예측

이것이 만족되지 않으면 성능↓

하지만, Sequential data는 iid를 따르지 않으므로, 데이터 인스턴스 간의 **contextual information**을 잘 파악하는 것이 중요하다.

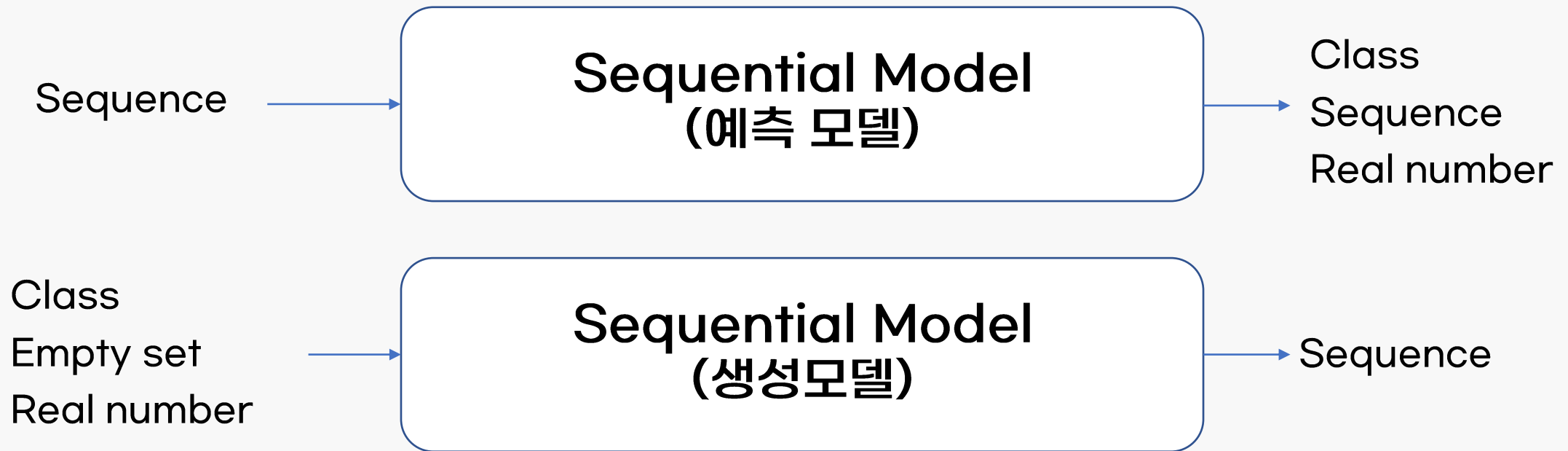
따라서, 데이터 sequence에 집중하여 이것을 효율적으로 처리할 수 있는 sequence model이 따로 필요하게 된 것!

■ CHAPTER 02

Sequential Model

Sequential Model

Sequential Model이란, Sequential data를 입력 혹은 출력으로 사용 하는 모델을 뜻한다.

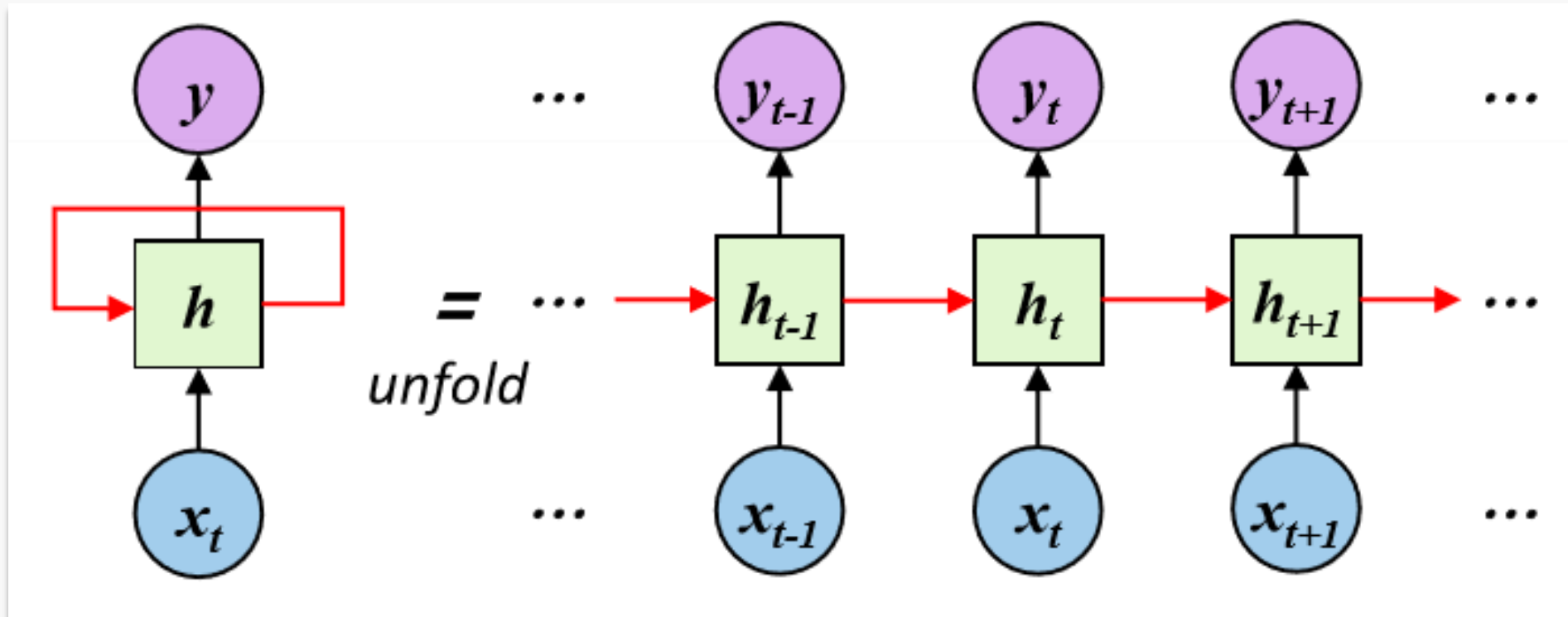


■ CHAPTER 03

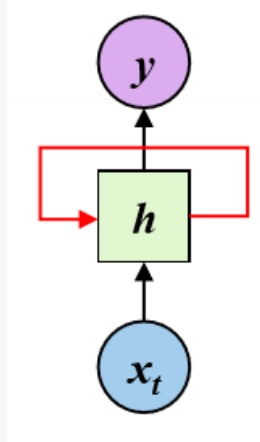
RNN

RNN(Recurrent Neural Network)

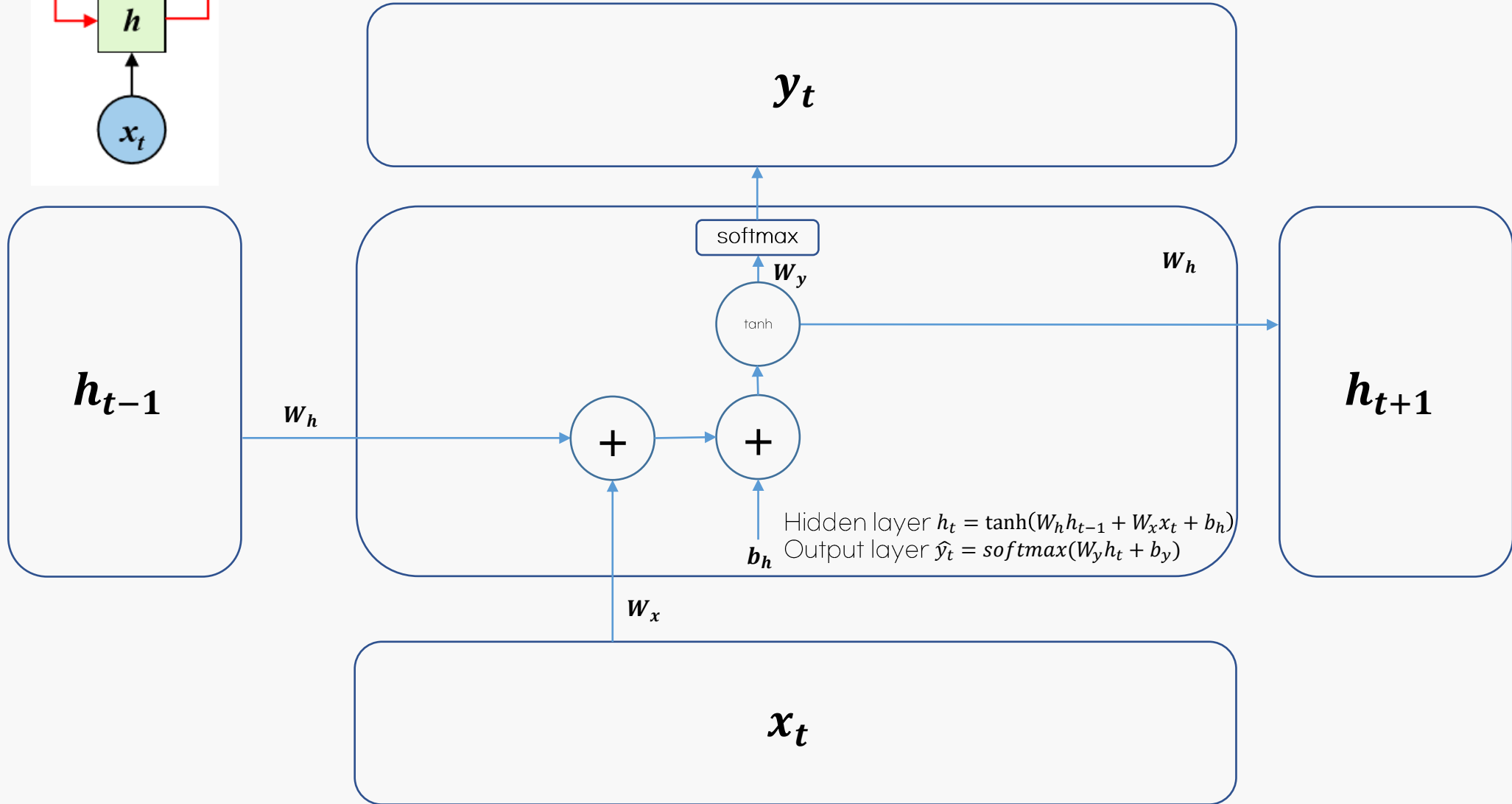
순환신경망, 메모리를 사용하여 sequential data를 처리하는 신경망.



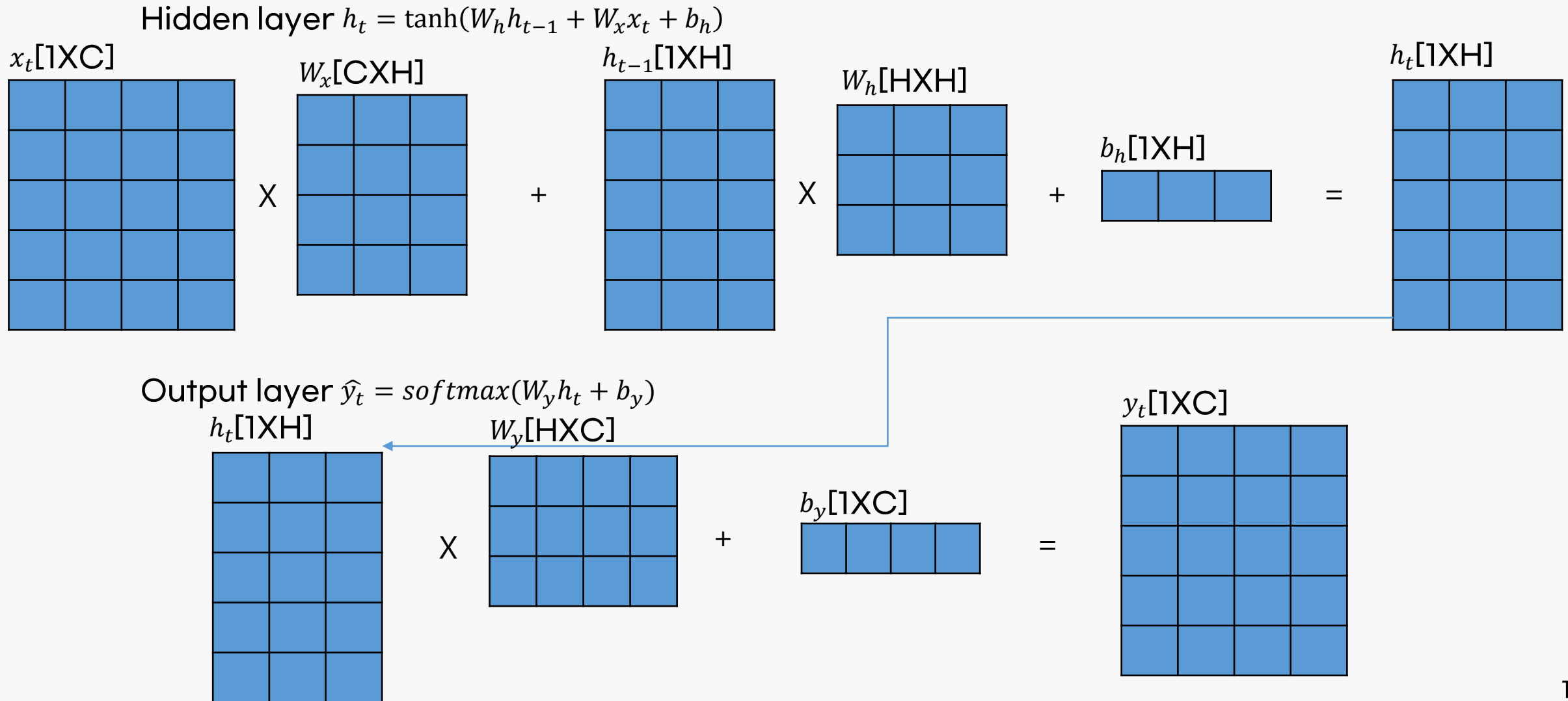
- 은닉상태(hidden state): 시각 t 일 때의 상태, 현재 시점 t 에서의 메모리 셀이 갖고 있는 값.
- Cell: 여러 개의 뉴런이 존재하지만, 뉴런을 모두 표시하지 않고 하나의 셀로 layer를 표시.



RNN의 feed forward(순전파)

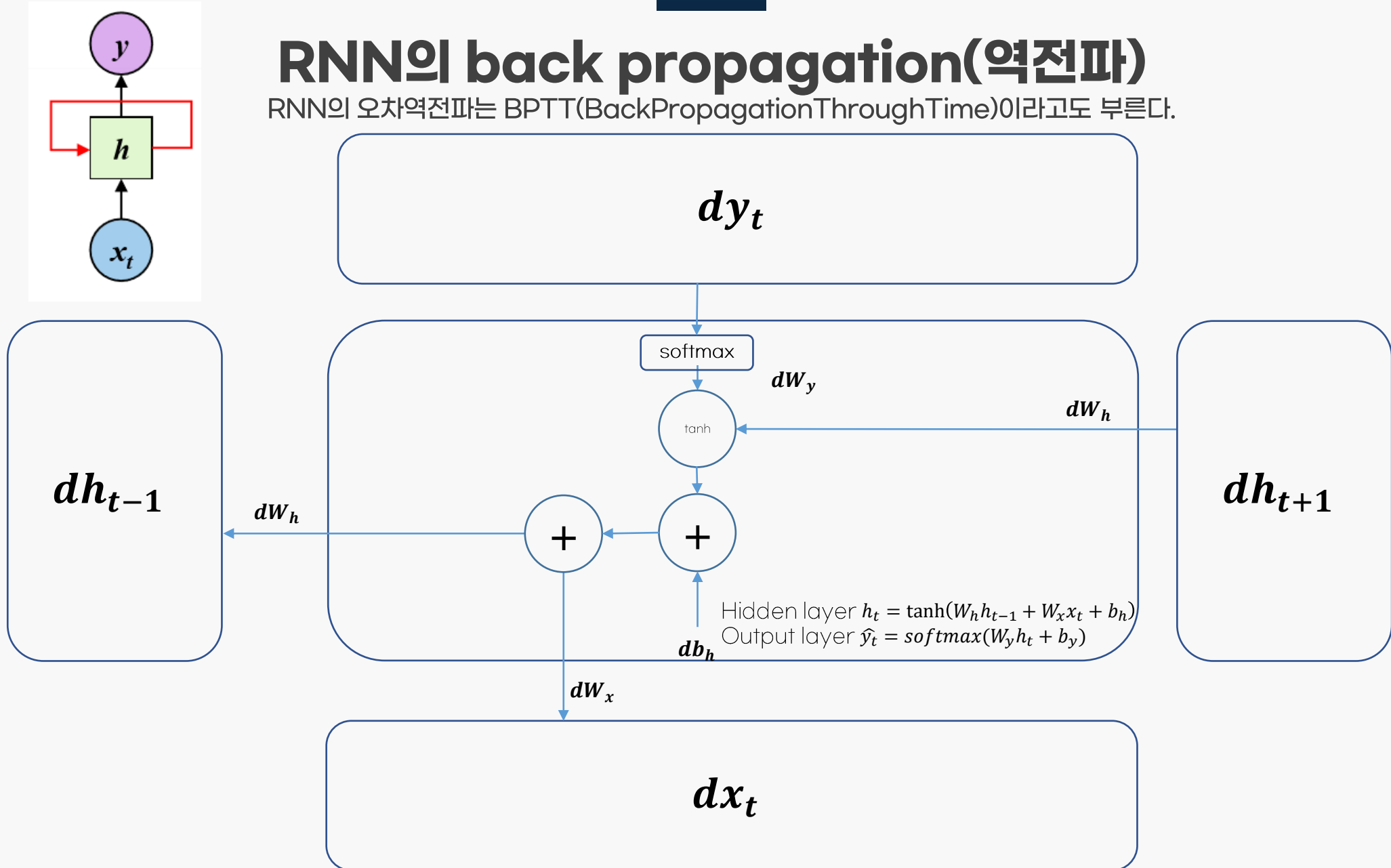


RNN의 feed forward(순전파)



RNN의 back propagation(역전파)

RNN의 오차역전파는 BPTT(BackPropagationThroughTime)이라고도 부른다.

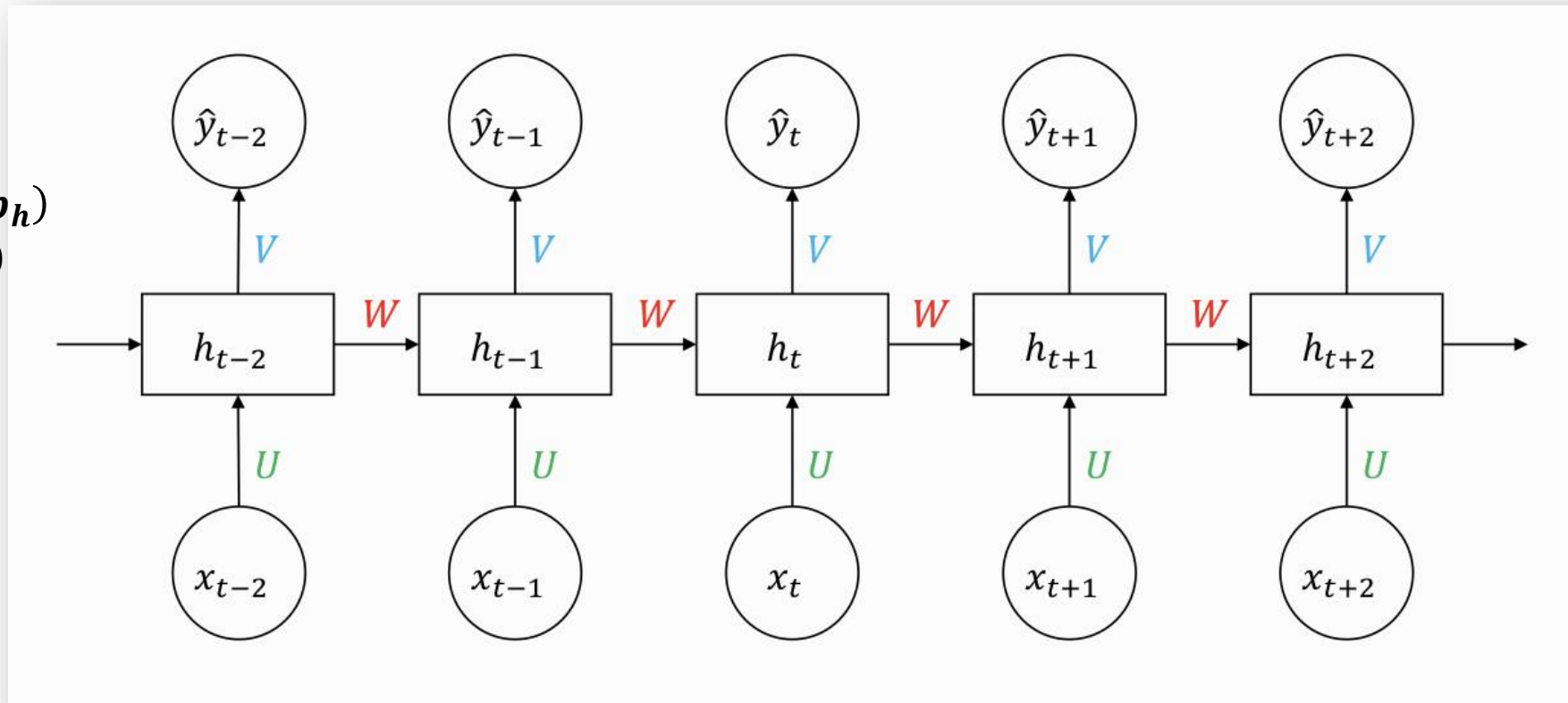


RNN의 back propagation(역전파)

RNN의 오차역전파는 BPTT(BackPropagationThroughTime)이라고도 부른다.

$$h_t = \tanh(Ux_t + Wh_{t-1} + b_h)$$

$$\hat{y}_2 = \text{softmax}(Vh_t + b_y)$$



다대다 예시

$$L = (y - \hat{y})^2$$

$$L_2 = (y_2 - \hat{y}_2)^2$$

$$\frac{\partial L_2}{\partial V} = \frac{\partial L_2}{\partial \hat{y}_2} \frac{\partial \hat{y}_2}{\partial V}$$

$$\frac{\partial L_2}{\partial W} = \frac{\partial L_2}{\partial \hat{y}_2} \frac{\partial \hat{y}_2}{\partial h_2} \frac{\partial h_2}{\partial W} + \frac{\partial L_2}{\partial \hat{y}_2} \frac{\partial \hat{y}_2}{\partial h_2} \frac{\partial h_2}{\partial h_1} \frac{\partial h_1}{\partial W} + \frac{\partial L_2}{\partial \hat{y}_2} \frac{\partial \hat{y}_2}{\partial h_2} \frac{\partial h_2}{\partial h_1} \frac{\partial h_1}{\partial h_0} \frac{\partial h_0}{\partial W} = \sum_{i=0}^2 \frac{\partial L_2}{\partial \hat{y}_2} \frac{\partial \hat{y}_2}{\partial h_2} \left(\prod_{j=i+1}^2 \frac{\partial h_j}{\partial h_{j-1}} \right) \frac{\partial h_i}{\partial W}$$

RNN의 계산 정리

$$h_t = \tanh(W_x x_t + W_h h_{t-1} + b_h)$$

$$\hat{y}_2 = \text{softmax}(W_y h_t + b_y)$$

다대다 예시

$$L = (y - \hat{y})^2$$

$$L_2 = (y_2 - \hat{y}_2)^2 \quad y_2 = \text{output sequence의 두번째 요소}$$

$$\frac{\partial L_2}{\partial W_y} = \frac{\partial L_2}{\partial \hat{y}_2} \frac{\partial \hat{y}_2}{\partial W_y}$$

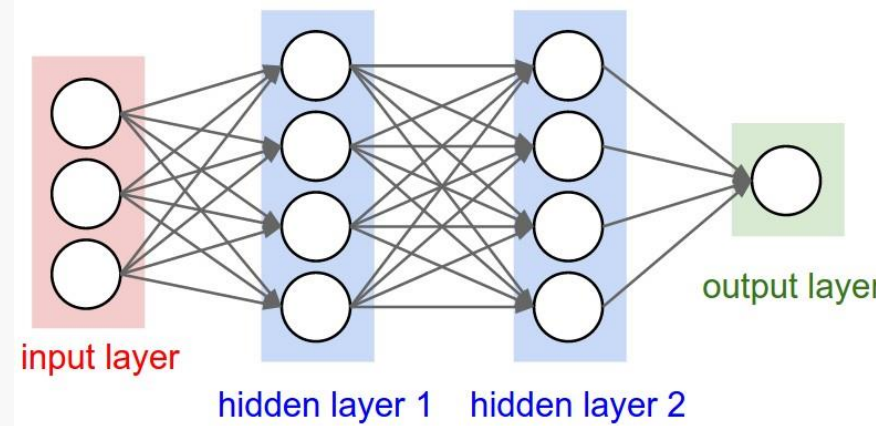
$$\frac{\partial L_2}{\partial W_h} = \frac{\partial L_2}{\partial \hat{y}_2} \frac{\partial \hat{y}_2}{\partial h_2} \frac{\partial h_2}{\partial W_h} + \frac{\partial L_2}{\partial \hat{y}_2} \frac{\partial \hat{y}_2}{\partial h_2} \frac{\partial h_2}{\partial h_1} \frac{\partial h_1}{\partial W_h} + \frac{\partial L_2}{\partial \hat{y}_2} \frac{\partial \hat{y}_2}{\partial h_2} \frac{\partial h_2}{\partial h_1} \frac{\partial h_1}{\partial h_0} \frac{\partial h_0}{\partial W_h} = \sum_{i=0}^2 \frac{\partial L_2}{\partial \hat{y}_2} \frac{\partial \hat{y}_2}{\partial h_2} \left(\prod_{j=i+1}^2 \frac{\partial h_j}{\partial h_{j-1}} \right) \frac{\partial h_i}{\partial W_h}$$

$$\frac{\partial L_2}{\partial W_x} = \frac{\partial L_2}{\partial \hat{y}_2} \frac{\partial \hat{y}_2}{\partial h_2} \frac{\partial h_2}{\partial W_x} + \frac{\partial L_2}{\partial \hat{y}_2} \frac{\partial \hat{y}_2}{\partial h_2} \frac{\partial h_2}{\partial h_1} \frac{\partial h_1}{\partial W_x} + \frac{\partial L_2}{\partial \hat{y}_2} \frac{\partial \hat{y}_2}{\partial h_2} \frac{\partial h_2}{\partial h_1} \frac{\partial h_1}{\partial h_0} \frac{\partial h_0}{\partial W_x} = \sum_{i=0}^2 \frac{\partial L_2}{\partial \hat{y}_2} \frac{\partial \hat{y}_2}{\partial h_2} \left(\prod_{j=i+1}^2 \frac{\partial h_j}{\partial h_{j-1}} \right) \frac{\partial h_i}{\partial W_x}$$

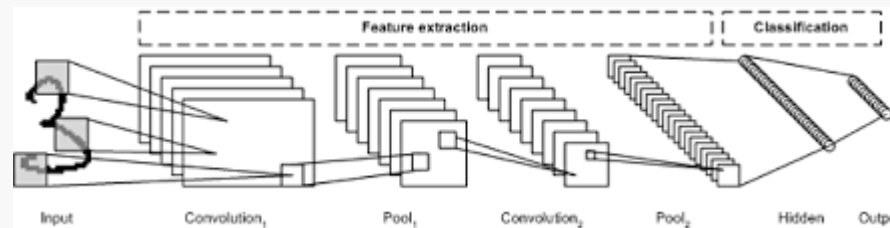
FCN vs CNN vs RNN

Feed Forward

FCN(Fully Connected Layer)는 layer간에 모든 노드들의 값이 다음 레이어의 노드들에 반영



CNN은 weight sharing과 weight localization으로 데이터의 특징을 효과적으로 추출

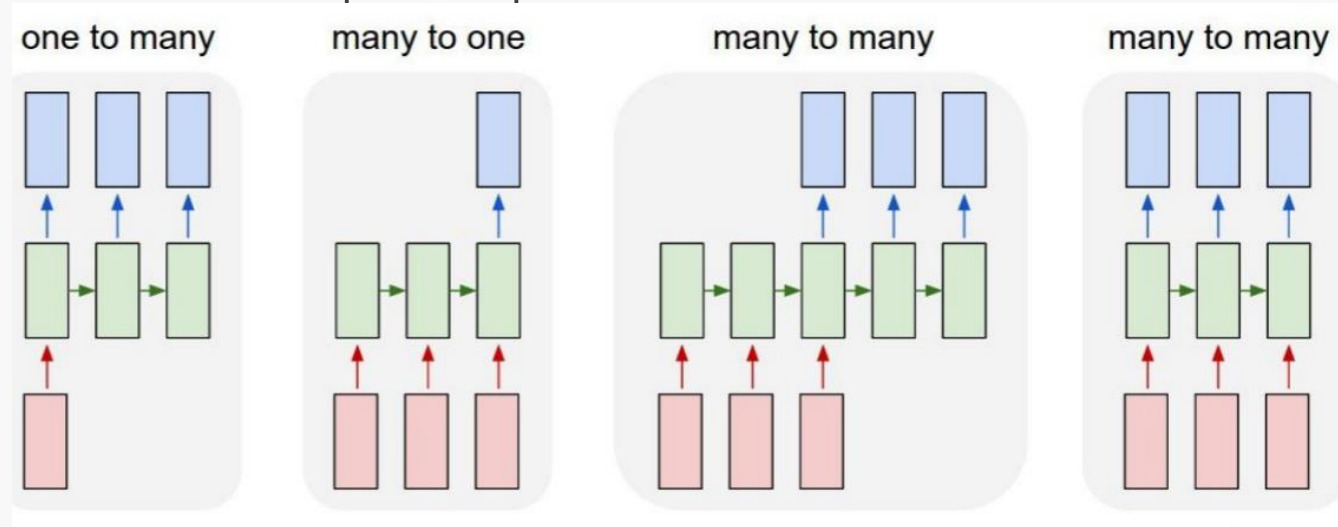


Recurrent

Hidden state에서 데이터 집합 sequence의 정보를 기억하여, sequence data를 효과적으로 처리

RNN의 구조 예시

RNN에는 input과 output의 형태에 따라 다양한 구조가 존재합니다.



1. **One to many** : 이미지 캡셔닝, 정해진 이미지에 문장을 생성
2. **Many to One** : 감성 분석, 연속적 단어인 문장이 긍정적 또는 부정적 감정인지 판정
3. **Many to Many** : 기계 번역, 한 언어의 문장을 읽고 다른 언어의 문장 생성
4. **Many to Many(Synchronized)** : 동영상 분류, POS(Part of Speech)

RNN 모델 예시



A person is walking along a beach with a big dog

ratings		reviews	label
0	5	배송빠르고 곳	1
1	2	택배가 엉망이네용 저희집 밑에층에 말도없이 놔두고가고	0
2	5	아주좋아요 바지 정말 좋아서2개 더 구매했어요 이가격에 대박입니다. 바느질이 조금 ...	1
3	2	선물용으로 빨리 받아서 전달했어야 하는 상품이었는데 머그컵만 와서 당황했습니다. 전...	0
4	5	민트색상 예뻐요. 옆 손잡이는 거는 용도로도 사용되네요 ㅎㅎ	1

영어 ▼		한국어 ▼
I arrived at the bank on the corner. I arrived at the bank after crossing the river. I crossed the river. Then I sat down on the bank. No. This sheep is already very sickly. Make me another. Not so small that. Look! He has gone to sleep...	×	나는 모퉁이에 있는 은행에 도착했다. 나는 강을 건너 후 강둑에 도착했다. 강을 건너 강둑에 앉았다. 아니, 이양은 이미 병약해 다른 양으로 만들어 줘 그렇게 작진 않아요, 보세요! 그는 잠들었어요...

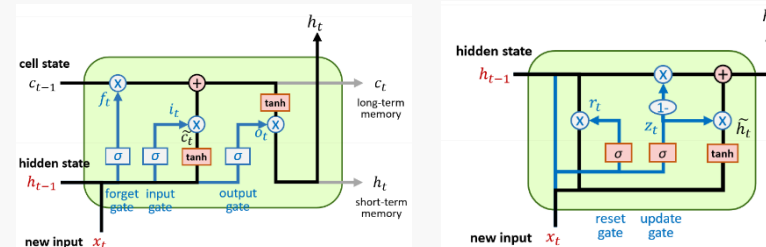
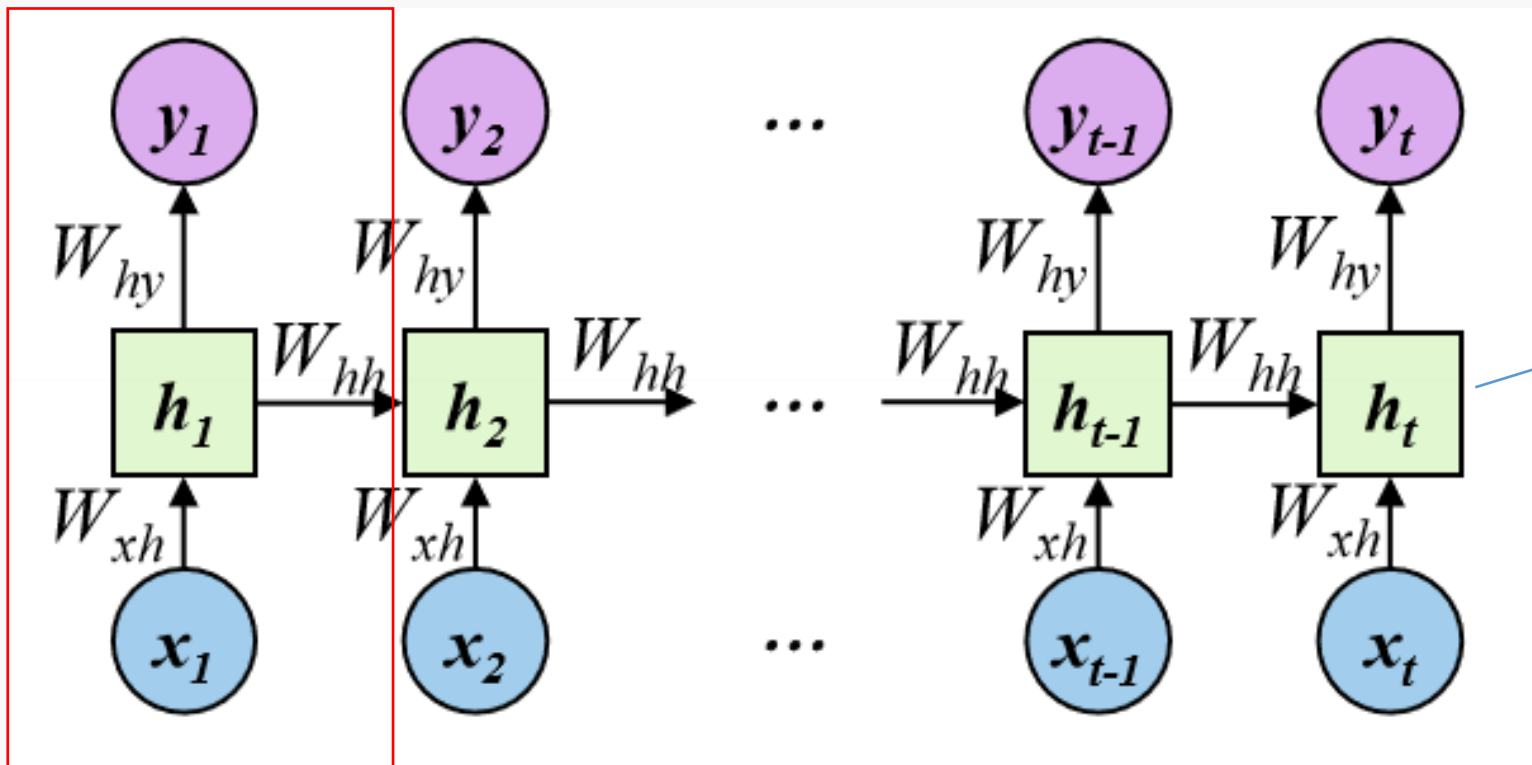
■ CHAPTER 03

한계점

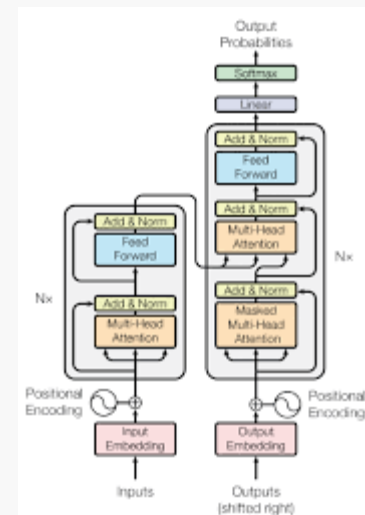
한계점

1. RNN은 BPTT에 따른 기울기 소실문제.,
2. 병렬화 불가능 문제.

LSTM, GRU



Transformer





Thank You :)