

Deep Learning

RNN 모델 개요

(Recurrent Neural Network)

강사 양석환



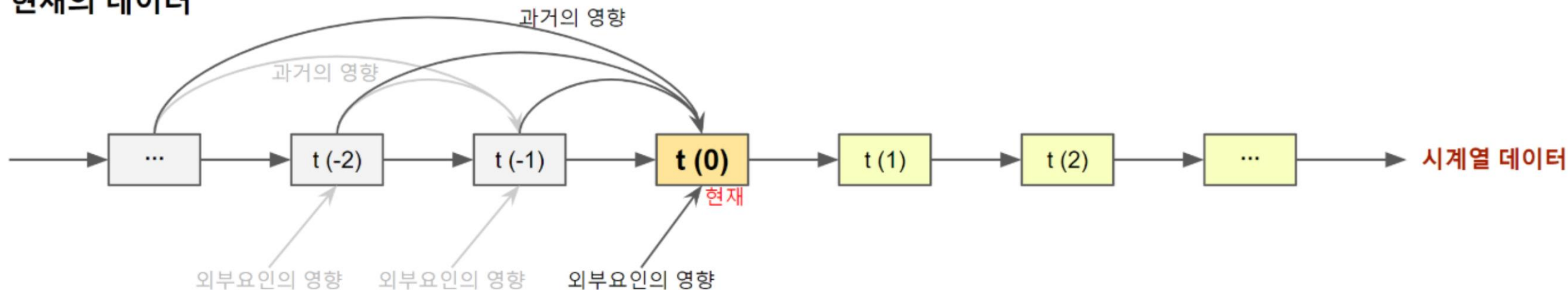
시계열 데이터의 처리



- 시간의 순서대로 일정한 주기에 따라 측정, 저장된 데이터
- 활용 분야
 - 데이터를 기반으로 한 예측 분야에 가장 많이 활용됨
 - 간혹 장애 검출에도 활용할 수 있다고 이야기하고 있으나 결국 데이터 기반 예측임
 - 입력 데이터를 순서대로 살펴보면 조금씩 어긋나는 데이터 발견 가능
 - 어긋나는 데이터의 범위가 특정 기준을 벗어나는 시점이 장애 발생 시점
 - 데이터의 변동을 기반으로 분석, 예측한 결과를 장애 검출, 장애 예측으로 표현하는 것뿐
 - 시계열 데이터는 시간에 따른 변화를 보고 패턴을 분석하여 이후에 어떤 데이터를 얻을 것인지 예측하는 것이 가장 큰 목적이자 활용분야

- 시계열 데이터는 어떤 영향을 주고 받으면서 구성되는가?
 - 시간의 흐름에 따른 데이터는 주변 환경과 특정 이벤트로부터 많은 영향을 받음
 - 따라서 단순히 데이터의 변화 패턴을 분석하는 것만으로는 예측이 어려움
 - 또한 어떤 값과 어떤 변수를 중심으로 분석하는가에 따라 서로 다른 성능변화, 결과의 차이 발생

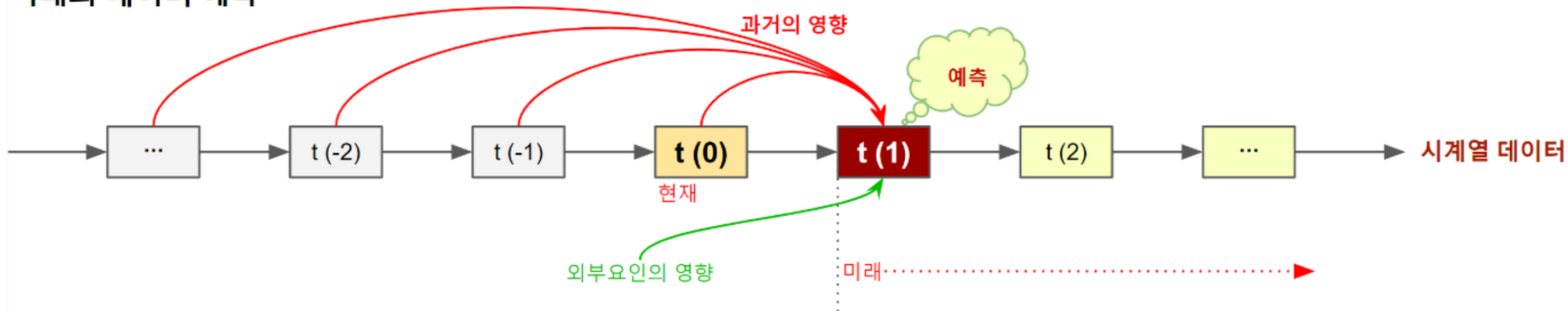
현재의 데이터



- 시계열 데이터를 이용한 예측은 어떤 개념인가?

- 현재($t(0)$)의 데이터는 과거($t(-m) \sim t(-1)$)의 데이터가 누적된 결과
- 가장 가까운 미래($t(1)$)의 데이터는 과거와 현재($t(-m) \sim t(-1) + t(0)$)의 데이터가 누적된 결과
- 미래($t(n)$)의 데이터는 ($t(-m) \sim t(0) \sim t(n-1)$)의 데이터가 누적된 결과
- 여기에 각 시점에서 적용되는 특성이벤트(외부요인)의 반영으로 예측

미래의 데이터 예측



• 모델 제안의 1단계

- 시계열 데이터란 시간의 순서대로 일정한 주기에 따라 측정, 저장된 데이터
 - 과거의 데이터가 현재, 미래의 데이터에 영향을 미침
 - 즉, 과거의 데이터를 가지고 와서 현재, 미래의 데이터에 적용시켜야 함
 - 이를 위해서는 과거의 데이터를 기억하고 있어야 함
 - 과거 데이터를 저장, 참조할 수 있는 메모리 효과를 가진 모델이 필요함

• 모델 제안의 2단계

- 현재의 데이터를 확인하기 위하여 과거의 데이터를 순차적으로 살펴보고

→ $t(1)$ 미래의 데이터를 확인하기 위하여 과거+현재의 데이터를 순차적으로..

→ $t(n)$ 미래의 데이터를 확인하기 위하여...

→ 알고자 하는 시점의 데이터를 기준으로 과거의 데이터를 순서대로 확인 반복

→ 즉, 동일한 과정, 모델이 계속적으로 반복되는 순환 과정을 가진 모델 필요

→ 순환신경망 (RNN, Recurrent Neural Network) 제안됨

RNN 모델의 구조와 이해



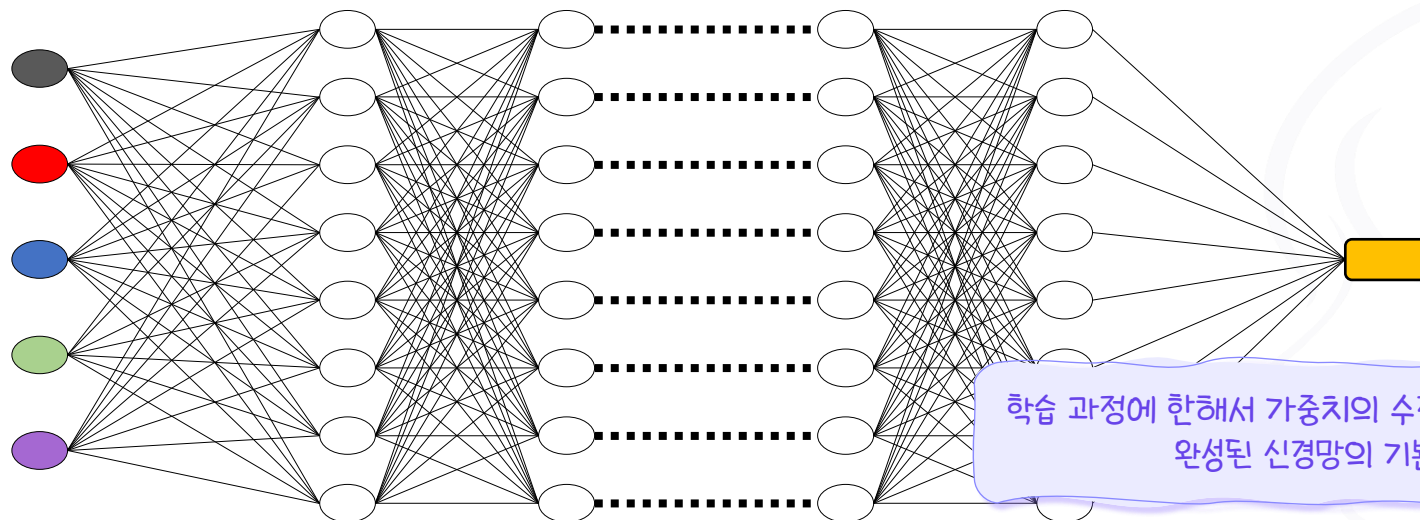
- RNN (Recurrent Neural Network, 순환 신경망)
 - 전체 네트워크 안에서 **순환적으로 데이터를 처리**하는 신경망 모델
 - 과거의 데이터를 끊임없이 참조하여 현재의 데이터를 학습하는 모델
 - 시간의 흐름에 따라 과거의 데이터의 특징과 패턴을 반영하여 현재의 데이터를 학습하는 모델
 - 순차열, 즉 **순서가 있는 일련의 값**을 처리하는 것에 특화된 모델



- 기존의 신경망

- Feed Forward 방식

- 데이터 처리의 흐름이 한 방향



역전파라대?

학습 과정에 한해서 가중치의 수정, 갱신을 위하여 뒤로 돌아가기는 하지만
완성된 신경망의 기본적인 데이터의 흐름은 순방향

• 기존의 신경망의 단점

- 한 방향으로만 데이터를 처리하므로 시계열 데이터의 처리가 어려움
- 시계열 데이터는 신경망의 학습이 완료된 후에도 실제 사용 도중에 지속적으로 과거의 데이터를 참조하고 활용하여야 함
- 학습 과정에서도 시계열 데이터의 성질, 패턴을 제대로 학습할 수 없음
- 순환 신경망(Recurrent Neural Network, RNN) 등장 이유
(앞에서의 설명과 연관됨)

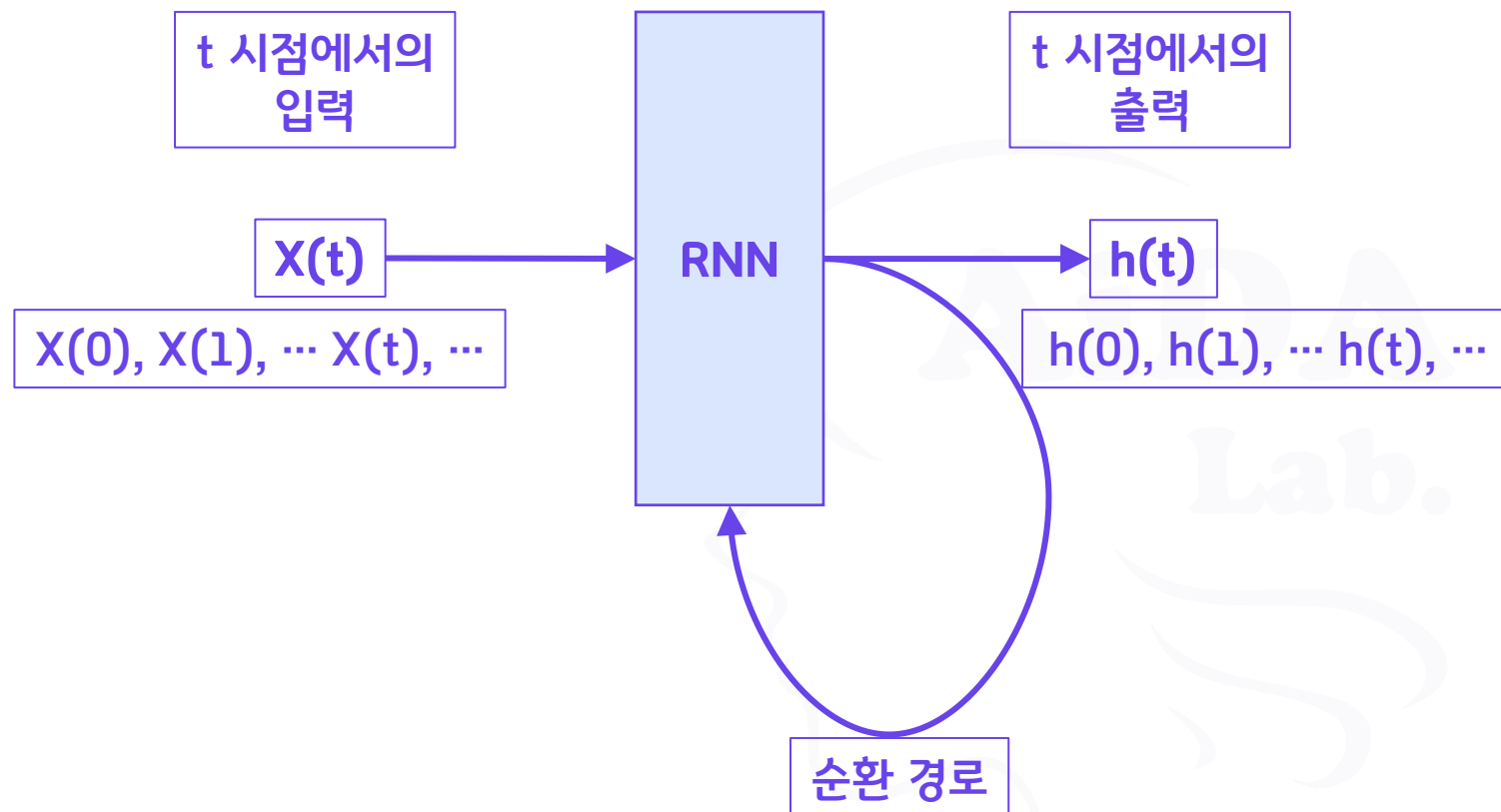


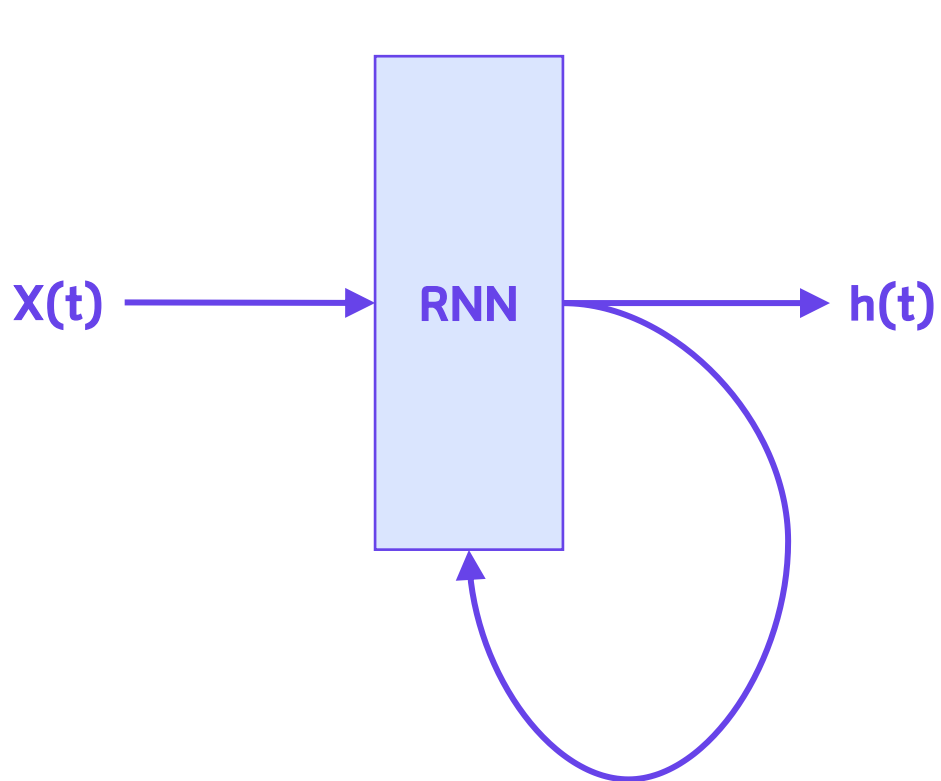
- RNN (Recurrent Neural Network, 순환 신경망)
 - 전체 네트워크 안에서 **순환적으로 데이터를 처리**하는 신경망 모델
 - 순차열, 즉 **순서가 있는 일련의 값을 처리하는 것에 특화된 모델**
 - 순환 처리를 하기 위해서는 **닫힌 경로(=순환하는 경로)**가 필수
 - 경로가 닫혀 있기 때문에 → 데이터가 순환하고, 데이터의 저장에 가능함
 - 순환하기 때문에 → 끊임없이 데이터가 갱신됨



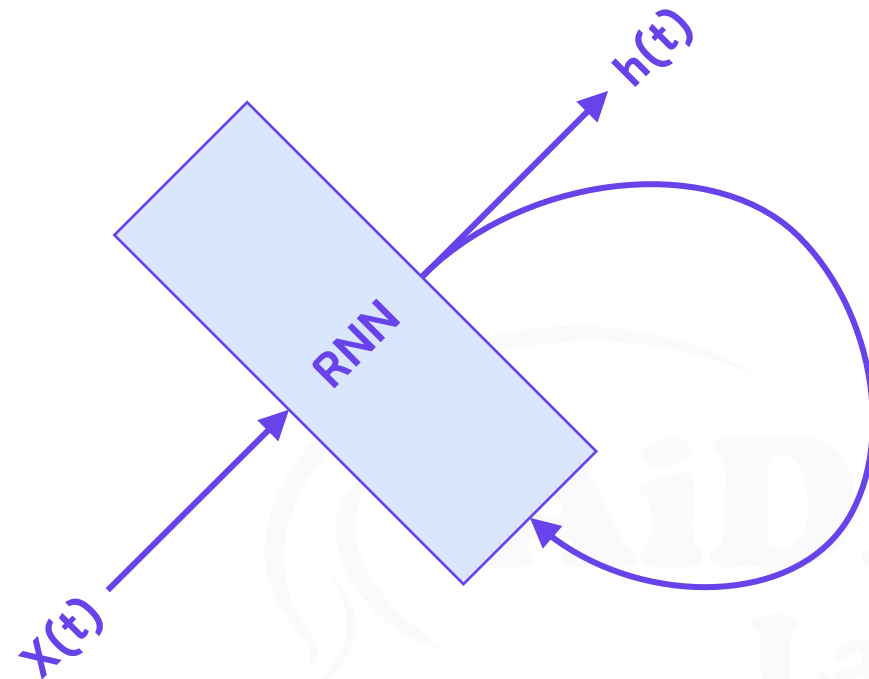
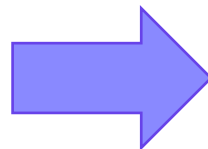
- RNN의 기본 단위: RNN 계층

- 순환경로를 포함한다.
- 계속적인 참조가 반영된다.
- 과거의 정보를 기억한다.



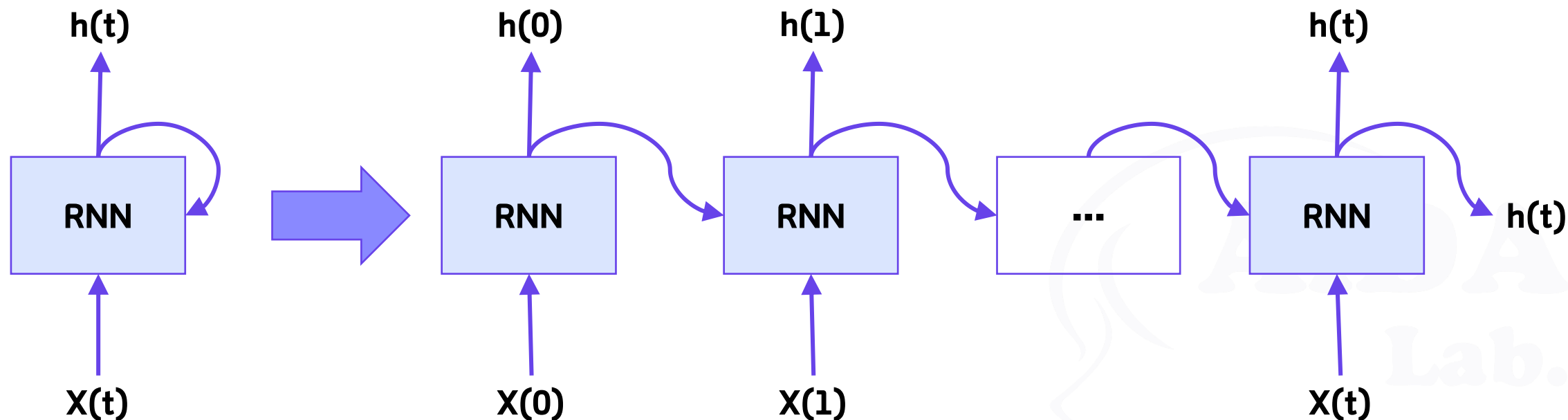


시간 흐름에 따라 계속 이어지므로
세로로 길어짐 !!!

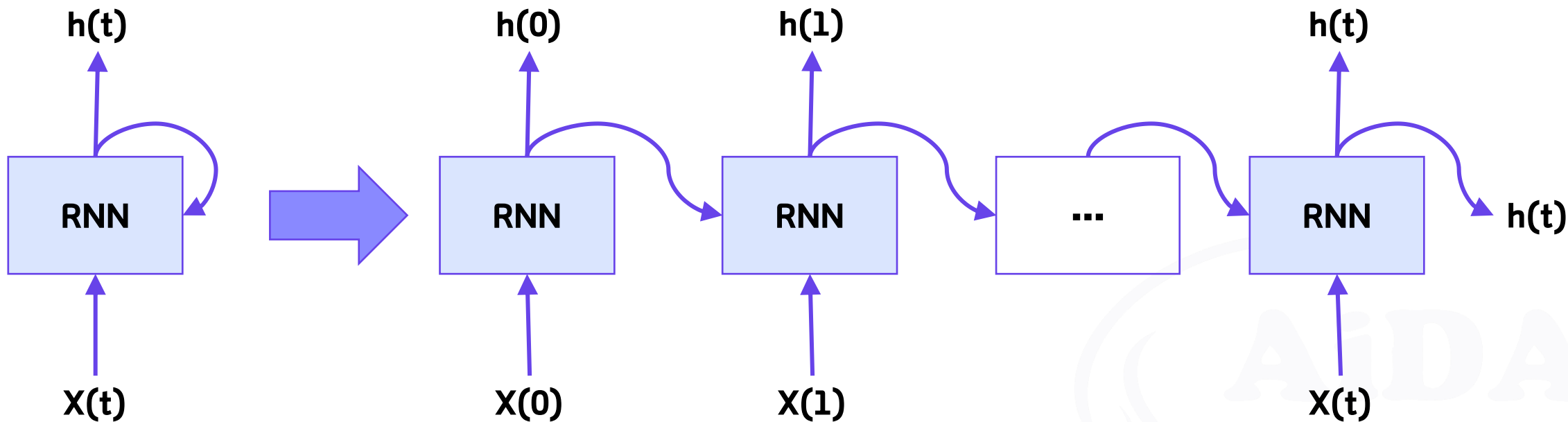


표현하기 편하게 돌려서 쓴다.

- 순환 구조를 알기 쉽게 펼쳐보면

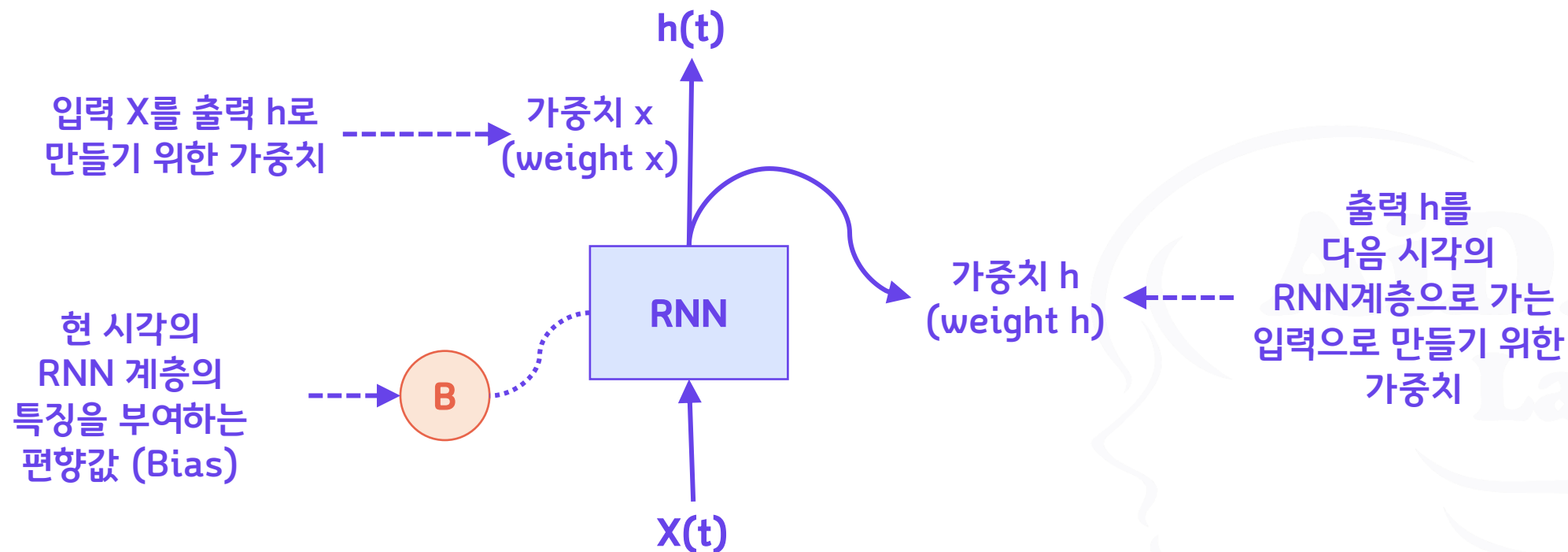


- 데이터는 시간이 흐르는 방향으로 나열됨 (인덱스 t 는 시간이 아니라 **시각**)

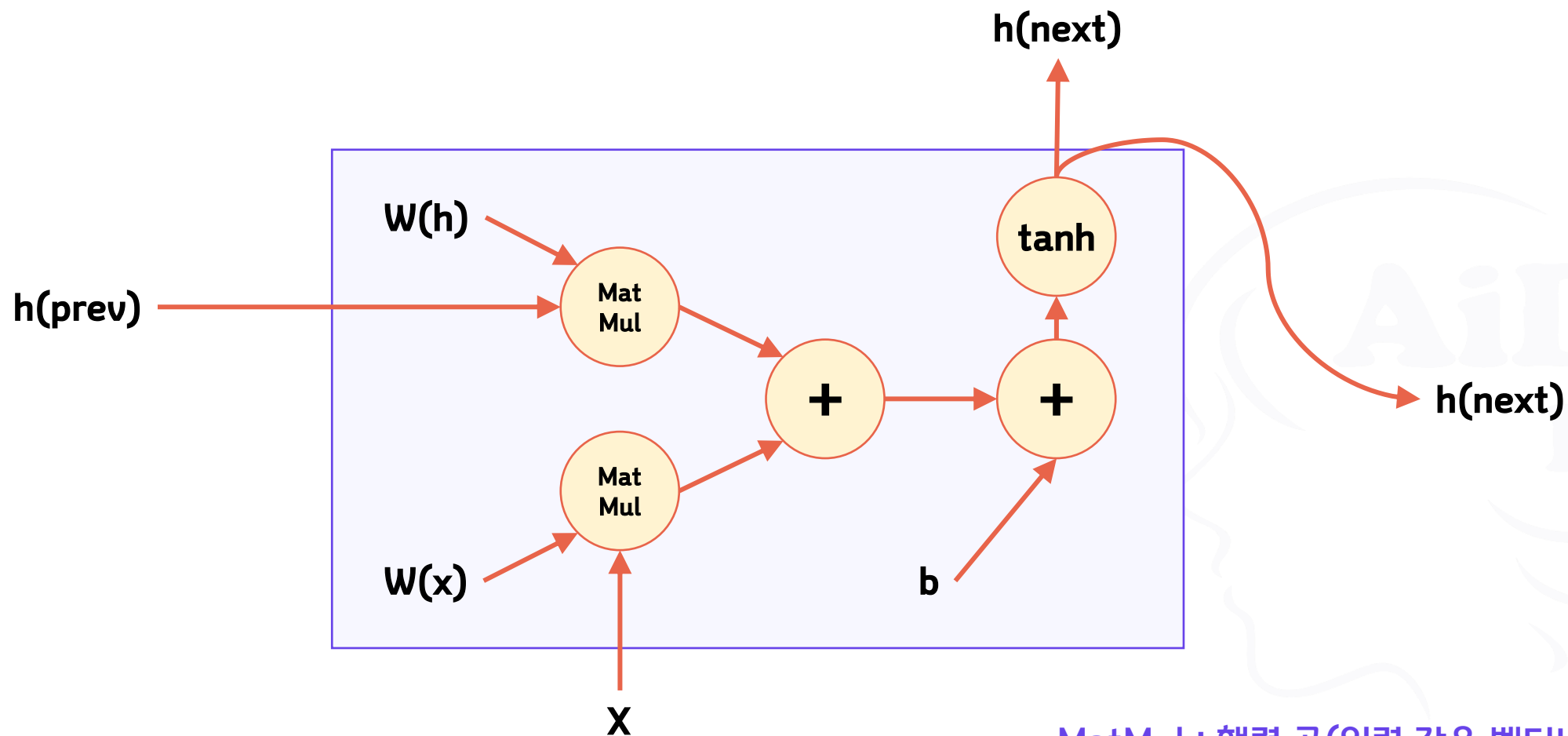


- 각 시각의 RNN 계층은 그 계층으로의 입력과 1개 전의 RNN 계층으로부터의 출력을 입력으로 받음
- 두 개의 입력을 기반으로 현 시각의 출력을 계산함
- 각 RNN 계층에서의 출력 계산에는 기존 신경망과 동일하게 경로 별 가중치, 편향 값이 포함됨

- 경로별 가중치와 편향치의 적용

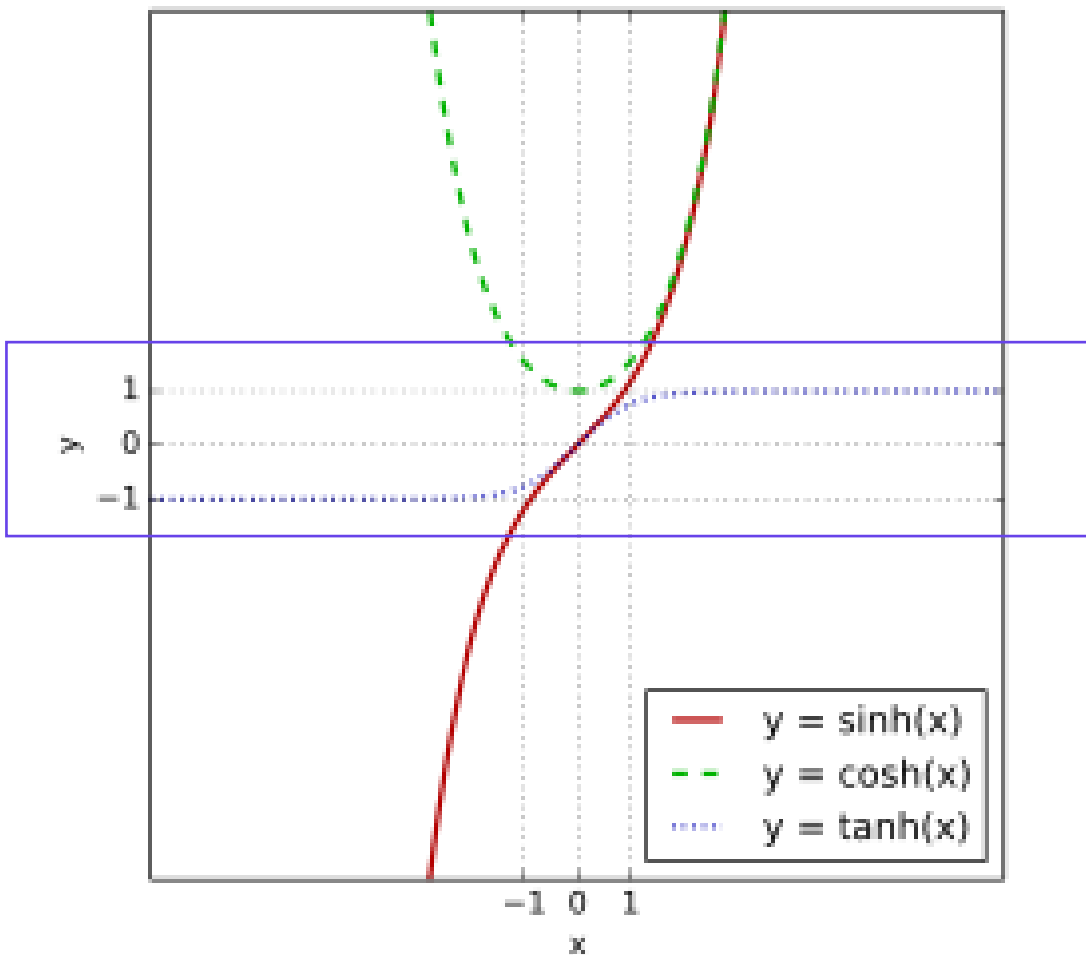


- RNN 계층은 기존 신경망과 동일하다 → 어떻게 처리되나?



MatMul : 행렬 곱(입력 값은 벡터니까)

- 참고



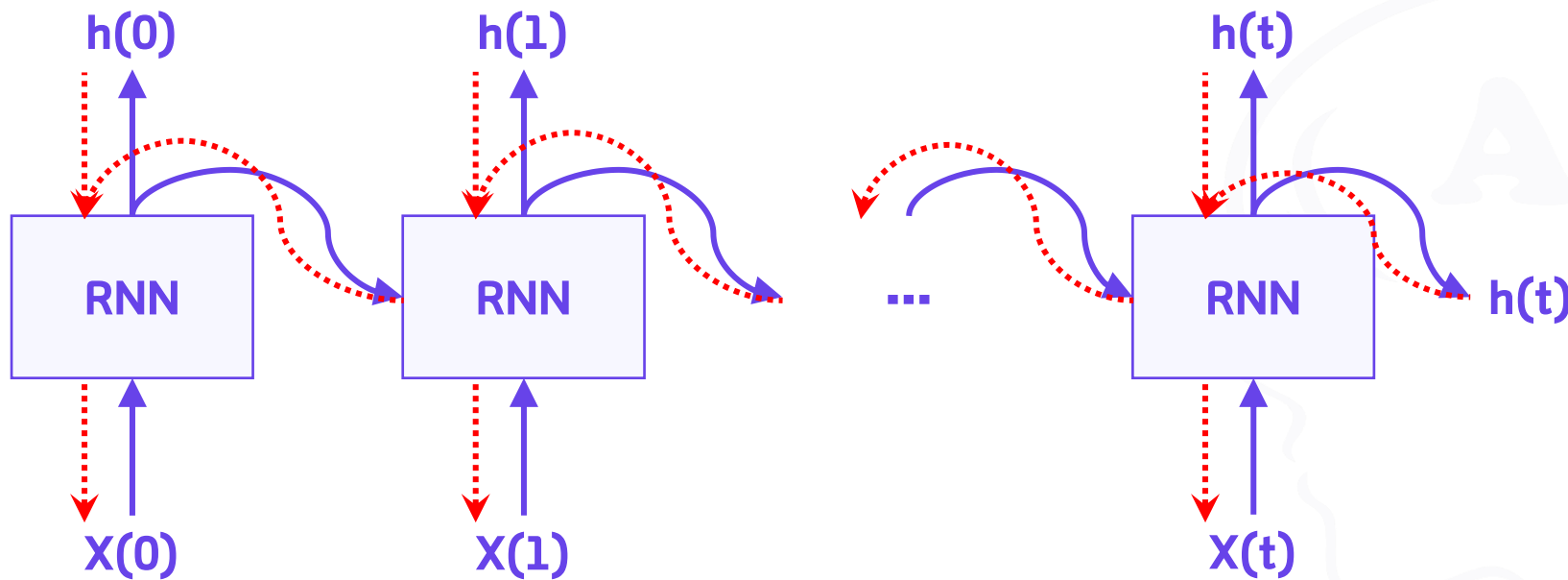
\tanh : 쌍곡 탄젠트 (Hyperbolic Tangent)

이렇게 생긴 함수다
(활성화 함수의 역할)

- RNN 모델의 학습 방법

- RNN 계층은 가로로 펼쳐 놓은 신경망과 동일하다고 가정할 수 있음

→ 학습 방법도 동일하게 적용할 수 있음(오류 역 전파 방식 등) → BPTT 방법



BPTT: Back Propagation Through Time

- BPTT (Back Propagation Through Time)의 문제점

- 긴 시계열 데이터를 학습할 때

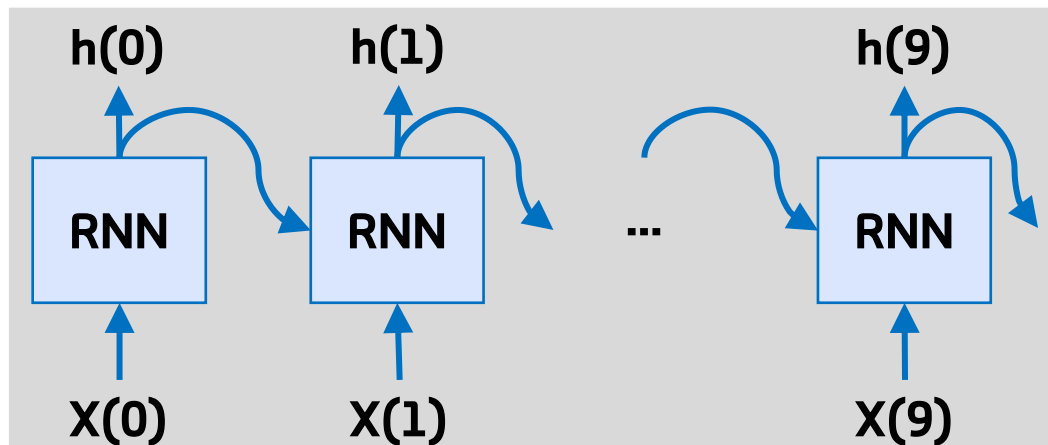
- 시계열 데이터의 시간 크기가 커질수록 BPTT가 소비하는 컴퓨팅 자원도 비례하여 증가함
 - 시간의 크기가 커질수록 역 전파 시의 비율 조정을 위한 기울기가 불안정해짐

BPTT를 이용하여 기울기를 구할 때, 매 시각의 RNN 계층의 중간 데이터를 메모리에 유지해 두어야 함
→ 시계열 데이터가 길어질 수록 계산 량 및 메모리의 사용량이 증가함

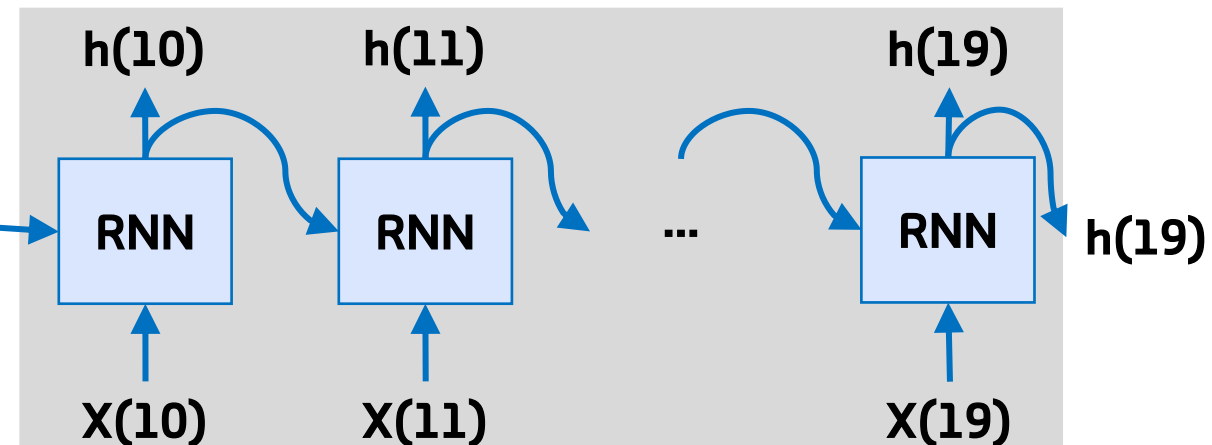
- 개선을 위하여 Truncated BPTT 기법 제안

• Truncated BPTT

- 시간축의 방향으로 길어진 신경망을 적당한 지점에서 잘라내어 여러 개의 작은 신경망으로 만듦
- 잘라낸 작은 신경망에서 BPTT를 수행함
- 주의점
 - 신경망을 잘라낼 때 역전파의 연결만 절단해야 함 (순전파의 연결은 반드시 유지)
 - 순전파의 연결이 사라지면 네트워크 자체가 성립되지 않음
- 역전파가 연결된 RNN 계층의 모임을 **블록**이라고 하여 다른 블록과 구분하여 처리함



첫 번째 블록의 순 전파와 역 전파



두 번째 블록의 순 전파와 역 전파

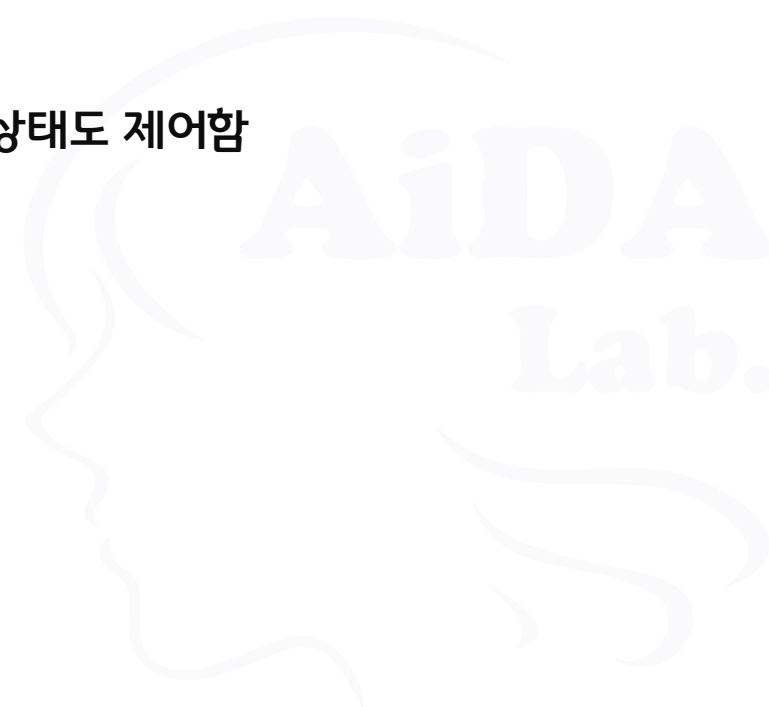
- 시계열 데이터의 학습에서 장기 의존 관계의 학습이 어렵다
 - BPTT에서 기울기 소실 또는 기울기 폭발 발생
 - 기울기 소실: 역 전파 처리 시 기울기 값이 점점 작아지다가 사라지는 현상
 - 기울기 폭발: 역 전파 처리시 기울기 값이 점점 커지다가 수직에 가까워지는 현상
 - 게이트 구조를 적용하여 개선 가능



- 기본적인 RNN 모델을 개선한 다양한 RNN 모델이 존재함

- 게이트가 추가된 RNN 모델:

- 게이트라는 구조를 활용하여 시계열 데이터의 장기 의존 관계를 학습할 수 있는 모델
- 게이트를 통하여 입력 및 출력의 상태(범위)를 제어함
- 데이터 기억을 위한 메모리 셀을 추가한 형태이며 데이터의 기억 여부, 상태도 제어함
- LSTM (Long-Short Term Memory networks)
 - RNN의 장기 의존성 문제를 해결할 뿐만 아니라 학습 또한 빠르게 수렴함
- GRU (Gated Recurrent Unit)
 - LSTM의 간소화된 버전이라고 볼 수 있음



RNN 모델의 활용



- **활용분야 간의 연관성**

- 우리가 글을 읽거나 대화를 할 때, 지금 사용한 하나의 단어만이 의미가 있는가?
- 우리는 글을 읽거나 대화를 할 때, 이전의 맥락을 이해하면서 현재 사용된 단어를 이해함
- 같은 단어인데 다른 의미를 가지는 경우도 이런 맥락을 이용하여 문제없이 정확하게 이해함
- 수많은 단어로 이루어진 언어의 표현은 시간의 흐름에 따라 각 단어들이 배치, 연결됨
- 따라서 시간과 순서(순차) 정보를 잘 활용할 수 있는 RNN이 많이 사용됨

- RNN 모델의 주요 사용 분야
 - 기상데이터 분석 및 예측, 주가 정보 예측 (시계열 데이터)
 - 자연어 처리, 번역, 언어 모델링 등 (순차 열 데이터)
- RNN 모델의 사용 분야에 CNN 등의 모델을 적용한다면?
 - 시간의 흐름, 순서에 따른 맥락과 각 요소(단어) 사이의 인과관계를 처리 가능한가?
 - 가능하기는 하겠지만 매우 많은 노력과 자원이 소요됨

- Recurrent Neural Network(순환 신경망)과 Recursive Neural Network(재귀 신경망)을 같은 것이라고 설명하는 경우가 많음
- 순환 신경망과 재귀 신경망은 서로 다른 모델임
 - 영문 번역 시 Recurrent와 Recursive를 동일하게 **순환**의 의미로 번역하여 발생한 문제
 - 일본 도서의 번역서에서 자주 발견됨(국내 도서에도 존재함)

- 차이점

- 재귀(Recursive) 신경망은 순환(Recurrent) 신경망의 또 다른 일반화 버전이라고 보면 됨
- 순환(Recurrent) 신경망은 체인 형태의 계산 그래프를 사용
- 재귀(Recursive) 신경망은 트리 형태의 계산 그래프를 사용

- 일반적으로 말하는 RNN은 순환 신경망을 의미함



- **CNN (Convolutional Neural Network, 합성곱 신경망)**
 - 딥러닝 모델에서 가장 쉽게 접근할 수 있는 분야
 - 이미지처럼 여러 값이 격자형태로 구성된 입력 데이터를 처리하는 것에 특화된 모델
 - 이미지, 영상의 인식, 분류 등에서 현재까지 가장 뛰어난 결과를 보여줌
 - 이미지, 영상 분야의 다양한 모델(YOLO 등)이 대부분 CNN을 기반으로 고안된 모델임
 - 입력된 이미지보다 너비와 높이가 더 큰 이미지로 손쉽게 확장 가능

- RNN (Recurrent Neural Network, 순환 신경망)
 - 순차 열, 즉 순서가 있는 일련의 값을 처리하는 것에 특화된 모델
 - CNN이 더 큰 이미지로 쉽게 확장이 가능한 것처럼 RNN도 더 긴 순차열로 쉽게 확장 가능
 - 여기서 말하는 더 긴 순차 열이란 순차열에 특화되지 않은 신경망으로는 처리하는 것이 비현실적이라고 할 만큼 긴 데이터를 말함
 - 가변 길이의 순차 열 데이터도 처리 가능
 - 이미지 데이터도 결국 각 픽셀에 대한 수치가 배열의 형태로 모여 있는 것이기때문에 어떤 모델을 사용하더라도 딥 러닝의 적용은 가능함. 단지 그 효율성의 차이가 중요함

THANK
YOU

