

# TIME SERIES DATA ANALYSIS

**RNN · seq2seq · Transformer**

# **PART I**

# **TIME SERIES DATA ANALYSIS**

# TIME SERIES DATA ANALYSIS

---

## ◆ 전통 통계 기반 시계열 데이터 분석 방법론

- 이동평균법(Moving Average) : 구간을 정하고 구간 이동하며 구간다마 평균 구해서 미래 예측
- 지수평활법(Exponential Smoothing) : 이동평균법+가중치(최신 데이터 가중치 ▲, 과거 데이터 ▼)
- **ARIMA Model (Autoregressive Integrated Moving Average)** : 단별량 시계열 예측, 과거 정보 이용  
미래예측
- SARIMA Model (Seasonal ARIMA) : 계절적 특성을 반영한 ARIMA 모델
- Binary Variable Model (해당 시점당 1, 나머지 0 변환) : 각 변수를 0과 1로 표현한 모델

# TIME SERIES DATA ANALYSIS

---

## ◆ 전통 통계 기반 시계열 데이터 분석 방법론

- Trigonometric Model (sine과 cosin함수 조합으로 표현)
  - 계절적 변동을  $\sin() + \cosin()$  조합해서 표현하는 모델
- Growth Curve Mode
- Time Series Regression With Autocorreation ( 에러항이 시점별 correlation이 있다는 가정하에 )
  - 에러항이 서로 독립이다  $\Rightarrow$  시계열은 서로 연관이 있음  $\rightarrow$  에러항은 서로 관계 있음
- Autoregressive Integrated Moving Average Exogenous(ARIMAX) (x변수 이용하여 시계열 Y예측)

# TIME SERIES DATA ANALYSIS

---

## ◆ 머신러닝 기반 시계열 데이터 분석 방법론

- Support Vector Machine/Regression
- Random Forest
- Boosting
- Gaussian Process
- Hidden Markov Model (HMM)

# TIME SERIES DATA ANALYSIS

---

## ◆ 인공지능 기반 시계열 데이터 분석 방법론

- RNN (1986)
- LSTM (1997)
- GRU (2014)
- Seq2Seq (NIPS 2014)
- Seq2Seq with attention (ICLR 2015)
- CNN and Variants (2016)
- Transformer (NIPS 2017)
- GPT-1(2018), BERT(2019), GPT-3(2020), GPT-4(2023)

# TIME SERIES DATA ANALYSIS

## ◆ 시계열 데이터(Time Series Data)






- 시간의 흐름에 따라 순서대로 관측되어 시간의 영향을 받게 되는 데이터

Time	X	Y
09:00	1	1
10:00	2	3
11:00	3	6
12:00	4	10
13:00	5	15

시계열 단변량 데이터

Time	X1	X2	X3	Y
09:00	1	98	2	1
10:00	2	71	4	3
11:00	3	93	7	6
12:00	4	80	5	10
13:00	5	95	9	15

시계열 다변량 데이터

Time	X	Y
09:00		working
09:01		working
09:02		working
09:03		working
09:04		working

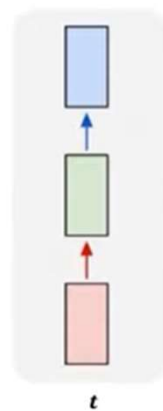
시계열 이미지 데이터

# TIME SERIES DATA ANALYSIS

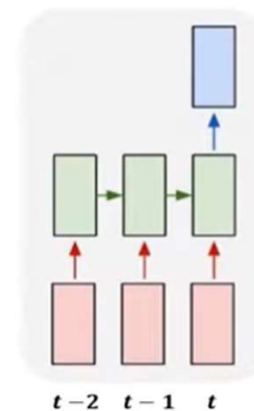
## ◆ 인공지능 기반 시계열 데이터 분석 방법론 - RNN

### ▪ RNN (Recurrent Neural Network)

- 이전 시점 정보들을 반영하여 시계열 데이터 모델링에 적합한 인공신경망 모델



[ DNN ]



[ RNN ]



# TIME SERIES DATA ANALYSIS

## ◆ 인공지능 기반 시계열 데이터 분석 방법론 - RNN

### ▪ DNN 경우

1번째 시점 데이터만 사용

Time	X1	X2	X3	Y
09:00	1	98	2	1

09:00

x1  
x2  
x3

$W_{xh}$

$W_{hy}$

y

입력층  
 $X_1$

은닉층  
 $h_1 = f(W_{xh}X_1)$

출력층  
 $y_1 = f(W_{hy}h_1)$

# TIME SERIES DATA ANALYSIS

## ◆ 인공지능 기반 시계열 데이터 분석 방법론 - RNN

### ▪ DNN 경우

2번째 시점 데이터만 사용

Time	X1	X2	X3	Y
09:00	1	98	2	1
10:00	2	71	4	3

10:00

x1  
x2  
x3

$W_{xh}$

$W_{hy}$

y

입력층

$x_2$

은닉층

$h_2 = f(W_{xh}x_2)$

출력층

$y_2 = f(W_{hy}h_2)$

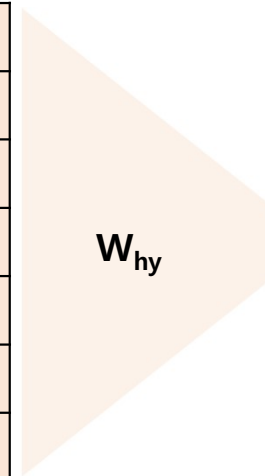
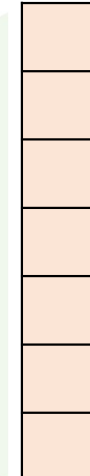
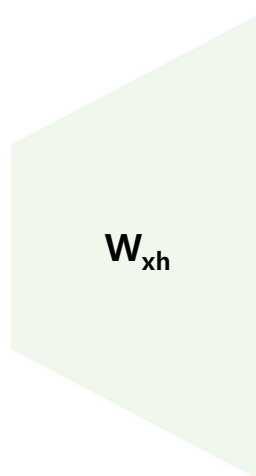
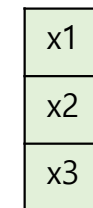
# TIME SERIES DATA ANALYSIS

## ◆ 인공지능 기반 시계열 데이터 분석 방법론 - RNN

### ▪ DNN 경우

3번째 시점 데이터만 사용

Time	X1	X2	X3	Y
09:00	1	98	2	1
10:00	2	71	4	3
11:00	3	93	7	6



입력층  
 $X_3$

은닉층  
 $h_3 = f(W_{xh}X_3)$

출력층  
 $y_3 = f(W_{hy}h_3)$

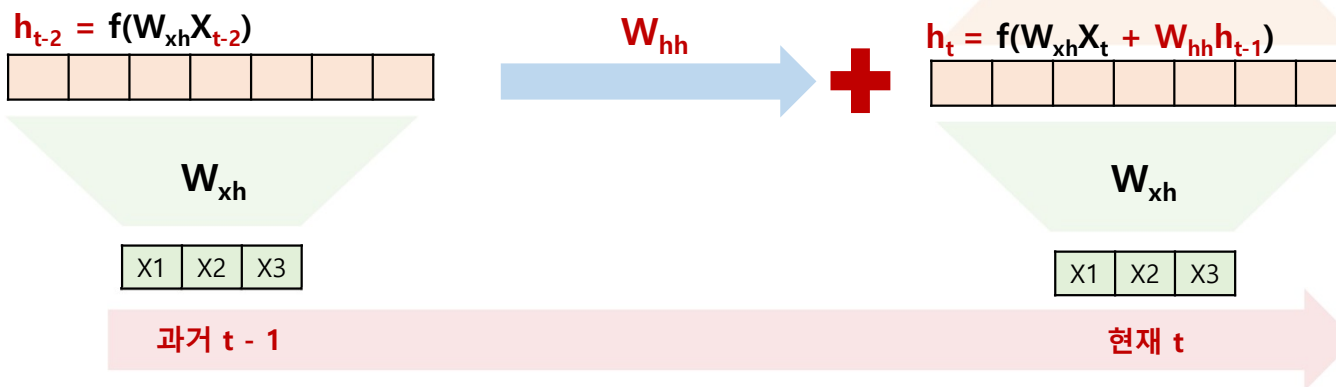
# TIME SERIES DATA ANALYSIS

## ◆ 인공지능 기반 시계열 데이터 분석 방법론 - RNN

### ▪ RNN 경우

이전 시점(09:00) 결과를 현재 시점에 반영

	Time	X1	X2	X3	Y
t-1	09:00	1	98	2	1
t	10:00	2	71	4	3



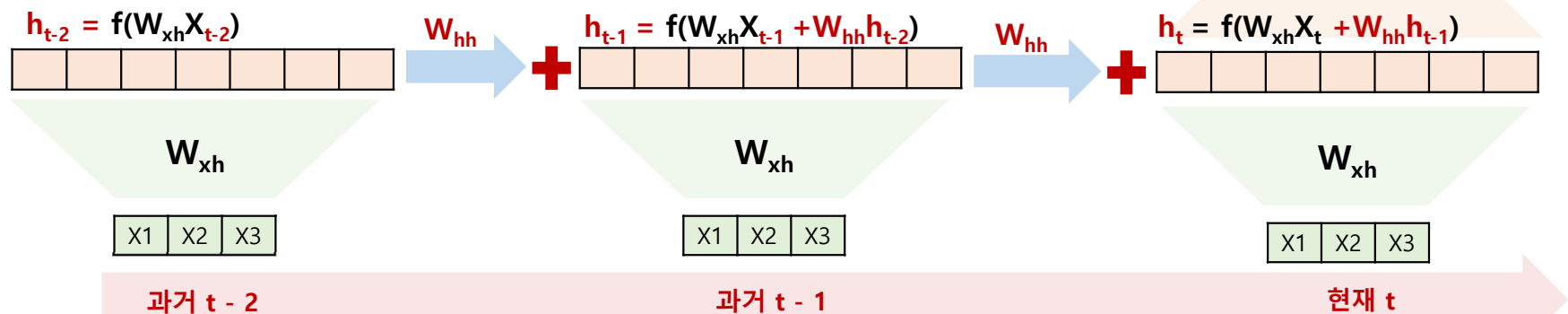
# TIME SERIES DATA ANALYSIS

## ◆ 인공지능 기반 시계열 데이터 분석 방법론 - RNN

### ▪ RNN 경우

이전 시점(09:00), 이전 시점(10:00) 결과를 현재 시점에 반영

	Time	X1	X2	X3	Y
t-2	09:00	1	98	2	1
t-1	10:00	2	71	4	3
t	11:00	3	93	7	6



# TIME SERIES DATA ANALYSIS

## ◆ 인공지능 기반 시계열 데이터 분석 방법론 - RNN

### ▪ RNN 경우

이전 시점 결과 저장 한 것 → Hidden State

	Time	X1	X2	X3	Hidden State
t-2	09:00	1	98	2	$h_{t-2} = f(W_{xh}X_{t-2})$
t-1	10:00	2	71	4	$h_{t-1} = f(W_{xh}X_{t-1} + W_{hh}h_{t-2})$
t	11:00	3	93	7	$h_t = f(W_{xh}X_t + W_{hh}h_{t-1})$

가중치 파라미터

$W_{xh}, W_{hh}$

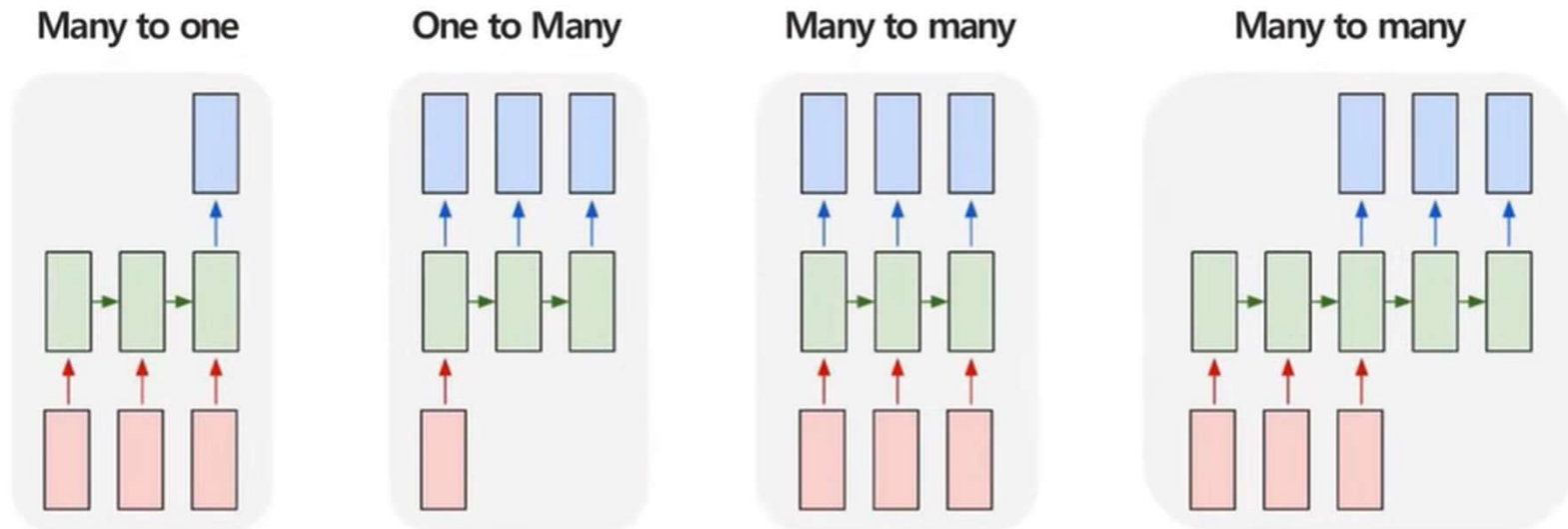
모든 시점에서 동일

$$y_t = f(W_{hy}h_t)$$

# TIME SERIES DATA ANALYSIS

## ◆ 인공지능 기반 시계열 데이터 분석 방법론 - RNN

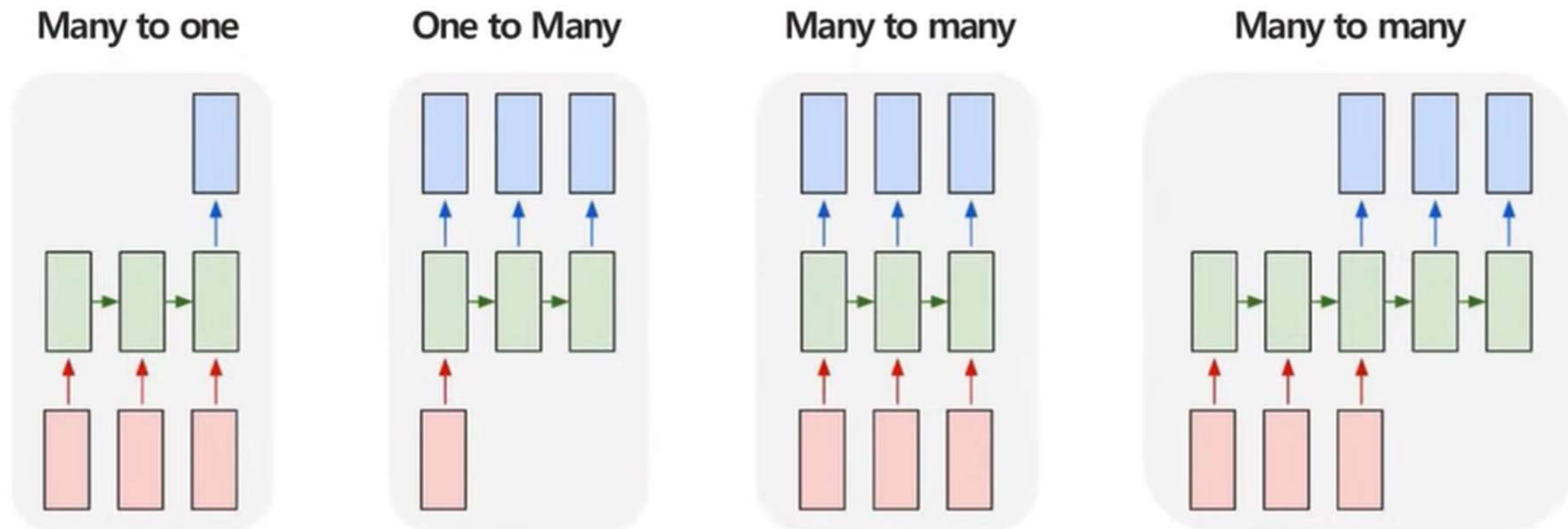
### ▪ 다양한 RNN 구조



# TIME SERIES DATA ANALYSIS

## ◆ 인공지능 기반 시계열 데이터 분석 방법론 - RNN

### ▪ 다양한 RNN 구조



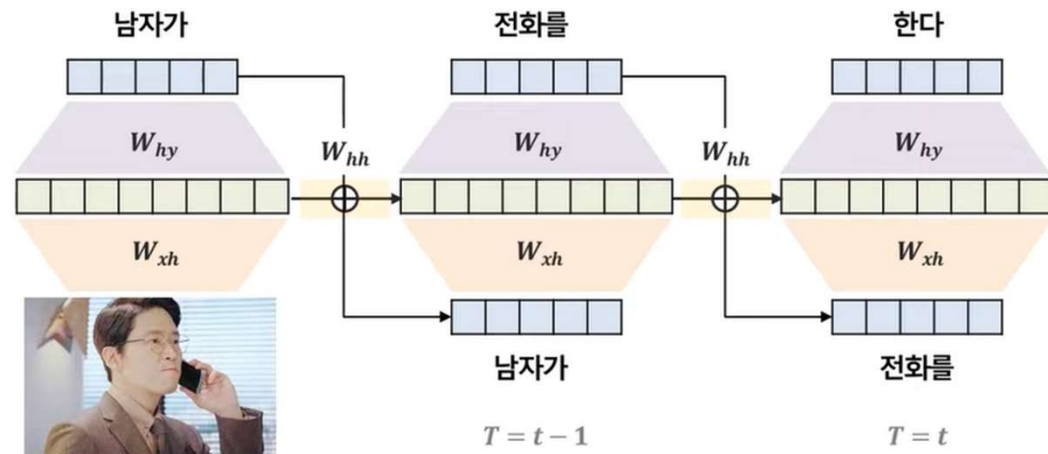


# TIME SERIES DATA ANALYSIS

## ◆ 인공지능 기반 시계열 데이터 분석 방법론 - RNN

### ▪ 다양한 RNN 구조 - One to Many

- 단일 시점  $X \rightarrow$  순차적인  $Y$  예측
- 예: 이미지 데이터 >>> 이미지에 대한 정보 글 생성 : 이미지 캡셔닝

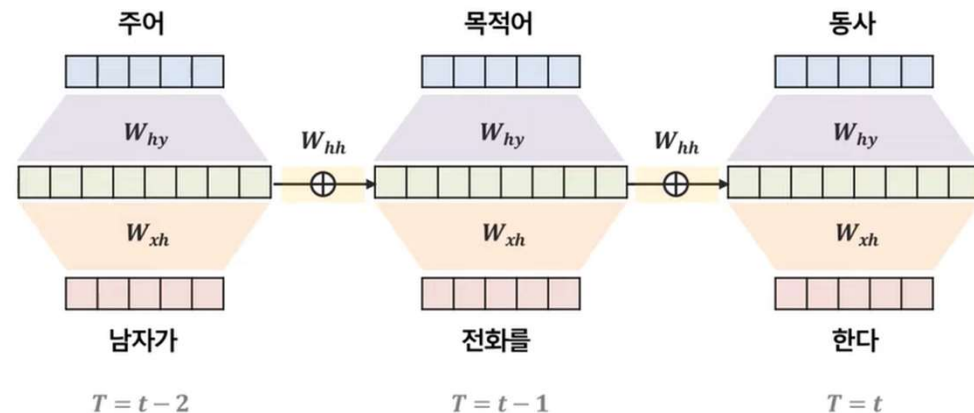


# TIME SERIES DATA ANALYSIS

## ◆ 인공지능 기반 시계열 데이터 분석 방법론 - RNN

### ▪ 다양한 RNN 구조 - Many to Many

- 순차적인  $X \rightarrow$  순차적인  $Y$  예측
- 예: 문장에서 각 단어의 품사 예측 POS(Part Of Speech) Tagging

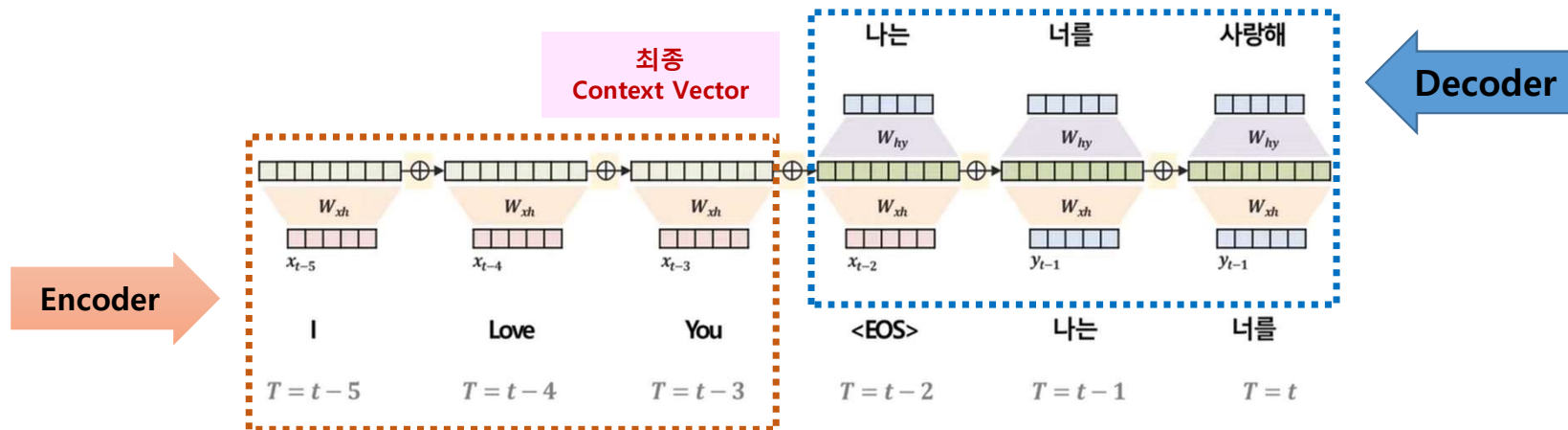


# TIME SERIES DATA ANALYSIS

## ◆ 인공지능 기반 시계열 데이터 분석 방법론 - RNN

### ▪ 다양한 RNN 구조 - Many to Many

- 순차적인  $X \rightarrow$  순차적인  $Y$  예측
- 구성 : Encoder + Context Vector + Decoder  $\rightarrow$  Sequence To Sequence : Seq2Seq
- 예: 영어 문장 >>> 한글 문자 번역



# TIME SERIES DATA ANALYSIS

---

## ◆ 인공지능 기반 시계열 데이터 분석 방법론 - RNN

### ▪ 다양한 RNN 구조 - Many to Many : seq2seq

- Encoder            역할 : 입력 시퀀스를 하나의 벡터 표현으로 압축 즉, Context Vector
- Context Vector    역할 : 입력 시퀀스의 함축 데이터
- Decoder            역할 : 벡터 표현을 통해서 출력 시퀀스 생성

❖ **입력과 출력의 시퀀스 길이는 다를 수 있음**

- 문제점 : 입력 시퀀스를 하나의 벡터로 압축하는 과정에서 **입력 시퀀스의 정보가 일부 손실**