

# Sequence Modeling

강사 : 백병인

[pi.paek@modulabs.co.kr](mailto:pi.paek@modulabs.co.kr)

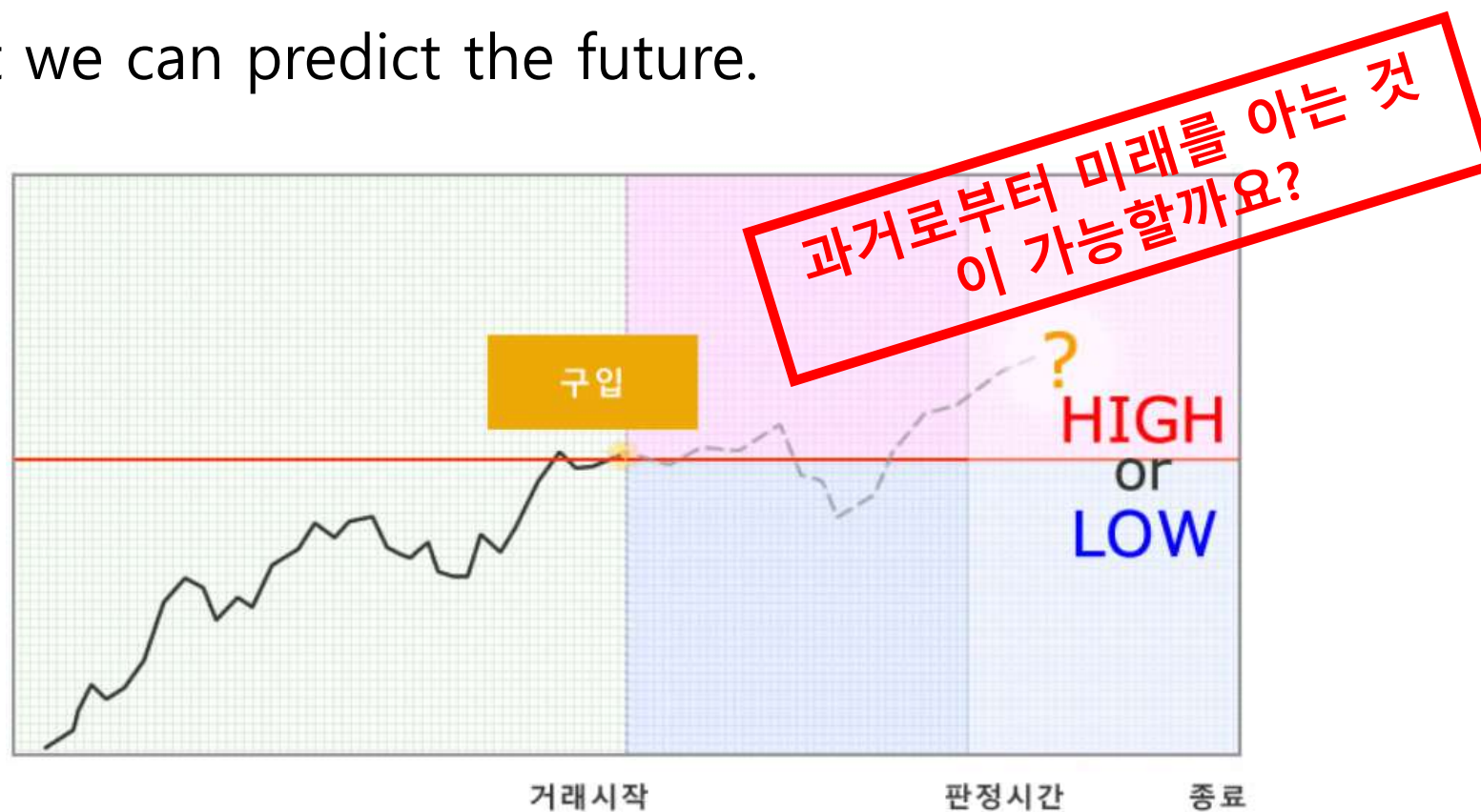
모두의연구소 Research Scientist



2019 모두의연구소

# We want to predict next step.

- so that we can predict the future.



# 두 가지 유형 (1)

## • Sequence Prediction

### • (예시)

- 사용자가 방문한 웹페이지 정보를 바탕으로 다음 방문할 웹페이지 예측(혹은 구매 예측)
- DNA의 나선형 단백질 순서 모델링
- 자연어처리(Natural Language Processing, NLP)

## • Time-Series Forecasting

### • (예시)

- 지금까지의 주가변동곡선을 바탕으로 다음 주가변동 예측
- 특정지역의 기후데이터를 바탕으로 내일의 온도변화 예측
- 공장 센서데이터 변화 이력을 토대로 이상 발생 예측

Q) 위 두 가지 유형의 본질적 차이는 무엇일까요?



## 두 가지 유형 (2)

- Sequence Prediction

- 심볼(Symbol)의 리스트를 대상으로 다음 단계를 예측합니다.

- A, B, C, ??

- Time-Series Forecasting

- 숫자(Numeric)의 리스트를 대상으로 다음 단계를 예측합니다.

- 1, 2, 3, ??

# Sequence와 Series

- Sequence (수열)

- 2, 4, 6, 8, 10, ... (등차수열)
- 2, 3, 5, 8, 13, ... (피보나치수열)
- 그러나 Sequence는 엄밀히, 숫자의 연속이 아니라 심볼의 연속이다.
- 단지 우리가 Numeric Sequence에 익숙할 뿐이다.

$$\{a_i\}_{i=1}^n$$

- Series (급수)

- $1+1/2+1/4+1/8+\dots$  (무한등비급수)
- 급수는 더하기(+) 연산을 전제로 정의된다.
- 급수도 수열이긴 하다. 숫자 수열은 모두 초기값과 변화량의 급수로 표현 가능하다.

$$S_n = \sum_{r=1}^n a_r$$



# Numerical Series Tests

- 예측 가능하십니까?

- 4, 8, 16, 32, ?

- 54, 49, ?, 39, 34

- 4, 3, 5, 9, 12, 17, ?

# 만약 숫자가 아니라 문자라면?

- 그래도 예측 가능하십니까?
  - D, H, P, ?
  - I ate ???
  - A, C, G, T, ?
  - R, E, T, O, QW, QU, ?
- 문자는 기호(Symbol)일 뿐 그 자체로 의미나 값을 가지지 않습니다.
- 그러므로 외부적 정보가 추가되지 않으면 sequence만으로는 다음 단계를 예측할 수 없습니다.



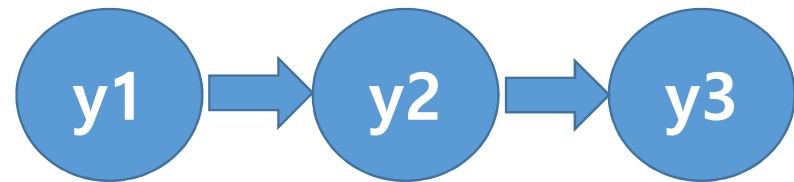
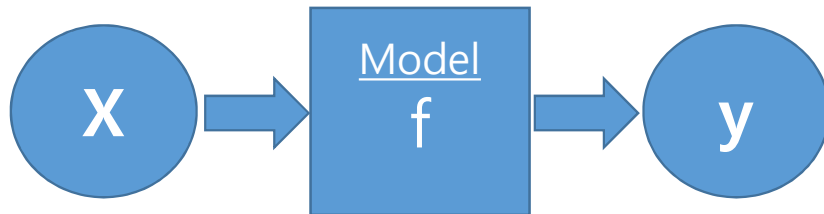
우선, 숫자 기반의 시계열 데이터를 예측하는  
**Time Series Analysis**에 대해 알아보시다.





# Time-Series Forecasting의 전제 (1)

- ML Functional Modeling
- Time Series Analysis



일반적인 Machine Learning은  $X$ 와  $y$ 의 관계를 규명하는 모델  $f$ 를 통해  $y$ 를 예측한다.  
그러나 Time-Series 계열은 시간에 따른  $y$  값의 변화에만 초점을 맞춘다.

# Univariate / Multivariate

- Univariate

시간	단가(원)
09:00	27.2
09:30	27.2
10:00	28.1
10:30	28.1
11:00	28.5

- Multivariate

시간	전압(V)	전류(A)	저항( $\Omega$ )	온도( $^{\circ}\text{C}$ )
01:01	27.2	9.7	0.385	293
01:02	27.2	9.7	0.385	293
01:03	28.1	9.4	0.387	296
01:04	28.1	9.3	0.388	295
01:05	28.5	9.1	0.379	297

**Time-Series 방법론으로 충분히  
분석할 수 있을까?**



# Time-Series Forecasting의 전제 (2)

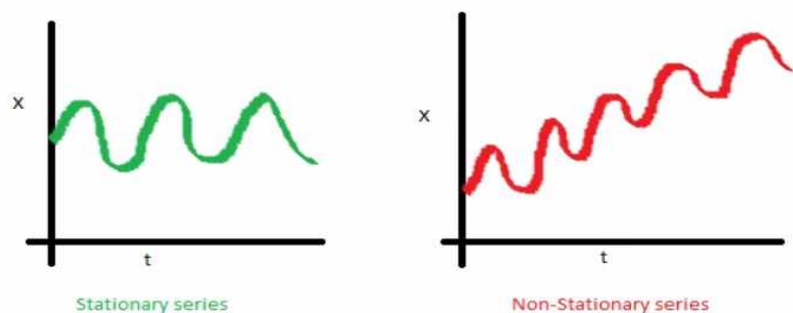
- 과거로부터 미래를 아는 것이 정말 가능한 건가요??

과거 데이터에서 반복되는 패턴을 발견할 수 있으면,  
그 패턴이 미래에도 반복된다는 가정 하에 가능합니다.

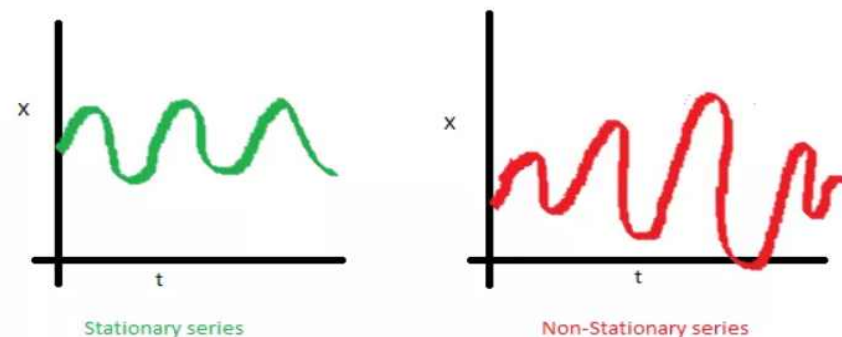
- 그래서, 원칙적으로 안정적인 시계열(Stationary Time-Series)를 전제합니다.

# Stationary Time-Series

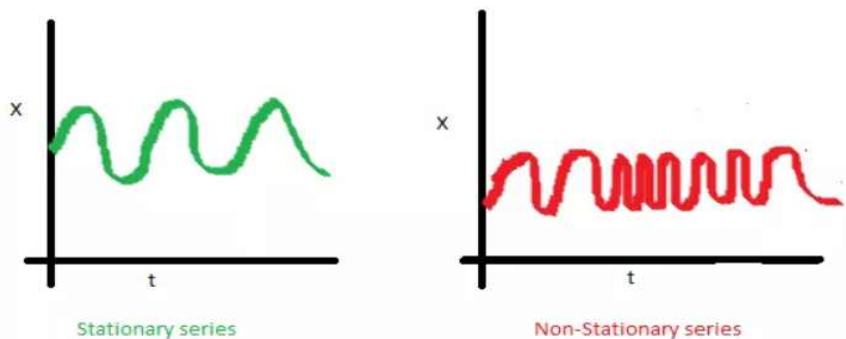
1. 시간의 추이와 관계 없이 평균이 불변



2. 시간의 추이와 관계 없이 분산이 불변



3. 두 시점 간의 공분산이 기준시점과 무관



**Stationary하지 않은 문제라도,  
어떻게든 Stationary하게 변환해서 풀어야 한다.**

(참고) [http://www.dodomira.com/2016/04/21/arima\\_in\\_r/](http://www.dodomira.com/2016/04/21/arima_in_r/)  
에서 Analyticsvidhya 를 인용한 내용



2019 모두의연구소

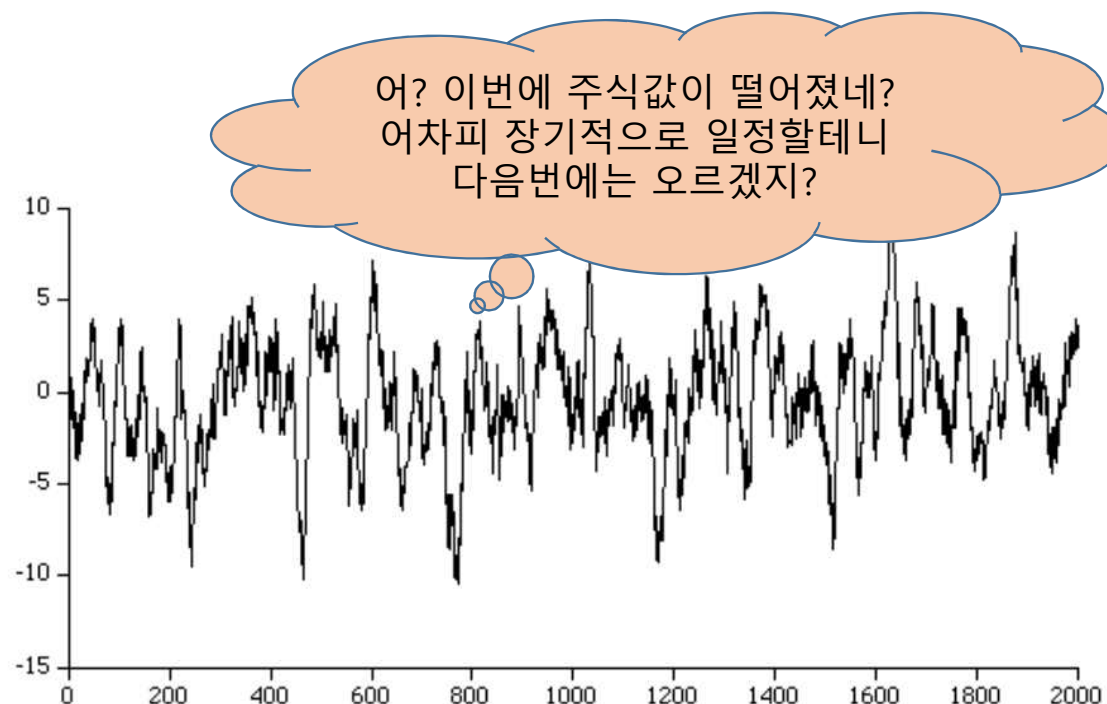
# Univariate (1) AR

- 자기회귀(AutoRegressive)

- AR(1)

$$X(t) = \{a * X(t-1) + c\} + u * e(t)$$

- AR(p)는 p step만큼의 이전 시점 값들을 고려한다.



AR은 시점  $t$ 의 예측값이 이전 시점의 값(X)들의 (가중)평균값과 비슷할 것이라고 가정한다.  
시간의 흐름에 따라 데이터가 일정한 평균적 흐름을 유지하는 경향을 모델링한다.

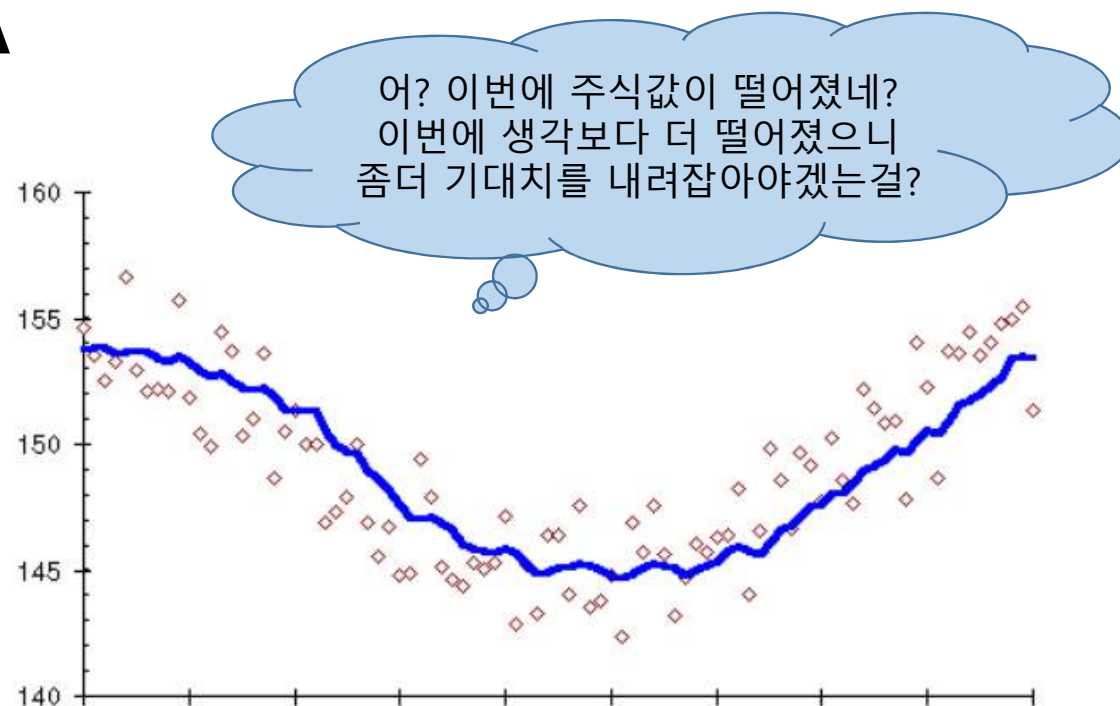
# Univariate (2) MA

- 이동평균(Moving Average)

- MA(1)

$$X(t) = \{a * e(t-1) + c\} + u * e(t)$$

- MA(q)는 q step만큼의 이전 시점 값들을 고려한다.



MA는 시점 t의 예측값이 이전 시점의 예측오차(e)들의 (가중)평균값과 비슷할 것이라고 가정한다. 시간의 흐름에 따라 데이터가 변화한다고 가정하고 그 추세를 모델링한다.

# Univariate (3) ARMA & ARIMA

- ARMA : AutoRegressive Moving Average
- $\text{ARMA}(p, q) = \text{AR}(p) + \text{MA}(q)$
- $\text{ARMA}(1, 1) \quad X(t) = \{a \cdot X(t-1)\} + \{b \cdot e(t-1)\} + c + u \cdot e(t)$
- $\text{ARMA}(2, 2) \quad X(t) = \{a_1 \cdot X(t-1) + a_2 \cdot X(t-2)\} + \{b_1 \cdot e(t-1) + b_2 \cdot e(t-2)\} + c + u \cdot e(t)$
- ARIMA : AutoRegressive **Integrated** Moving Average
- $\text{ARIMA}(p, d, q)$  : d-step difference =  $\text{ARIMA}(p, q)$
- $\text{ARIMA}(1, 2, 1) \quad a \cdot [X(t) - X(t-1)] - [X(t-1) - X(t-2)] = \{b \cdot X(t-1)\} + \{c \cdot e(t-1)\} + d + u \cdot e(t)$

**ARMA**는 거시적으로 일정한 흐름 속에서 미시적으로 발생하는 변화를 모델링한다.  
**ARIMA**는 거시적인 추세 변화까지 모델링할 수 있도록 **ARMA**를 개선한다.



# Component & Feature

- TSA Components
  - Level
  - Trend
  - Seasonality
- ARIMA(p, d, q)의 p, d, q는 component에 해당합니다.
- ARIMA는 feature를 가지지 않습니다.

Time Series Analysis에서 time t는 feature가 되지 않습니다.  
Q) 시간 t가 feature가 된다면 문제의 정의가 어떻게 바뀔까요?

