

시계열 DATA SCIENCE LAB 스터디

시계열 1팀
5기 임낙준 5기 정승연 5기 조신형

CONTENTS

01 시계열의 개념

02 시계열 예측 모델

03 주식 가격 예측 대회

01

시계열 개념

- 01. 시계열 정의
- 02. 시계열 특성1 - 자기상관성
- 03. 시계열 특성2 - 계절성
- 04. 시계열 특성3 - 정상성

01 시계열의 개념

01 시계열개념

02 시계열 특성

03 푸리에 변환

시계열 데이터 정의

- 시계열 데이터 정의: 일정한 시간동안 수집되어, 순서 있게 나열된 데이터 셋의 집합
- 시계열 데이터 예시:
 - POS (Point of sales) 자료 (불규칙적인 시차)
 - 일일 코스피 주식가격
 - 월별/분기별/연도별 범죄발생 수 등 (규칙적인 시차)

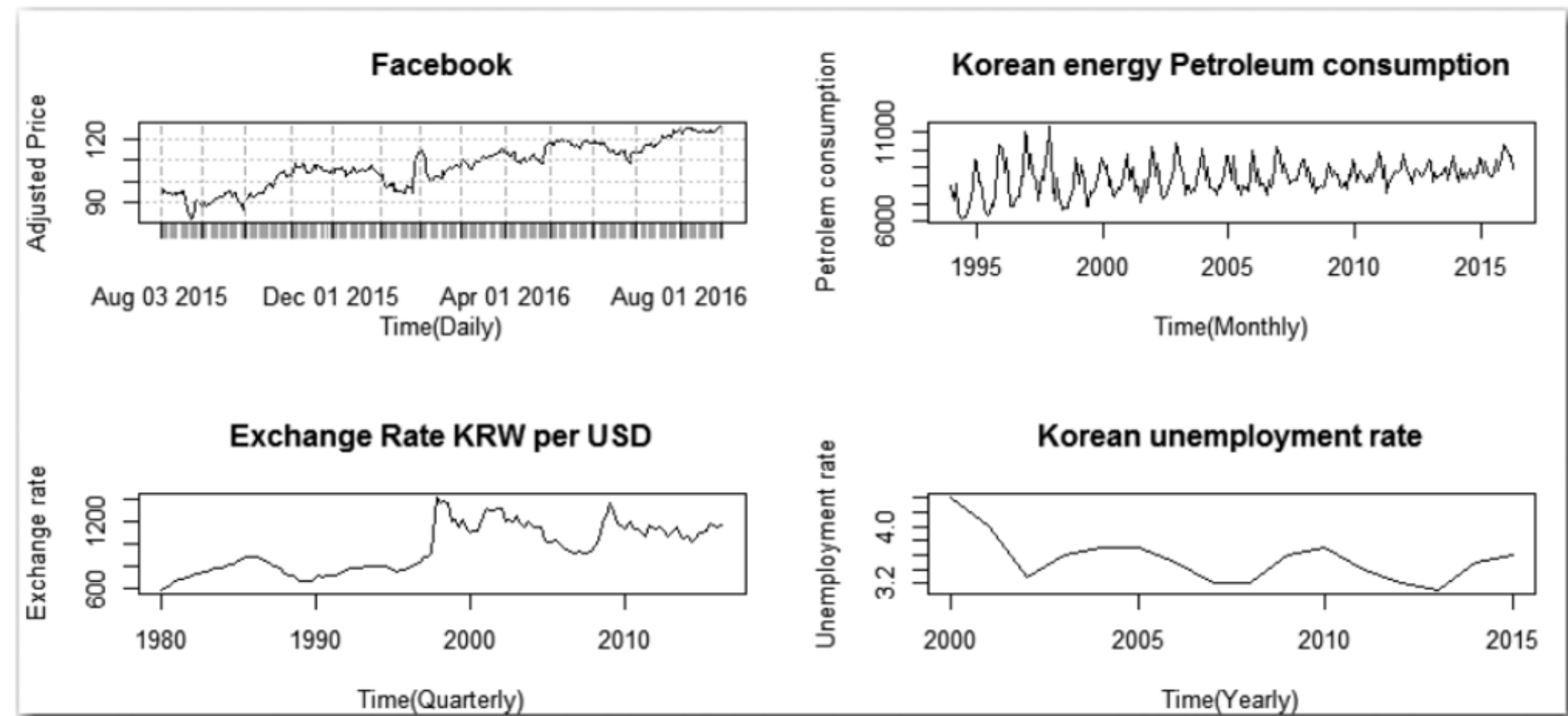
01 시계열의 개념

01 시계열개념

02 시계열 특성

03 푸리에 변환

시계열 데이터 정의



- 시계열 분석의 목적: 시계열이 갖고 있는 숨겨진 규칙을 발견해 이를 모형화하고, 또 추정된 모형을 통하여 미래의 값을 예측(forecasting) 하는 것.

01 시계열의 개념

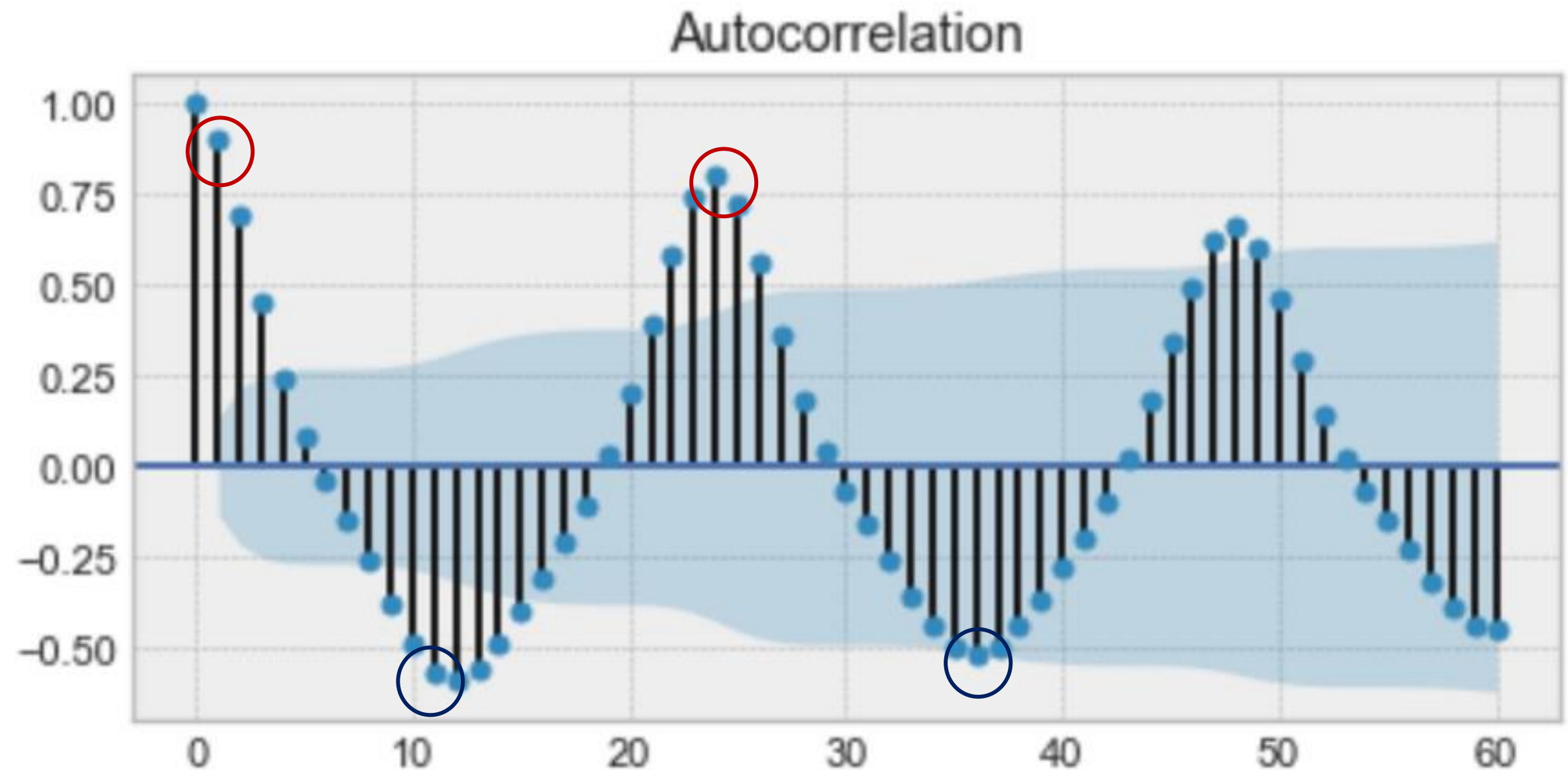
01 시계열 개념

02 시계열 특성

03 푸리에 변환

시계열 특성 1 - 자기상관성 (Autocorrelation)

- 자기상관(AUTOCORRELATION)은 시계열의 시차 값(LAGGED VALUES)사이의 선형 상관관계



01 시계열의 개념

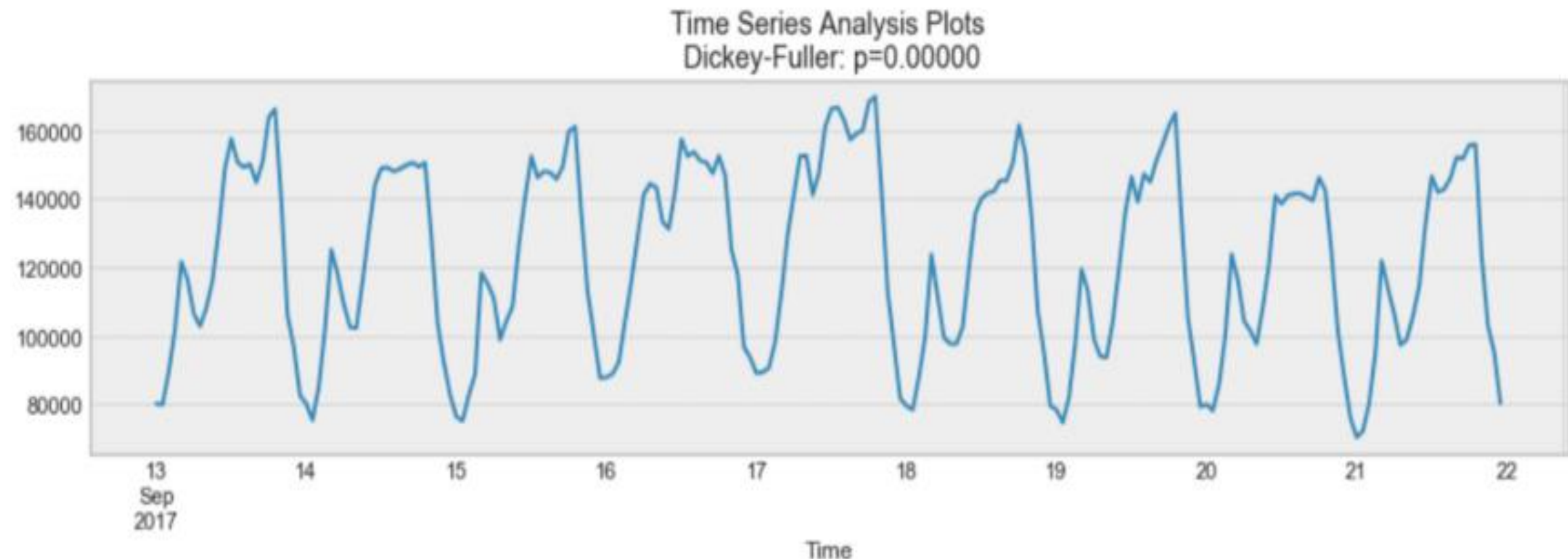
01 시계열 개념

02 시계열 특성

03 푸리에 변환

시계열 특성 2 -계절성 (Seasonality)

- 해마다 어떤 특정한 때나 1주일마다 특정 요일에 나타나는 것 같은 계절성 요인이 시계열에 영향을 줄 때 *계절성(seasonality)* 패턴이 나타난다고 말함.
- Cf. 주기성 (Cycle)



01 시계열의 개념

01 시계열 개념

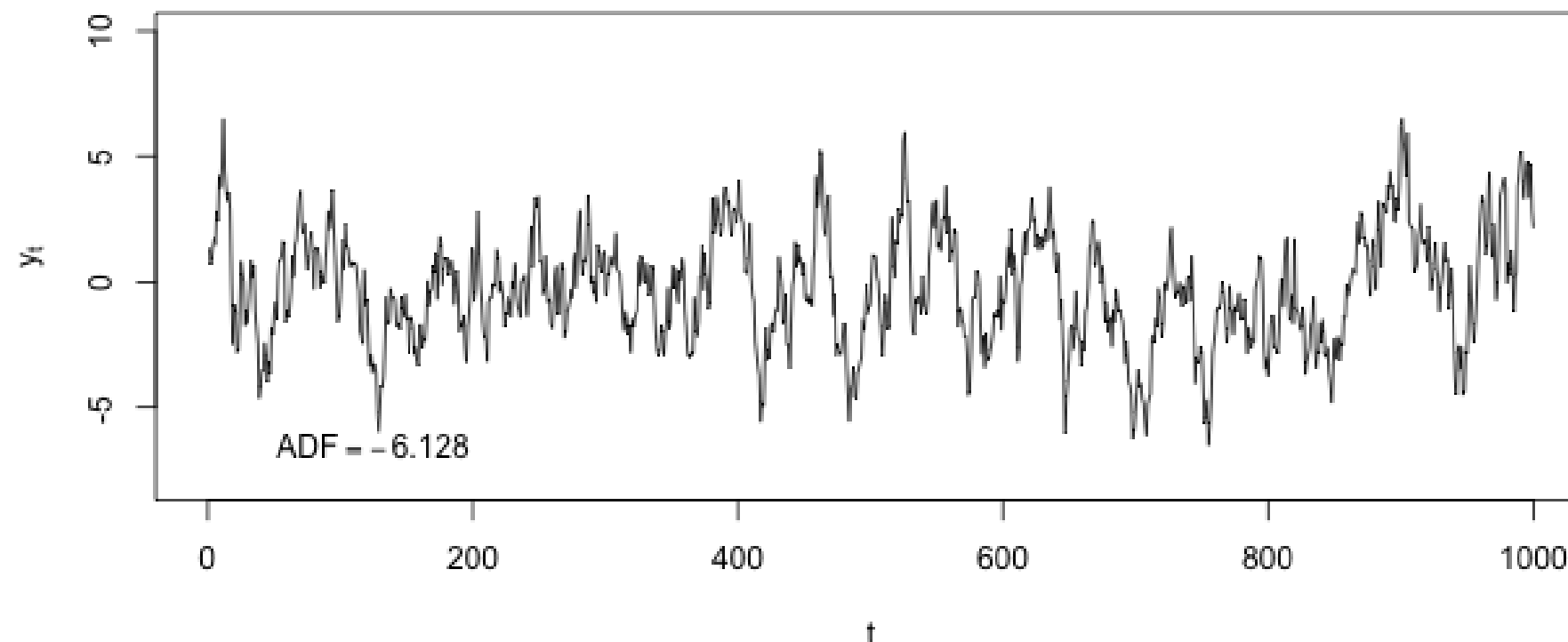
02 시계열 특성

03 푸리에 변환

시계열 특성 3 -정상성 (Stationary)

- 평균과 분산 같은 통계적 특성이 시간에 대해 일정한 성질
- 정상시계열(Stationary)
평균과 표준편차가 일정하다는 조건이 선행되어야 분석이 가능
- 비정상시계열
차분이나 log함수를 씌워 정상시계열로 변환 후 분석을 해야 함.

Stationary Time Series



01 시계열의 개념

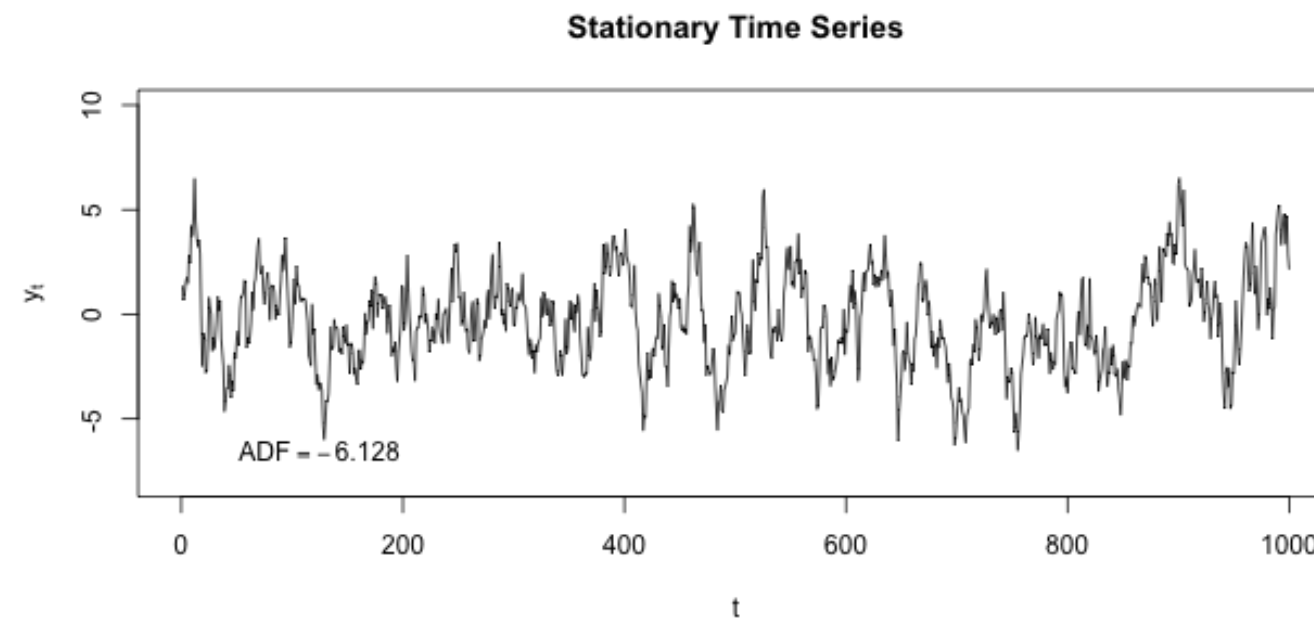
01 시계열 개념

02 시계열 특성

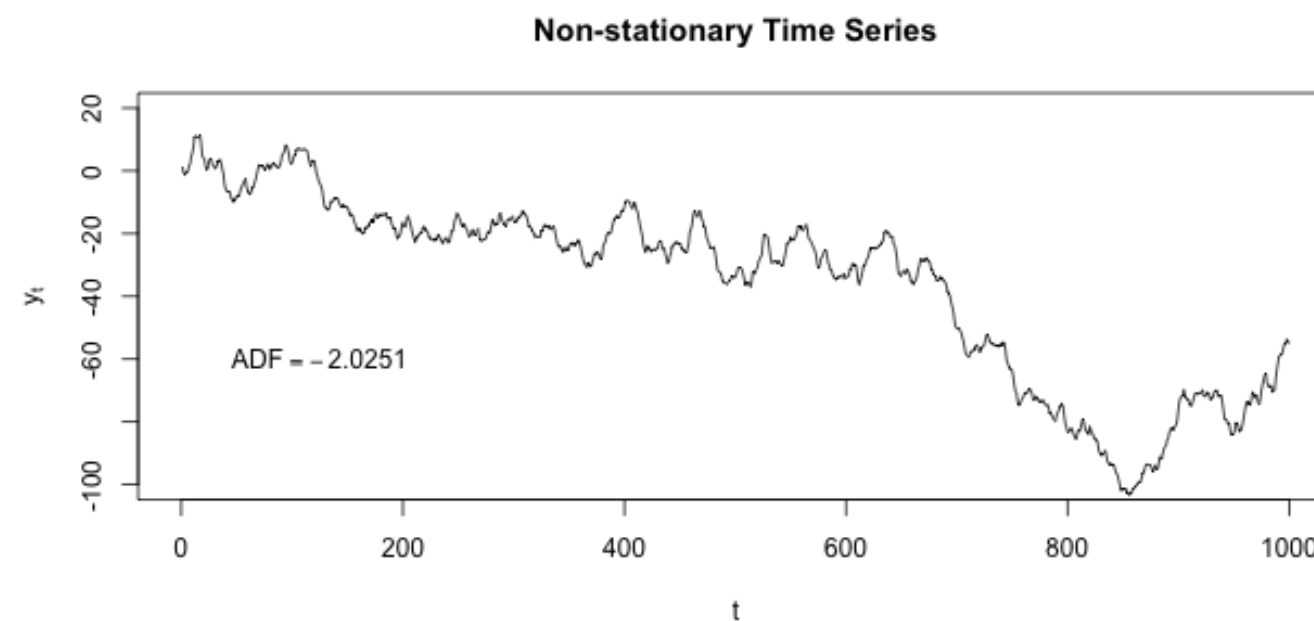
03 푸리에 변환

시계열 특성 3 -정상성 (Stationary)

- 정상 시계열 데이터 vs 비정상 시계열 데이터



평균을 기준으로 움직이는 패턴. 데이터의 움직임 또한 일정 수준을 넘어가지 않고 있음.



평균값 중심이 아니며 변동 폭 역시 다양함.

01 시계열의 개념

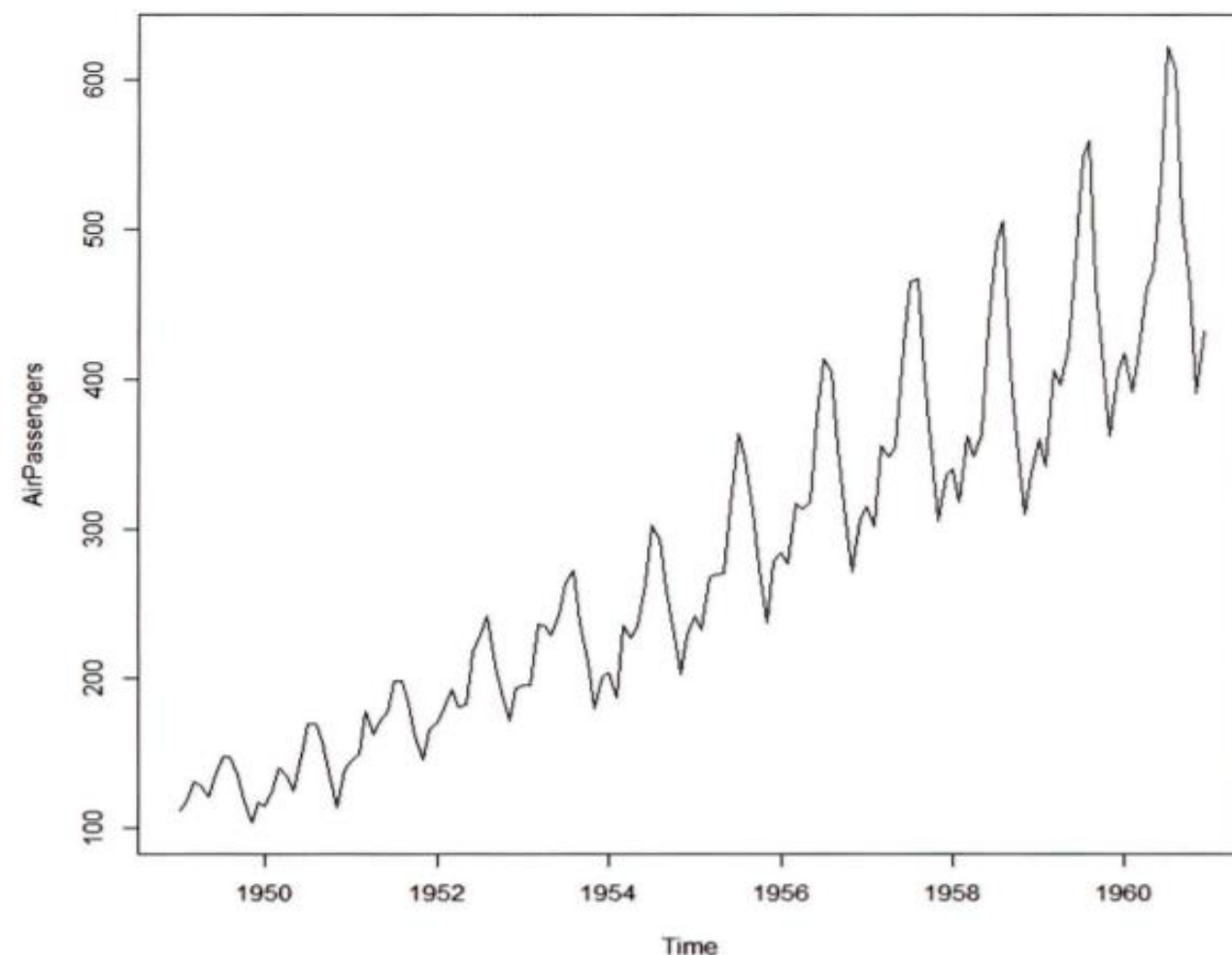
01 시계열 개념

02 시계열 특성

03 푸리에 변환

시계열 특성 3 -정상성 (Stationary)

- 정상성 검정: 디키 풀러 (*Dickey-Fuller*) 검정으로 정상성 검정



<정상적이지 않은
데이터셋의 예시 (연도별
비행기 탑승객 수)>

01 시계열의 개념

01 시계열 개념

02 시계열 특성

03 푸리에 변환

푸리에 변환 (Fast Fourier Transform)

- 푸리에 변환(FOURIER TRANSFORM): 주기 T 를 가지고 반복되는(CYCLIC) 모든 함수 $Y(T)$ 는 주파수와 진폭이 다른 몇 개의 사인 함수의 합으로 나타낼 수 있음.

이 사인 함수의 진폭을 구하는 과정을 푸리에 변환(FOURIER TRANSFORM)이라고 함.

$$\hat{f}(\xi) = \int_{-\infty}^{\infty} f(x) e^{-2\pi i x \xi} dx$$

* 푸리에 변환(FT)는 TIME-DOMAIN 함수를 FREQUENCY DOMAIN으로 분해함.

이때 FREQUENCY에 역수를 취함으로써 시계열 데이터의 주기를 구할 수 있음.

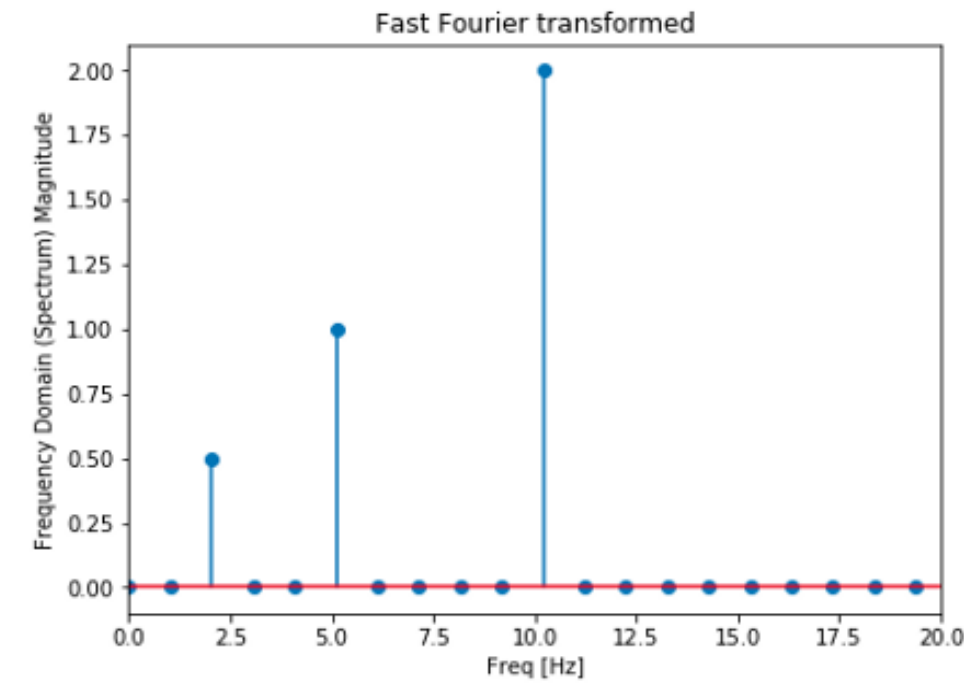
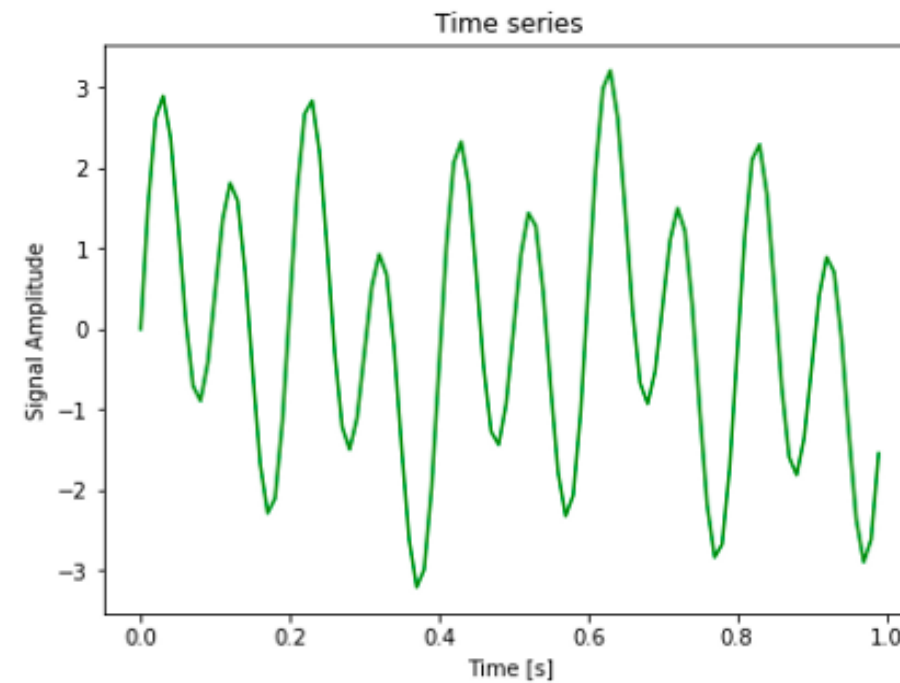
01 시계열의 개념

01 시계열 개념

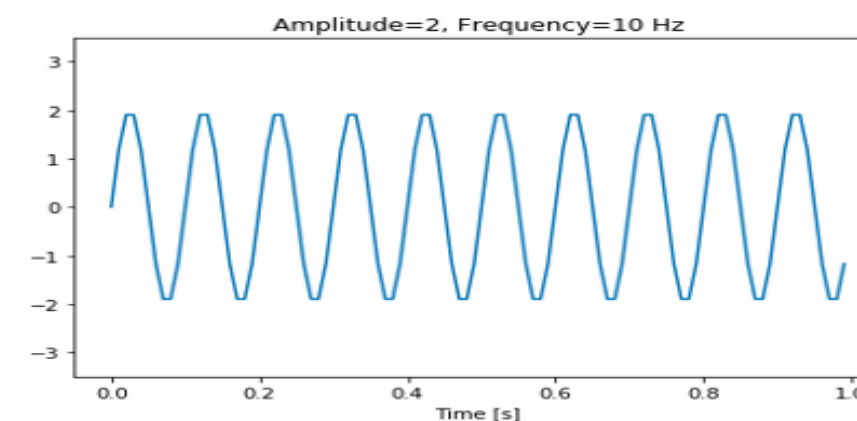
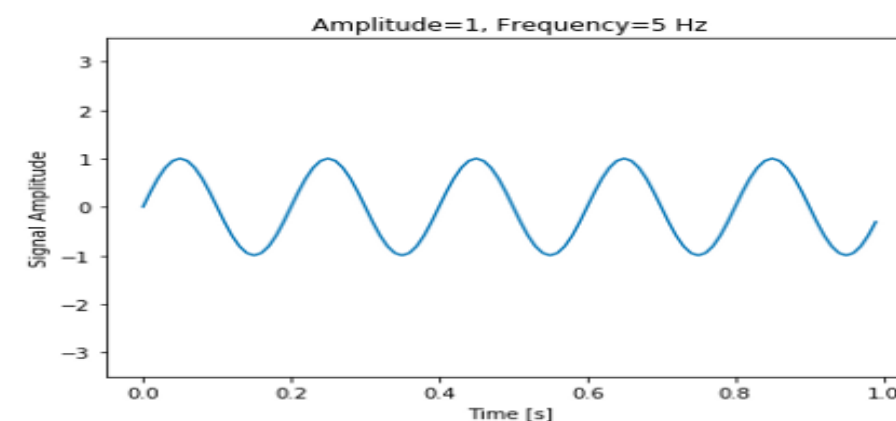
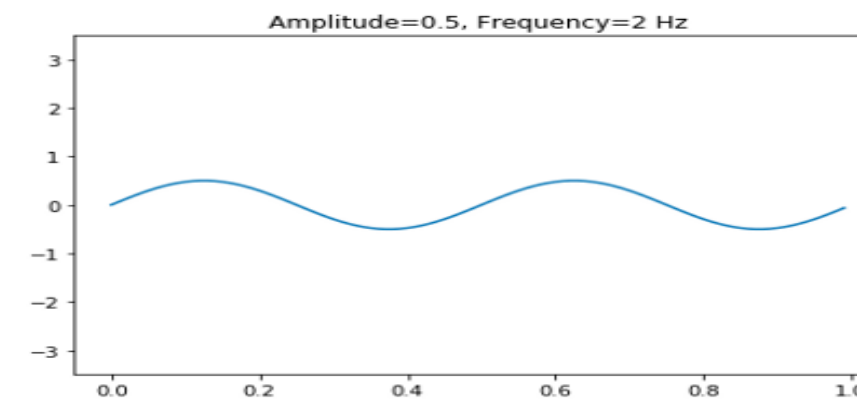
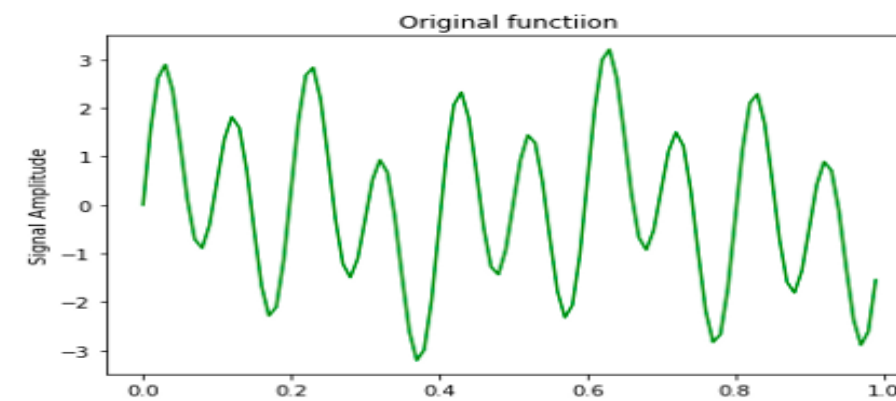
02 시계열 특성

03 푸리에 변환

푸리에 변환 (Fast Fourier Transform)



주파수(Frequency)
3개: 2, 5, 10 Hz
Amplitudes 3개:
0.5, 1, 2



02

시계열 예측 모델

- 01. 통계 모델
- 02. 머신러닝/딥러닝
- 03. Fbprophet

02 시계열 예측 모델

01 통계 모델

02. 머신러닝/딥러닝

03. Fbprophet

통계모델: AR, MA, ARIMA, SARIMA, VAR 등

시계열을 잘 설명할 수 있는 모형을 만들고,
모형의 모수를 찾는다. (회귀모형과 비슷한 아이디어)

구체적인 방법(간략히)

1. 시계열의 **정상성**을 확보
2. ACF(자기상관함수), PACF(편자기상관함수), 시도표 등을 통해 데이터에 적합한 **모형과 모형의 파라미터를 선택**
3. 선택한 모형의 **모수**를 찾는다 (손계산 x, 컴퓨터 패키지 사용)
4. 찾아낸 모수를 바탕으로 미래 시계열 **예측**

02 시계열 예측 모델

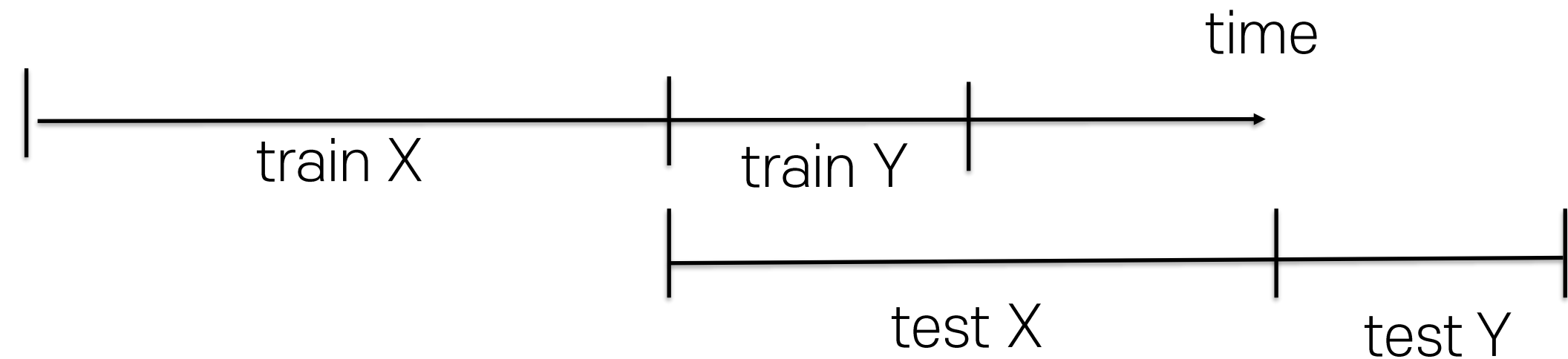
01. 통계 모델

02 머신러닝/딥러닝

03. Fbprophet

머신러닝

- 주로 클러스터링과 트리 기반 방법론으로 시계열의 예측과 분류문제를 다룸.



- 시계열의 특성에 맞게 train dataset과 test dataset 정의
- 다수 시계열 모델링 대회에서 **XGboost**는 전통적인 통계모델보다 우수한 성능을 보여줌.
- 시계열 예측 -> 시계열 **회귀** 또는 **분류**
ex) 과거 12개월 간의 성적으로 미래 3개월 이후의 자퇴 여부 예측 -> 분류문제

02 시계열 예측 모델

01. 통계 모델

02 머신러닝/딥러닝

03. Fbprophet

딥러닝

- RNN, LSTM, Sequence to Sequence
- 딥러닝은 정상성을 요구하지 않고, ARIMA 모델의 차수 계절성에 따른 파라미터를 고르는 방법을 걱정할 필요가 없음.
- 단, Dataset에 따라 딥러닝 모델이 만능은 아님.

02 시계열 예측 모델

01. 통계 모델

02. 머신러닝/딥러닝

03 Fbprophet

Fbprophet



Facebook에서 만든
시계열 분석 라이브러리



직관적인 파라미터 조정



ARIMA 모델이 요구하는
데이터 전처리가 불필요



빠른 훈련, 쉬운 사용

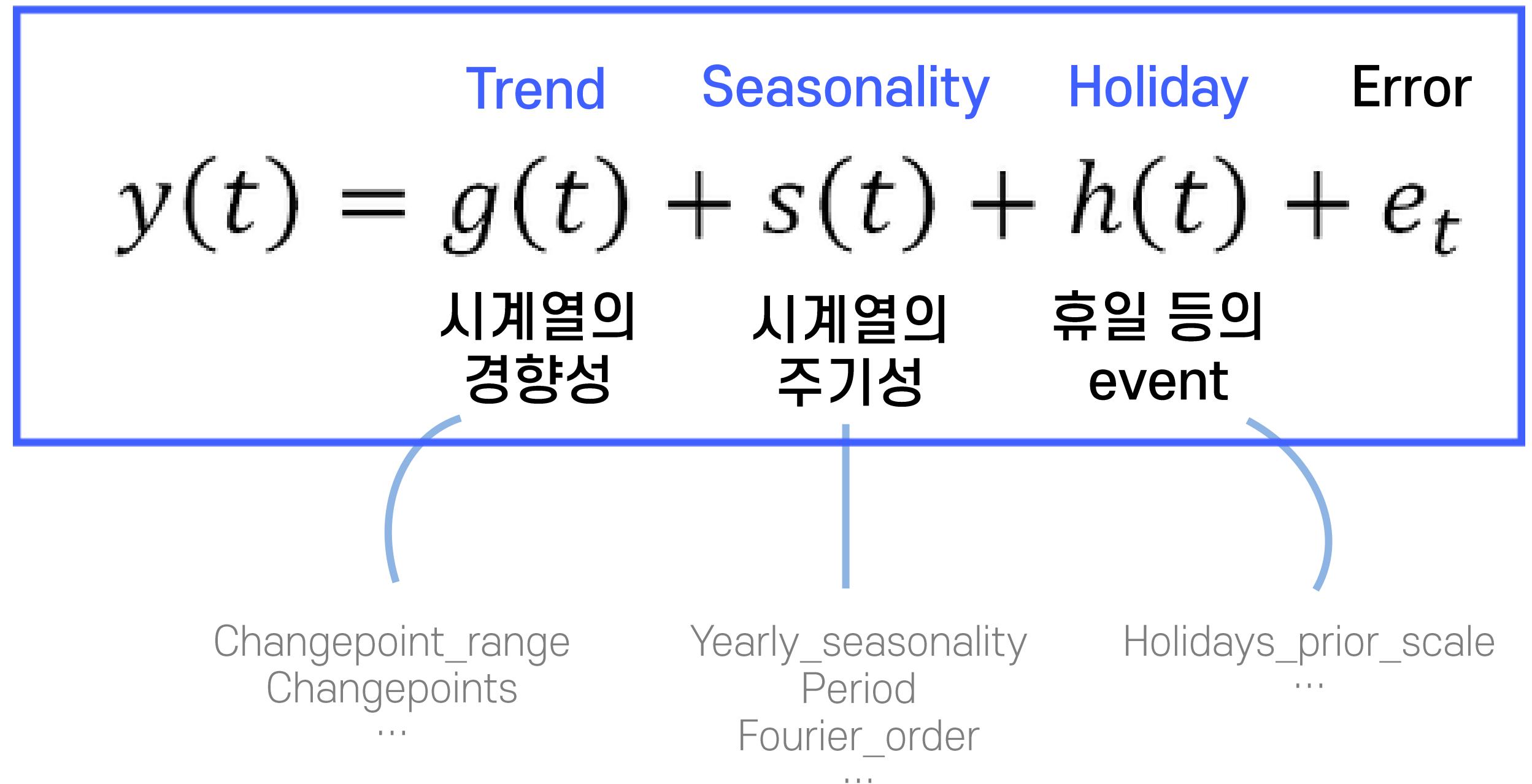
02 시계열 예측 모델

01. 통계 모델

02. 머신러닝/딥러닝

03 Fbprophet

Fbprophet의 구성 요소



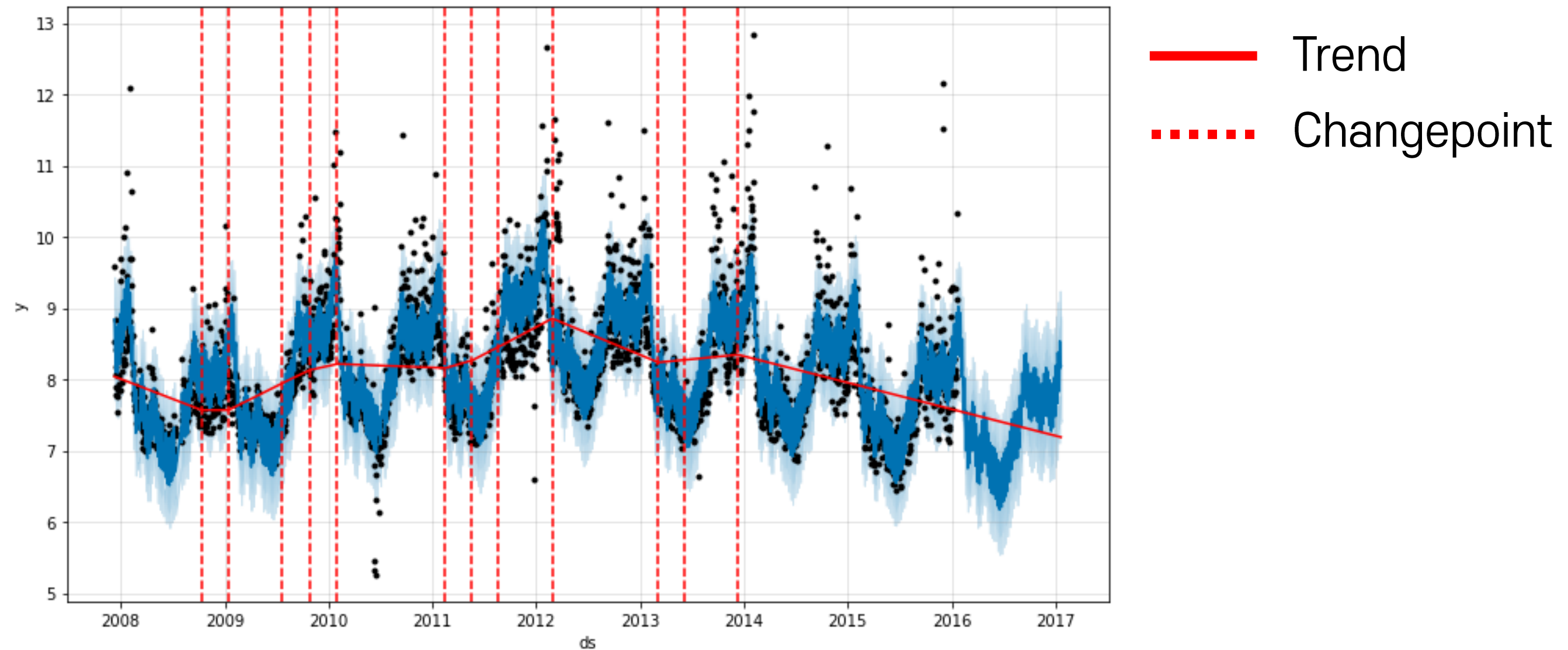
02 시계열 예측 모델

01. 통계 모델

02. 머신러닝/딥러닝

03 Fbprophet

Parameter : ^{Trend} $g(t) + s(t) + h(t) + e_t$



Changepoint_range : Changepoint의 개수
Changepoint_prior_scale : Trend의 유연성(fitting 조절)
Changepoints : Changepoint 날짜 지정

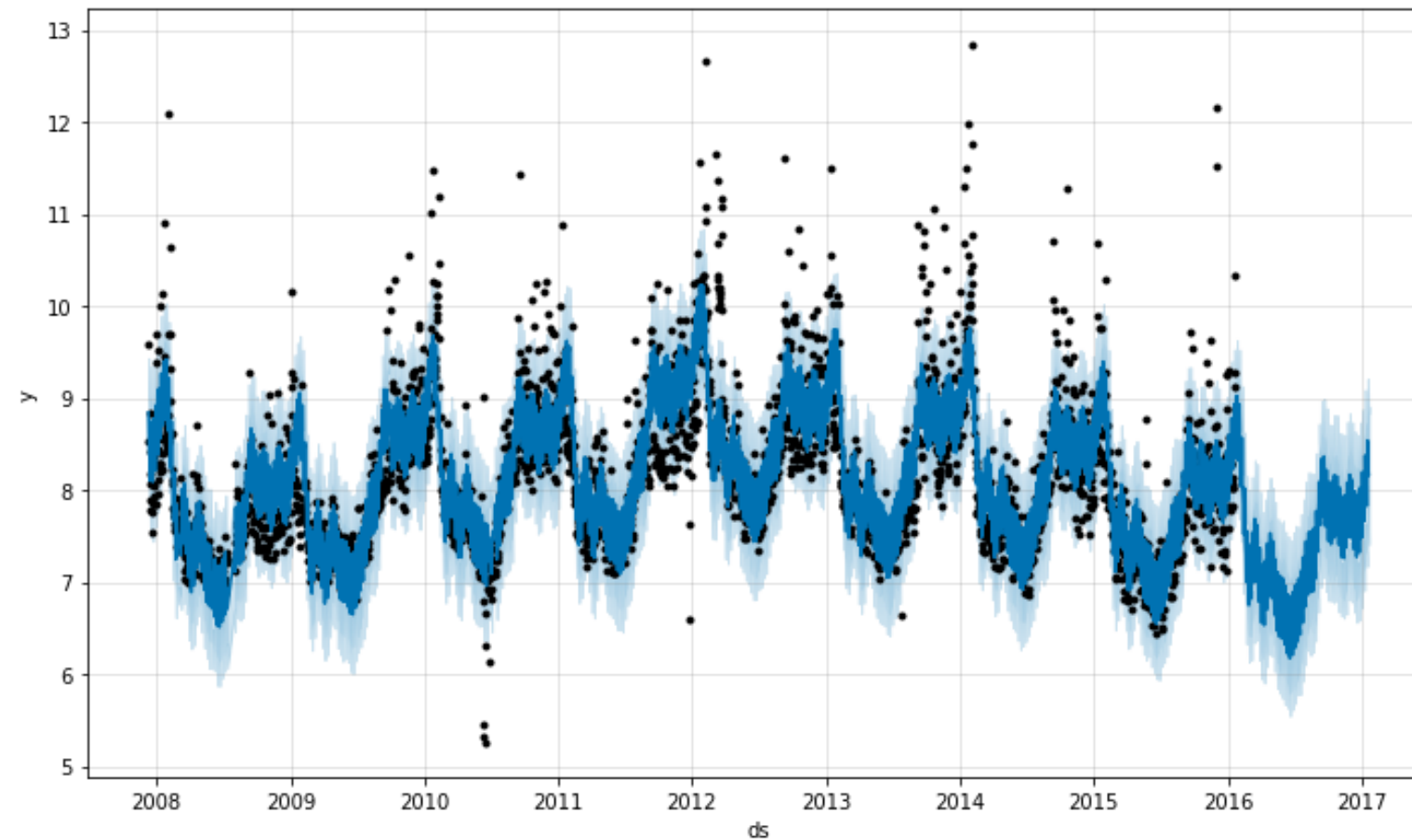
02 시계열 예측 모델

01. 통계 모델

02. 머신러닝/딥러닝

03 Fbprophet

Parameter : $g(t) + \overset{\text{Seasonality}}{s(t)} + h(t) + e_t$



Fourier series를
이용해 주기 패턴을 추정

Yearly(weekly, daily)_seasonality : 연(주, 일) 주기

add_seasonality(period, : 주기의 기간

fourier_order, : fit faster changing cycles

prior_scale) : 추정된 seasonality의 반영 강도

02 시계열 예측 모델

01. 통계 모델

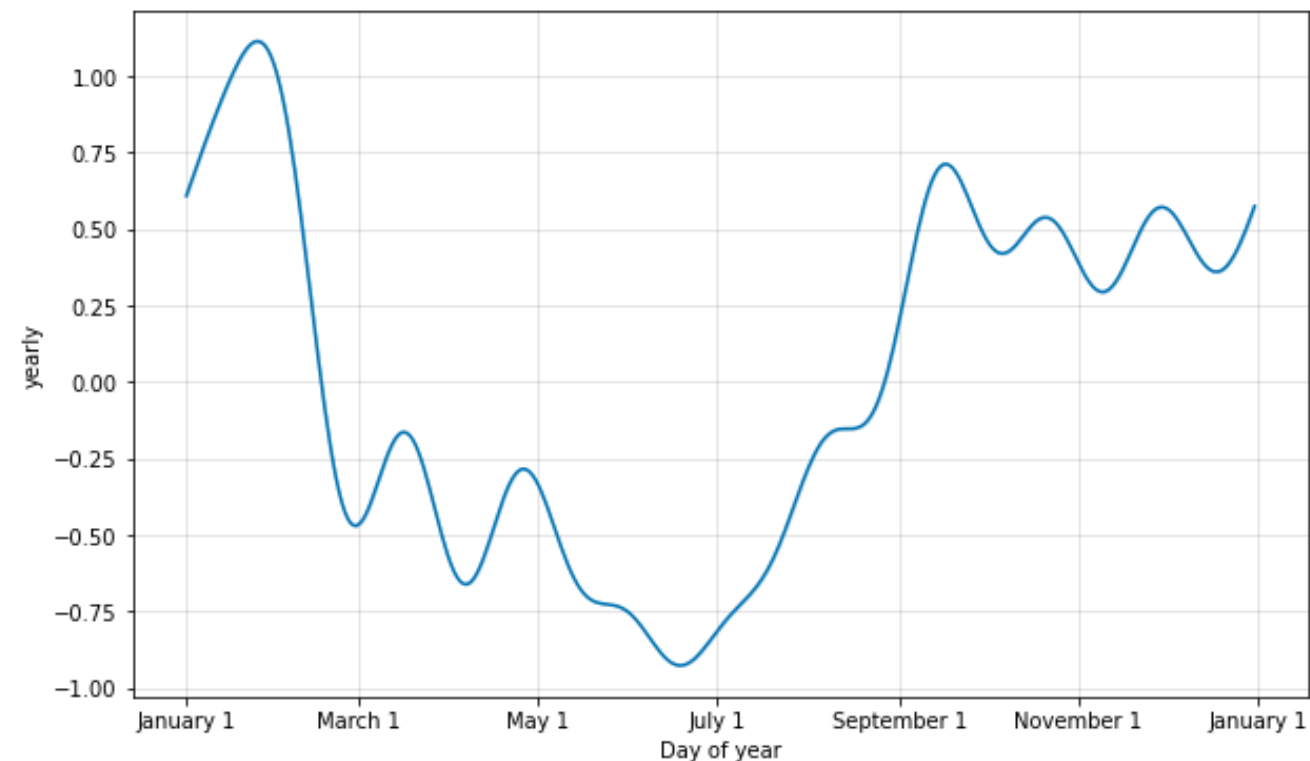
02. 머신러닝/딥러닝

03 Fbprophet

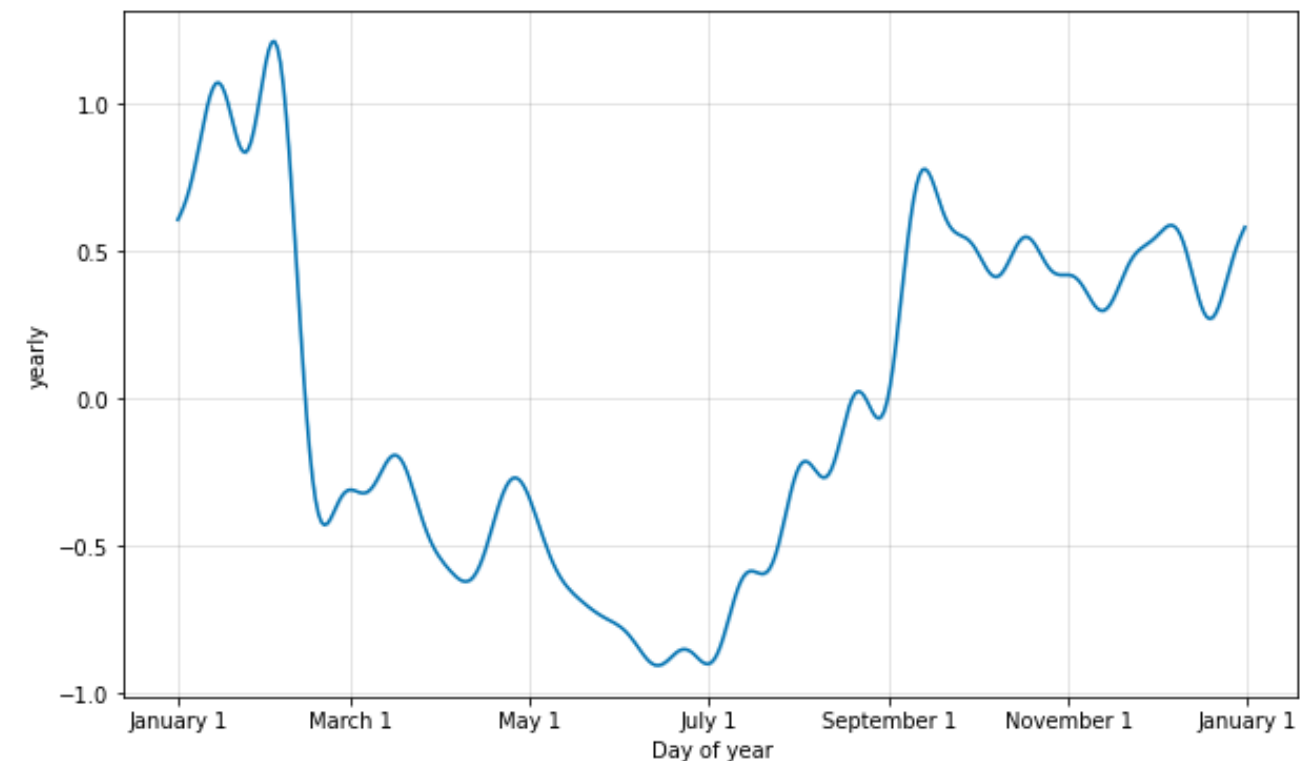
Parameter : $g(t) + \overset{\text{Seasonality}}{s(t)} + h(t) + e_t$

add_seasonality(period, **fourier_order**, prior_scale)

fourier_order = 10



fourier_order = 20



02 시계열 예측 모델

01. 통계 모델

02. 머신러닝/딥러닝

03 Fbprophet

Parameter : $g(t) + s(t) + \overset{\text{Holiday}}{h(t)} + e_t$

lower(upper)_window : holiday의 범위

add_country_holidays(country_name='KR') : holiday(한국) 추가

holiday_prior_scale : holiday의 반영 강도

03

주식 종가 예측 대회

01. FBPROPHET

02. LSTM

03 주식 종가 예측

01 Fbprophet

02. LSTM

주식 종가 예측 대회



국내 주식 종가 예측 대회 (DACON 주관)

- 주가: 정상성X
- 비트코인 대회에서의 많은 Fbprophet 사용



Fbprophet 시도!

03 주식 종가 예측

01 Fbprophet

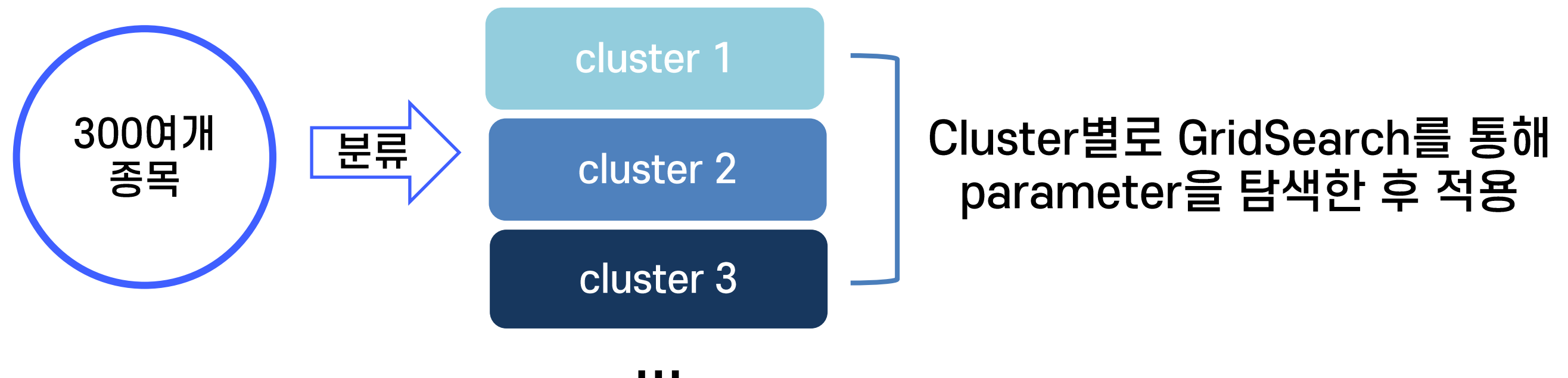
02. LSTM

Fbprophet 적용



비트코인 트레이더 경진대회
(DACON 주관)

GridSearch를 통한 최적 **parameter** 탐색
-> 모든 코인 종목에 동일 파라미터 적용

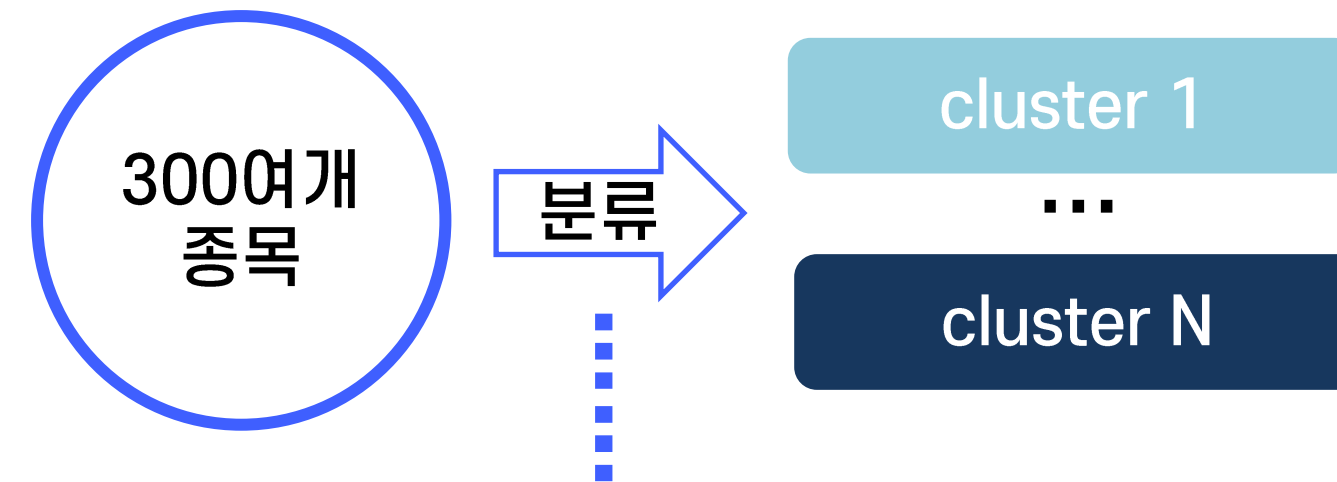


03 주식 종가 예측

01 Fbprophet

02. LSTM

Fbprophet 적용



분류기준: 산업군 (네이버 금융에서 주식 종목의 산업군 크롤링)

은행



자동차

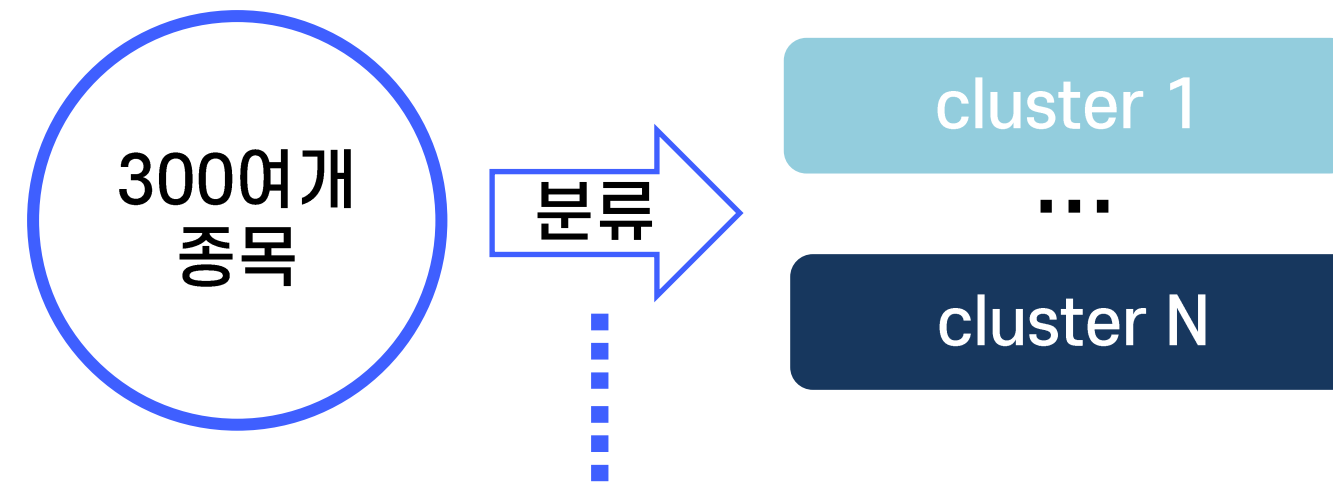


03 주식 종가 예측

01 Fbprophet

02. LSTM

Fbprophet 적용



결과: 7.14 (NMAE*100)

- 산업군 기준 클러스터링의 한계
- 파라미터 범위 잘못 설정

분류기준: 산업군 (네이버 금융에서 주식 종목의 산업군 크롤링)

은행



자동차



03 주식 종가 예측

01 Fbprophet

02. LSTM

Fbprophet 적용

```
Prophet(seasonality_mode='multiplicative',  
yearly_seasonality=False,  
weekly_seasonality=False,  
daily_seasonality=False,  
changepoint_prior_scale=0.5,  
changepoint_range=0.95)
```

결과
6.61, 6위

Changepoint_prior_scale = 0.5
- 변칙적인 주가 -> 유연하게 fit

ChangePoint_range = 0.95
- 주가의 빠른 변동성 포착

03 주식 종가 예측

01 Fbprophet

02. LSTM

LSTM 모델 적용

주식 종목 하나에 대해,

1. Rolling Window 방법으로, 주가를 MinMaxScaling (-1, 1)

2. 은닉 노드 개수 : 20, 은닉층 1개를 사용한 LSTM 모델로
모델 훈련 및 예측

- 은닉 노드의 개수는 늘릴수록 성능은 좋아지고,
은닉층의 개수는 늘릴수록 과적합(overfitting)이 쉽게 발생하여 은닉층은 딱
1개로 결정함.

3. inverse_transformation으로 scaling 된 값을 원래 단위로
변환

- 이 과정을 모든 주식 종목(376개)에 적용

결과

7.63387, 20위

감사합니다

시계열 1팀