시계열자료분석팀

5팀

문병철
손재민
주혜인
한지원
홍현경

INDEX

1

ARIMA

정의

결정적 vs 확률적

결정적(Deterministic)

Random의 의미와 반대

EX) $\beta_0 + \beta_1 t$*에 동일한* t *대입*

↓

동일한 추세 반환

확률적(Stochastic)

인접한 자료 간

확률적 구조에 의해 형성

EX) 인접한 시계열간

강한 상관관계로

추세가 존재하는 경우

정의

결정적 vs 확률적

그렇다면
계절성에서는 말이지 ~

계절성

결정적 계절성 → 조화회귀로 추정

$$a_0 + \sum_{j=1}^{k} \left(a_j \cos(\lambda_j t) + b_j \sin(\lambda_j t)\right)$$

1주차 클린업 참고!

확률적 계절성 → SARIMA 모형으로 추정

정의

ARIMA 모형

: 자기회귀누적이동평균과정

d차 차분한 시계열 X_t가
ARMA(p,q) 과정

=

ARIMA(p,d,q)

차분과 ARMA 모형이 결합된 모델

추세(Trend)+ARMA(p,q)형태가 존재할 때 사용

정의

ARIMA 모형

: 자기회귀누적이동평균과정

ARIMA(p,d,q)

$$\phi(B)(1-B)^d X_t = \theta(B) Z_t$$

$$\left(1-\phi_1 B-\phi_2 B^2-\cdots-\phi_p B^p\right)(1-B)^d X_t=\left(1+\theta_1 B+\theta_2 B^2+\cdots+\theta_q B^q\right) \quad Z_t$$

ARIMA의 장점 !

여전히 선형과정

차분까지 포함해 간결한 표현 가능

정의

ARIMA 모형

: 자기회귀누적이동평균과정

ARIMA(p,d,q)

ARIMA에서의 I란...?

“ Integration ”

$$\phi(B)(1-B)^d X_t = \theta(B) Z_t$$

$$\phi(B)(1-B)X_t = \theta(B) Z_t$$

$$\left(1-\phi_1 B-\phi_2 B^2-\cdots-\phi_p B^p\right)(1-B)^d X_t=\left(1+\theta_1 B+\theta_2 B^2+\cdots+\theta_q B^q\right) \ Z_t$$

$$Y_t = (1-B)X_t = X_t - X_{t-1}$$

$$X_t = X_{t-1} + Y_t = (X_{t-2} + Y_{t-1}) + Y_t = \cdots = X_0 + \boxed{\sum_{j=1}^{t} Y_j}$$

여전히 선형과정

차분까지 포함해 간결한 표현 가능

X_t의 관점에서 Y_t의 누적합

이게 바로 ARIMA의 I ~

적합절차

모형의 적합절차

1. 시계열/ACF 그래프로 정상성 파악

비정상성을 띈다면 !

2. 차분을 통해 정상화

3. 모수의 추정(model estimation)

4. 모형의 진단(model diagnostics)

5. 예측모형(forecast model)으로 선택

적합절차

모형의 적합절차

1. 시계열/ACF 그래프로 정상성 파악

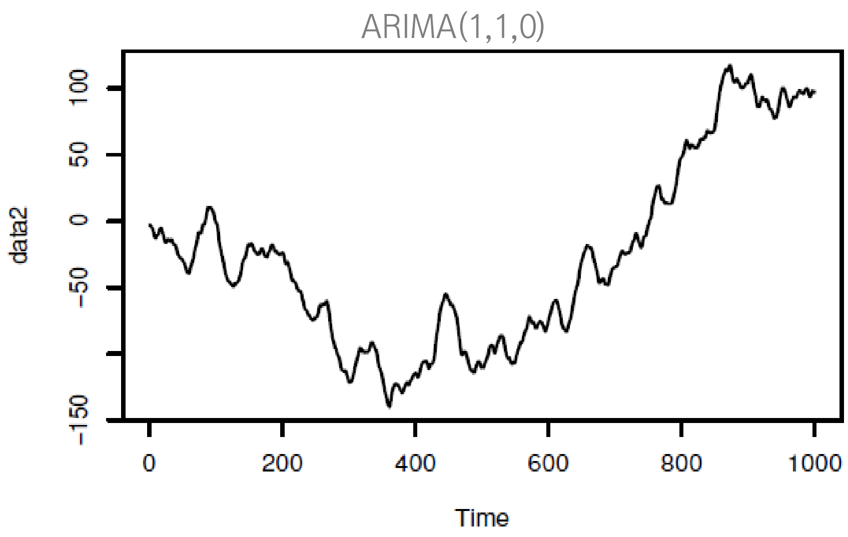
비정상성을 띈다면 !

2. 차분을 통해 정상화
3. 모수의 추정(model estimation)
4. 모형의 진단(model diagnostics)
5. 예측모형(forecast model)으로 선택

적합절차

모형의 적합절차

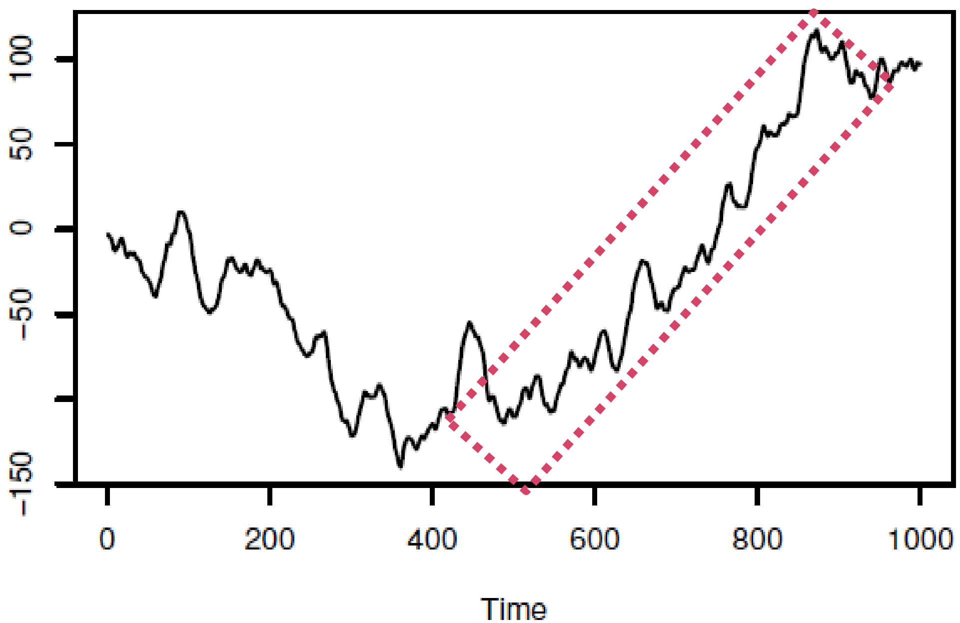
시계열 Plot



시계열/ACF 그래프로 정상성 파악

적합절차

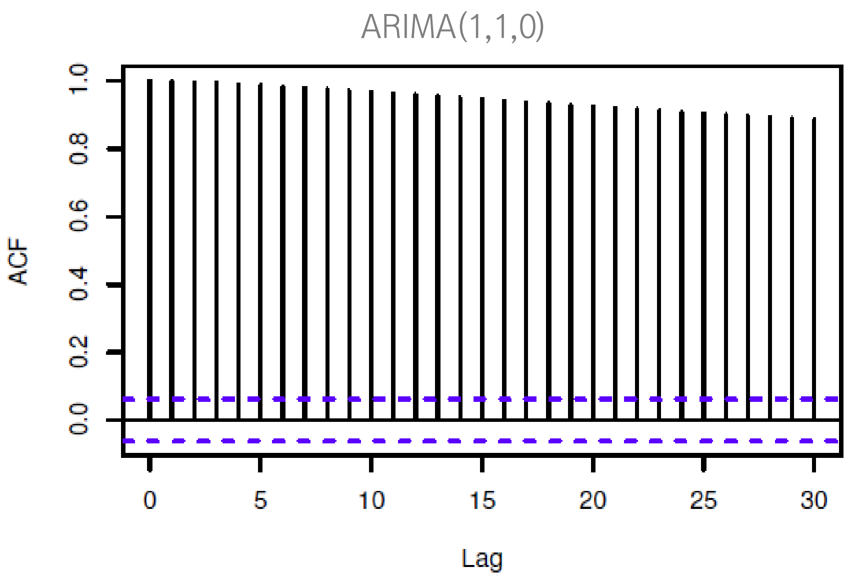
시계열 그래프로 추세가 존재함을 확인



적합절차

모형의 적합절차

ACF Plot



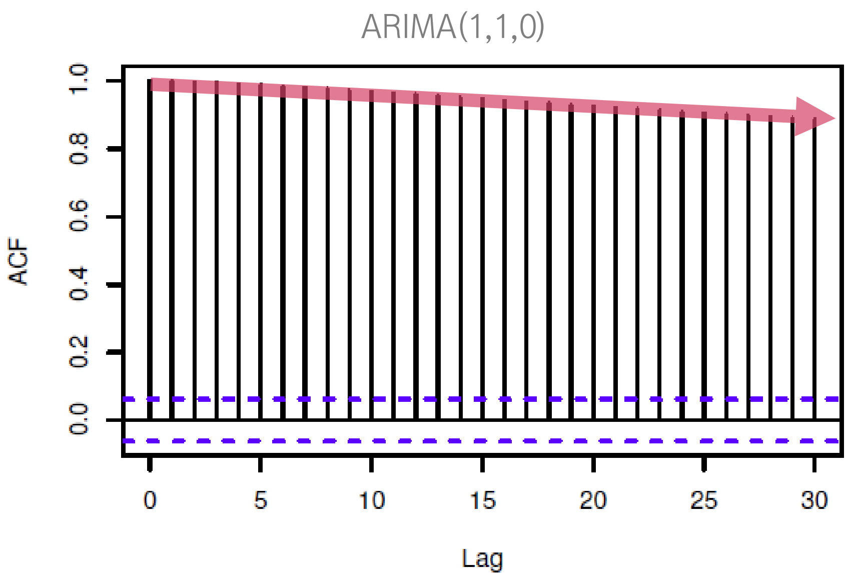
시계열/ACF 그래프로 정상성 파악

적합절차

모형의 적합절차

ACF Plot

ACF가 천천히 감소하고 있구나 ~



추세가 존재한다면, ACF는 점점 감소

적합절차

모형의 적합절차

추세가 존재한다면 → **차분**을 통해 정상화

이 때,

차분의 차수 d가 1,2를 넘어가면 **과대차분** 위험

과대차분(overdifferencing)을 조심하세요

과대차분이란 말 그대로 차분을 과하게 하는 것 입니다. 이미 정상화가 되어있음에도 불구하고 차분을 또 시도하는 경우, 정상시계열의 선형 결합은 다시 정상 시계열이 되어 "정상성"자체에는 문제가 없습니다. 그러나, 지나친 차분은 ACF를 복잡하게 만들거나 분산이 커지는 문제가 생깁니다.

⇨ 이 성질은 모형 선택시에도 유용하게 사용됩니다. (분산의 크기 비교를 통해 적절한 모형 선택)

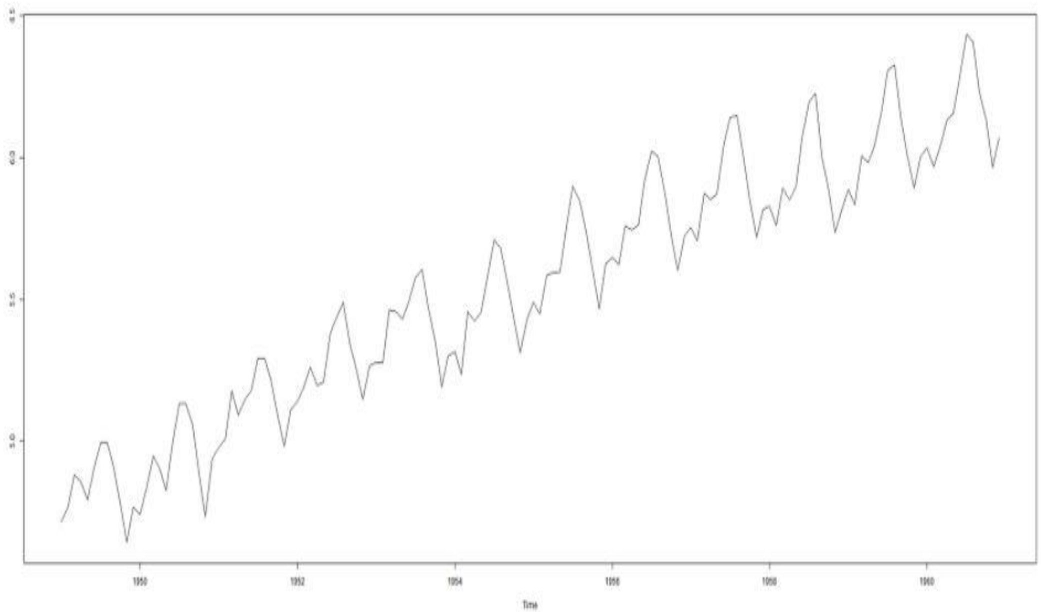
시계열 1주차 클린업 참고 !

2

SARIMA

● **계절자기회귀이동평균 모형**

SARIMA 모형



추세와 계절성이 모두 존재할 때 사용가능한 모형

계절자기회귀이동평균 모형

SARIMA 모형

SARIMA(p,d,q) x (P,D,Q)s

$$\phi(B)\Phi(B^s)(1-B)^d(1-B^s)^D X_t$$

$$=$$

$$\theta(B)\Theta(B^s)Z_t\ , Z_t \sim WN(0,\sigma^2)$$

추세와 **계절성**을 한 번에 제거

추세와 계절성

한번에 처리하는 나란 놈

어쩌면 . . .

조금 . . . 멋있을지도 ?

계절자기회귀이동평균 모형

IDEA

SARIMA는 또 뭐냐구..? 예시를 통해 알아보자 ~

EX)

월 / 연도	Month1	Month2	…	Month12
Year1	Y_1	Y_2	…	Y_{12}
Year2	Y_{13}	Y_{14}	…	Y_{24}
⋮	⋮	⋮	⋮	⋮
Year r	$Y_{1+12(r-1)}$	$Y_{2+12(r-1)}$	…	$Y_{12+12(r-1)}$

⋮

주기가 s=12인 시계열 데이터가 존재

계절자기회귀이동평균 모형

IDEA

STEP 1)

Column을 고정하고,

시계열들이 ARMA(P,Q)모형을 따른다 가정

EX)

연도 \ 월	Month1	Month2	…	Month12
Year1	Y_1	Y_2	…	Y_{12}
Year2	Y_{13}	Y_{14}	…	Y_{24}
⋮	⋮	⋮		⋮
Year r	$Y_{1+12(r-1)}$	$Y_{2+12(r-1)}$	…	$Y_{12+12(r-1)}$

$$Y_{j+12t} - \Phi_1 Y_{j+12(t-1)} - \cdots - \Phi_P Y_{j+12(t-P)}$$
$$\shortparallel$$
$$U_{j+12t} + \Theta_1 U_{j+12(t-1)} + \cdots + \Theta_Q U_{j+12(t-Q)}$$
$$t = 0,1,\ldots,11$$

주기가 s=12인 시계열 데이터가 존재

계절자기회귀이동평균 모형

IDEA

STEP 1)

Column을 고정하고,

시계열들이 ARMA(P,Q)모형을 따른다 가정

EX)

연도 \ 월	Month1	Month2	…	Month12
Year1	Y_1	Y_2	…	Y_{12}
Year2	Y_{13}	Y_{14}	…	Y_{24}
⋮	⋮	⋮		⋮
Year r	$Y_{1+12(r-1)}$	$Y_{2+12(r-1)}$	…	$Y_{12+12(r-1)}$

$$Y_{j+12t} - \Phi_1 Y_{j+12(t-1)} - \cdots - \Phi_P Y_{j+12(t-P)}$$

$$\parallel$$

$$U_{j+12t} + \Theta_1 U_{j+12(t-1)} + \cdots + \Theta_Q U_{j+12(t-Q)}$$

$$t = 0,1,\ldots,11$$

특성함수로 나타내면,

$$\Phi(B^{12})Y_t = \Theta(B^{12})U_t$$

$$U_t \sim WN(0, \sigma_U^2)$$

계절자기회귀이동평균 모형

IDEA

$$\Phi(B^{12})Y_t = \Theta(B^{12})U_t$$

$$U_t \sim WN(0,\sigma_U^2)$$

Month1	Month2	…	Month12
Y_1	Y_2	…	Y_{12}

STEP 2)

한 주기 내에서도 서로 Correlation이 존재할 수 있음

오차항 U_t가 백색잡음이 아닌 ARMA(p,q)모형을 따른다고 가정

계절자기회귀이동평균 모형

IDEA

$$\Phi(B^{12})Y_t = \Theta(B^{12})U_t$$

$$U_t \sim WN(0,\sigma_U^2)$$

Month1	Month2	⋯	Month12
Y_1	Y_2	⋯	Y_{12}

STEP 2)

한 주기 내에서도 서로 Correlation이 존재할 수 있음

오차항 U_t가 백색잡음이 아닌 ARMA(p,q)모형을 따른다고 가정

이렇게 표현할 수 있겠다 !

$$\phi(B)U_t = \theta(B)Z_t \, , Z_t \sim WN(0,\sigma^2)$$

계절자기회귀이동평균 모형

IDEA

SARIMA는 또 뭐냐구? 예시를 통해 알아보자 ~

EX)

STEP 1 + STEP 2 표현

월 / 연도	Month1	Month2	…	Month12
Year1	Y_1	Y_2	…	Y_{12}
Year2	Y_{13}	Y_{14}	…	Y_{24}
⋮	⋮			⋮
Year r	$Y_{1+12(r-1)}$	$Y_{2+12(r-1)}$		$Y_{12+12(r-1)}$

고심이는 앞의 두 과정을 합쳐보았어 ~

$$\Phi(B^{12})Y_t = \Theta(B^{12})\phi^{-1}(B)\theta(B)Z_t$$

$$\phi(B)\Phi(B^{12})Y_t = \theta(B)\Theta(B^{12})Z_t, Z_t \sim WN(0,\sigma^2)$$

주기가 s=12인 시계열 데이터가 존재

● 계절자기회귀이동평균 모형

IDEA

SARIMA는 또 뭐냐구? 예시를 통해 알아보자 ~

EX)

STEP 1 + STEP 2 표현

연도 \ 월	Month1	Month2	⋯	Month12
Year1	Y_1	Y_2	⋯	Y_{12}
Year2	Y_{13}	Y_{14}	⋯	Y_{24}
⋮	⋮			⋮
Year r	$Y_{1+12(r-1)}$	$Y_{2+12(r-1)}$		$Y_{12+12(r-1)}$

고심이는 앞의 두 과정을 합쳐보았어 ~

$$\Phi(B^{12})Y_t = \Theta(B^{12})\phi^{-1}(B)\theta(B)Z_t$$

$$\phi(B)\Phi(B^{12})Y_t = \theta(B)\Theta(B^{12})Z_t\,, Z_t \sim WN(0,\sigma^2)$$

주기가 s=12인 시계열 데이터가 존재

추세 존재시, d차 차분까지 포함 가능!

$$Y_t = (1-B)^d(1-B^{12})^D X_t$$

계절자기회귀이동평균 모형

SARIMA모형의 종류

SARIMA

- 순수 SARIMA → 계절성만을 고려하는 모델
- 승법 SARIMA → 계절 이외 요소의 연관성까지 고려하는 모델

순수 SARIMA

정의

ARIMA(0, 0, 0) (P, D, Q)s

$$\Phi(B^s)(1-B^s)^D X_t = \Theta(B^s)Z_t$$

$$\Phi(\mathrm{B}^s) = (1-\Phi_1 B^s - \Phi_2 B^{2s} - \cdots - \Phi_P B^{Ps})$$

$$\Theta(\mathrm{B}^s) = (1+\Theta_1 \mathrm{B}^s + \Theta_2 \mathrm{B}^{2s} + \cdots + \Theta_Q \mathrm{B}^{Qs})$$

$$\mathrm{Z_t} \sim WN(0, \sigma^2)$$

✓계절성만을 고려하는 모델

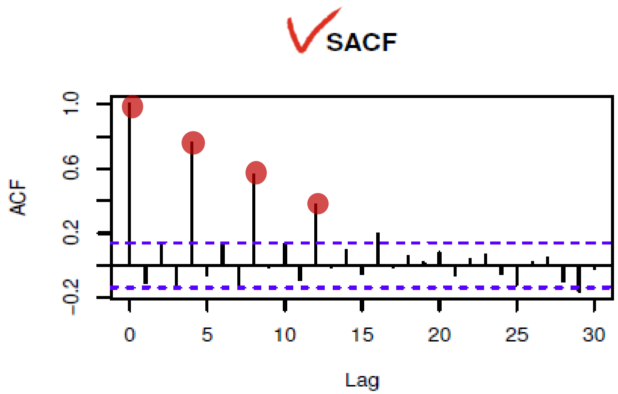
D번의 계절차분, 과거 주기의 P개의 관측치, 과거 주기의 Q개의 오차항

→ 현재의 관측치 설명

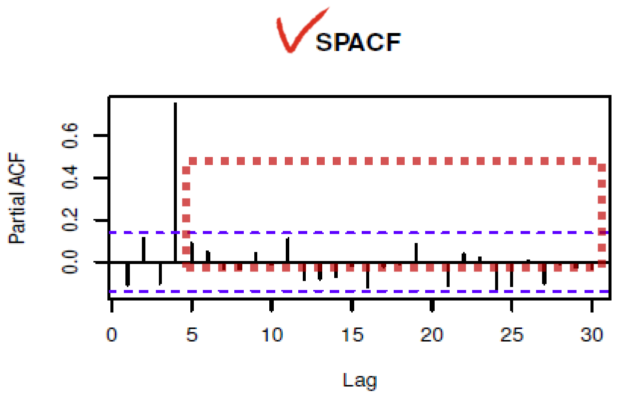
순수 SARIMA

ACF, PACF

ARIMA(0,0,0) X $(1,0,0)_4$



각 주기에서 감소



첫번째 주기 이후로 절단

각 주기에 해당하는 시차의 패턴을 통해 파악

순수 SARIMA

순수 SARIMA 모형의 단점

비계절 요소에 대해서는 고려하지 않아 사용이 제한적

비계절 요소의 연관성까지 고려하는

승법 SARIMA 사용

승법 SARIMA

정의

ARIMA(p, q, d) (P, D, Q)s

$$\phi(B)\Phi(B^s)(1-B)^d(1-B^s)^D X_t = \theta(B)\Theta(B^s)Z_t$$
$$\Phi(B^s) = (1-\Phi_1 B^s - \Phi_2 B^{2s} - \cdots - \Phi_P B^{Ps})$$
$$\Theta(B^s) = \left(1+\Theta_1 B^s + \Theta_2 B^{2s} + \cdots + \Theta_Q B^{Qs}\right)$$
$$\phi(B) = \left(1-\phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p\right)$$
$$\theta(B) = \left(1+\theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q\right)$$
$$Z_t \sim WN(0, \sigma^2)$$

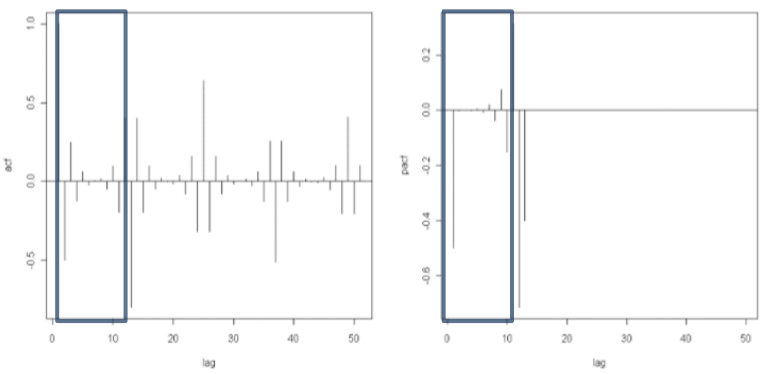
✓계절 이외 요소의 연관성까지 고려하는 모델

오차가 백색잡음인 순수 SARIMA와는 달리, 승법 SARIMA의 경우 오차가 ARMA

승법 SARIMA

ACF, PACF

ARIMA(1, 0, 0) (1, 0, 0)s



계절적 요소 순수 ARIMA와 동일하게 파악

비계절적 요소 한 주기 내의 ACF와 PACF를 통해 파악

3

이분산 시계열모형

3 이분산 시계열모형

모형의 필요성

정의

: 조건부 분산을 시간의 함수로 표현하는 시계열 모형

ARIMA 등의 **전통적 시계열 모형**

등분산 가정 + 평균 부분의 움직임에 관심

BUT !

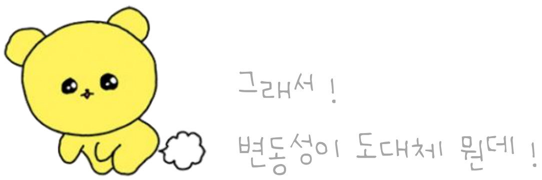
수익률, 이자율, 환율 등의 시계열 자료에는

시간에 따라 **변동성**이 변하는 성질 존재

시간에 따라 변화하는 변동성을 설명할 모형 필요

모형의 필요성

변동성(Volatility)



경제학 분야에서 위험(risk)을 측정하는 수단으로 사용

변동성은 곧 조건부 분산

강한 변동성

미래 값의 분산(변화 가능한 폭)이

현재 상황에 크게 의존하면서 결정될 때 나타나는 현상

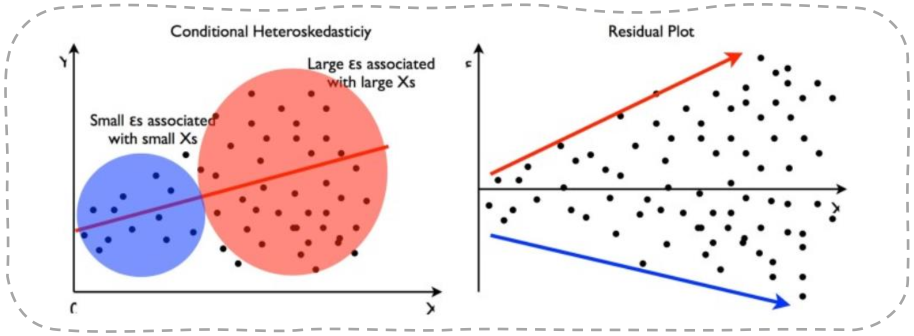
자료의 빈번한 급락 또는 급상승과 관련 !

모형의 필요성

이분산성 (Heteroscedasticity)

분산이 시간이나 관측치별로

독립적으로 나타나지 않거나 일정하지 않은 경우



이분산성

비조건부 이분산성

일반적인 구조적 변동성의 변화가

이전 기간의 변동성과 관련이 없는 경우

조건부 이분산성

일정하지 않은 변동성이

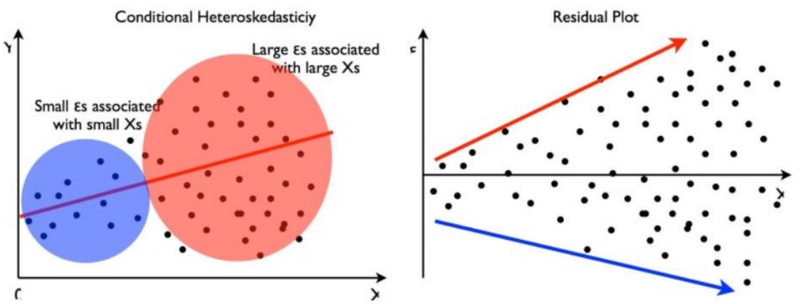
이전 시점의 변동성과 관련된 경우

모형의 필요성

이분산성 (Heteroscedasticity)

분산이 시간이나 관측치별로

독립적으로 나타나지 않거나 일정하지 않은 경우



이분산성

비조건부 이분산성

일반적인 구조적 변동성의 변화가
이전 기간의 변동성과 관련이 없는 경우

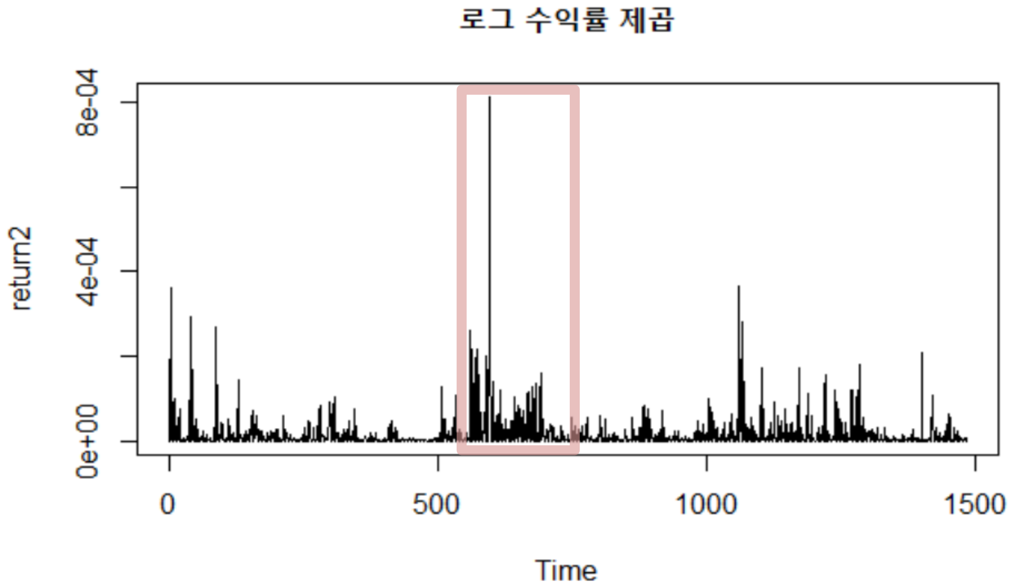
조건부 이분산성

ARCH/GARCH의 CH는
조건부 이분산성을 의미!

일정하지 않은 변동성이
이전 시점의 변동성과 관련된 경우

모형의 필요성

변동성 집중(Volatility Clustering)

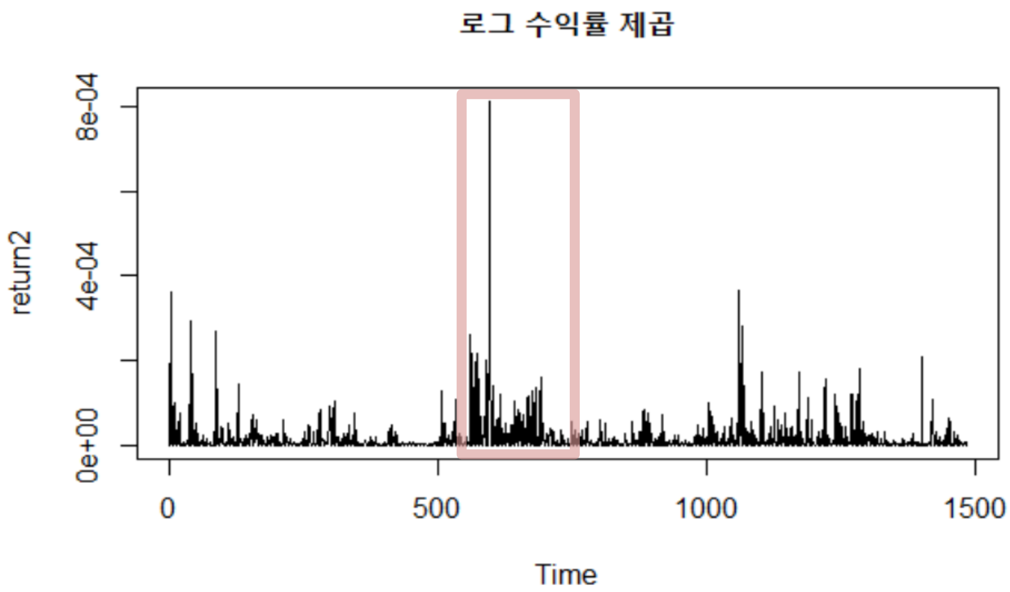


✓한 번의 큰 충격으로 일정 시간 동안 변동성이 유지되는 현상

한번 나타난 큰 변화가 당분간 큰 변화 유지, 작은 변화는 당분간 작은 변화 유지

모형의 필요성

변동성 집중(Volatility Clustering)



변동성 집중에 의한 변화 폭 → log나 √ 등의 자료 변환으로 상쇄 불가

오차의 제곱에 대한 시계열 모형 적용 : ARCH, GARCH

ARCH(Auto Regressive Conditional Heteroscedasticity)

IDEA

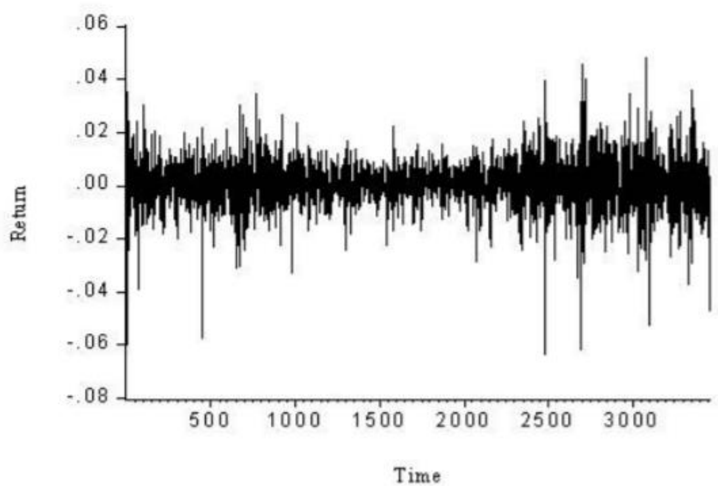
변동성 집중 문제 발생

→ 오차항을 과거 시점의 오차항으로 설명

EX)

〈 NYSE 수익률 시계열 자료 〉

모델 적합 후 오차항만 남은 상황



ARCH(Auto Regressive Conditional Heteroscedasticity)

IDEA

변동성 집중 문제 발생

→ 오차항을 과거 시점의 오차항으로 설명

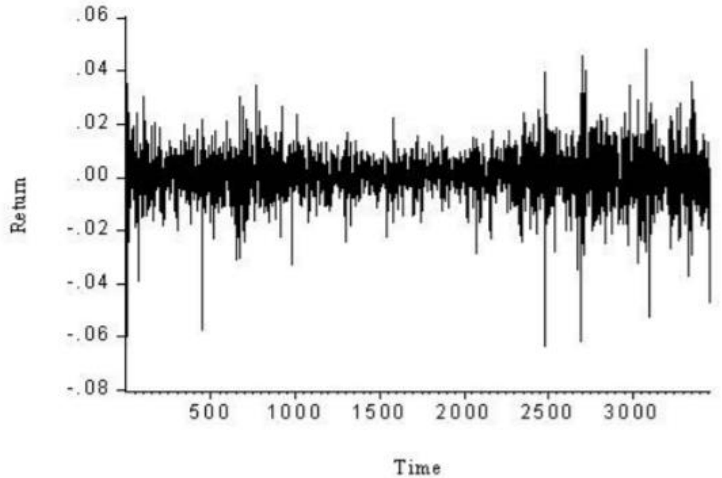
$$Var(\varepsilon_t) = \sigma_t^2 = \alpha_0 + \alpha_1 \sigma_{t-1}^2$$

변동성(오차항의 분산)을 어제의 분산으로 설명

EX)

〈 NYSE 수익률 시계열 자료 〉

모델 적합 후 오차항만 남은 상황



ARCH(Auto Regressive Conditional Heteroscedasticity)

IDEA

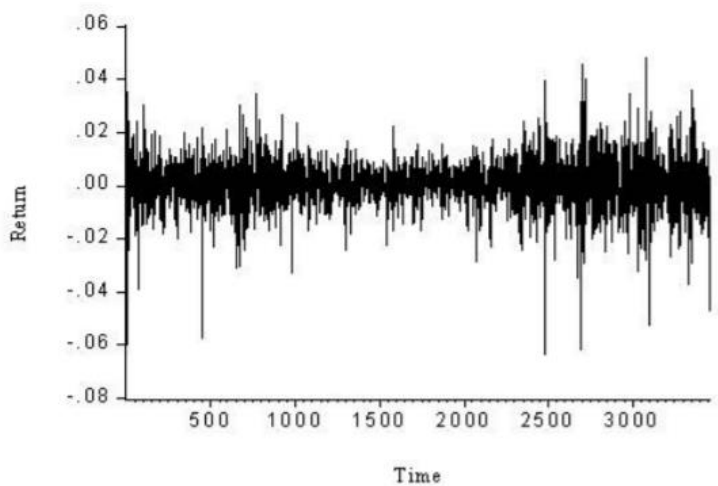
변동성 집중 문제 발생

→ 오차항을 과거 시점의 오차항으로 설명

EX)

〈 NYSE 수익률 시계열 자료 〉

모델 적합 후 오차항만 남은 상황



$$Var(\varepsilon_t) = \sigma_t^2 = \alpha_0 + \alpha_1 \sigma_{t-1}^2$$

변동성(오차항의 분산)을 어제의 분산으로 설명

백색잡음

$$\epsilon_t = w_t \sqrt{\alpha_0 + \alpha_1 \varepsilon_{t-1}^2}$$

ARCH(Auto Regressive Conditional Heteroscedasticity)

IDEA

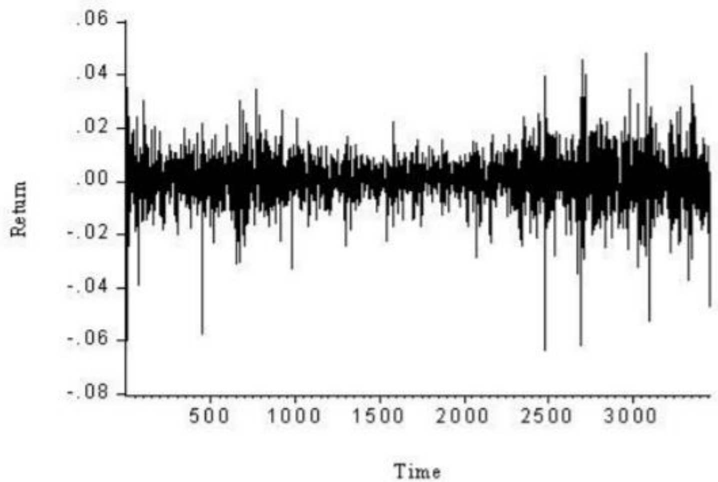
변동성 집중 문제 발생

→ 오차항을 과거 시점의 오차항으로 설명

EX)

〈 NYSE 수익률 시계열 자료 〉

모델 적합 후 오차항만 남은 상황



$$Var(\varepsilon_t) = \sigma_t^2 = \alpha_0 + \alpha_1 \sigma_{t-1}^2$$

변동성(오차항의 분산)을 어제의 분산으로 설명

백색잡음

$$\epsilon_t = w_t \sqrt{\alpha_0 + \alpha_1 \varepsilon_{t-1}^2}$$

양 변 제곱

$$\epsilon_t^2 = w_t^2(\alpha_0 + \alpha_1 \varepsilon_{t-1}^2) = w_t^2 \alpha_0 + w_t^2 \alpha_1 + \varepsilon_{t-1}^2$$

과거의 오차항을 이용하여 AR 형태의 함수로 오늘의 오차항 설명

- ARCH(Auto Regressive Conditional Heteroscedasticity)

ARCH(p)

오차의 변동성이 자기회귀적으로 변함을 설명하는 비선형모델

ARCH(p)

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \cdots + \alpha_p \varepsilon_{t-p}^2$$

where,

$$Z_t \sim iidN(0,1),\ \varepsilon_t = \sigma_t Z_t$$

$$\alpha_0 > 0, \alpha_j \geq 0, j = 1,2,\ldots,p$$

t시점 오차항의 변동성을 **p시점 전까지의 오차항의 제곱**으로 설명

$$E(\varepsilon_t^2 | \varepsilon_{t-1}^2, \ldots) = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \cdots + \alpha_p \varepsilon_{t-p}^2$$

→ 과거의 오차항으로 변동성을 표현 가능

- ARCH(Auto Regressive Conditional Heteroscedasticity)

ARCH(p)

오차의 변동성이 자기회귀적으로 변함을 설명하는 $\boxed{\text{비선형모델}}$

ARCH(1) : $\varepsilon_t = \sigma_t z_t \,, \sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2$

양 변 제곱, 재귀적으로 계산

$$\varepsilon_t^2 = \sigma_t^2 z_t^2 = (\alpha_o + \alpha_1 \varepsilon_{t-1}^2) z_t^2 = \left(\alpha_0 + \alpha_1(\sigma_{t-1}^2 z_{t-1}^2)\right) z_t^2$$

$$= \left(\alpha_0 + \alpha_1\left((\alpha_0 + \alpha_1 \varepsilon_{t-2}^2) z_{t-1}^2\right)\right) z_t^2$$

$$= \cdots$$

$$= \alpha_0 \sum_{j=0}^{n} \alpha_1^j \, z_t^2 z_{t-1}^2 \ldots z_{t-j}^2 + \alpha_1^{n+1} z_t^2 \ldots z_{t-n}^2$$

- ARCH(Auto Regressive Conditional Heteroscedasticity)

ARCH(p)

오차의 변동성이 자기회귀적으로 변함을 설명하는 비선형모델

ARCH(1) : $\varepsilon_t = \sigma_t z_t, \sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2$

양 변 제곱, 재귀적으로 계산

$$\varepsilon_t^2 = \sigma_t^2 z_t^2 = (\alpha_o + \alpha_1 \varepsilon_{t-1}^2) z_t^2 = \left(\alpha_0 + \alpha_1(\sigma_{t-1}^2 z_{t-1}^2)\right) z_t^2$$

$$= \left(\alpha_0 + \alpha_1\left((\alpha_0 + \alpha_1 \varepsilon_{t-2}^2) z_{t-1}^2\right)\right) z_t^2$$

$$= \cdots$$

$$= \alpha_0 \sum_{j=0}^{n} \alpha_1^j z_t^2 z_{t-1}^2 \ldots z_{t-j}^2 + \alpha_1^{n+1} z_t^2 \ldots z_{t-n}^2$$

z_t에 대해 비선형적인 형태

ARCH(Auto Regressive Conditional Heteroscedasticity)

추정과 검정

추정 최대가능도추정법(MLE) 활용

검정

Lagrange Multiplier(LM)

$H_0: \ \alpha_1 = \cdots = \alpha_p = 0$ → 기각 시 ARCH 효과 존재

ε_t가 $iid \ N(0,1)$을 따른다는 귀무가설 하에서

검정통계량 nR_ε^2 이 점근적으로 $\chi^2(p)$를 따른다는 사실을 활용

Ljung-Box Q 검정

ε_t^2이 AR(p)모형을 따름을 이용, ε_t^2의 자기상관계수의 유의성을 판단

오차항의 정규성 검정

QQ-plot, Jarque-Bera test

- ARCH(Auto Regressive Conditional Heteroscedasticity)

모형의 문제점

진"모형" 아님

$$\text{ARCH(p)} = \sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \cdots + \alpha_p \varepsilon_{t-p}^2$$

p↑ ~ 추정량의 **정확도 감소**

&

ARCH 모형의 **정상성 조건**을 만족시키지 못할 가능성↑

일반화된 모델 GARCH 도입

"generalized"

GARCH(Generalized ARCH)

GARCH 모형

GARCH(p,q)

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2$$

$$= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \cdots + \alpha_p \varepsilon_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_q \sigma_{t-q}^2$$

$$\alpha_0 > 0, \alpha_j \geq 0, \beta_j \geq 0$$

$$Z_t \sim iidN(0,1),\ \varepsilon_t = \sigma_t Z_t$$

t시점 오차항의 변동성을

p시점 이전의 **오차항들의 제곱**과 q시점 이전의 **변동성**으로 설명한 모형

GARCH(Generalized ARCH)

GARCH 모형

GARCH(1,1)

$$\varepsilon_t = \sigma_t z_t \,, \sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2$$

$$= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta(\alpha_0 + \alpha_1 \varepsilon_{t-2}^2 + \beta \sigma_{t-2}^2)$$

$$= \alpha_0(1 + \beta + \beta^2 + \cdots) + \alpha_1(\varepsilon_{t-1}^2 + \beta \varepsilon_{t-2}^2 + \beta^2 \varepsilon_{t-3}^2 + \cdots) + \beta^\infty \sigma_0^2$$

$$= \cdots = \frac{\alpha_0}{1-\beta} + \sum_{j=1}^{\infty} \alpha_1 \beta^{j-1} \varepsilon_{t-j}^2 = \text{ARCH}(\infty)$$

장점 !

모수의 절약

GARCH(1,1)은 ARCH(∞)으로 표현 가능

4

시계열과 머신러닝

시계열 자료

시계열자료와 머신러닝

Machine learning

Time Series Data

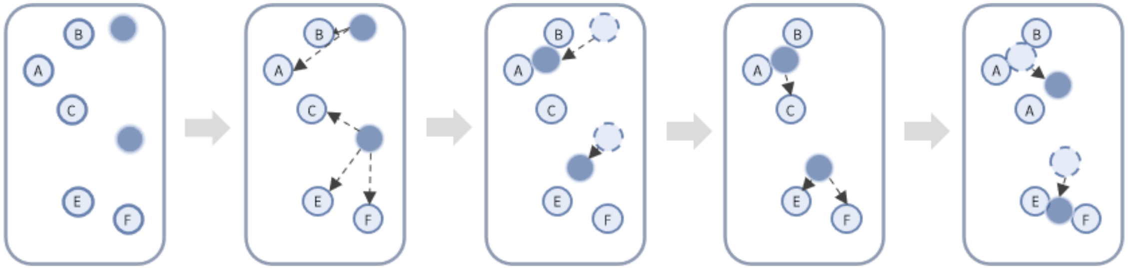
시계열자료의 시간적 특징 반영

클러스터링

정의

IDEA 유사한 집단은 가까이, 이질적인 집단은 멀리!

ex) k=2 일 때 k-means clustering



중심점을 이동하였지만
데이터들의 중심점
소속 변경이 없으면
군집화 완료

유사도 및 거리를 기반으로

비슷한 특징을 보이는 구성요소들끼리 **소집단(cluster)**형성

클러스터링

정의

IDEA 유사한 집단은 가까이, 이질적인 집단은 멀리!

ex) k=2 일 때 k-means clustering

시계열자료의 경우

관측값의 **순서(기록된 시간)**자체가 중요

중심점을 이동하였지만
데이터들의 중심점
소속 변경이 없으면
군집화 완료

시계열의 순서 정보를 보존하며

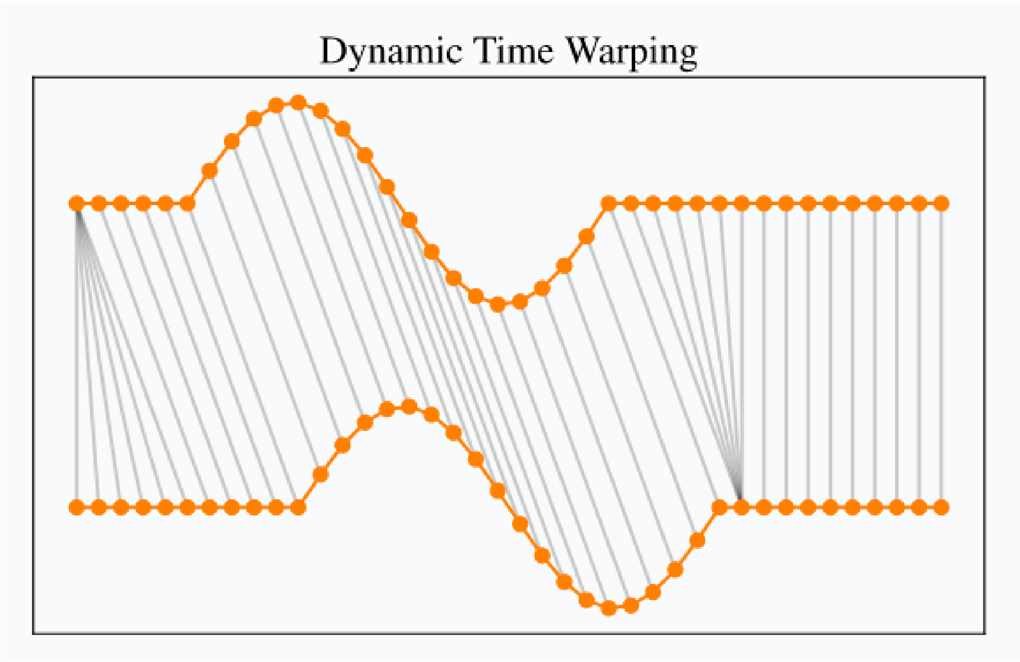
두 시계열의 **시각적 형태를 비교**하는 거리지표 필요

유사도 및 거리를 기반으로

비슷한 특징을 보이는 구성요소들끼리 **소집단(cluster)**형성

시계열 거리 측정

동적시간왜곡(DTW)



음성인식에 주로 사용되는 방법으로

두 sequence간의 유사도를 측정하는 방법 중 하나

동적시간왜곡(DTW)

DTW의 장점

X \ Y	1	1	2	3	3
1	$\lvert 1-1 \rvert = 0$	$0+\lvert 1-1 \rvert = 0$	$0+\lvert 1-2 = 1 \rvert$	$1+\lvert 3-1 \rvert = 3$	$3+\lvert 1-3 \rvert = 5$
2	$0+\lvert 1-2 \rvert = 1$	1	0	1	2
3	$1+\lvert 1-3 \rvert = 3$	3	1	0	0
3	$3+\lvert 1-3 \rvert = 5$	5	2	0	0
2	$5+\lvert 1-2 \rvert = 6$	6	2	1	1
1	$6+\lvert 1-1 \rvert = 6$	6	3	3	3

자세한 과정은 시계열 3주차 교안을 참고해주세요^_^

패턴의 진행, 두 시계열의 **모양 비교**에 초점

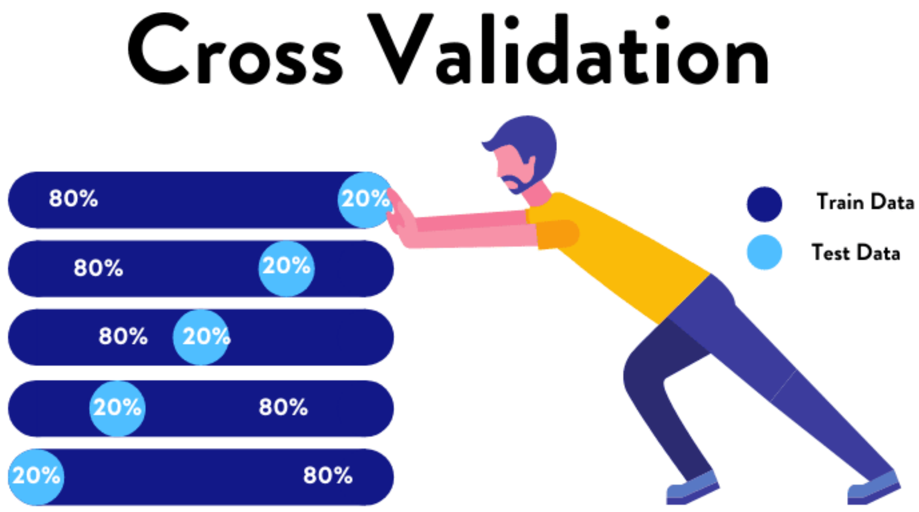
→ 두 sequence의 길이가 달라도 거리 측정 가능

ex) 나노초 단위의 시계열 vs 수천 년 단위의 시계열

교차검증(Cross Validation)

정의

과적합을 방지하기 위한 방법



ex) 5-fold cross validation

Train set을 다시 train/test set으로 나누어 모델 평가

교차검증(Cross Validation)

시계열 분석의 목적

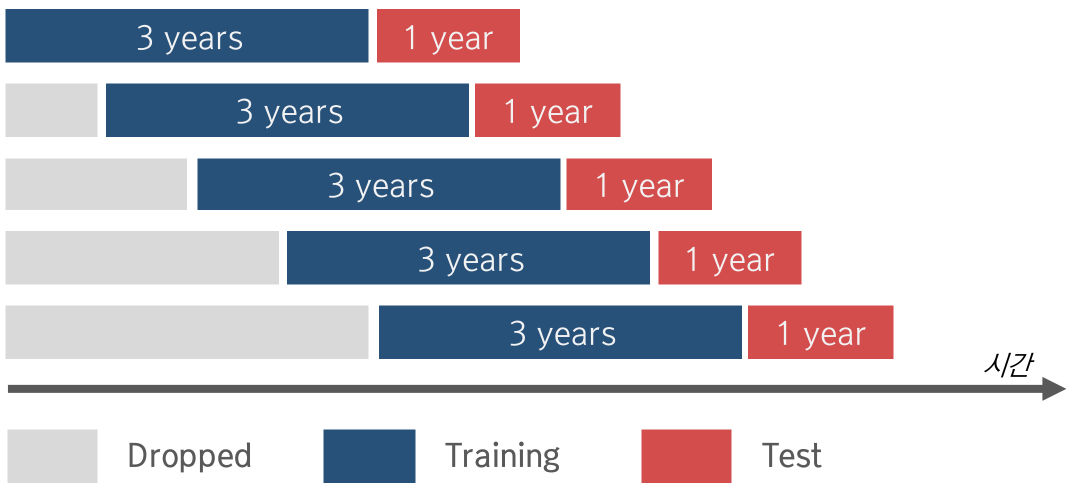
: 과거시점의 정보를 활용하여 미래 예측

"시간"이라는 속성을 보존하는 교차검증방식 필요

교차검증(Cross Validation)

Blocked Time series CV

Fixed rolling window forecast



동일한 사이즈의 윈도우 내에서 일정한 비율의 train/test 분리

교차검증(Cross Validation)

Blocked Time series CV

Fixed rolling window forecast

EX) Dataset : [1,2,3,4,5]

Training : [1,2] Test : [3]

Training : [2,3] Test : [4]

Training : [3,4] Test : [5]

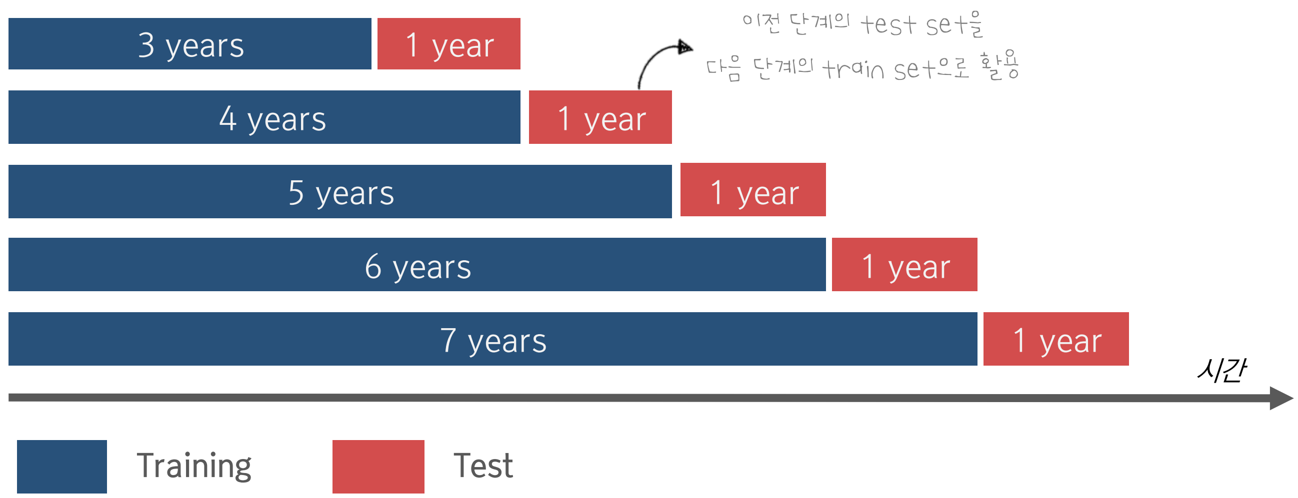
각 단계에서 예측오차 추정

동일한 사이즈의 윈도우 내에서 일정한 비율의 train/test 분리

교차검증(Cross Validation)

Time series CV

Expanding window forecast



누적하며 이동하는 윈도우 내에서 일정한 크기의 test set 분리

교차검증(Cross Validation)

Time series CV

Expanding window forecast

EX) Dataset : [1,2,3,4,5]

- Training : [1] Test : [2]
- Training : [1,2] Test : [3]
- Training : [1,2,3] Test : [4]
- Training : [1,2,3,4] Test : [5]

각 단계에서 예측오차 추정

누적하며 이동하는 윈도우 내에서 일정한 크기의 test set 분리

THANK YOU