

Chapter. 03

시계열 예측모형만들기

ARIMA 모델을 적용한 예 측 모형 구축 프로세스

FAST CAMPUS
ONLINE

직장인을 위한
파이썬 데이터분석
강사. 주세민

Chapter. 03

시계열 예측모형만들기

IARMA/ARIMA 모델을 이야기하기에 앞서 복습

• AR (Auto Regressive) 모델

- 이번기의 결과는 이전기의 결과에 영향을 받는 모델
- 외부 충격이 길게 반영되는 Long memory 모델

$$\text{AR}(1): Y_t = \delta + \phi Y_{t-1} + \varepsilon_t$$

$$\text{AR}(2): Y_t = \delta + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t$$

$$\vdots$$

$$\text{AR}(n):$$

• MA (Moving Average) 모델

- 이번기의 결과는 이전기의 결과와 상관이 없음
- 외부 충격이 일정기간만 지속되고 없어지는 Short memory 모델

$$\text{MA}(1): Y_t = \mu + \varepsilon_t + \theta \varepsilon_{t-1}$$

$$\text{MA}(2): Y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}$$

$$\vdots$$

$$\text{MA}(n):$$

IAR + MA = ARMA(p, q), ARIMA(p, k, q)

- ARMA (Auto Regressive Moving Average) 모델
 - 예를들어, ARMA(2,2) 모델은 AR(2)+MA(2) 모델임

$$Y_t = \delta + \underbrace{\phi_1 Y_{t-1} + \phi_2 Y_{t-2}}_{\text{AR}} + \underbrace{\varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}}_{\text{MA}}$$

- ARIMA (Auto Regressive Integrated Moving Average) 모델
 - ARMA 모델의 원계열 Y_t 를 차분하여 Y'_t 로 변환한 모형
 - 예를 들어, ARIMA(2,1,2) 모델은 원계열을 1번 차분하고 AR(2)+MA(2)을 진행한 모델임

$$Y'_t = \delta + \phi_1 Y'_{t-1} + \phi_2 Y'_{t-2} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}$$

$$Y'_t = Y_t - Y_{t-1} \quad (1차 차분)$$

IARIMA 모델을 적용할 때 주의점

- ARIMA (p, k, q) 모델 : AR(p), Integrated(k), MA(q)
 - 라이브러리에 입력할때 Y_t 를 입력하면, Stationary해질 때까지 알고리즘이 $Y_t', Y_t'', \dots, Y_t^{(k)}$ 와 같이 차분
 - 차분하지 않는 경우, 설명력이 매우 높은 모형이 생성됨 → Training데이터에서만 정확도가 높은 잘못된 결과일 확률이 매우 높음
 - 시계열의 단순 차분값을 활용하기 보다 변동율로 변환하기 위해서 원계열에 log를 취하거나 $\Delta \log$ (log difference)를 취하는 경우도 많음.

원데이터		log (원데이터)		$\Delta \log$ (원데이터)
200	→	log(200)	→	NAN
300	→	log(300)	→	log(300)-log(200)
200	→	log(200)	→	log(200)-log(300)
100	→	log(100)	→	log(100)-log(200)
...	

IARIMA 모델을 이용한 예측 모형 만드는 순서 (다른 시계열 예측도 동일한 순서로 진행)

안정성 검토

Y_t 가 안정적이지 않으면, ΔY_t 가 안정적인지 확인.
→ 보통 1~2번의 차분으로 안정적인 시계열이 됨
ARIMA(p, 1, q) 또는 ARIMA(p, 2, q) 선정

데이터 특성에 맞는 모형 결정 (AR차수와 MA차수)

PACF peak와 ACF peak의 개수로 AR, MA계수 선정
PACF의 peak이 p개, ACF의 peak이 q개 이면,
ARIMA (p, k, q) 모델 선정

학습

특정 시점 이전 데이터(Training set)로 학습

평가

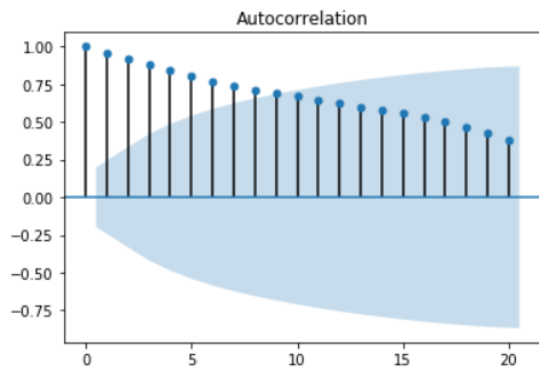
특정 시점 이후 데이터(Test set)로 평가

IAR, MA 차수 결정

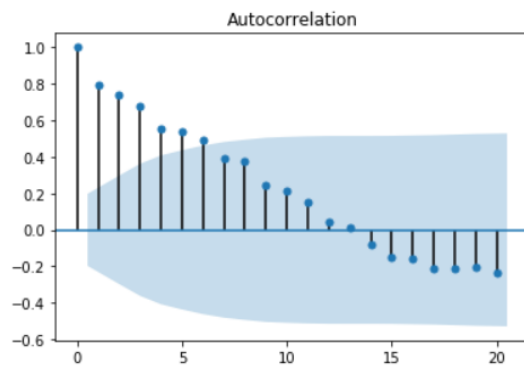
ACF

PACF

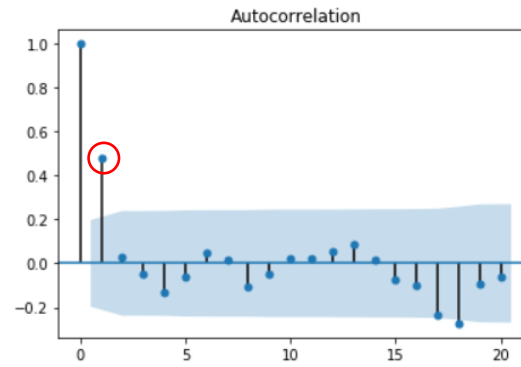
AR(1)



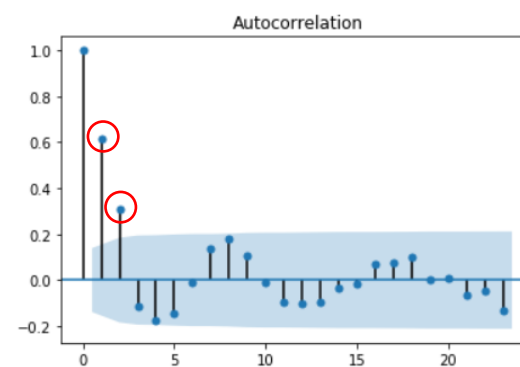
AR(2)



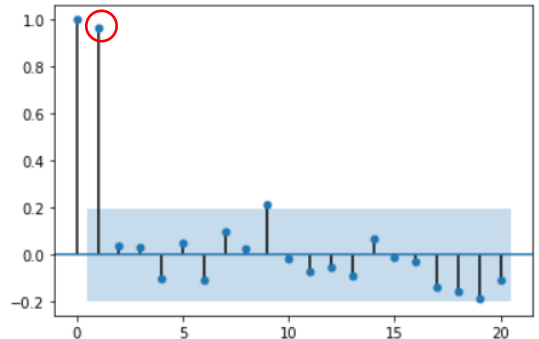
MA(1)



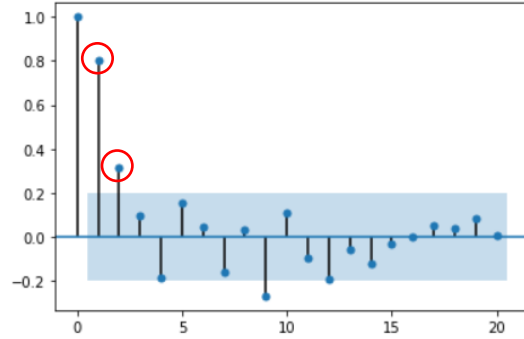
MA(2)



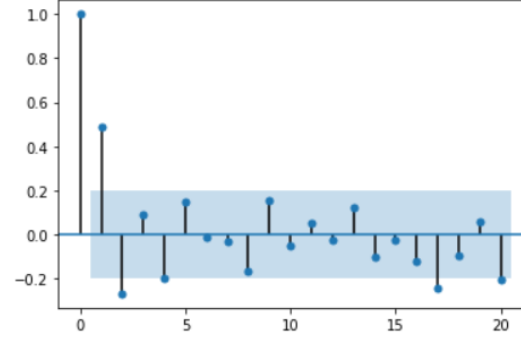
Partial Autocorrelation



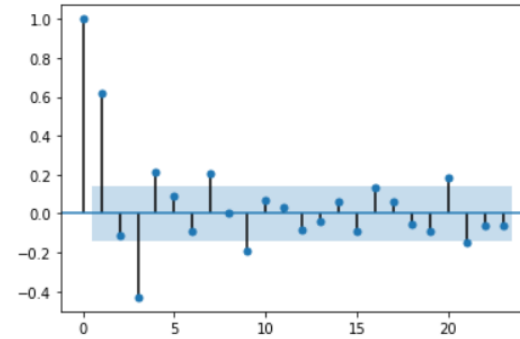
Partial Autocorrelation



Partial Autocorrelation



Partial Autocorrelation

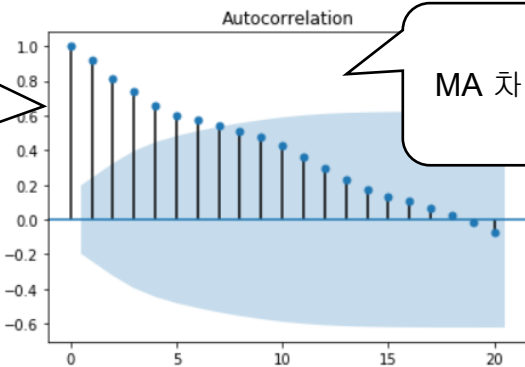


IARMA 차수 결정

ACF

명확히 AR성격을 가지고 있음

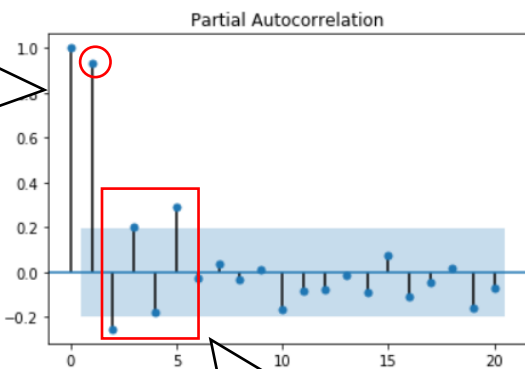
ARMA(1,1)



MA 차수는 모름

PACF

AR차수는 알 수 있다.

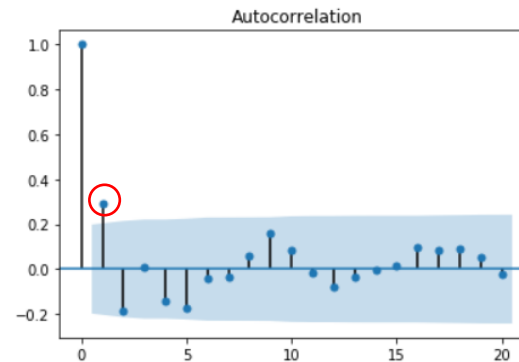


신뢰구간 밖으로 튀는 값 있음 : MA 성격

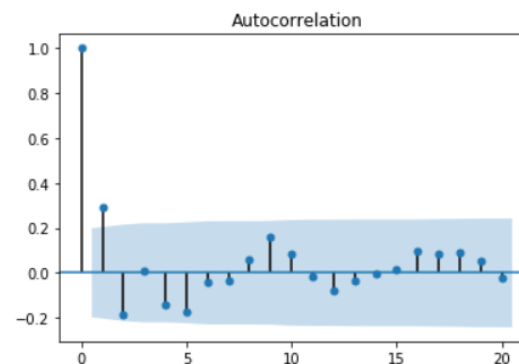
y_t

Δy_t

차분한 ARMA(1,1)을 ACF, PACF

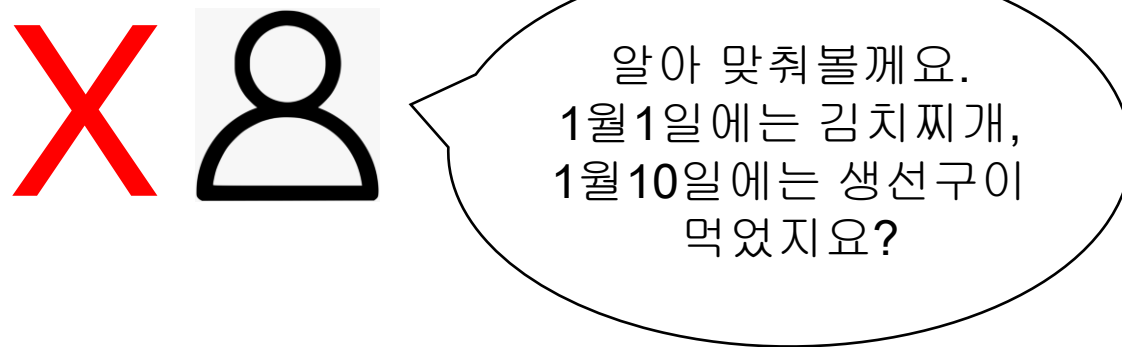
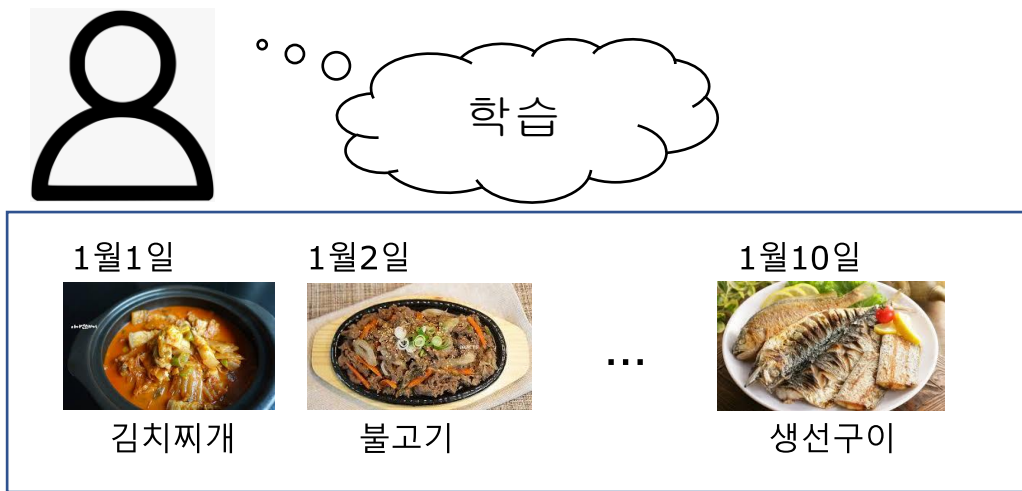


MA 차수=1

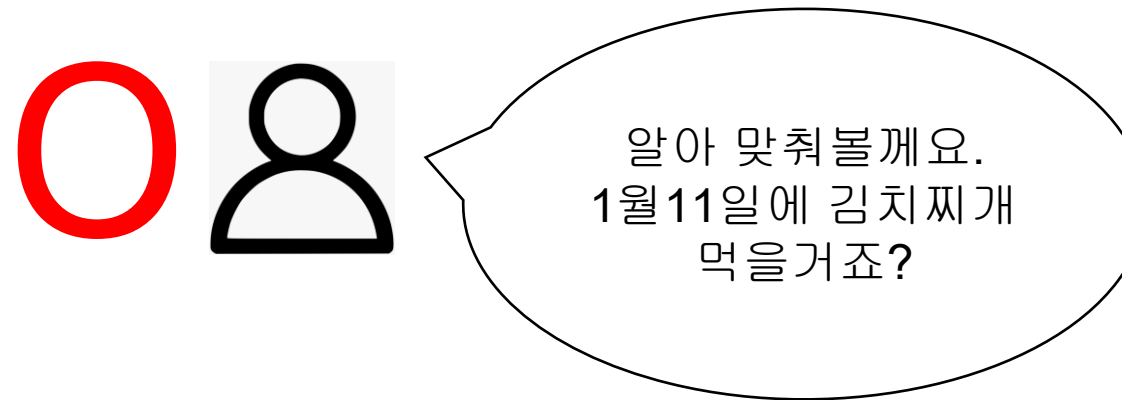
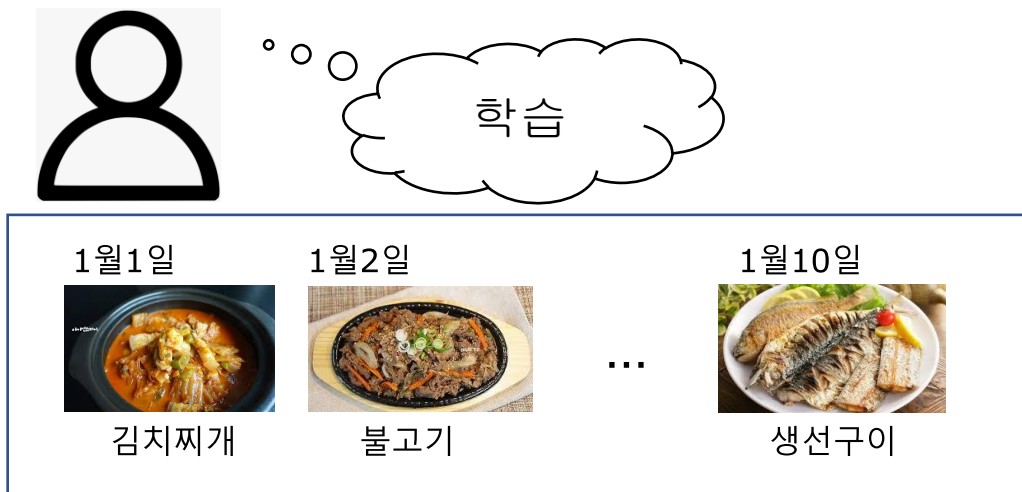


I 학습 및 평가

1. 학습 데이터를 평가에 사용하는 경우



2. 학습 데이터와 테스트 데이터를 분리하여 사용하는 경우



I 학습 및 평가

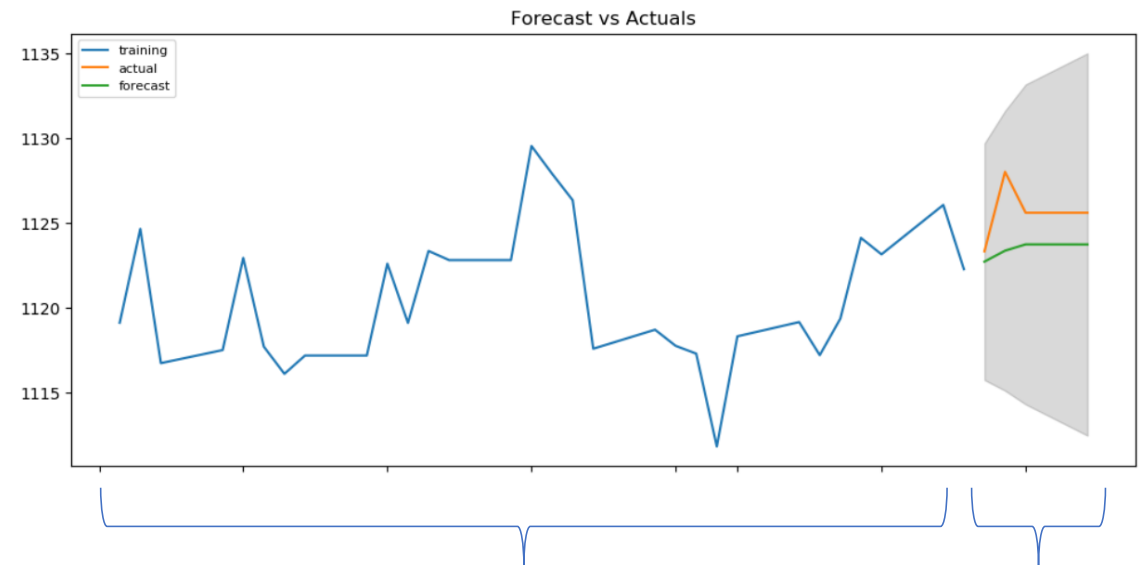
- 다 지나간 학습데이터를 맞춰보는게 목적이 아니라면 테스트 데이터를 분리해서 사용해야..

1. 학습 데이터를 평가에 사용하는 경우



AR 프로세스만 잘 학습해서 이전 관측치를 그 다음기에 그대로 예측하는 경향이 있음 (딤러닝도 똑같음)
→ “변동”을 예측하도록 보완하든지,
아예 **Level**보다는 변동을 예측 (실무에서 주로 사용)

2. 학습 데이터와 테스트 데이터를 분리하여 사용하는 경우



학습데이터

테스트 데이터

I 다음 강의에서 다룰 내용

- ARIMA (p,k,q) 모형 검토
- 예측 모형 만들기
- 시각화