

ARIMA 모델

개요

- ARIMA 는 Auto Regressive Intergrated Moving Average 로 시계열 데이터의 패턴을 파악하여 통계적 기법으로 미래값을 예측하는 모델이다. ARIMA 모델 이전 예측모델인 AR 과 MA 를 합친 후 차분하는 과정을 추가한 모델로 이전의 모델들과는 달리 비정상성을 가진 데이터도 차분을 통해 정상성을 가지게 한 후 예측이 가능하게 하였다.

구성요소 및 파라미터

- 구성요소
 - AR(Auto Regressive) : 과거 값들이 현재 값에 미치는 영향
 - I(Intergrated) : 시계열 데이터가 정상성을 가질 수 있게 차분을 적용
 - MA(Moving Average) : 과거 오차들이 현재 값에 미치는 영향
- 파라미터
 - P : 자기회귀(Auto Regressive)되는 항의 수를 의미
이전의 관측 값들이 현재의 값에 얼마나 영향을 미치는지 나타냄
 - D : 차분 차수를 의미
시계열의 비정상성을 제거하고 데이터의 평균이나 분산이 시간에 따라 일정하도록 함
시계열 데이터의 추세(Trend)를 제거하기 위해 사용됨
 - Q : 이동평균(Moving Average)되는 항의 수를 나타냄
예측 값이 예측 오차에 따라 조정되는 정도를 의미하고 시계열데이터의 잡음을
모델링하기 위해서는 Q 값을 설정하여 오차정도를 보정할 수 있

수식

AR모델 수식

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \epsilon_t$$

MA모델 수식

$$y_t = c + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q} + \epsilon_t$$

ARIMA모델 수식

$$\hat{y}_t = c + \phi_1 \hat{y}_{t-1} + \dots + \phi_p \hat{y}_{t-p} + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q} + \epsilon_t$$

- ARIMA의 수식은 기존 예측모델인 AR 모델과 MA 모델의 수식을 합친 모양으로
c, ϕ , θ 는 모델에 넣는 학습데이터에 의해 정해지게 되며 p, q가 높아질수록 모델의 복잡성은 증가하고
d 차분의 경우 비정상성을 없애기 위한 방법으로 최대 2차 차분까지만 하는것이 바람직하다.

계산과정 예시

계산예시1

P, D, Q = 1, 1, 1 (base model)

시간	TagA	1차차분
1	10	-
2	12	2
3	15	3
4	14	-1
5	16	2
6	18	2
7	20	2
8	19	-1
9	22	3
10	24	2

ARIMA모델 계산과정

$$\hat{y}_t = c + \phi_1 \hat{y}_{t-1} + \dots + \phi_p \hat{y}_{t-p} + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q} + \epsilon_t$$

c=0.5, $\phi_1=0.7$, $\theta_1=-0.4$, $\epsilon_t = 0.3$ 으로 추정(모델 계산)

$$\hat{y}_t = c + \phi_1 \hat{y}_{t-1} + \theta_1 \epsilon_{t-1} \quad \text{P와 Q가 1이니 1개 전까지만 확인}$$

$$\hat{y}_{11} = c + \phi_1 \hat{y}_{10} + \theta_1 \epsilon_{10}$$

$$\hat{y}_{11} = 0.5 + 0.7 \times 2 + (-0.4) \times 0.3$$

$$\hat{y}_{11} = 1.78$$

차분된 값을 다시 변환

$$\hat{y}_{11} = \hat{y}_{10} + 1.78$$

$$\hat{y}_{11} = 25.78$$

- Step size를 늘려 더 미래의 예측을 진행할때는 기존에 예측했던 11번째 데이터까지 사용하여 파라미터를 추정하고 예측하는 재귀적 예측이 수행되게 됨

계산예시2

P, D, Q = 2, 2, 2

시간	TagA	1차차분	2차차분
1	10	-	-
2	12	2	-
3	15	3	1
4	14	-1	-4
5	16	2	3
6	18	2	0
7	20	2	0
8	19	-1	-3
9	22	3	4
10	24	2	-1

ARIMA모델 계산과정

$$\hat{y}_t = c + \phi_1 \hat{y}_{t-1} + \dots + \phi_p \hat{y}_{t-p} + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q} + \epsilon_t$$

c=0.5, $\phi_1=0.7$, $\phi_2=-0.3$, $\theta_1=-0.4$, $\theta_2=0.2$, $\epsilon_t = 0.2$ 으로 추정(모델 계산)

$$\hat{y}_{11} = c + \phi_1 \hat{y}_{10} + \phi_2 \hat{y}_9 + \theta_1 \epsilon_{10} + \theta_2 \epsilon_9$$

$$\hat{y}_{11} = 0.5 + 0.7 \times (-1) + (-0.3) \times 4 + (-0.4) \times 0.1 + 0.2 \times (-0.2)$$

$$\hat{y}_{11} = -1.48$$

1차분된 값을 다시 변환

$$\hat{y}_{11} = 2 + (-1.48) = 0.52$$

2차분된 값을 다시 변환

$$\hat{y}_{11} = 24 + 0.52 = 24.52$$

- P와 Q를 통해 얼마이전의 값까지 사용할 것인지 정하고 2차차분을 했을 때는 다시 반환하는 과정도 2번 수행되게 됨