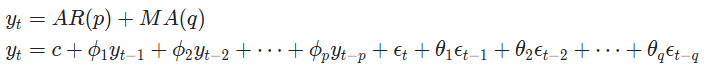
ARIMA모델

# 단일 변량 시계열 모델종류

하나의 변수에 대한 시계열 데이터를 분석하는 모델은 기본적으로 AR, MA, ARMA, ARIMA모델이 있는데  
ARIMA이전의 모델들의 특성들과 구성을 살펴보고자 함

AR 모델 : 자기상관성을 시계열 모형으로 구성한 것으로, 변수의 과거 관측값의 선형 결합을 통해  
 변수의 미래값을 예측하는 모델이다. 이전의 관측값이 이후의 관측값에 영향을 준다는  
 아이디어에 기반하고 있고 파라미터는 p를 이용하여 AR(p)의 형태로 나타난다.

  
  
 MA모델 : 예측 오차를 이용하여 미래값을 예측하는 모델이다. 파라미터는q를 이용하여  
 MA(q)의 형태로 나타난다.   
  
  
 ARMA모델 : AR(p)모형과 MA(q)모형이 합쳐진 모델이다. AR모형의 정상성, MA모형의 가역성  
 조건을 만족해야한다. 파라미터는 p, q를 사용하여 ARMA(p,q)의 형태로 나타난다.

ARMA(1,0) = AR(1), ARMA(0,1) = MA(1)과 동일함

ARIMA모델 : ARMA모델에서 차분(d)를 추가적으로 적용시킨 모델로서 위의 모델들은 데이터가   
 정상적이었을때만 사용된다는 조건이 있지만 ARIMA모델은 차분을 통해 비정상시계열에도  
 바로 적용할 수 있다는 장점이 있다. I는 Intergrated를 의미하고 누적을 의미함  
 ARIMA(p,0,0) = AR(p), ARIMA(0,0,q) = MA(q)와 동일함

# ARIMA모델 파라미터 설명

p : 자기회귀(Auto Regressive)되는 항의 수를 의미   
자기회귀란 현재 시점의 데이터가 이전 시점의 데이터와 관련이 있는 경우를 의미함  
몇 개의 이전 시점 데이터를 사용할 것인지 정하는 파라미터  
예를 들어 P=1은 바로 직전 시점의 데이터만을 사용하는 자기회귀 모델을 의미하고  
p=2는 두 시점 이전까지의 데이터를 사용하는 자기회귀 모델을 나타낸다.

d : 차분 차수를 의미  
시계열의 비정상성을 제거하고 데이터의 평균이나 분산이 시간에 따라 일정하도록 함  
시계열 데이터의 추세를 제거하기 위해 데이터 간의 차이를 계산하고   
d는 몇번의 차이를 적용할 것인지를 의미함  
값이 클수록 시계열의 추세를 더 강하게 제거함

q : 이동평균(Moving Average)되는 항의 수를 나타냄  
현재 시점의 데이터가 과거 시점의 랜덤한 오차 항에 의해 영향을 받는 경우를 나타냄  
q는 몇 개의 이전 시점의 오차항을 고려할 것인지를 결정함  
예측 값이 예측 오차에 따라 조정되는 정도를 의미하고 시계열데이터의 잡음을   
모델링하기 위해서는 q값을 설정하여 오차정도를 보정할 수 있음

# ARIMA모델의 장점

적용가능성 : ARIMA는 차분을 통해 비정상 시계열 데이터를 처리할 수 있는 기능이 있어 다양한 시계열 데이터에 적합함

단기 예측 : 트렌드, 자기상관성, 변동성을 반영하여 데이터의 패턴을 학습하고 가까운 미래를 예측하는데 적합

적은 데이터 필요 : 다른 모델들과 비교했을 때 상대적으로 적은 데이터로도 학습 및 예측이 가능