

데이터 사이언스 실무

시계열 분석: prophet

유재명

ARIMA

ARIMA

- **AR(p)**: p 이전 기간의 **값**의 영향이 존재
- **MA(q)**: q 이전 기간의 **오차**의 영향이 존재
- **ARMA(p, q)** = $AR(p) + MA(q)$
- **ARIMA(p, d, q)** = $ARMA(p, q) + d차 차분$
- **SARIMA** = $ARIMA(p, d, q) + 계절 효과$

Dynamic Linear Model

Dynamic Linear Model

- Linear Model: $Y = X\beta + \epsilon$ 꼴의 관계를 찾음
- Dynamic Linear Model: X 가 Y 의 과거가 되는 점이 차이
- AR과 거의 비슷하지만 추정 방법에 차이가 있음
 - DLM: 최소제곱법(OLS)
 - AR: 최대우도법(MLE)

prophet

prophet

- Facebook이 공개
- R, Python에서 사용 가능
- 추세, 계절효과, 추세 변화 탐지 등이 자동
- 휴일 추가 가능
- ARIMA와 달리 간격이 일정할 필요 X

prophet 모형

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t$$

- $y(t)$: 시계열 데이터
- $g(t)$: 트렌드
- $s(t)$: 계절 효과
- $h(t)$: 휴일 효과
- ϵ_t : 오차

트렌드

- 선형 트렌드와 로지스틱 트렌드 2가지가 가능
- 트렌드가 변하는 점(change point)을 자동 추정
- sparsity parameter(τ): 트렌드가 얼마나 자주 변하는 지
 - $\tau = 0$: 트렌드가 불변
 - 커질 수록 자주 변함

휴일

- 휴일을 별도의 data.frame으로 넣어줄 수 있음

holiday	ds	lower_window	upper_window
휴일의 종류	날짜	~일 전까지 포함 예) 크리스마스 이브의 경우 크리스마스의 -1일	~일 후까지 포함 예) 추석의 경우 추석 +1일