

7.1~7.4 연습과제

▼ 7.1 시계열 문제

- 불규칙 변동 : 규칙적인 움직임과 달리 어떤 규칙성이 없어 예측 불가능하고, 우연적으로 발생하는 변동. ex. 전쟁, 홍수, 화재, 지진, 파업
- 추세변동 : 시계열 자료가 갖는 장기적인 추세변화. 장시적인 증가, 감소, 일정상태 유지. ex. 국내 총생산(GDP), 인구 증가율
- 순환변동 : 2-3년을 주기로 순환적으로 나타나는 변동. ex. 경기변동
- 계절 변동 : 계절적 영향과 사회적 관습에 따라가 1년 주기로 발생하는 것.
- 규칙적 시계열 : 트렌드와 분산이 불변
- 불규칙적 시계열 : 트렌드와 분산이 변환
- 최근에는 딥러닝이 더 좋은 성능

▼ 7.2 AR, MA, ARMA, ARIMA

▼ 7.2.1 AR

- autoregression
- 이전 관측 값이 이후 관측값에 영향을 준다는 아이디어. 자기 회귀 모델.
- ①은 시계열 데이터에서 현재 시점을 의미하며, ②는 과거가 현재에 미치는 영향을 나타내는 모수(Φ)에 시계열 데이터의 과거 시점을 곱한 것. 마지막으로 ③은 시계열 분석에서 오차 항을 의미하며 백색 잡음. 따라서 수식은 p 시점을 기준으로 그 이전의 데이터에 의해 현재 시점의 데이터가 영향을 받는 모형

$$\underbrace{Z_t}_{\text{①}} = \underbrace{\Phi_1 Z_{t-1} + \Phi_2 Z_{t-2} + \cdots + \Phi_p Z_{t-p}}_{\text{②}} + \underbrace{a_t}_{\text{③}}$$

Copyright © Gilbut, Inc. All rights reserved.

▼ 7.2.2 MA

- Moving Average(이동평균), 트렌드가 변화하는 상황에 적합한 회귀모델

- 윈도우라는 개념을 사용하는데, 시계열을 따라 윈도우 크기만큼 슬라이딩 된다고 하여 이동평균 모델이라고 한다.
- ①은 시계열 데이터에서 현재 시점을 의미하며, ②는 매개변수(θ)에 과거 시점의 오차를 곱한 것. 마지막으로 ③은 오차 항을 의미. 따라서 수식은 AR 모델처럼 이전 데이터의 '상태'에서 현재 데이터의 상태를 추론하는 것이 아닌, 이전 데이터의 오차에서 현재 데이터의 상태를 추론하겠다는 의미

$$\underbrace{Z_t}_{\text{①}} = \underbrace{\theta_1 a_{t-1} + \theta_2 a_{t-2} + \dots + \theta_p a_{t-p}}_{\text{②}} + \underbrace{a_t}_{\text{③}}$$

Copyright © Gilbut, Inc. All rights reserved.

▼ 7.2.3 ARMA

- AR+MA (자기회귀 이동평균). 연구기관에서 주로 사용
-

$$Z_t = a + \Phi_1 Z_{t-1} + \dots + \Phi_p Z_{t-p} + \theta_1 a_{t-1} + \dots + \theta_q a_{t-q} + a_t$$

Copyright © Gilbut, Inc. All rights reserved.

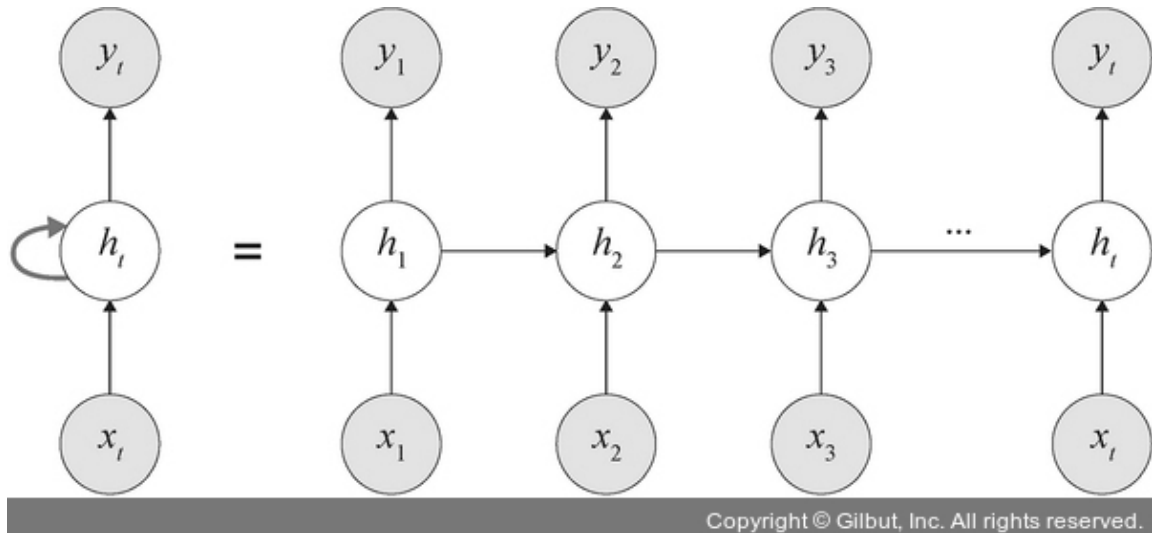
▼ 7.2.4 ARIMA

- autoregressive integrated moving average(자기 회귀 누적 이동 평균)
- ARMA와 달리 과거 데이터의 선형 관계뿐만 아니라 추세까지 고려한 모델.
- statsmodels 라이브러리 사용

•

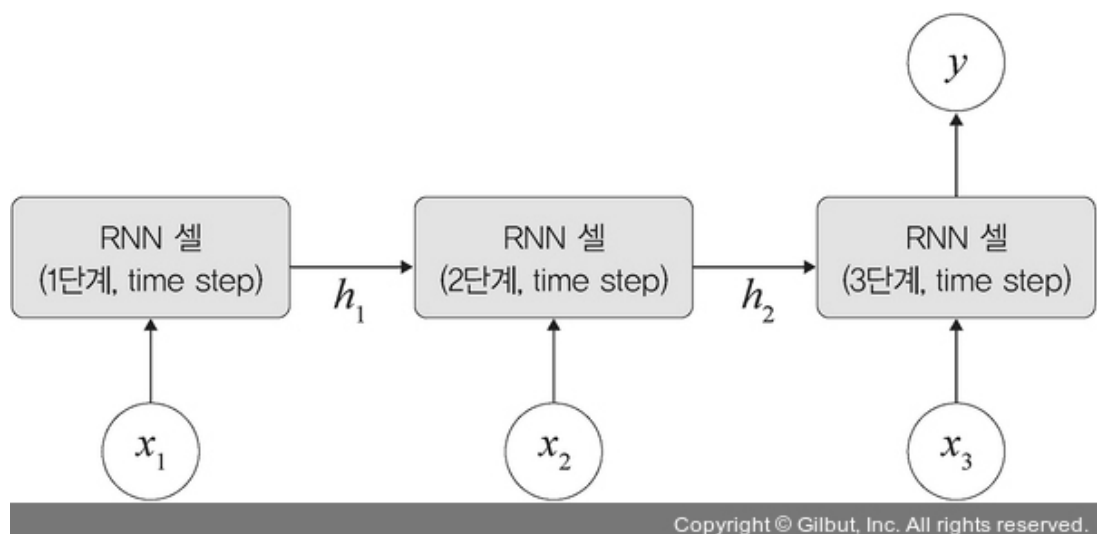
▼ 7.3 순환신경망 (RNN)

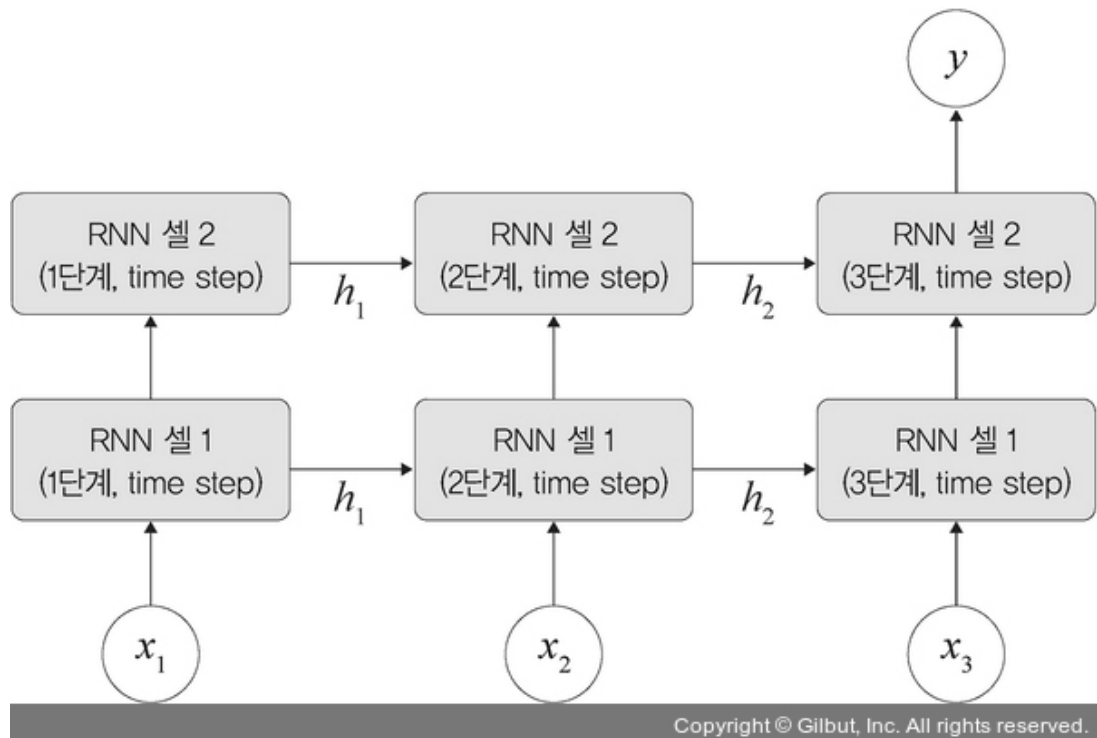
- 시간적으로 연속성이 있는 데이터를 처리하려고 고안된 인공신경망
- Recurrent(반복되는): 이전 은닉층이 현재 은닉층의 입력이 되면서 '반복되는 순환 구조를 갖는다'는 의미. '기억(현재까지 입력데이터를 요약한 정보)'을 하고 있음



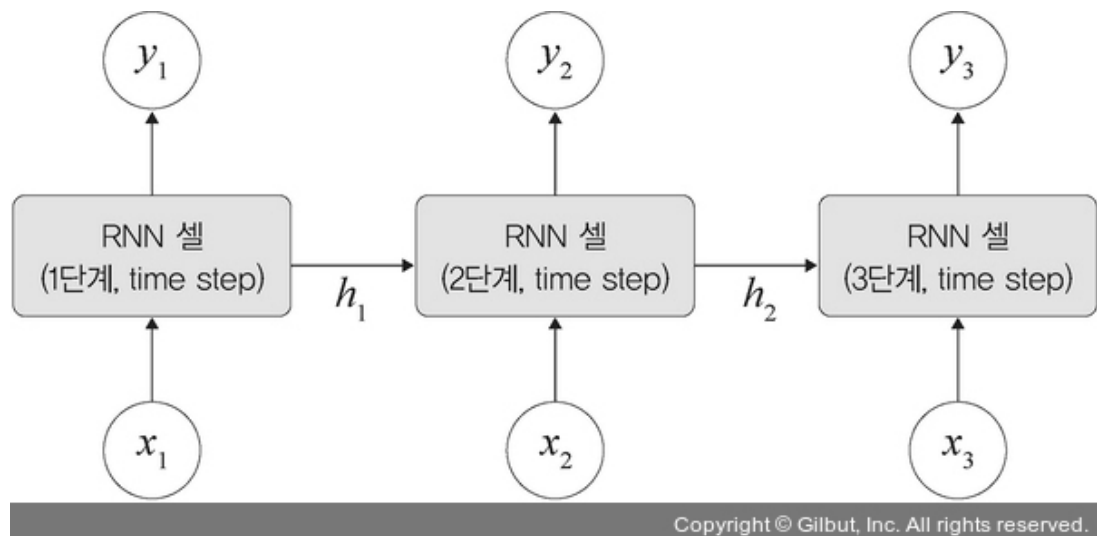
- 일대일 : 순환이 없기 때문에 rnn이라고 말하기 어려우며, 순방향 네트워크가 대표적
- 일대다 : 입력이 하나, 출력이 다수. 이미지를 입력해서 이미지에 대한 설명을 문장으로 출력하는 이미지 캡션이 대표적 사례
- 다대일 : 입력이 다수고 출력이 하나인 구조. 문장을 입력해서 긍부정을 출력하는 감성분석기에 사용

◦





- 다대다 : 입력과 출력이 다수. 언어를 번역하는 자동번역기가 대표적.



- RNN 셀은 RNN계층의 for loop구문을 갖는 구조

▼ 7.4 RNN구조

- 은닉층 : 하이퍼볼릭 탄젠트 활성화 함수 사용
- 출력층 : 소프트맥스 함수 사용

- 역전파 : BPTT(backpropagation through time) 사용 → 기울기 소멸 문제 발생.
→ 생략된 BPTT사용 → 근본적으로는 LSTM, GRU 사용