

RNN

(Recurrent Neural Network)

Mattias

Contents

- Concept
- Structure
- Applications
- Variants
- Example

Concept

- 기본적으로 DNN 구조는 모든 입력과 출력이 각각 독립적이라고 가정
- 그러나 이는 많은 상황에서 옳지 않은 접근이 됨
- 우리는 기억을 활용해 맥락을 살핌. 특정 이벤트가 일어났을 때 과거의 기억을 토대로 맥락을 파악하는 것을 생각할 수 있음.
- 가령 문장을 이해할 때, 우리는 한 단어 한 단어 따로 따로 의미를 이해하지 않고 앞 내용의 기억을 연결해 문장 전체의 맥락을 살피게 됨.

배가 고파서 ○○○를 열었다.

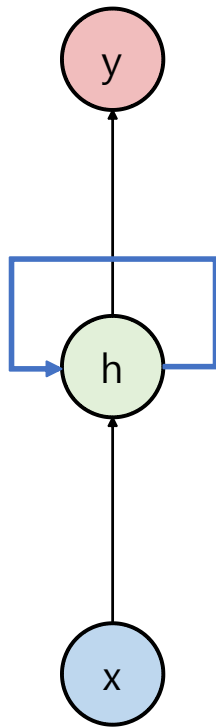
Concept

- 즉, 우리는 언어를 Sequence data로 받아들이고, 기존의 입력을 활용하여 전체 의미를 이해함.
- 기존의 DNN, CNN으로 Sequence data를 처리할 수 없던 한계를 해결하고자 등장한 것이 바로 RNN (Recurrent Neural Network)
- 여기서 Recurrent란 ‘순환하는, 재귀하는’의 의미로, 기존의 출력이 되풀이되어 입력되는 형태를 의미함

 [트친소 중]코딩노예 템실
@DevCode_in

어느 한 컴퓨터공학과 학생이 유명한 교수님을 찾아가 물었다.
"재귀함수가 뭔가요?"
"잘 들어보게. 옛날에 산 꼭대기에 현자가 있었어. 질문엔 모두 지혜롭게 대답해 주었지.
그런데 어느날, 그 선인에게 한 선비가 찾아와서 물었어.
"재귀함수가 뭔가요?"
"잘 들어보게. 옛날에 산 꼭대기..."

Structure – Recurrent Neural Network



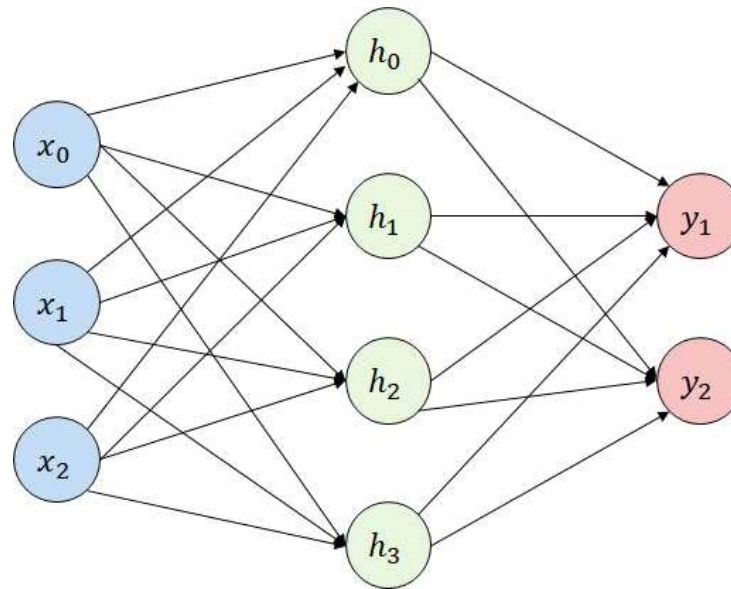
보통 이런 형태로 RNN을 표현

은닉층의 출력이
다시 은닉층으로 입력되는 구조,
“Recurrent”

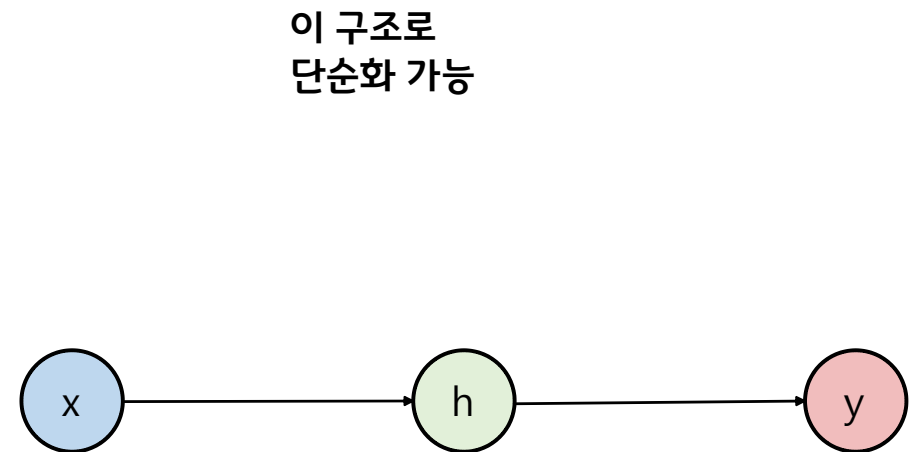
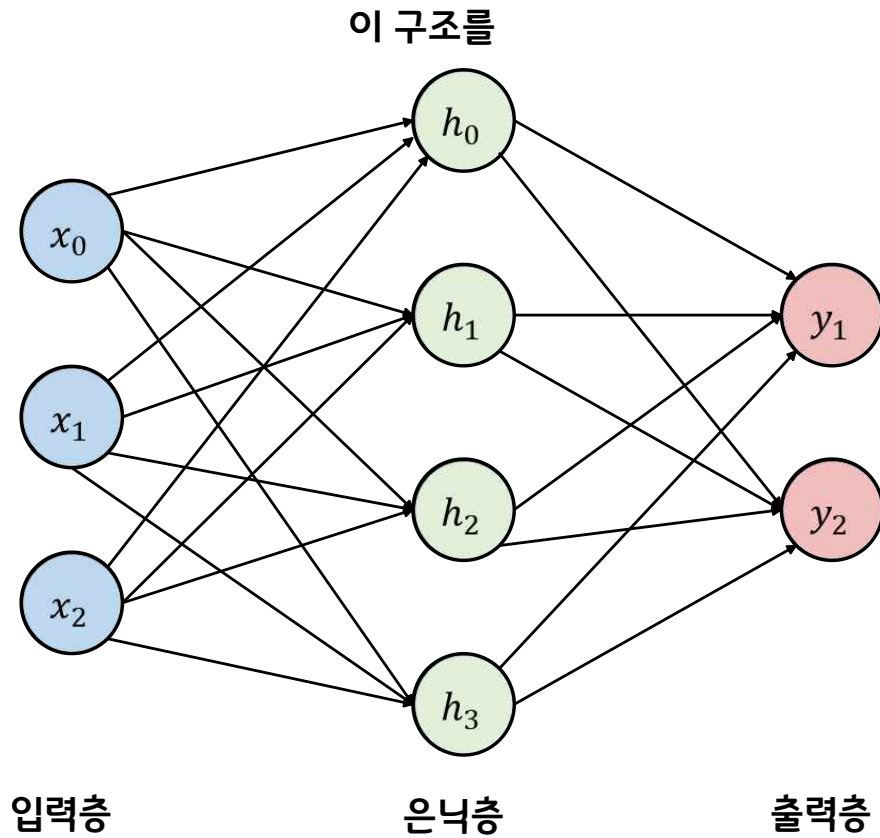
이런 구조 덕분에
‘기억’을 가질 수 있고,
시계열 데이터를 학습할 수 있음

어떻게 이런 형태가?

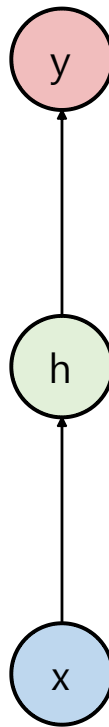
Structure – Neural Network



Structure – Neural Network

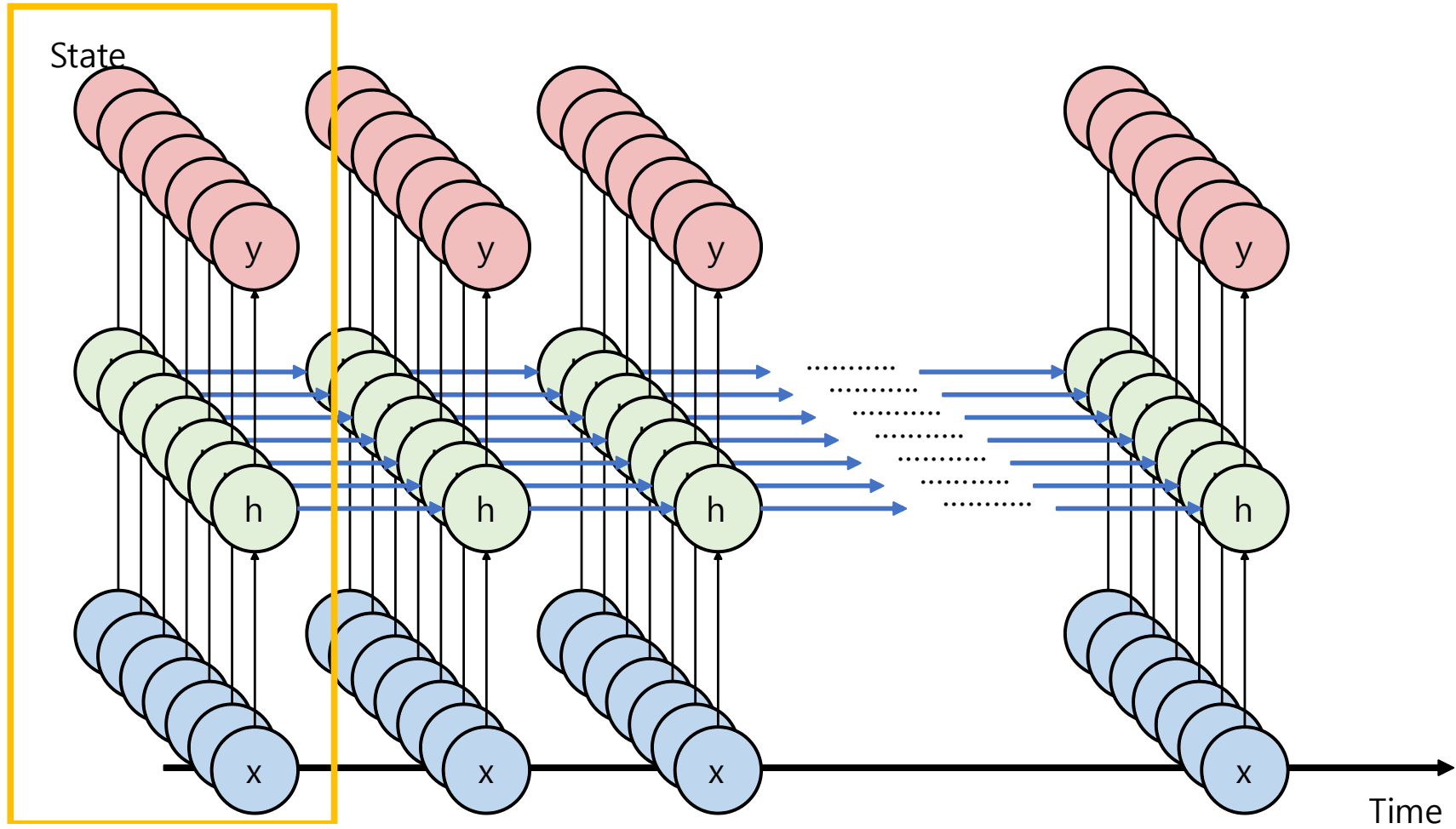


Structure – Neural Network

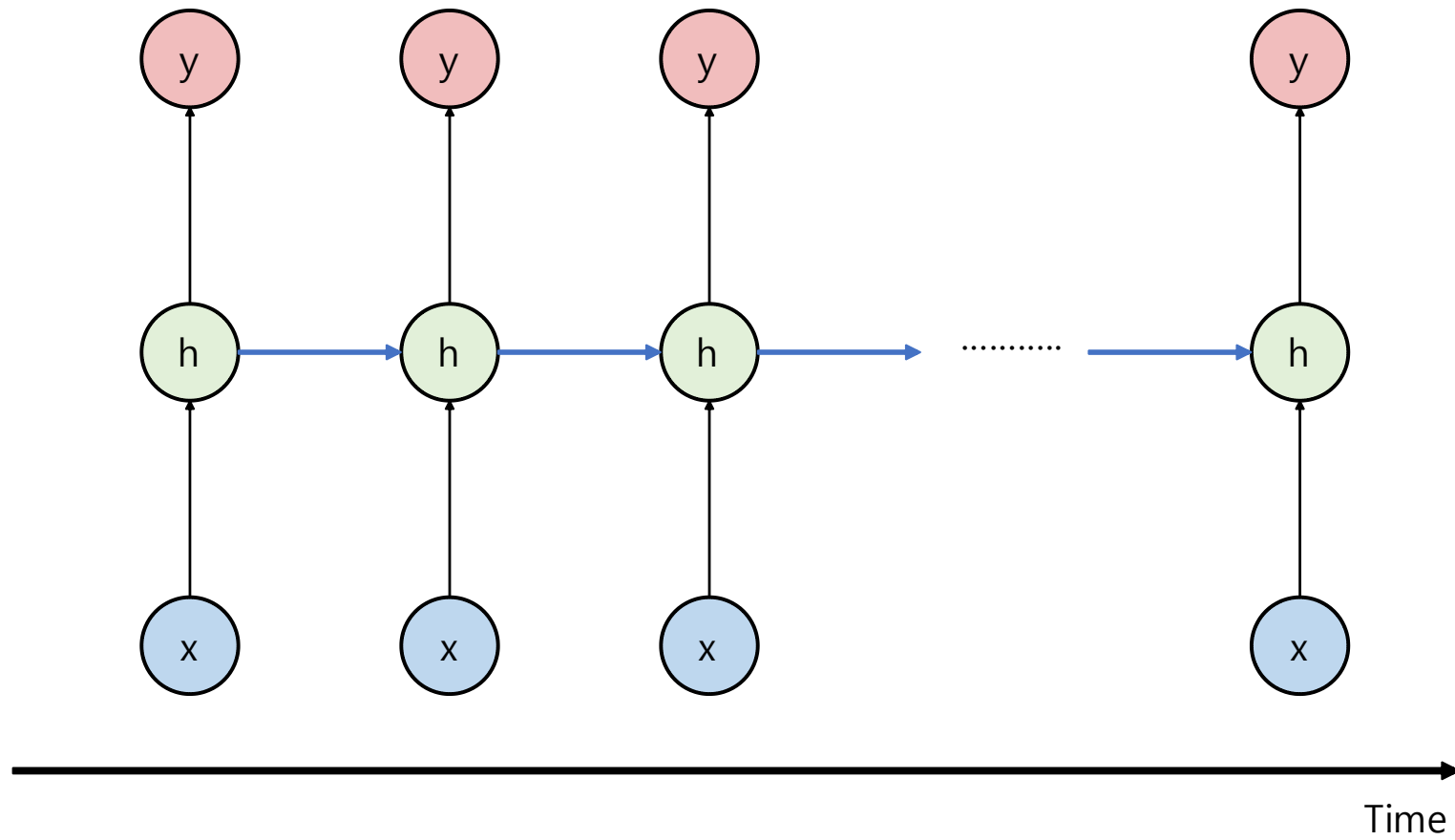


세워보았습니다

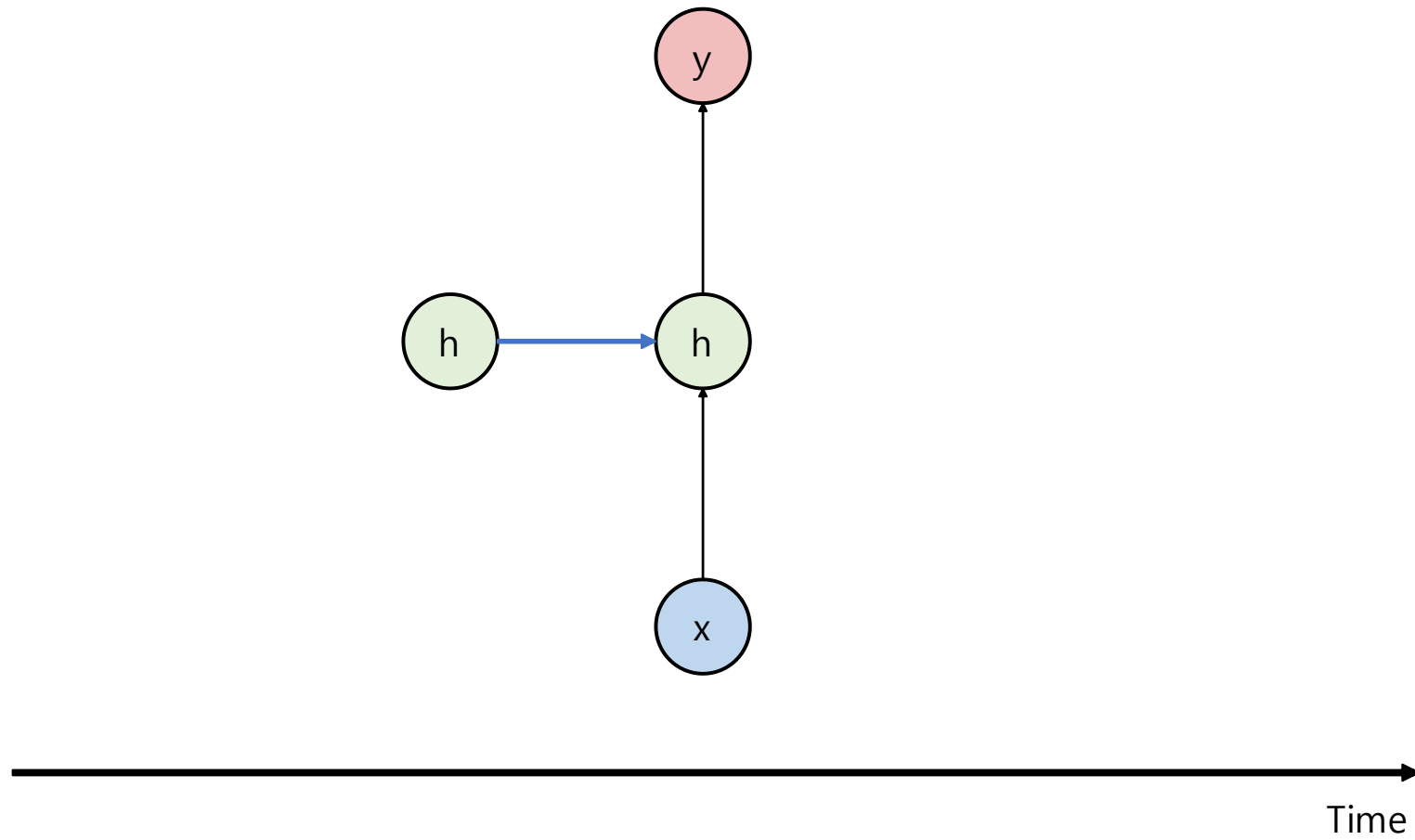
Structure – Recurrent Neural Network



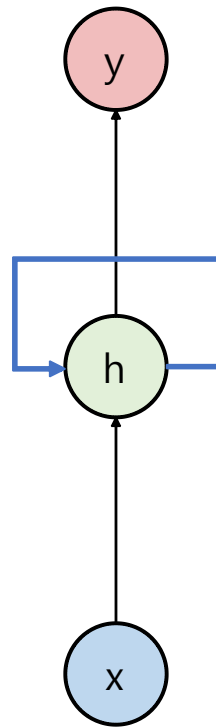
Structure – Recurrent Neural Network



Structure – Recurrent Neural Network



Structure – Recurrent Neural Network

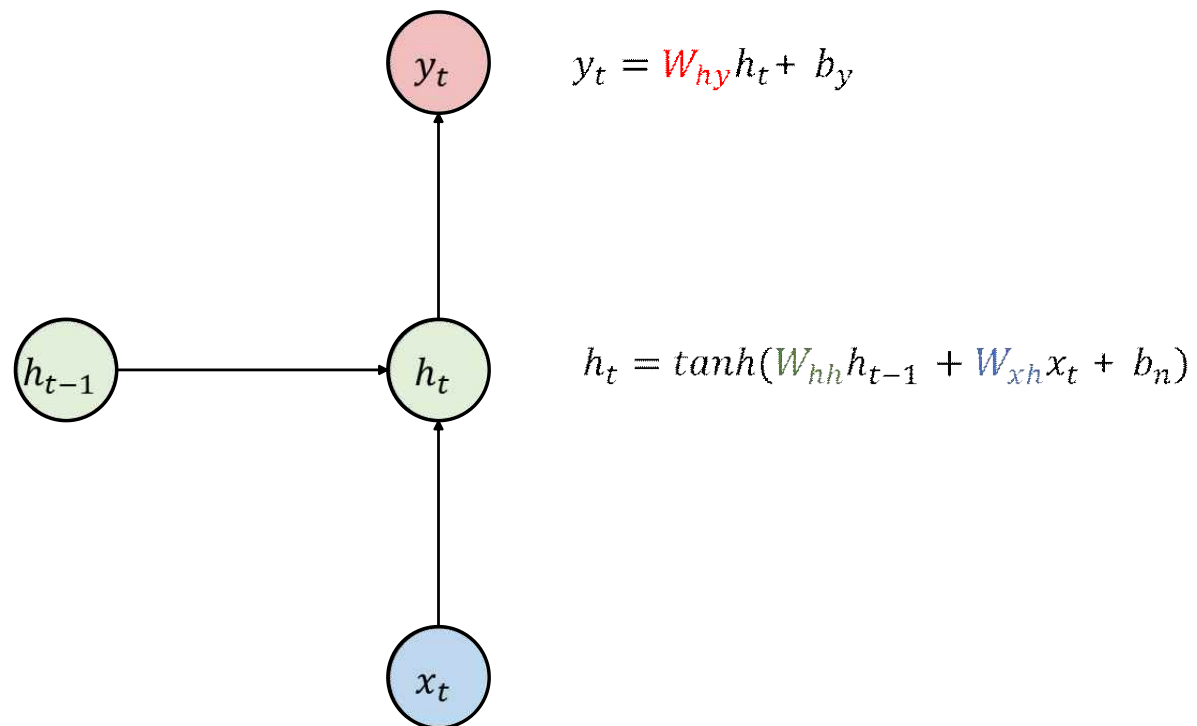


결과적으로 이러한 형태로
RNN 구조를 단순화할 수
있음

은닉층의 출력이
다시 은닉층으로 입력되는 구조,
“Recurrent”

이런 구조 덕분에
‘기억’을 가질 수 있고,
시퀀스 데이터를 학습할 수 있음

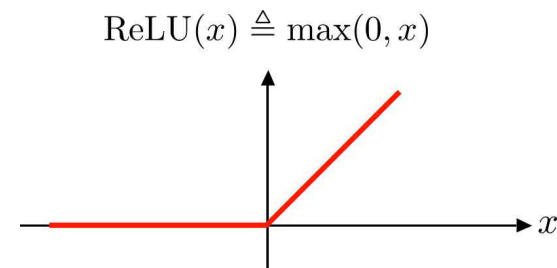
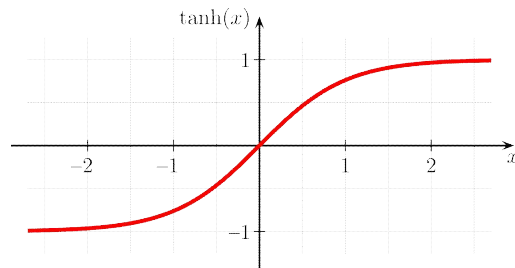
Structure – Recurrent Neural Network



Structure – Recurrent Neural Network

<Features>

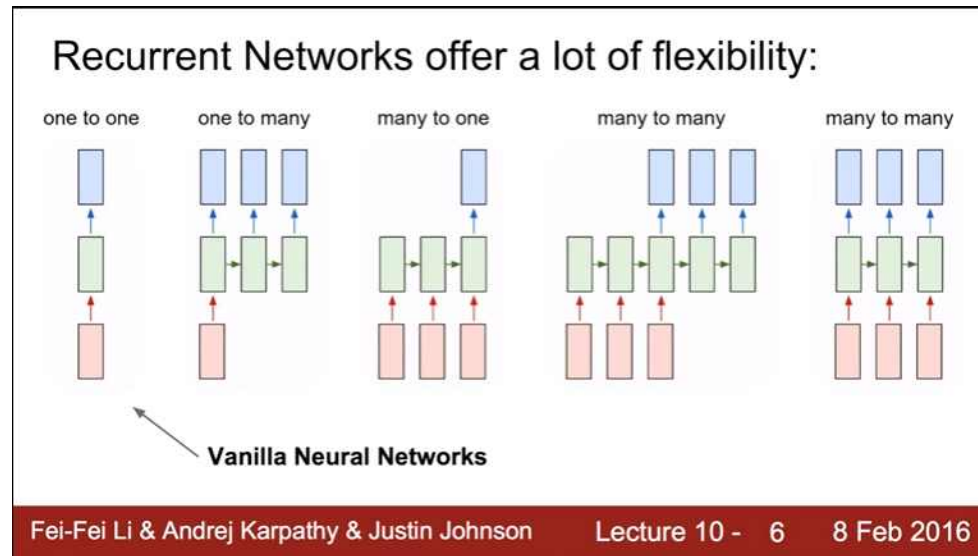
- 각 time step 마다 똑같은 함수와, 똑같은 parameter set이 사용됨. 이러한 방법을 통해, 학습해야하는 parameter의 수를 많이 줄여줄 수 있음.
(각 layer마다 전부 다 다른 기존의 DNN과는 다름)
- 보통 Activation Function은 tanh(hyperbolic tangent)나 ReLU(Rectified Linear Unit) 함수를 사용



Structure – Recurrent Neural Network

<Features>

- 입력값의 차원과 출력값의 차원이 각각 다를 수 있음.
- 또한, 입력층과 출력층의 구성을 자유롭게 바꿀 수 있음.

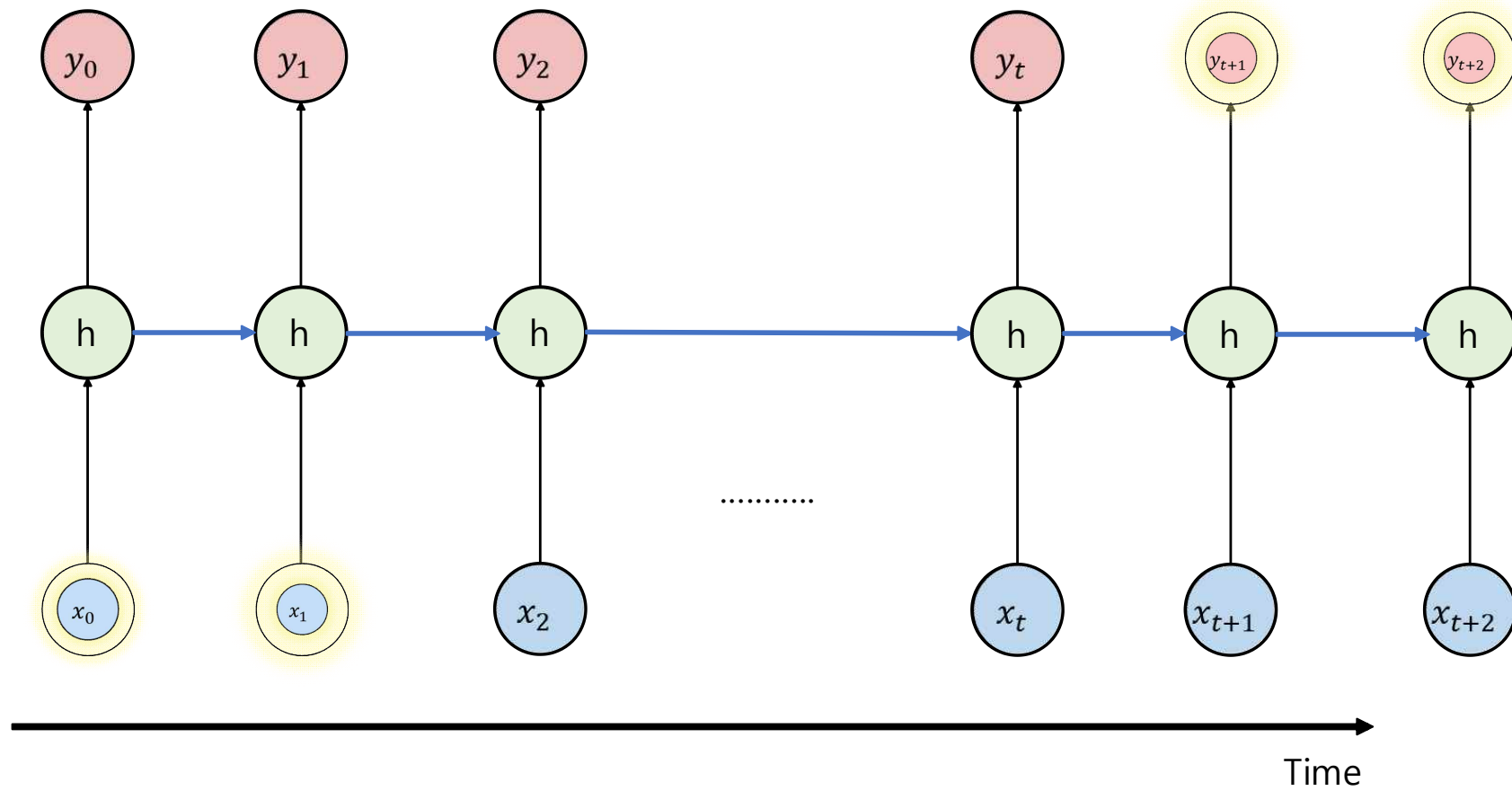


Structure – Recurrent Neural Network

<Features>

- 장기 의존성(Long-Term Dependency) 문제를 안고 있음
 - >입력값과 출력값의 거리가 멀수록 역전파시 gradient가 점차 줄어 학습 능력이 크게 저하되는 문제 (Vanishing Gradients)
ex) $0.2^{10} = 1.024e-07$
- 예) 나는 딥러닝을 3년간 눈만 뜨면 공부했다. 가끔 공부하다 배가 고플 때 백종원 짬뽕 끓이는 방법을 찾아보곤 했다. 긴 시간 끝에, 이제 나는 ○○○를 잘 한다. -> “짬뽕 끓이기”로 대답할 가능성이 높아짐
- 해결책으로 LSTM(Long-Short Term Memory) 모델이 주로 사용됨.
- 수학적 설명은 다음을 참고
: <http://aikorea.org/blog/rnn-tutorial-3/>

Structure – Long Term Dependency



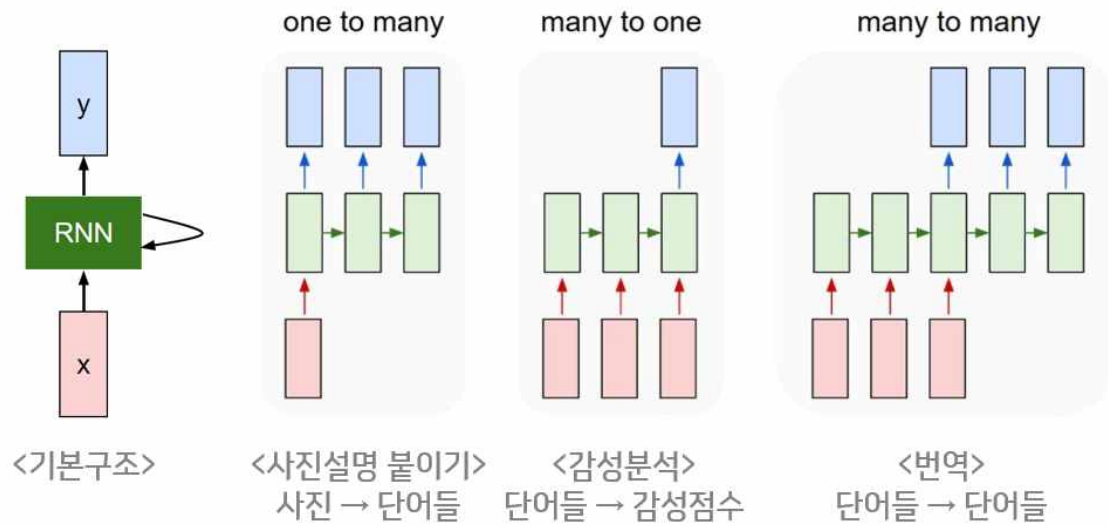
RNN Application

- 언어 모델링
- 음성 인식
- 기계 번역
- 챗봇
- 이미지/영상 자막 처리
- 이미지/음악/안무 생성기



- 특히 자연어 처리에 강점을 보임

RNN Application



RNN Application – Language Modeling

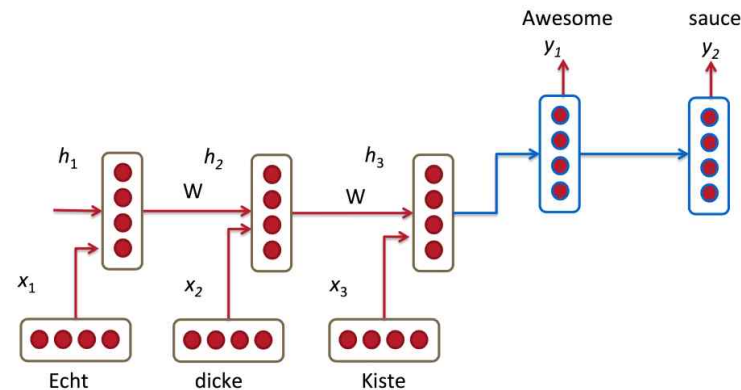
- 언어 모델은 주어진 문장에서 이전 단어들을 보고 다음 단어가 나올 확률을 계산해주는 모델

배가 고파서 ○○○를 열었다.

- 언어 모델은 어떤 문장이 실제로 존재할 확률이 얼마나 되는지 계산해주기 때문에, 자동 번역의 출력값으로 어떤 문장을 내보내는 것이 더 좋은지 (실생활에서 높은 확률로 존재하는 문장들은 보통 문법적/의미적으로 올바르기 때문) 알려줄 수 있음
- 문장에서 다음 단어가 나타날 확률을 계산해주는 주 목적 외의 부수적인 효과로 *생성 (generative)* 모델을 얻을 수 있음

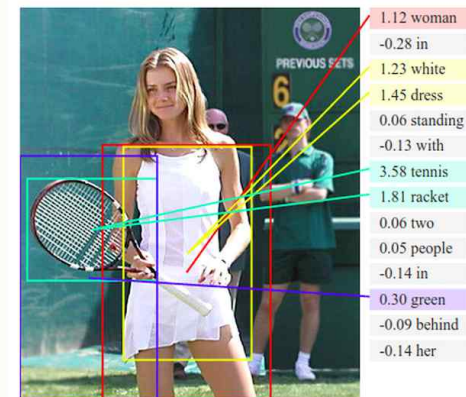
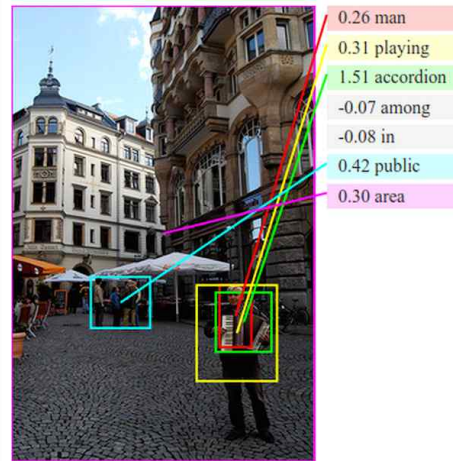
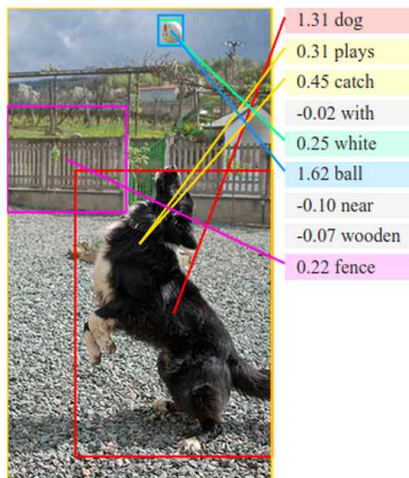
RNN Application – Machine Translation

- 기계 번역 문제는 입력이 단어들의 시퀀스라는 점에서 언어 모델링과 비슷하지만, 출력값이 다른 언어로 되어있는 단어들의 시퀀스라는 점에서 차이가 있음
- 네트워크 상에서의 중요한 차이점은, 입력값을 전부 다 받아들이고 다음에서야 네트워크가 출력값을 내보낸다는 점
- 번역 문제에서는 어순이 다른 문제 등이 있기 때문에 대상 언어의 문장의 첫 단어를 알기 위해선 번역할 문장 전체를 봐야 할 수도 있기 때문



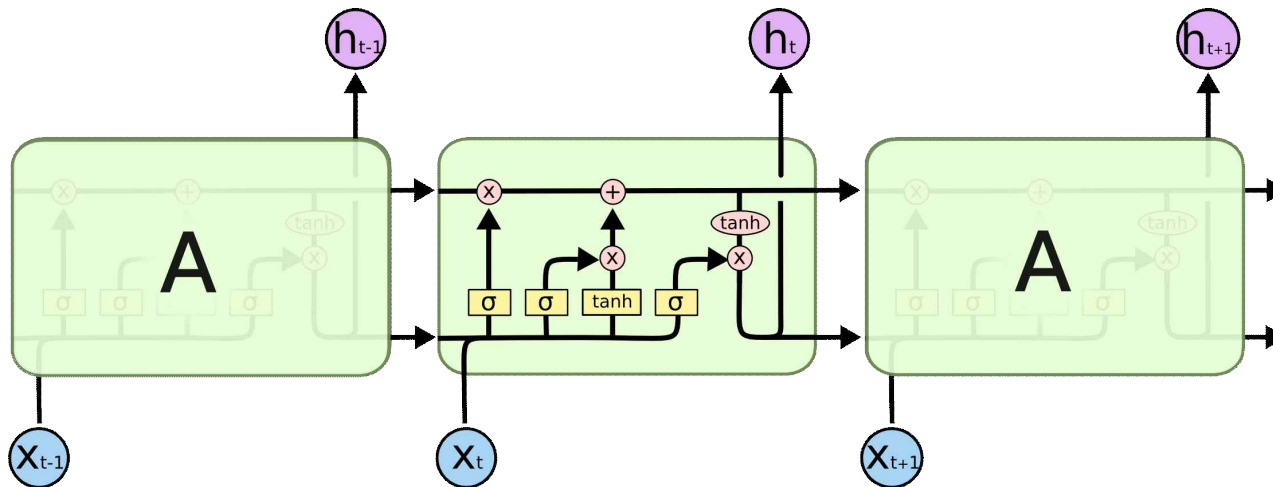
RNN Application – Image/Video Captioning

- CNN(Convolutional Neural Network)과 결합하여, 임의의 이미지를 텍스트로 설명해주는 시스템을 만들 수 있음
- 컴퓨터 비전을 사용해 이미지로부터 단어를 추출하고 매칭시켜, RNN으로 설명해주는 텍스트 생성



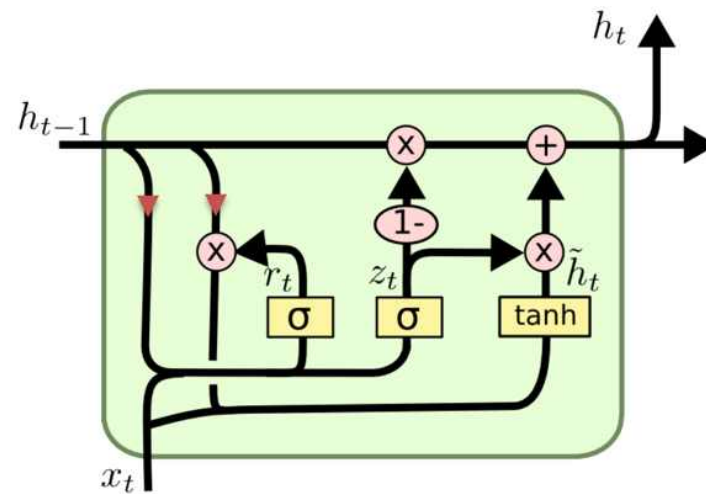
RNN Variants – LSTM(Long-Short Term Memory)

- 장기 의존성 문제를 해결하는 방안으로 제시됨.
- 새로운 정보를 각 Cell State마다 저장할지(기억할지), 잊을지를 결정하는 과정을 거침



RNN Variants – GRU(Gated Recurrent Unit)

- LSTM처럼 장기 의존성 문제를 해결하면서도, 게이트 일부를 생략함으로써 계산 복잡성을 크게 낮춤.
- 크게 Update gate와 Reset gate 두가지로 나뉘어져 있음.



RNN Variants – Attention Mechanism

- 장기 의존성 문제를 더욱 효과적으로 해결하는 방법으로, “중요한 부분에만 집중하게 만들자” 를 제시
- 또한 시각화가 용이함
- “So in summary forget RNN and variants. Use attention. Attention really is all you need!”
 - the fall of RNN / LSTM

<https://towardsdatascience.com/the-fall-of-rnn-lstm-2d1594c74ce0>

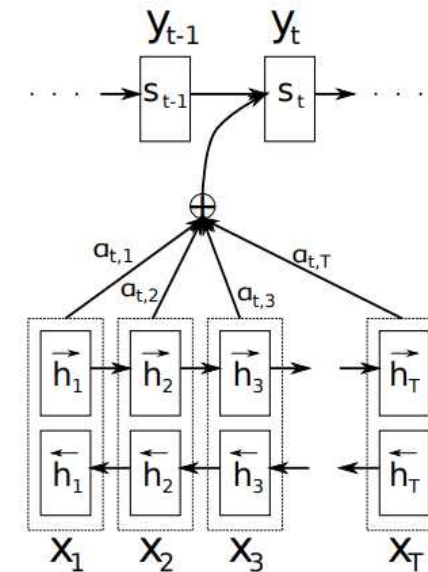


Figure 1: The graphical illustration of the proposed model trying to generate the t -th target word y_t given a source sentence (x_1, x_2, \dots, x_T) .

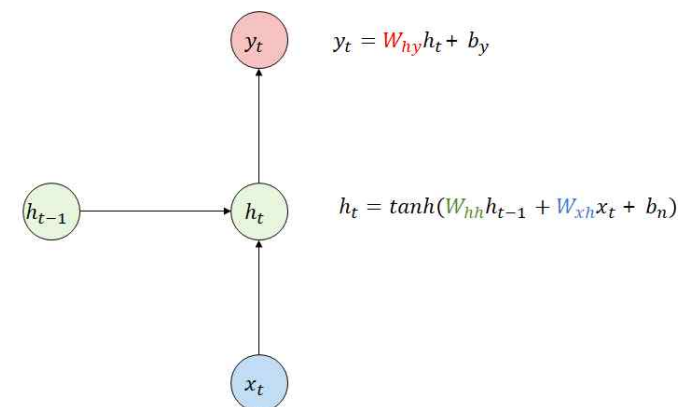
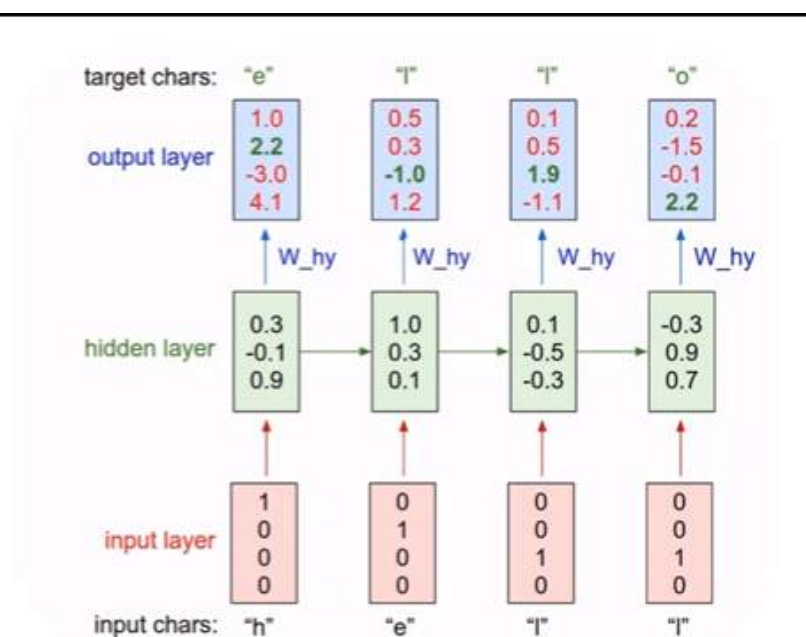
Example – Language Modeling

Character-level language model example

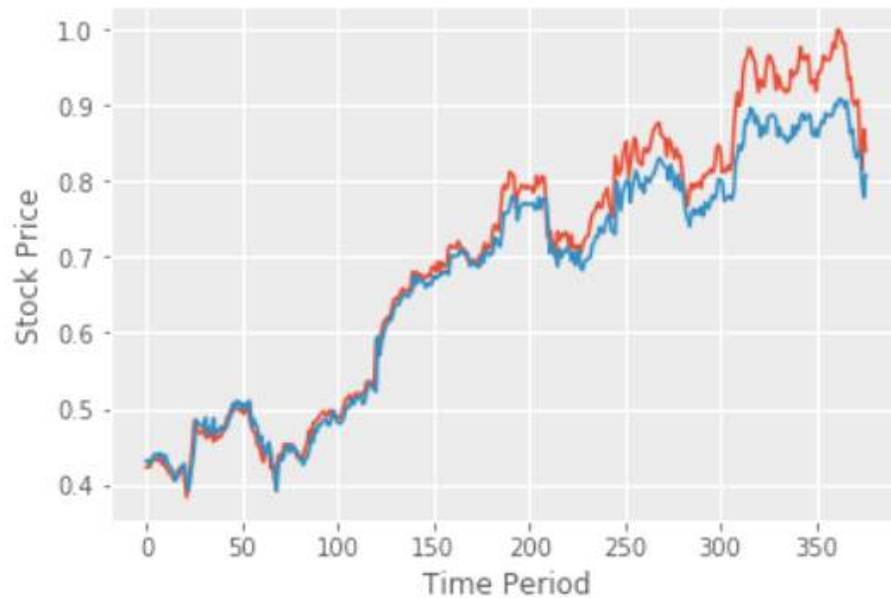
$$y_t = W_{hy}h_t$$

Vocabulary:
[h,e,l,o]

Example training
sequence:
“hello”



Example – Apple Stock Price Prediction



- 모멘텀 투자의 방법론을 적용해서, 얼마나 잘 예측할 수 있을까?
- 시가, 최저, 최고, 종가, 거래량, NASDAQ, Dow Jones, oil price, dollar index, S&P index, gold price, Russell 2000 데이터를 활용
- 결과는 좋지 않음
- 정보 부족? 하이퍼파라미터 튜닝? 모델 문제?



감사합니다!