

SMARCLE
AI study

RNN

Modeep

Team RNN

이용빈, 유지원, 김범주, 송지원, 이운서



Index

01

RNN 개념 1

시퀀스 모델과 RNN

02

RNN 개념2

LSTM / GRU / DRNN

03

pre-class quiz

& 쉬는시간

04

예제 1

로이터 뉴스 기사 분류

05

예제 2

영화리뷰 분석

06

실습

Fake News



RNN 개념 1

- 보통 모델과 시퀀스 모델의 데이터 차이
- 시퀀스 모델이란?
 - 시퀀스 데이터란?
 - 시퀀스 모델의 종류
 - One to many(일대다)
 - Many to one(다대일)
 - Many to many(다대다)
- RNN이란?
 - RNN 모델
 - RNN의 한계



데이터의 비교



1. 첫 번째 던지는 동전, 두 번째 던지는 동전의 앞면이 나올 확률

👉 내가 동전을 찰칵하면서 던지든, 비행기 타고 가면서 던지든, KTX 타고 가면서 던지든, 동전의 앞면이 나올 확률은 같다!

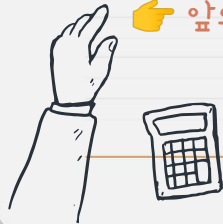
👉 과거는 알 필요 없고, 데이터 간의 순서도 상관없다!

2. “ J에게 이걸 갖다 줘. ”

↖ J는 지원이야.

👉 J가 누군데?

👉 앞의 데이터를 반영해야 하고, 순서가 중요하다!



무엇이 다를까요?

입력되는 데이터의 차이!

일반적인 입력 데이터의 가정

👉 독립 동일 분포 (IDD, Independent and Identically Distributed)

랜덤 확률변수의 집합에서 각각의 랜덤 확률변수들은 **독립적**이면서 동일한 분포를 가진다.

→ 각각의 사건이 다른 사건에 영향을 주지 않는다!

그동안은 과거 데이터를 알 필요 없는 경우만 다루었다.

그렇다면 순서가 중요하고 과거를 알아야 하는 데이터는?!

시퀀스 모델 등장!



시퀀스 모델이란?



말 그대로 **시퀀스 데이터**를 다룰 수 있도록 설계된 모델

시퀀스 데이터

- **순서가 있는 데이터**
 - Text: '문맥'이라는 순서
 - 시계열 데이터: '시간'이라는 순서
 - 음성, 영상, DNA 시퀀스

기존 모델은 모든 input 데이터가 독립적이라고 가정했다.

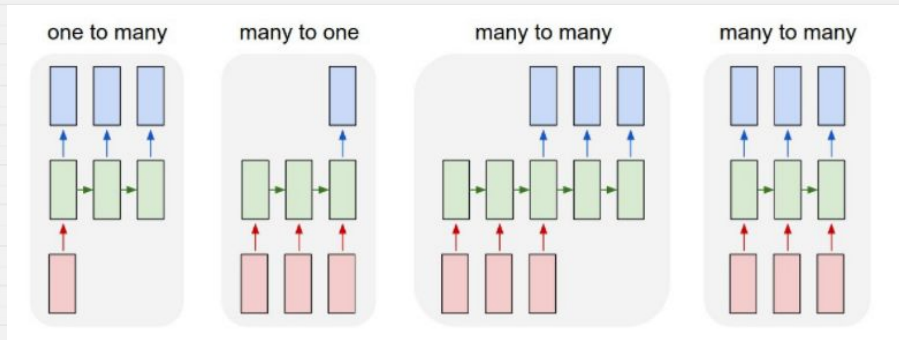
하지만 **데이터의 순서가 의미있을 경우, 순차적으로 과거 정보를 반영할 수 있어야 한다.**



시퀀스 모델의 종류

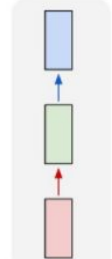


‘input 데이터 유형’과 ‘output 데이터 유형’에 따라 구분된다.



1. One to many(일 대 다)
2. Many to one(다 대 일)
3. Many to many(다 대 다)

one to one



※ one to one

- 그동안 많이 봐왔던 구조
- 시퀀스 모델이라고 보기 어렵다.



시퀀스 모델의 종류

One to many (일 대 다)

1. Image captioning

- Input: 이미지
- Output: 단어 sequence
- 이미지에서 설명 글 만들기



2. Music generation

- Input: 빈 집합, 정수, 음악 장르, 첫 음 등
- Output: 음표 sequence

Music generation

∅




시퀀스 모델의 종류

Many to one(다 대 일)

1. Sentiment classification(감정 분류)

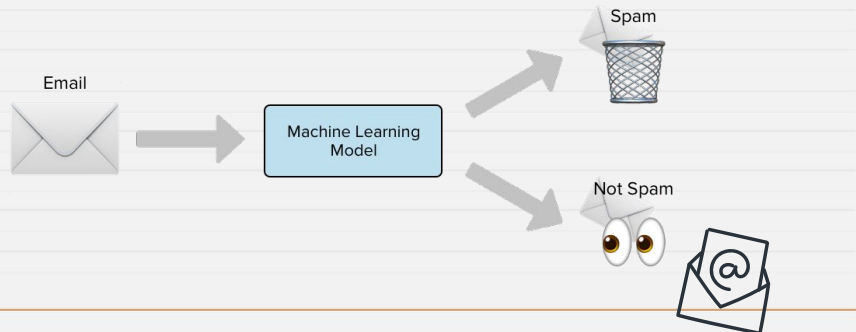
- Input: 단어 sequence
- Output: 별점 등
- 텍스트를 보고 감정의 긍정/부정 분류하기



Target	Aspect	Sentiment
LOC1	general	Positive
LOC1	price	None
LOC1	safety	None
LOC1	transit-location	None
LOC2	general	None
LOC2	price	Negative
LOC2	safety	None
LOC2	transit-location	Positive

2. Spam detection

- Input: 단어 sequence
- Output: 스팸인지 아닌지



시퀀스 모델의 종류

Many to many(다대다) - seq2seq

1. Machine Translation(기계 번역)

- Input: 단어 sequence
- Output: 단어 sequence

2. 챗봇

- Input: 단어 sequence
- Output: 단어 sequence

3. DNA sequence analysis(DNA 순서 분석)

- Input: DNA sequence
- Output: 염기 서열



ACC CCG CGA GAT ATT TTA TAT ATG TGC GCA

Generated sequence of words

Figure 3. Translating DNA sequence into sequence of words.

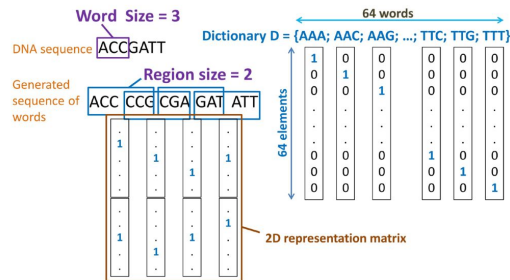


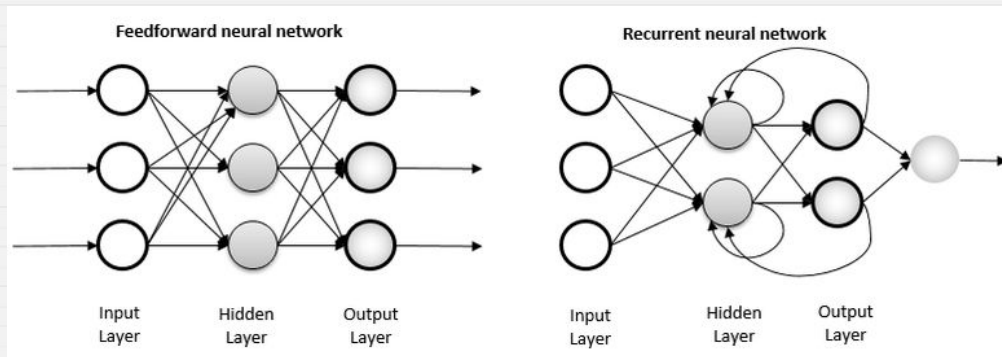
Figure 4. One-hot vector representation of DNA sequence.



RNN이란?

가장 대표적인 시퀀스 모델

RNN의 가장 큰 특징은 순환하는 구조라는 것



인접한 다음 노드로만 진행되는 Feed forward 방식과 비교하면,
RNN은 자기 자신이나 이전 노드로도 진행된다.



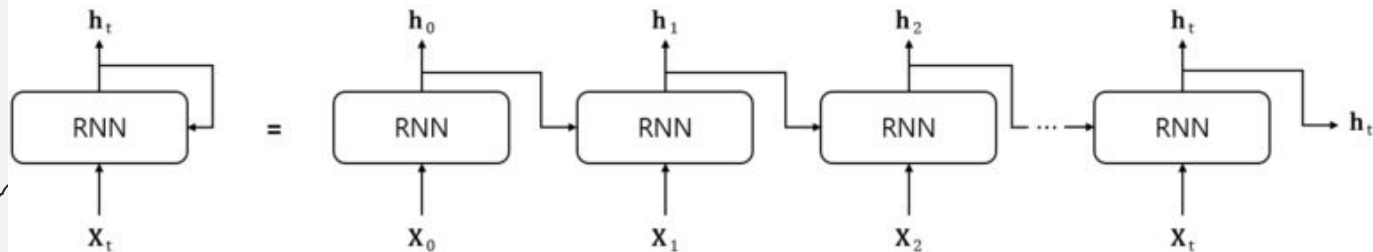
RNN이란?

RNN의 표기법(notation)

⁶⁶ Harry Potter and Hermione invented a new spell. ⁹⁹

$x^{(1)}$ $x^{(2)}$ $x^{(3)}$ $x^{(4)}$... $x^{(t)}$... $x^{(a)}$

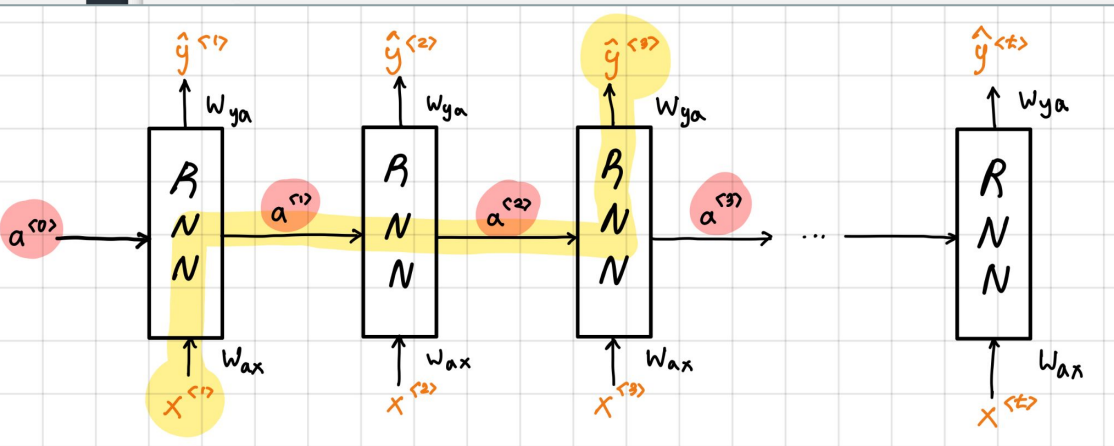
$y^{(1)}$ $y^{(2)}$ $y^{(3)}$ $y^{(4)}$... $y^{(t)}$... $y^{(a)}$



RNN이란?

RNN 알고리즘 - forward propagation

W_{ax} : W 는 x 를 구하기 위해 a 와 곱해진다.



* a 는 activation value

$$a^{(t)} = g_1 (W_{aa} a^{(t-1)} + W_{ax} x^{(t)} + b_a)$$

$$\hat{y}^{(t)} = g_2 (W_{ya} a^{(t)} + b_y)$$

Hyperbolic tangent(tanh)/ReLU
sigmoid

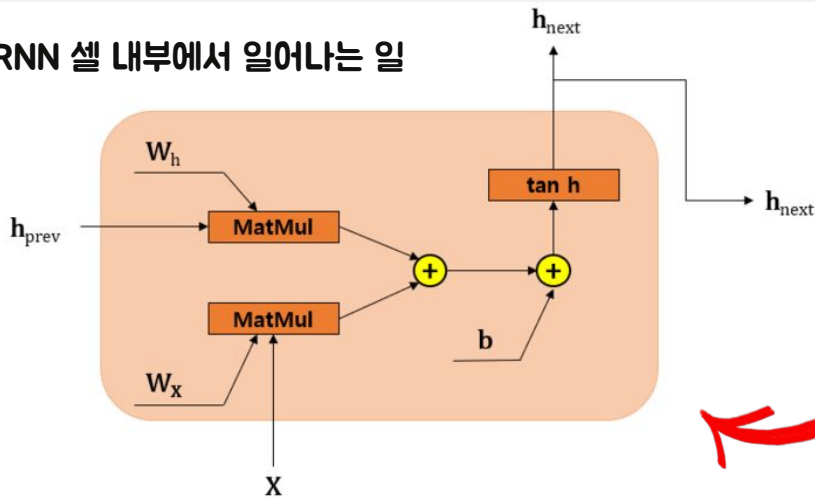
전 시점의 어떤 정보를 다음 시점으로 넘겨준다!
t 시점의 출력은 t-1 시점의 출력을 사용하여 계산된다.



RNN이란?

RNN 알고리즘 - forward propagation

RNN 셀 내부에서 일어나는 일



$$a^{<t>} = g_1 (w_{aa} a^{<t-1>} + w_{ax} x^{<t>} + b_a)$$
$$\hat{y}^{<t>} = g_2 (w_{ga} a^{<t>} + b_y)$$

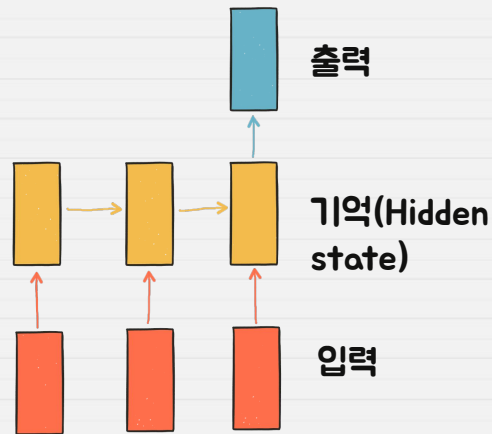
RNN이란?

Hidden state

: 현재 시점 이전의 데이터

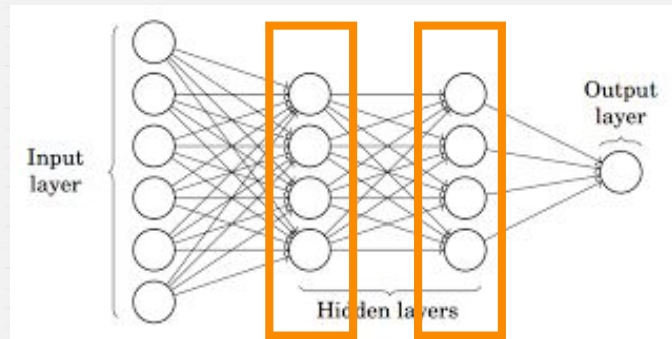
$$a^{<t>} = g_1 (w_{aa} a^{<t-1>} + w_{ax} x^{<t>} + b_a)$$
$$\hat{y}^{<t>} = g_2 (w_{ga} a^{<t>} + b_y)$$

‘기억’으로 여겨진다.

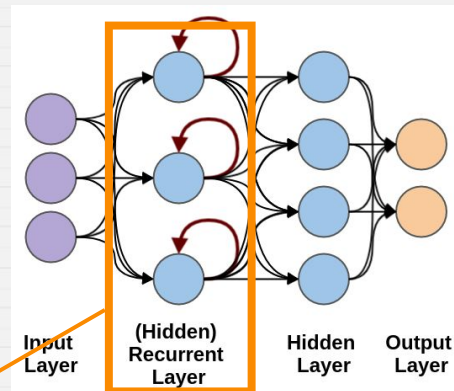


RNN이란?

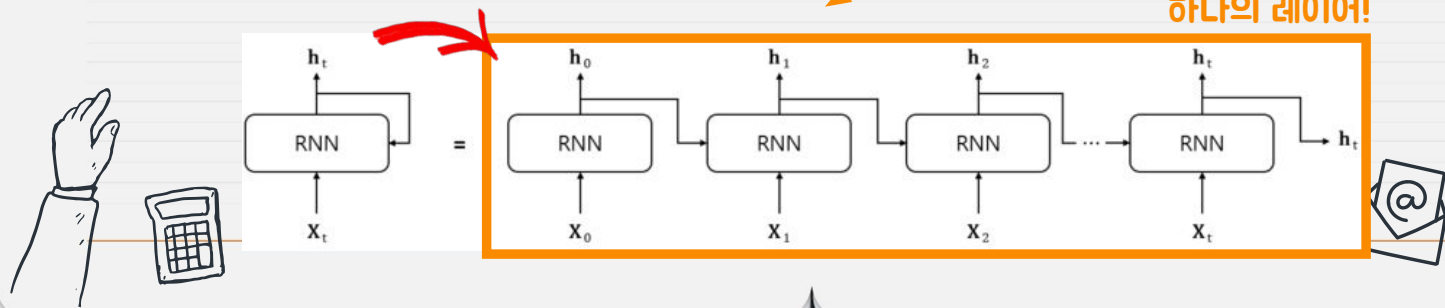
〈MLP〉



〈RNN〉



하나의 레이어!



RNN의 한계

RNN에서 최적화를 위해 역전파를 하면서
RNN layer(계층)이 길어질 수록 전달하는 gradient가 소실된다.

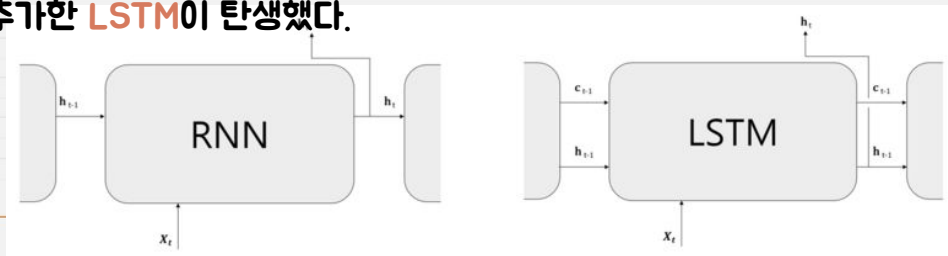
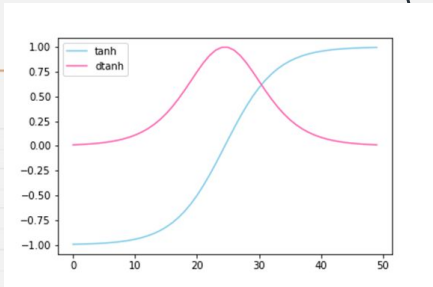
“지원이는 밥을 먹었다.

지원이는 청기를 올렸다가 백기를 올렸다가 앉았다가 일어섰다가 돌았다가 누웠다.”

➡ RNN은 '지원이는 밥을 먹었다'는 정보를 기억하지 못한다.

➡ 시간적으로 많이 떨어진 장기 의존 관계(Long Term)를 학습하기 어렵다.

이런 vanishing gradient problem을 극복하기 위한 방안으로
장기 기억이 가능하도록 기억 셀을 추가한 LSTM이 탄생했다.



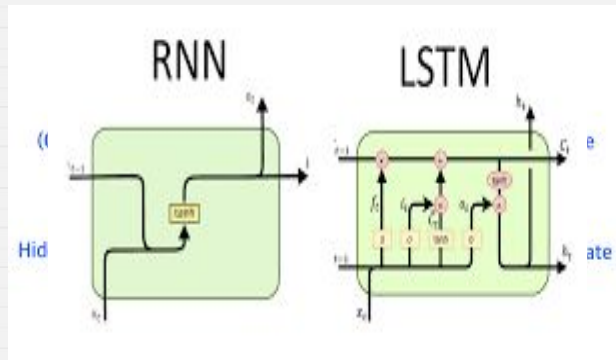
1. LSTM

Long short term **memory** / 장단기 기억, 장단기 **메모리**

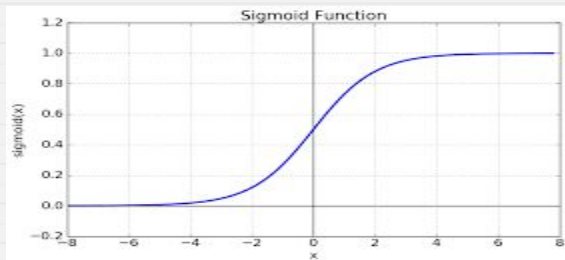
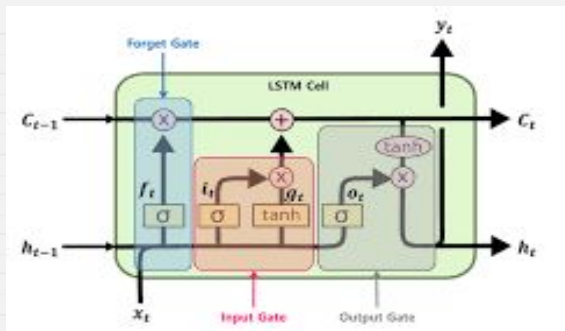
→ 기억...?

- **입력 게이트**
- 망각 게이트 → 핵심!!
- **출력 게이트**

불필요한 내용은 지우고, 기억해야 할 것들을
정하고 장, 단기적으로 저장



게이트의 역할



망각 게이트

- 과거의 정보를 잊기 위한 게이트
 - **시그모이드** 함수를 적용
- 출력 범위가 0~1
→ 0이면 이전 상태의 정보를 잊고,
1 이면 이전 상태의 정보를 기억

입력 게이트

- 현재 정보를 **입력**하기 위한 게이트

출력 게이트

- 결과값을 **출력**하기 위한 게이트

Sigmoid & Tanh



Sigmoid : 현재 정보를 저장할까 말까?

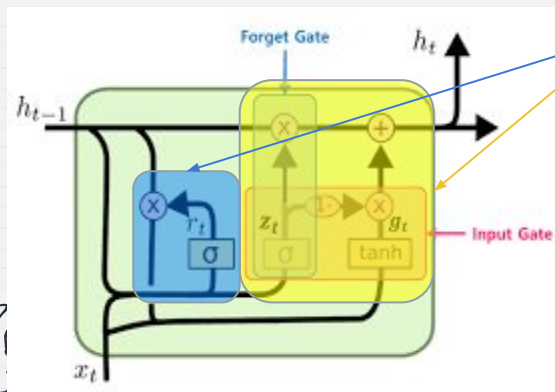
Tanh : 저장한다면 얼마나 저장할까?



2. GRU

Gated Recurrent Units -> LSTM과 기능은 유사, 구조 단순화

망각, 입력, 출력 3개의 게이트를 **Update(망각+입력)**, **Reset** 두 가지 게이트로 변경



게이트가 3->2

학습 가중치 down

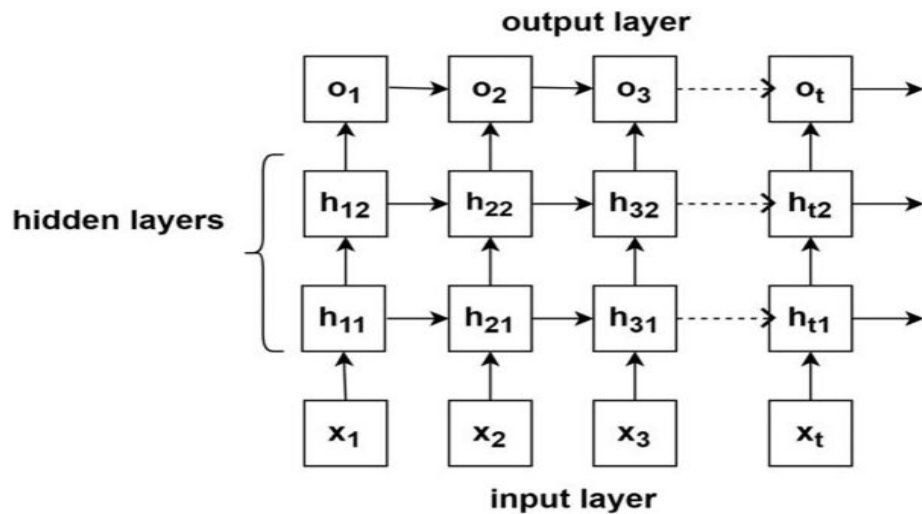
연산량 down

학습속도 fast

3. DRNN

Deep Recurrent Neural Networks

-> RNN을 여러 겹 쌓아 Deep 하게 만든 네트워크



DRNN 왜 쓰는건데?



긴 문장... 감당 가능해?

RNN : 충분히 **wide** 하거나 **deep** 하지 못하면 긴 문장은 예측 불가



층(layer) 중첩 이유(1)

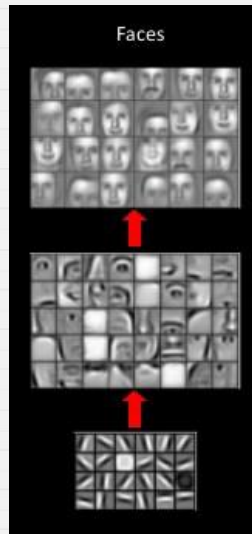
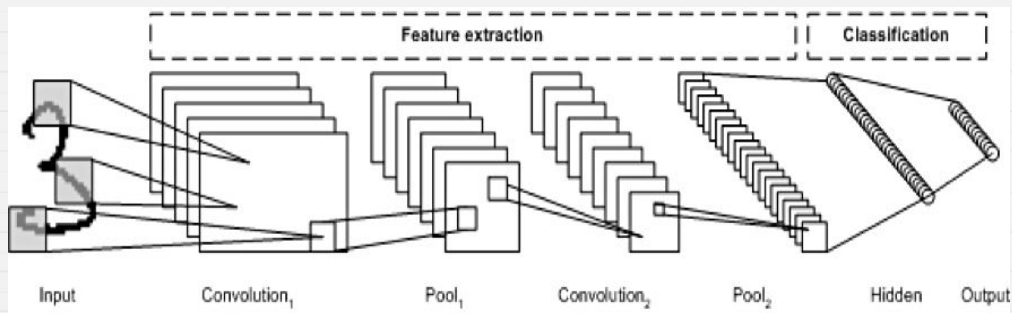
Layer를 여러 층 겹치면

Input에 가까운 layer는 edge(부분적) 특징을,

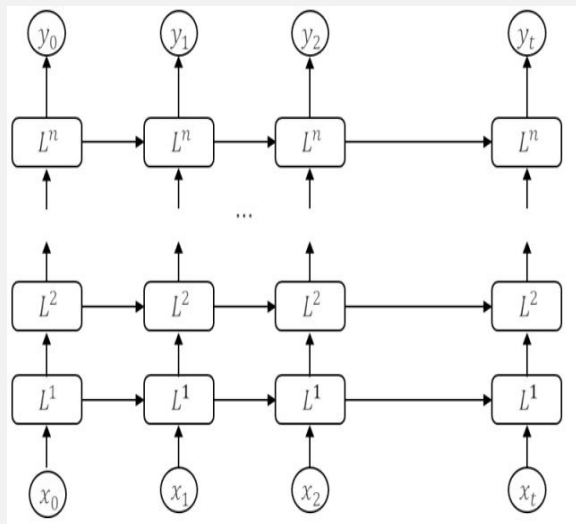
(Low-level : 윤곽->대각선, 수평선, 수직선, 원 등)

Output에 가까운 layer는 abstract(추상적) 특징을 추출

(High-level : 사람의 얼굴 등)



층(layer) 중첩 이유(2)



Semantic/의미적 : 문장이 나타내는 의미 분석

말 참 많은 : 말 => horse? speech?

Syntactic/구문적 : 문장 구성 성분들간의 관계 분석



정리

- 순차 데이터(sequential data)를 처리하기 위해서 RNN 사용
- Vanishing gradient 문제를 해결하기 위해 LSTM, GRU 사용
- 긴 문장 처리를 위해서는 layer 중첩 필요

RNN의 단점을 보완한 LSTM, GRU를 여러 겹 중첩시켜 DRNN을 만들어 사용하면 더 좋은 성능을 구현 할 수 있다!



STM을 이용한 로이터 뉴스 카테고리 분류하기

[데이터 불러오기 & 분석]

```
] #필요한 라이브러리 import
from keras.datasets import reuters #로이터 뉴스 데이터셋 불러오기
from keras.models import Sequential
from keras.layers import Dense, LSTM, Embedding
from keras.preprocessing import sequence
from keras.utils import np_utils
import numpy
import tensorflow as tf
import matplotlib.pyplot as plt
```

예제 1번

LSTM을 이용한 로이터 뉴스 카테고리 분석하기

colab

LSTM과 CNN의 조합을 이용한 영화 리뷰 분류하기.ipynb ☆

파일 수정 보기 삽입 런타임 도구 도움말 모든 변경사항이 저장됨

코드 + 텍스트

```
#필요한 라이브러리 불러오기
from keras.preprocessing import sequence
from keras.models import Sequential
from keras.layers import Dense, Dropout, Activation
from keras.layers import Embedding
from keras.layers import LSTM
from keras.layers import Conv1D, MaxPooling1D
from keras.datasets import imdb

import numpy
import tensorflow as tf
import matplotlib.pyplot as plt
```

예제 2번

**LSTM CNN을 이용한 영화 리뷰
반응 분석하기**

colab



Kaggle

upvote 눌러주세요 ~

[link](#)

A top-down view of a light blue desk. On the left is a white keyboard. In the bottom left is a small green succulent. In the top right is a brown pencil. In the center is a light blue sticky note with an orange one on top. On the right is a white piece of paper with a scalloped edge, held by an orange paperclip and secured with orange tape. The orange sticky note has the text "Do you have any questions?". The white paper has the text "Thank you for watching Team RNN" and a drawing of a hand holding a pen.

**Do you
have any
questions?**

**Thank you
for watching
Team RNN**

