CS231n, RNN

RNN

기존 Deep learning 모델은 하나의 input이 vector든 image든 어떠한 형태로 변형되어 hidden layer에 들어가고, 이에 따른 하나의 output이 (classification 등) 나오는 구조이다. (one to one) 하지만 input이 여러개고, output 또한 여러개가 나오는 구조를 만들고 싶다. 어떻게 해야 할까? 그리고 이러한 구조가 왜 필요한 것 일까?

- **One to Many (고정된(fixed) input, sequence output)**

  input이 image와 같은 vector가 들어갔는데, output이 특정 길이를 가지는 변수로 나올 때. (caption과 같은, 하나의 이미지로 그 이미지가 무엇을 하는지 설명을 할 때?)

- **Many to One (sequence input, fixed output)**

  input도 하나가 아니라 특정 사이즈의 여러개가 들어오는 구조. (Text, video등)

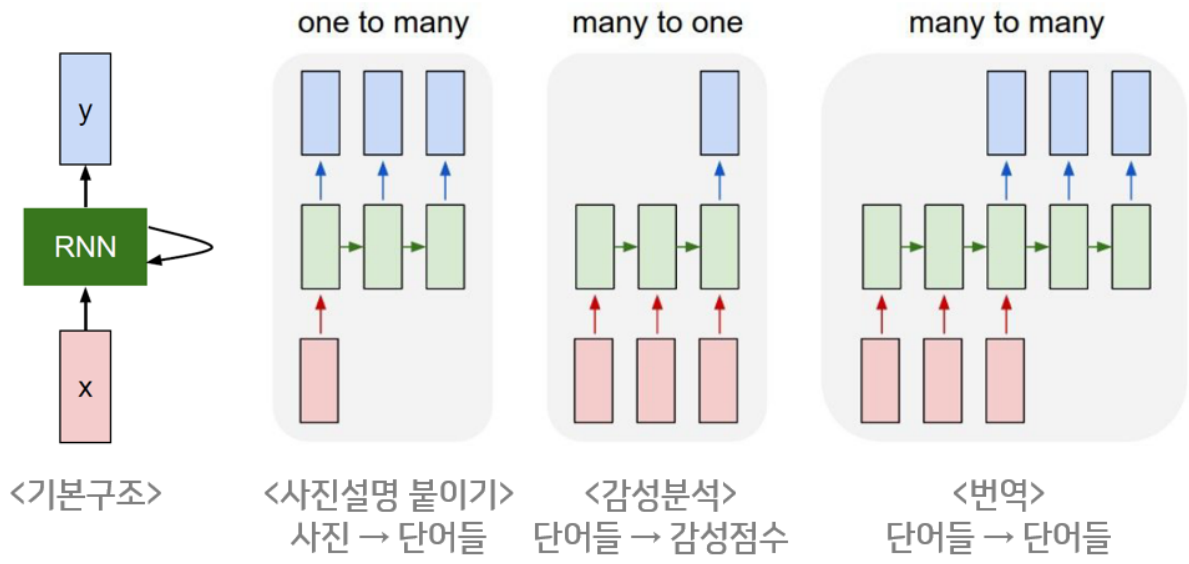
  video의 경우 연속된 frame이므로 여러개의 input이 들어간다고 볼 수 있고 이로써 특정 행동들을 추론하게 되는 결과를 얻고자 한다면 특정 행동이라는 하나의 결과를 보고자 하는 것 이므로, 하나의 output이 나오게 되는 many to one 구조를 하고 있다고 볼 수 있다.

- **Many to Many (sequence input, sequence output)**

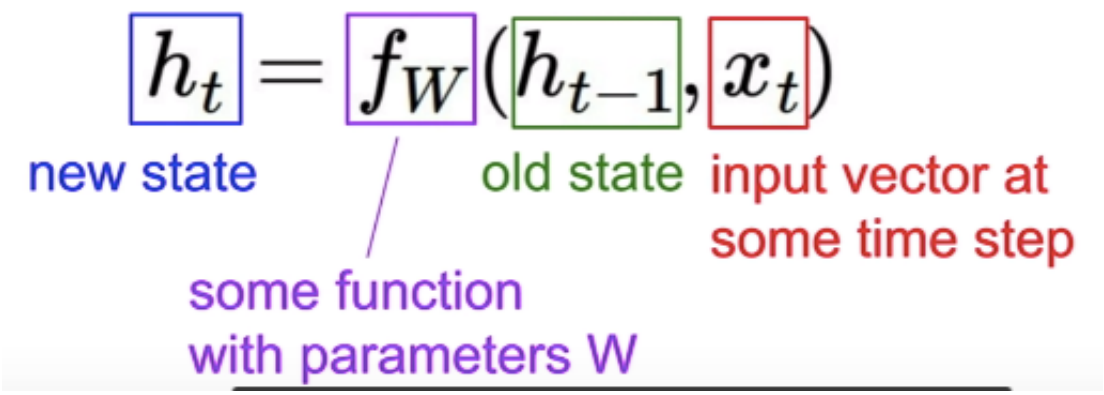
  machine translation에 이용됨. (input English output Korean, Seq to Seq...?) 이러한 Many to Many 구조는 크게 두 종류로 나뉘는데, translation과 같은 input과 output이 1-1 매칭이 되지 않는 경우와 video 분석의 경우, 하나의 input마다 어떠한 특정 output (즉, 행동에 대한 결과가 존재) 이 존재하므로 이러한 1-1 매칭이 되는 경우로 나눌 수 있겠다. (tranlation의 경우 각 단어를 이루는 철자 자체가 의미를 가진 것은 아니므로 결과에 대해 철자 자체가 1-1 대응이 안되기 때문에 이렇게 말을 할 수 있다고 생각함.)

RNN의 이용

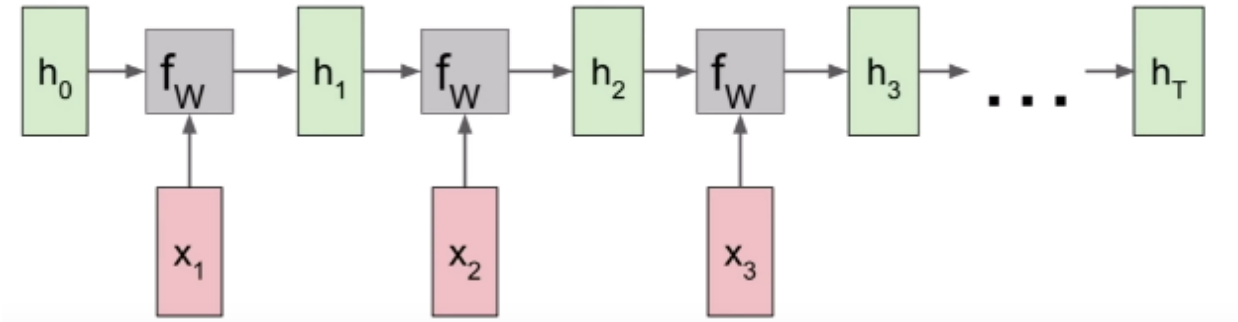
Mnist data라는 image를 input으로 사용하여 해당 image의 숫자가 무엇인지 판단하려고 한다. (뭐 이런...)



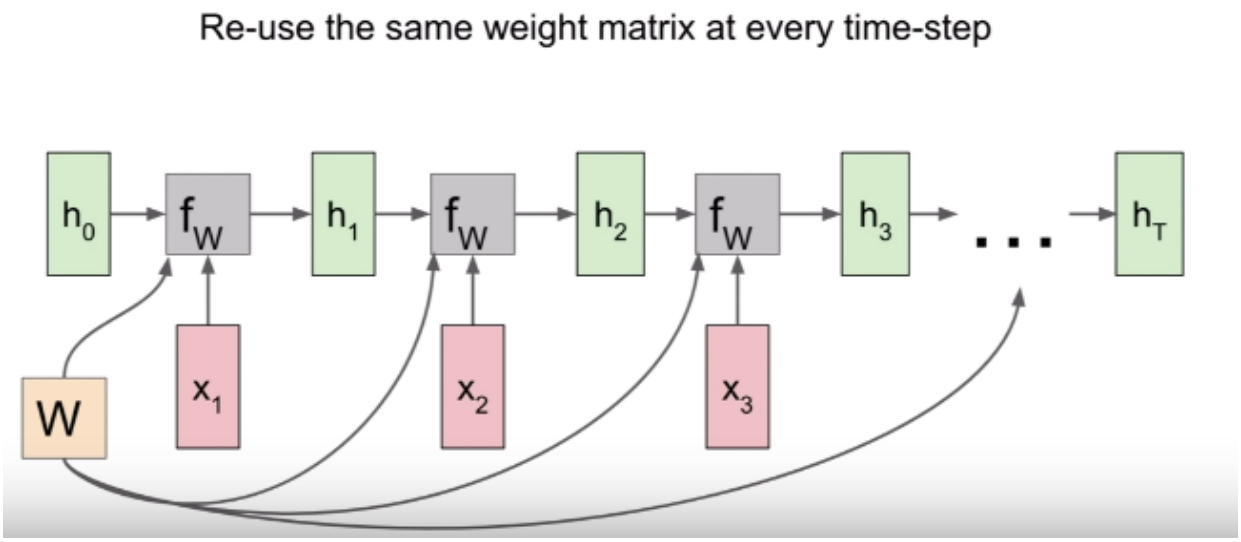
기본구조로 이해를 해보면 input 'x' 가 들어오고 input이 hidden layer인 'RNN' 으로 입성. input이 들어올때마다 hidden layer(internal hidden state)가 update. every step마다 output을 도출해 낸다.



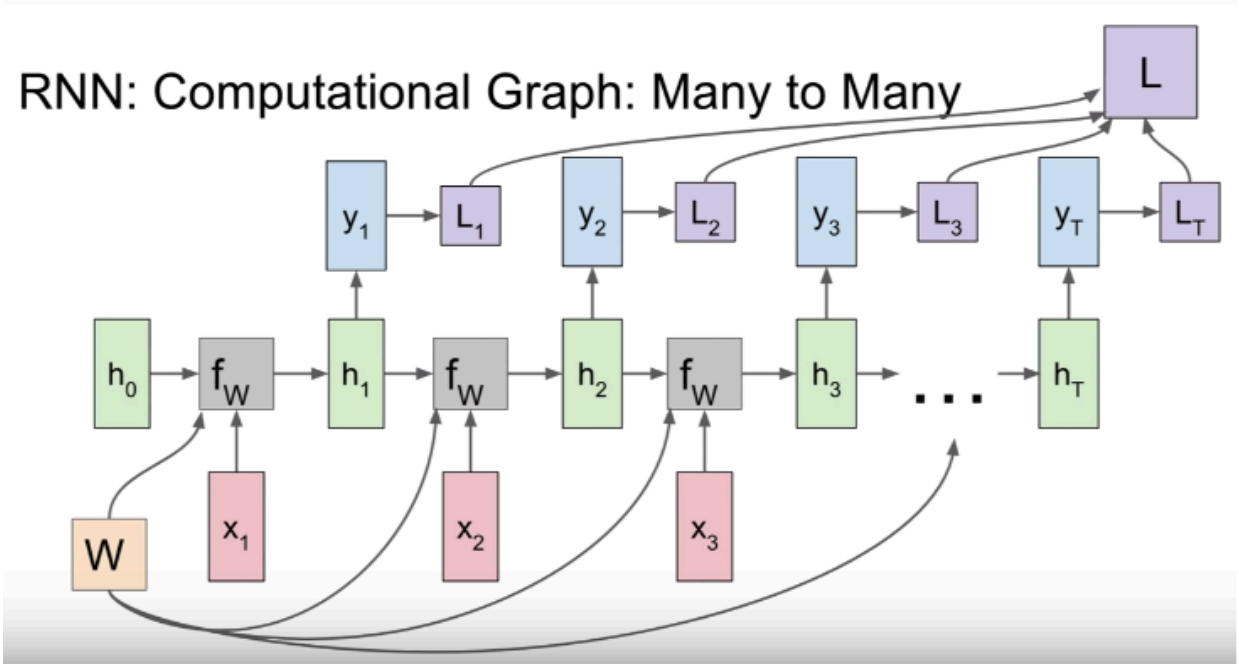
<RNN의 functional form, every step마다 update해준다.>



위와 같은 형태로 h_i 를 update해 나간다. 새로운 input이 들어올 때 마다 update되는 구조.



Weight는 계속해서 재사용된다. (time-step에서는)

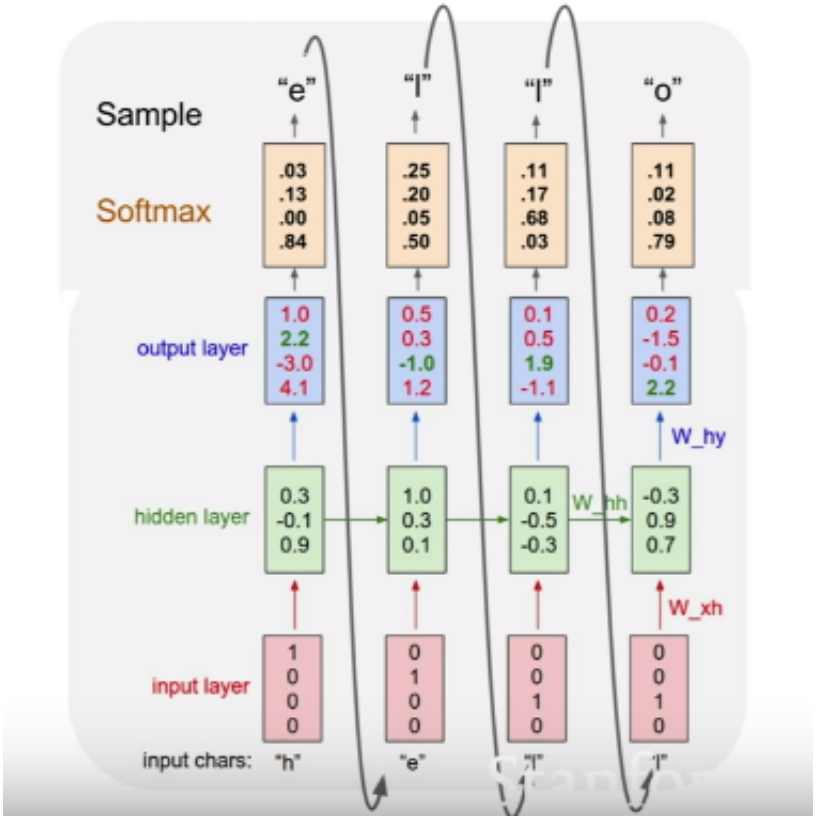


마지막으로 Loss의 sum... (큰 Loss) 역시 그림으로 이해하는게 좋다. input의 흐름으로 써, h를 update하고 계속해서 더 나은 결과를 얻어가는 과정이라 이해하면 되려나...?

RNN에서 encode와 decode 개념이 나오기 시작한다.

- Many to one encode input
- One to many decode output...?

[Example] Character-level Language Model



현재 'Hello' 라는 단어를 만들고 싶음. => h의 정답 label은 다음에 나올 철자인 'e'

이게 보면 sequnece가 길어지면 한 step이 (forward) 매우 매우 길어지고 훈련속도가 매우 매우 느려질 것 이다. 이러한 문제점을 해결 할 방안이 존재하는가? => truncated backpropagation이라는 해법을 내놓음 (backpropagation approximation 같은 건가) 이것도 batch를 나누어 하는 것 같은데. **(Markov 와는 다르다.)**

[Example] Image Captioning

CNN + RNN 구조를 가지고 있다. CNN의 결과를 RNN의 h_{0}로 쓰는 구조로 보인다.

Image Captioning with Attention

Performance를 좋게하기 위한 어떠한 방법 (Attention은 어디서 들어봤는데) input이 두개가 되는듯...? TMI) 문장이랑 이미지랑 FCN에서 합쳐버린다음에 학습을 시키는 듯 (captioning...)

Gradient Flow

Exploding gradients / Vanishing gradients 문제가 발생.

LSTM

have 2 hidden state. (cell state가 존재한다.) 4개의 gate가 존재 (기존 RNN에서 cell state가 추가된 상태...?)

Reference

- https://ratsgo.github.io/natural%20language%20processing/2017/03/09/rnnlstm/
- https://dreamgonfly.github.io/rnn/2017/09/04/understanding-rnn.html
- https://aikorea.org/blog/rnn-tutorial-1/