



Week5 Presentation

RNN (Recurrent Neural Network)

GDSC Hanyang ML/DL : Basic, Jaeseung Lee

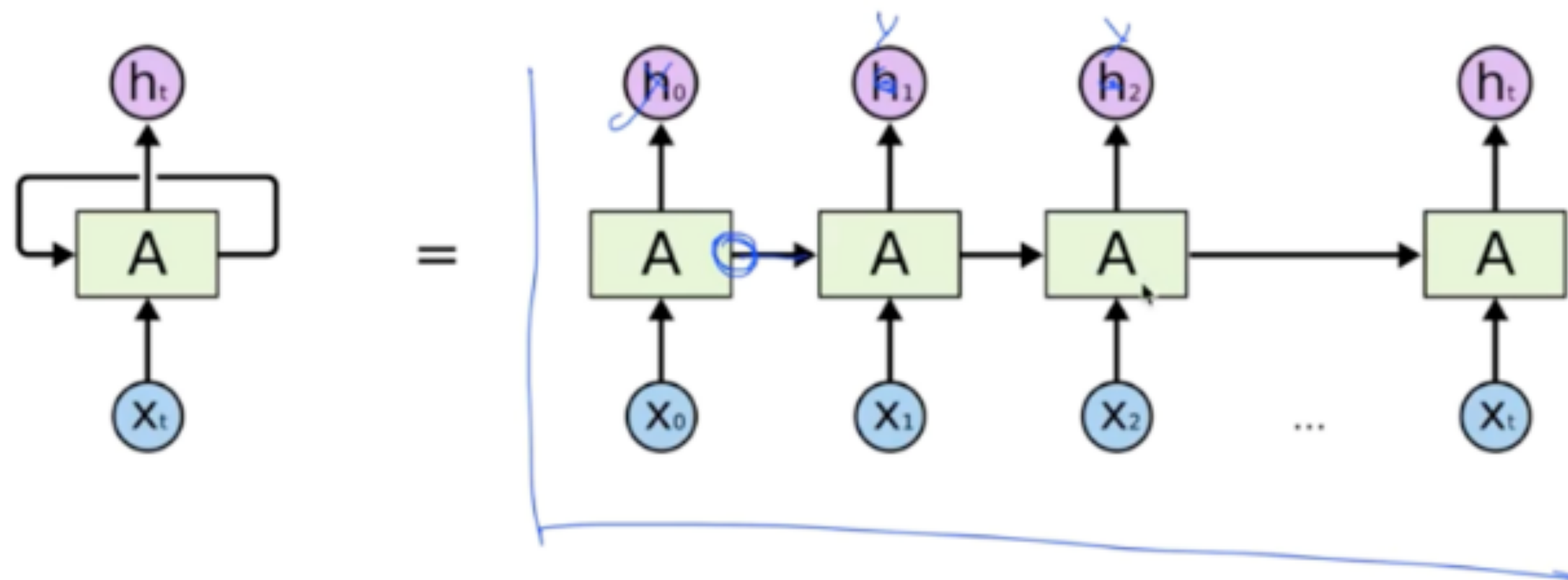
Index

- RNN
 - Basic concepts (one to one)
 - Types
 - Many to one
 - Many to many
 - Seq2Seq
 - Seq2Seq Attention

RNN

Basic concepts

- RNN (Recurrence Neural Network)
 - Sequence data를 효과적으로 처리할 수 있는 NN
 - Recurrence하게 이전 state를 통해 현재 state를 정의함



RNN

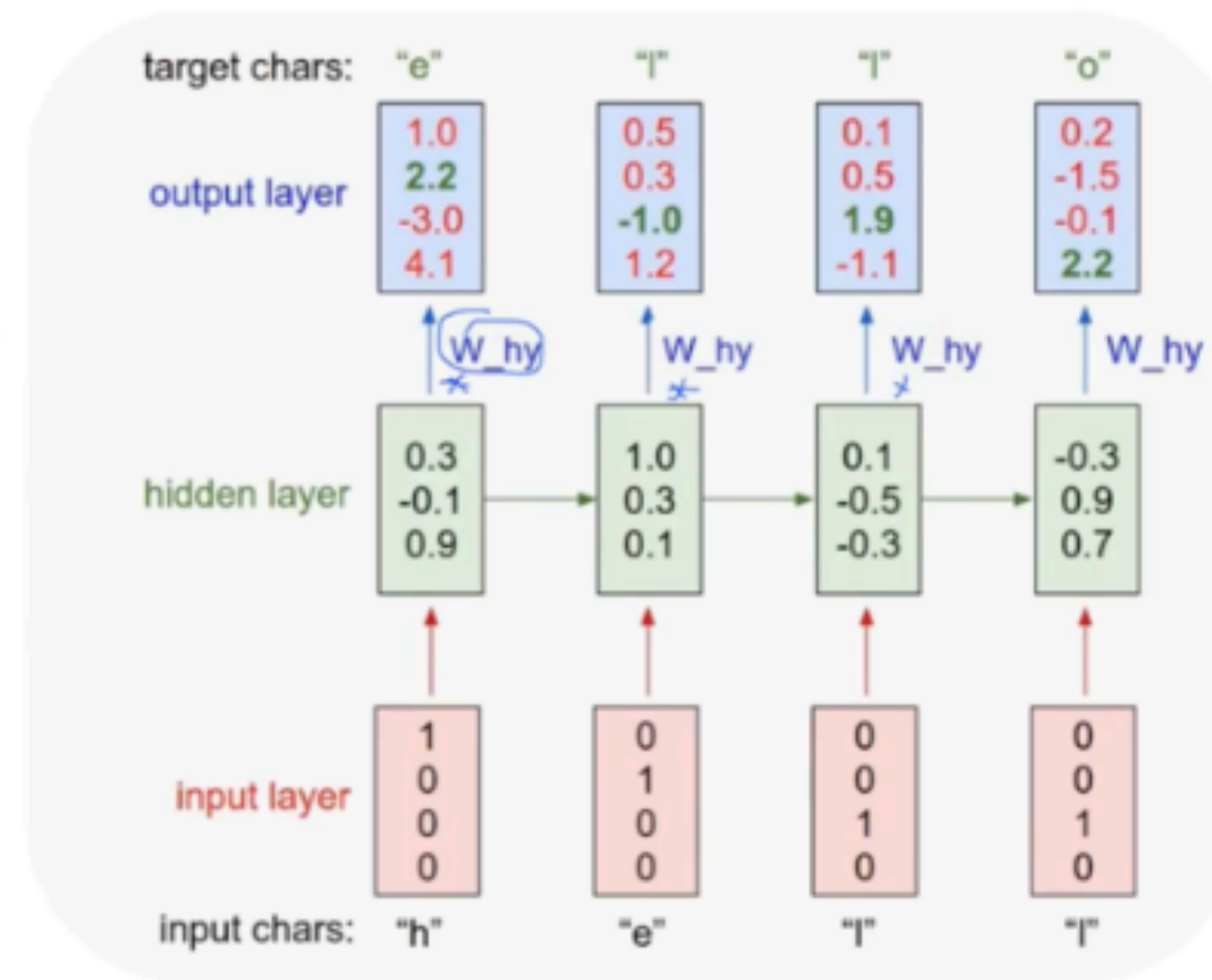
Basic concepts

Character-level
language model
example

$$y_t = W_{hy}h_t$$

Vocabulary:
[h,e,l,o]

Example training
sequence:
"hello"



Input layer (one-hot encoded)를 통해 hidden layer 값 도출

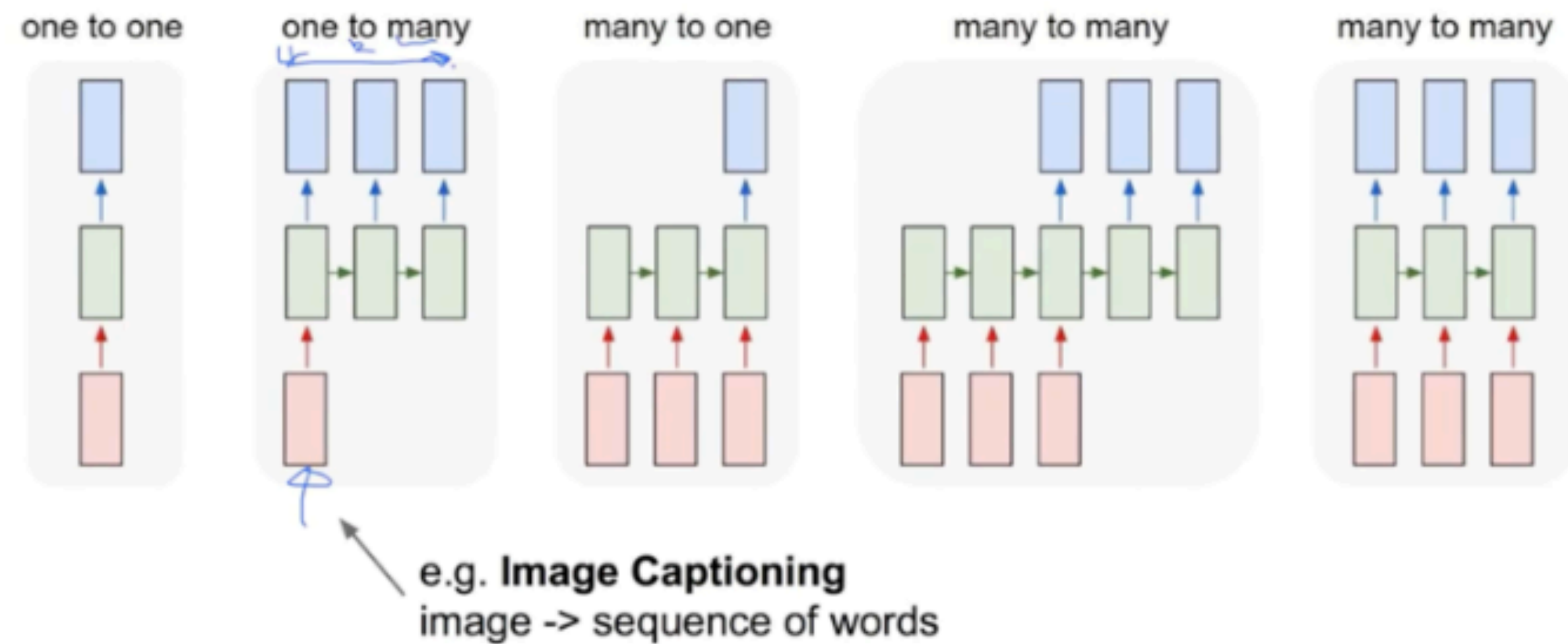
그 값들은 이전 값에 Weigth를 곱한 값과 현재 input의 Weigth를 곱한 값의 합

최종 output은 softmax를 통해 가장 큰 값을 가지는 인덱스를 나타냄.

RNN

Basic concepts

Recurrent Networks offer a lot of flexibility:



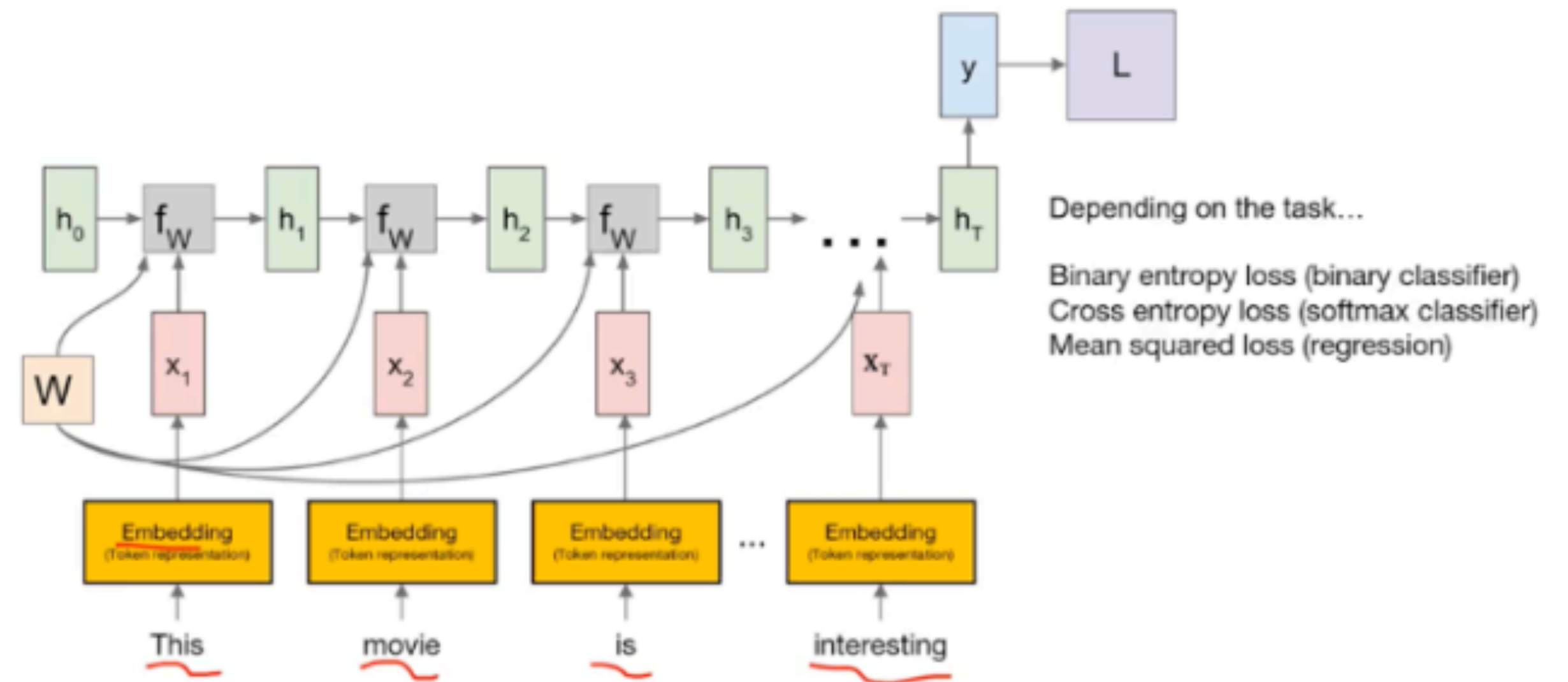
- RNN은 one to one 뿐만 아니라 Problem에 맞게 다양하게 구성가능하다 !

RNN Types

Many to one

- 자연어 처리에서 문장의 분위기 파악(긍정/부정)에 사용됨
- 문장을 word 단위로 tokenization (각각 단어로 쪼갬)
- Embedding layer : tokening node를 numeric vector로 바꾸어 학습 가능하게 함

What is “many to one”?



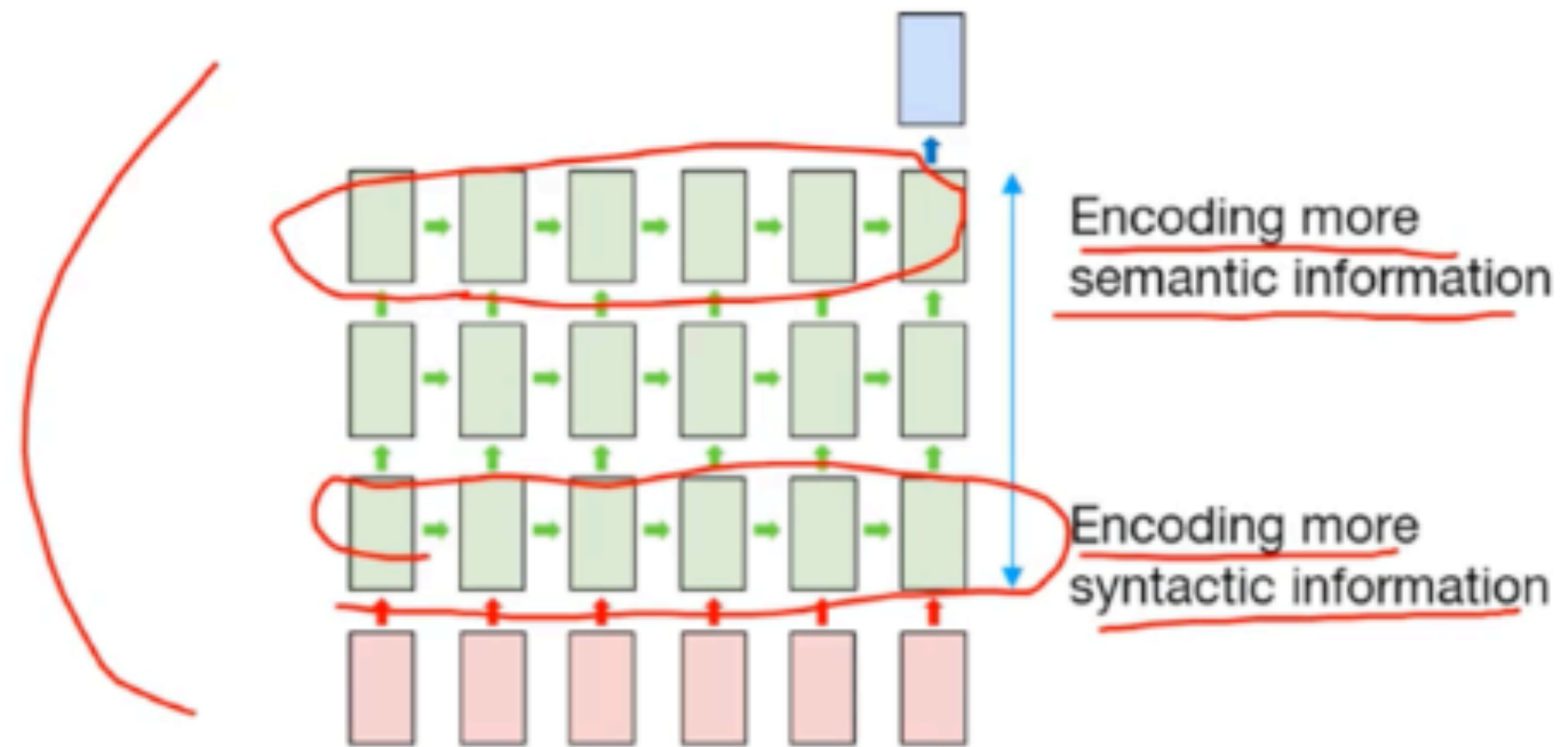
RNN Types

Many to one stacking

- Stacking
 - Recurrent layer를 여러개 쌓는 기법을 의미함.
- 자연어 처리에서 layer 의미
 - Input layer쪽은 syntactic
 - Output layer쪽은 semantic

What is “stacking”?

Besides, many works have shown that **different layers of deep RNNs encode different types of information.**

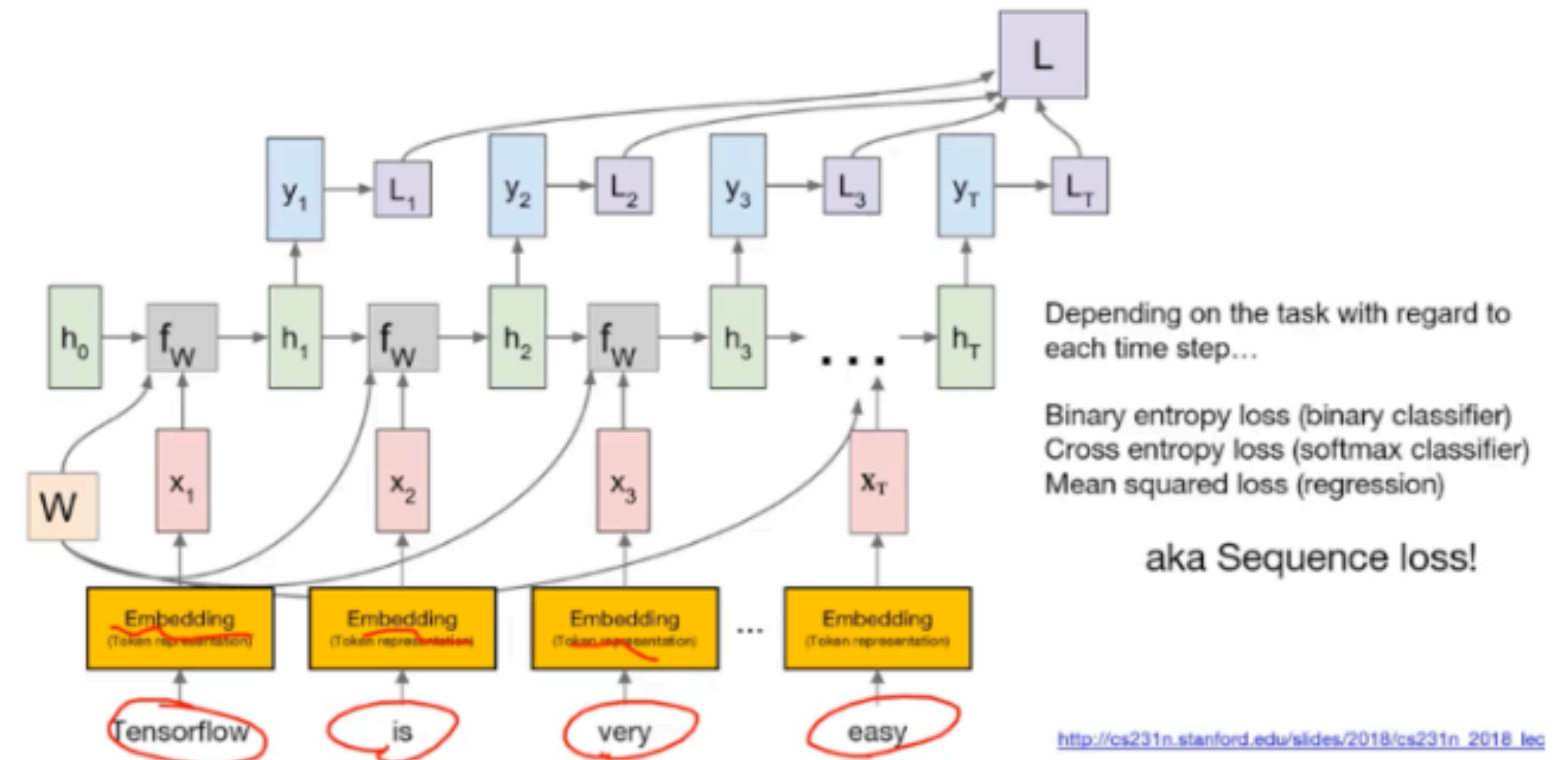


RNN Types

Many to many

- 자연어 처리에서 형태소 분석에 사용되고 있음.
- 각각의 토큰들이 모두 출력을 만들고, 토큰마다 loss를 계산 후 하나로 합침. 이를 sequence loss라고 함.
- Loss 계산 방식이 이전과 다르게 pad token을 고려해야 하여 masking 개념을 도입
 - Masking : pad는 학습 X

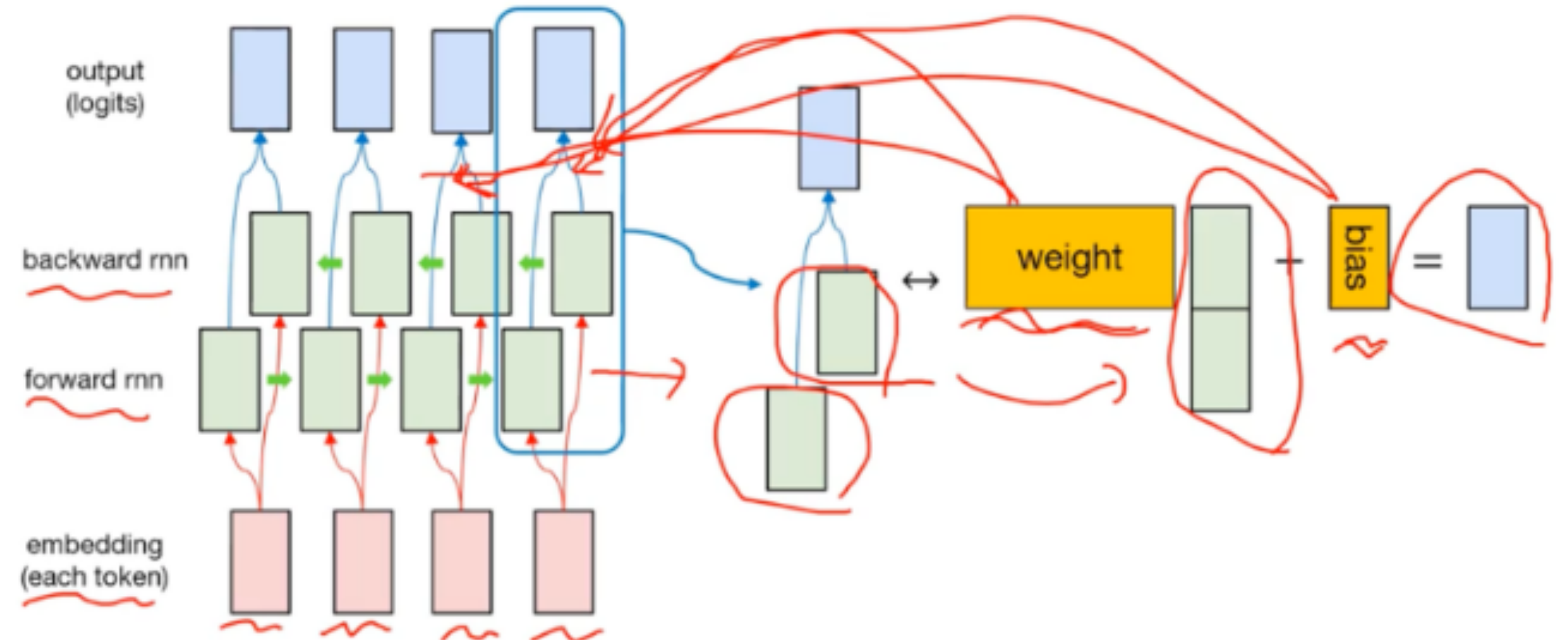
What is “many to many”?



RNN Types

Many to many bidirectional

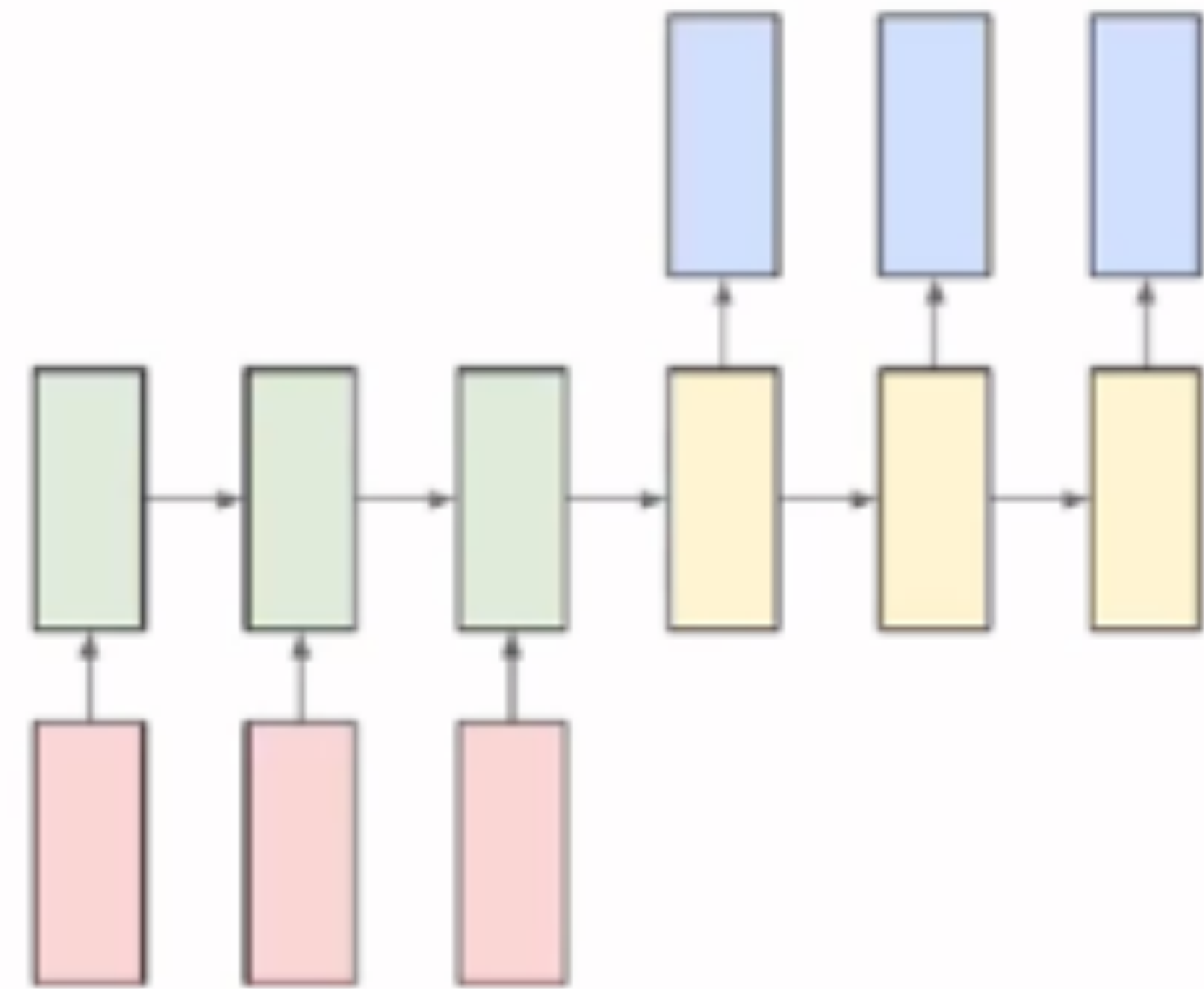
- RNN many to many에서는 단방향으로만 정보가 전달되기에 output layer에서 정보의 불균형이 존재한다.
- 그래서 이를 해결하기 위해 forward, backward layer를 만들어 정보 불균형을 없앴.



RNN Types

Seq2Seq

- Seq2Seq (Sequence to Sequence)
 - 이 방식은 Encoder/Decoder로 파트가 나뉨.
 - 자연어 처리에서 번역/Chatbot을 만들 때 많이 사용하는 방식
 - LSTM or GRU 방식을 통해 Encoder 학습
 - Decoder에는 FC를 통해 출력 Layer



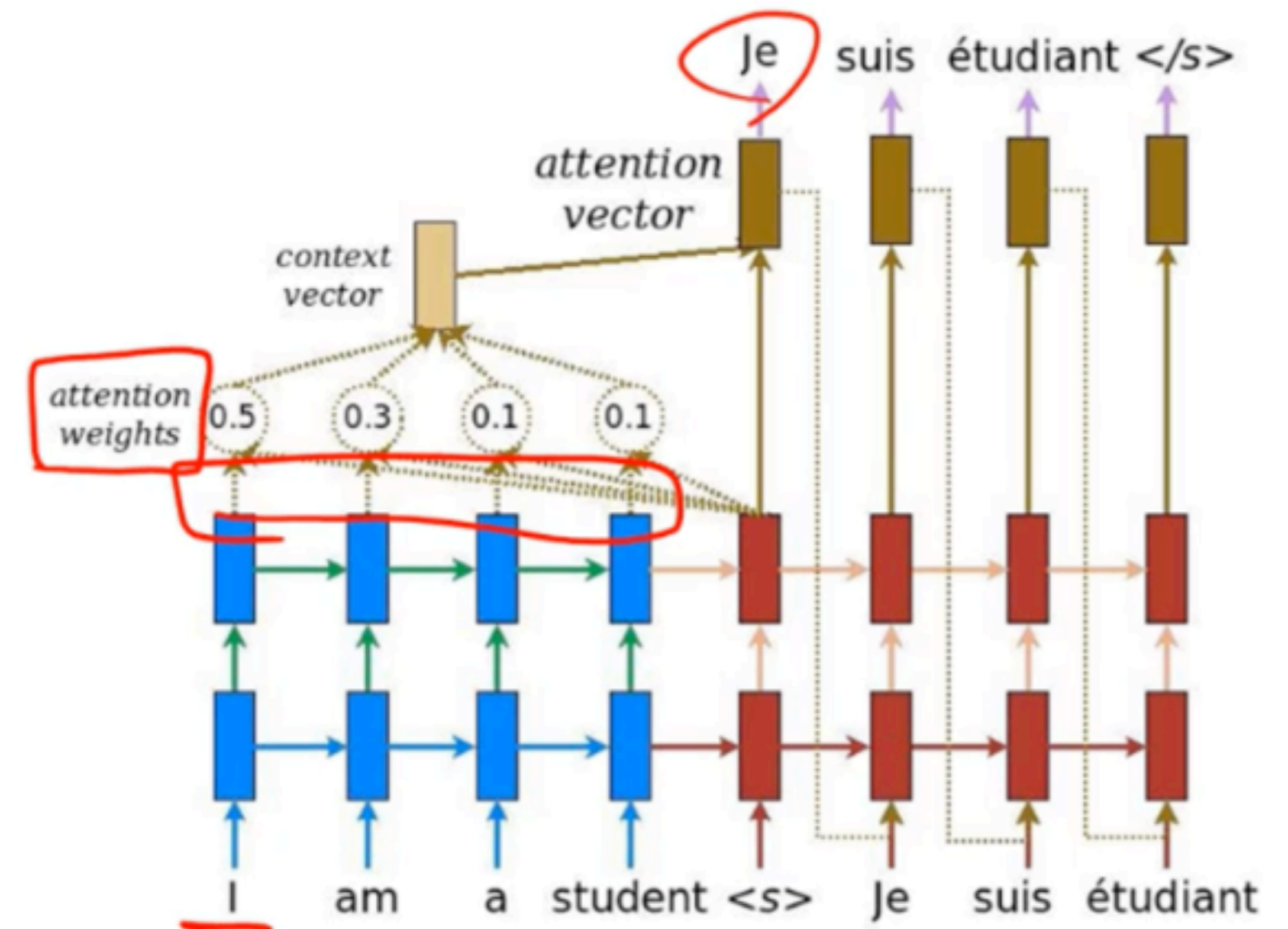
Seq2Seq

RNN Types

Seq2Seq Attention

- 원래 1개의 벡터를 통해 encoder의 정보를 abstract 했지만, 많은 정보를 나타내기에는 한계가 존재
- Attention method
 - Attention weight을 두어 가장 의미가 있는 단어에 가장 큰 값을 주어 집중시킴
- Decoder에 공식들과 함께 적용.

What is Attention?





Thank you for listening

RNN (Recurrent Neural Network)

GDSC Hanyang ML/DL : Basic, Jaeseung Lee