



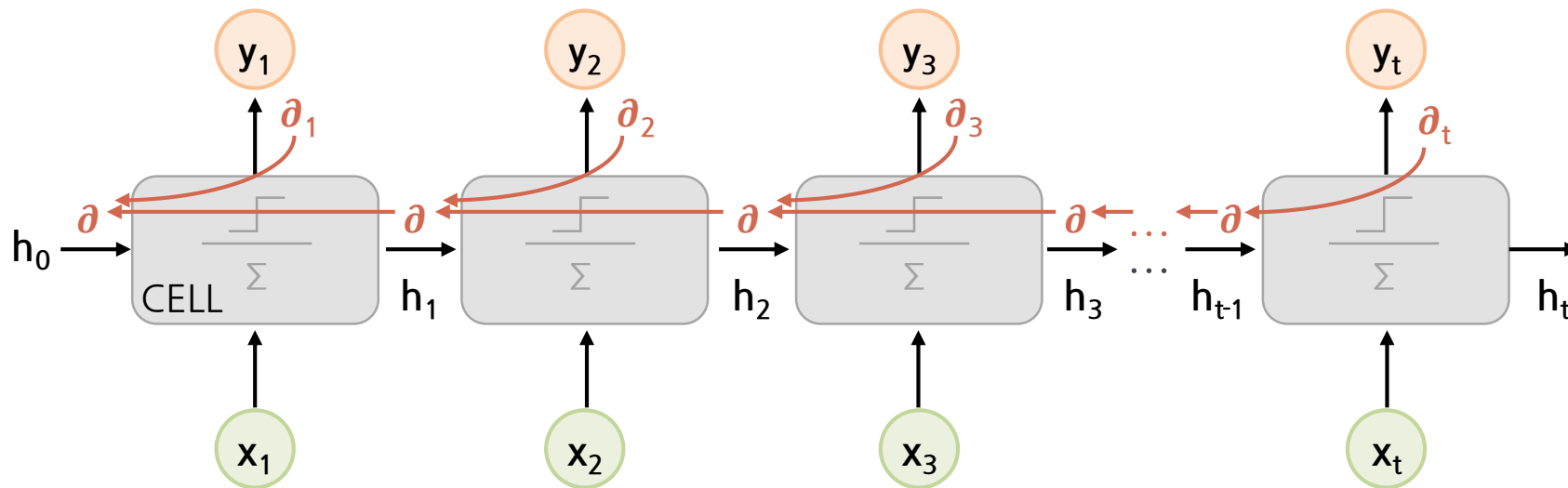
(3주차) Sequence To Sequence

● RNN의 문제점

RNN의 역전파 BPTT와 Gradient Vashing

- RNN에서의 역전파 방법인 BPTT(Back Propagation Through Time)는 Timstep의 마지막 단계에서 시작 단계까지 역전파 진행

역전파 진행 과정에서 Gradient Vanishing 발생 가능



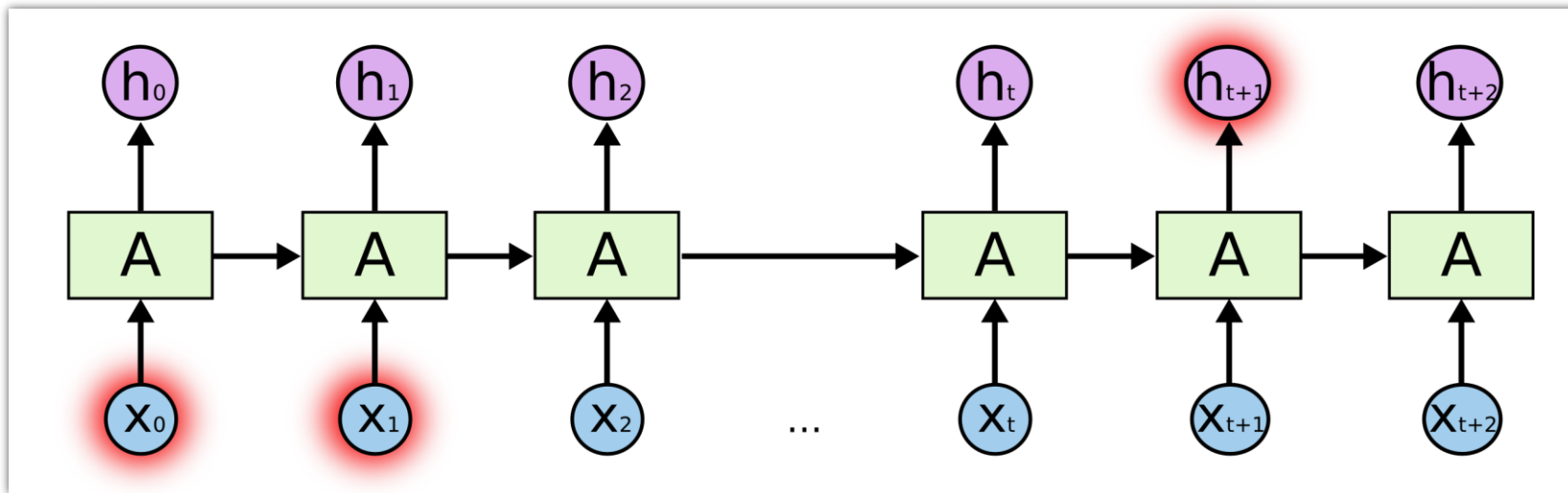
<출처: <https://excelsior-cjh.tistory.com/185>. 20. 9>

Sequence To Sequence

● RNN의 문제점

장기 의존성 문제(The Problem of Long-Term Dependencies)

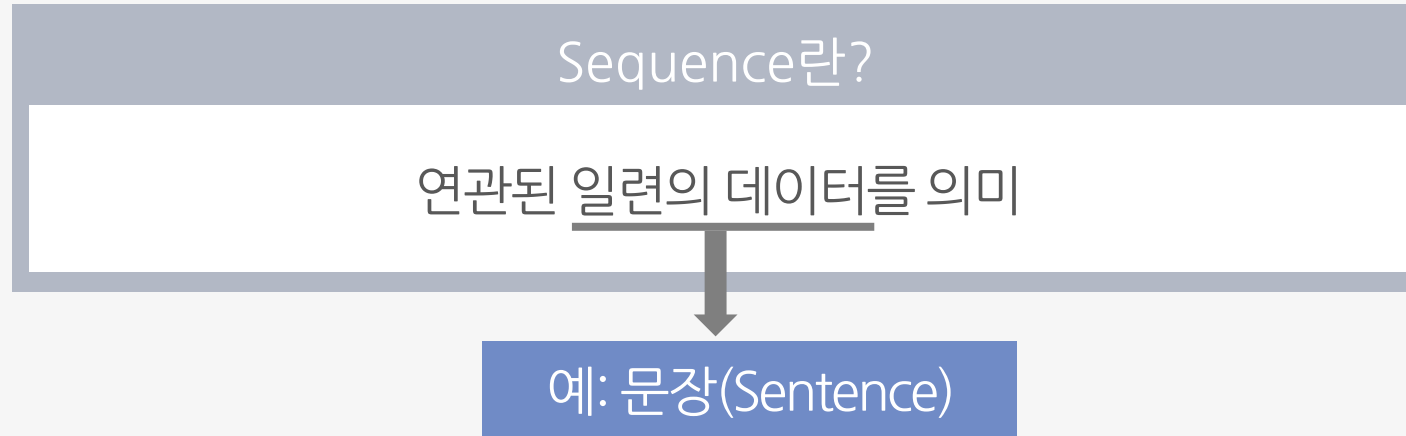
- 관련 정보와 그 정보가 필요한 곳의 거리가 먼 경우, 장기 Memory에 대한 학습능력 저하
- Time step의 크기가 큰 경우, Vanilla RNN은 성능 저하



<출처: <http://colah.github.io>. 20. 9>

Sequence To Sequence

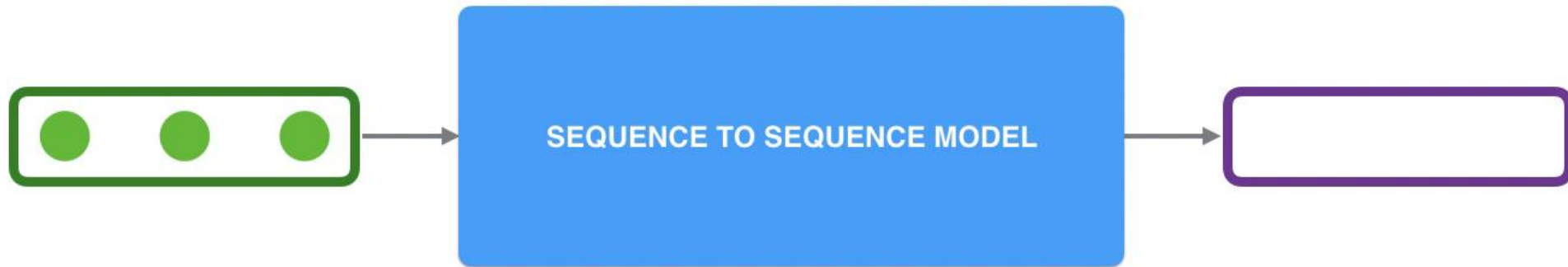
- Seq2Seq 개요



Sequence To Sequence

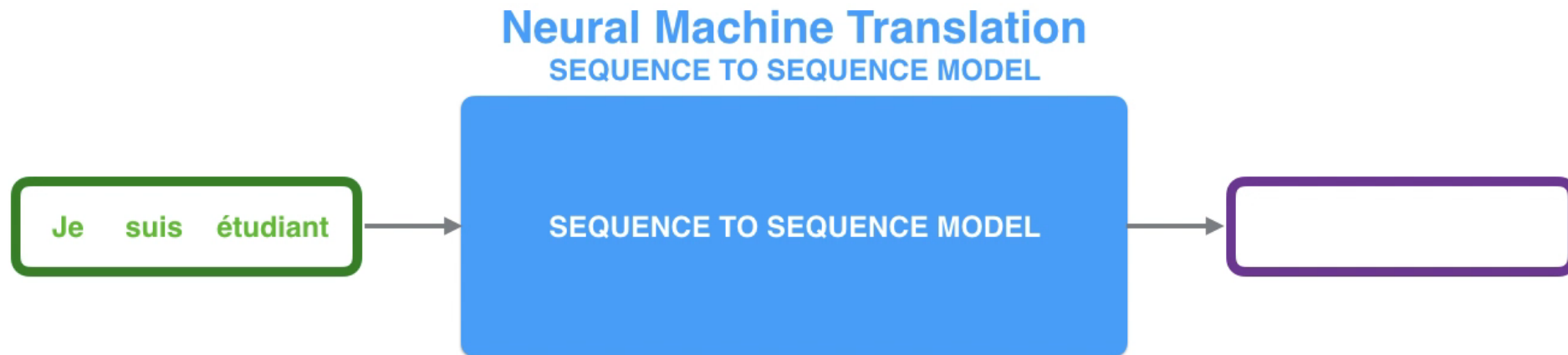
- Seq2Seq 매커니즘

- ▬ Seq2Seq 개요



Sequence To Sequence

- Seq2Seq 매커니즘
 - ▬ Seq2Seq 활용(NMT)



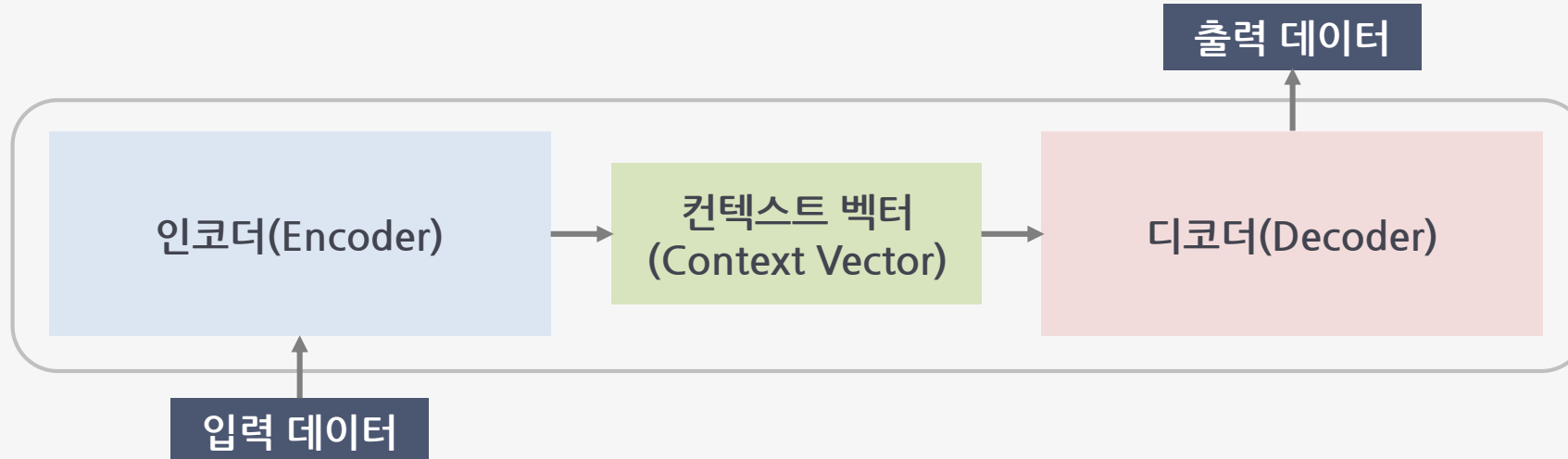
Sequence To Sequence

• Seq2Seq 개요

시퀀스 투 시퀀스(Sequence to Sequence)란?

RNN을 연결하여 하나의 Sequence 데이터를
다른 Sequence 데이터로 변환하는 모델

Seq2Seq는 Input Data의 처리를 위한 Encoder,
Output Data 출력을 위한 Decoder

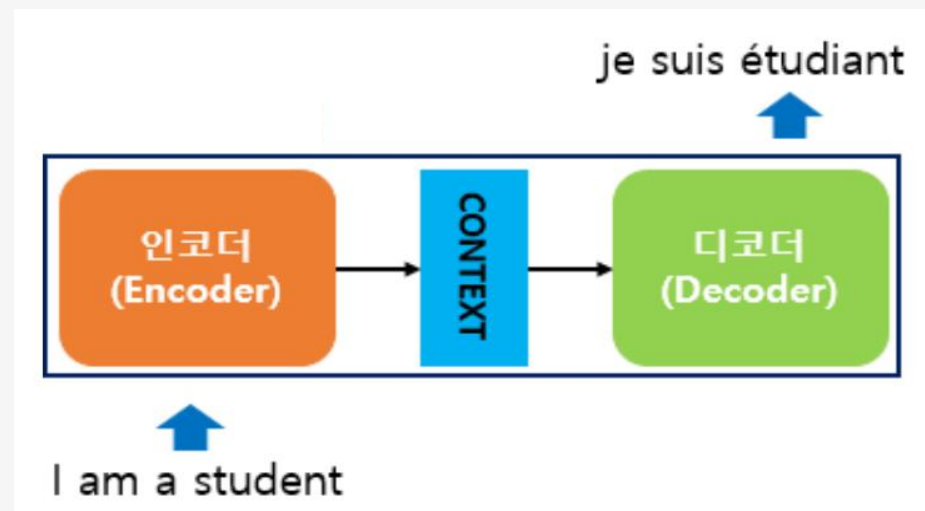
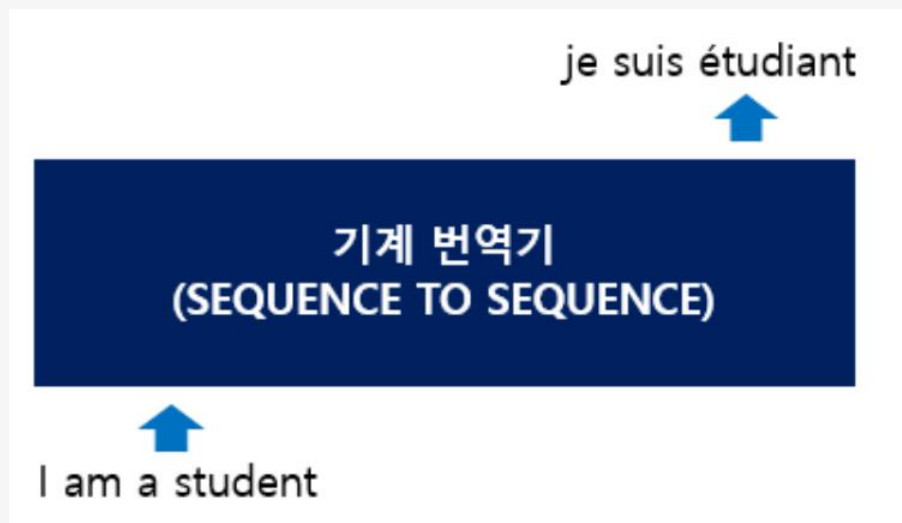


Sequence To Sequence

• Seq2Seq 개요

시퀀스 투 시퀀스(Sequence to Sequence)의 활용

seq2seq 모델 기반의 번역기가 'I am a student'라는 영어 문장을 'je suis étudiant'라는 프랑스 문장으로 번역하여 출력





• Seq2Seq의 구성

인코더(Encoder)란?

순차적으로 입력된 문장의 모든 단어들을 압축하여
컨텍스트 벡터(Context Vector) 생성

컨텍스트 벡터(Context Vector)란?

순차적으로 입력된 문장의 모든 단어들의 정보를 압축한 벡터
잠재 벡터(Latent Vector)라고도 함

디코더(Decoder)란?

입력된 컨텍스트 벡터(Context Vector)를 이용하여
출력 시퀀스 생성 및 출력



• Seq2Seq의 구조

Seq2Seq는 인코더(Encoder)와 디코더(Decoder)로 구성



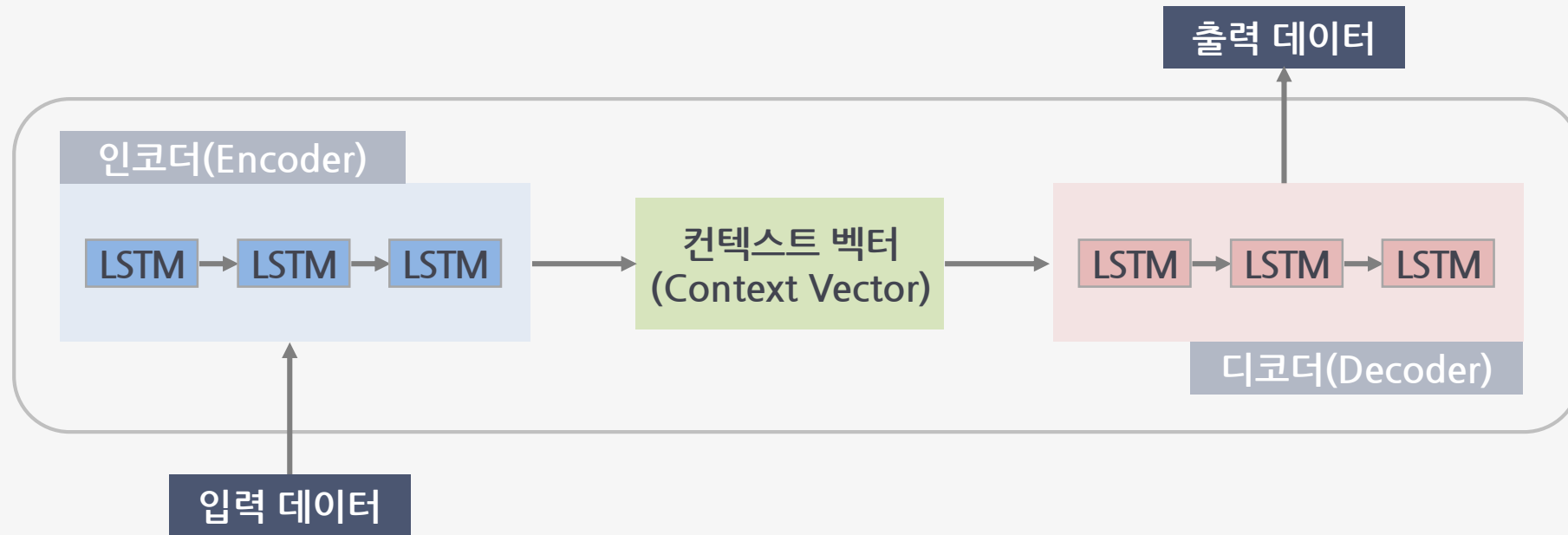
인코더(Encoder)는 여러 개의 LSTM이나 GRU Cell들로 구성

디코더(Decoder)도 여러 개의 LSTM이나 GRU Cell들로 구성

- Decoder는 전형적인 RNNLM(RNN Language Model)

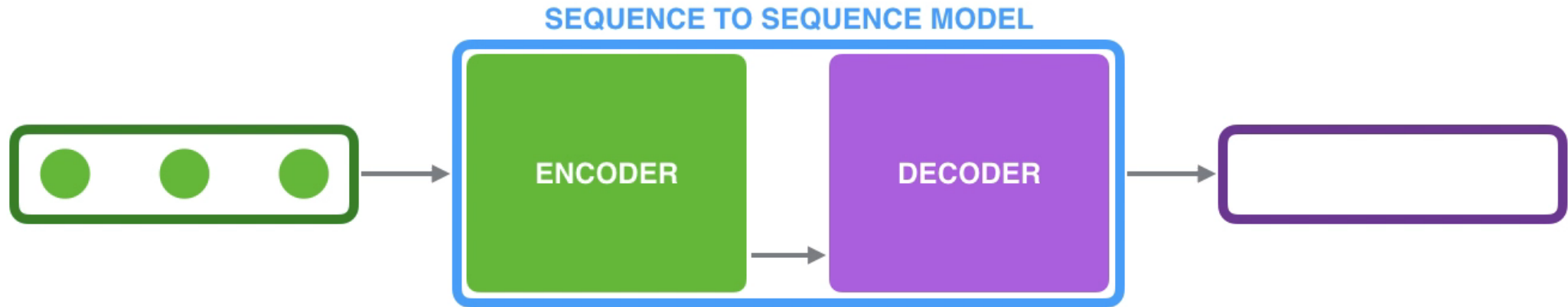
Sequence To Sequence

• Seq2Seq의 구조



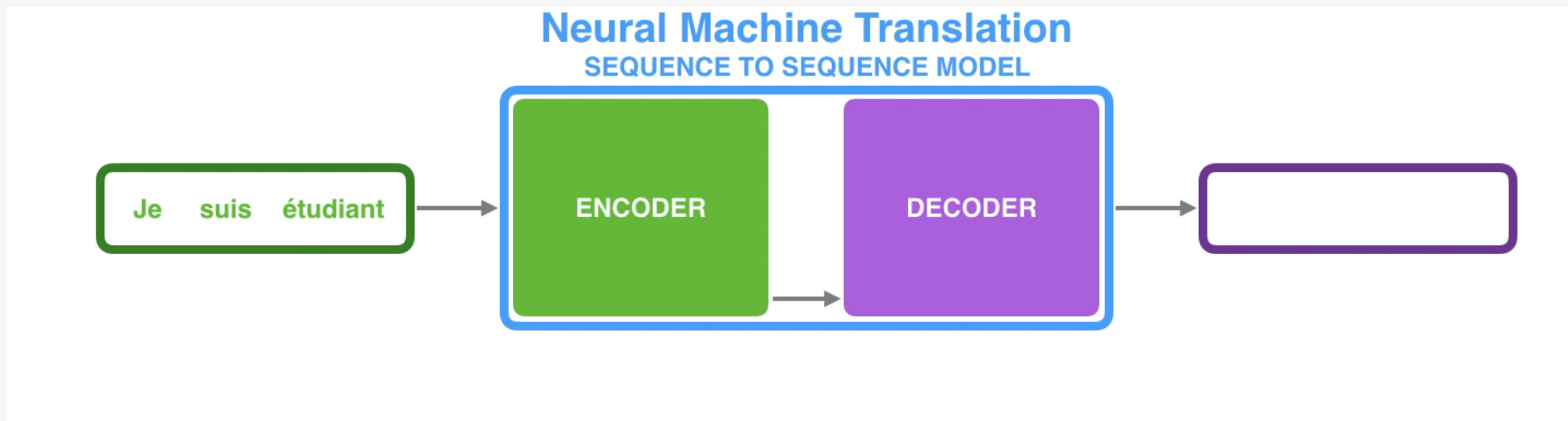
Sequence To Sequence

- Seq2Seq 매커니즘
 - ▬ Context Vector



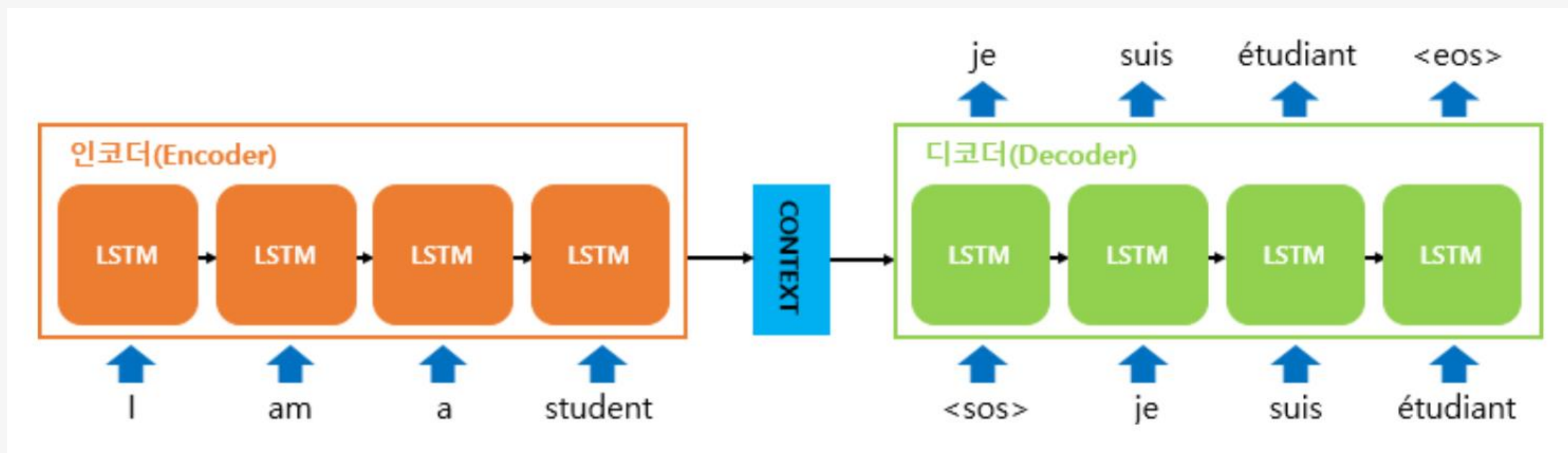
Sequence To Sequence

- Seq2Seq 매커니즘
 - ▬ Context Vector(NMT)



Sequence To Sequence

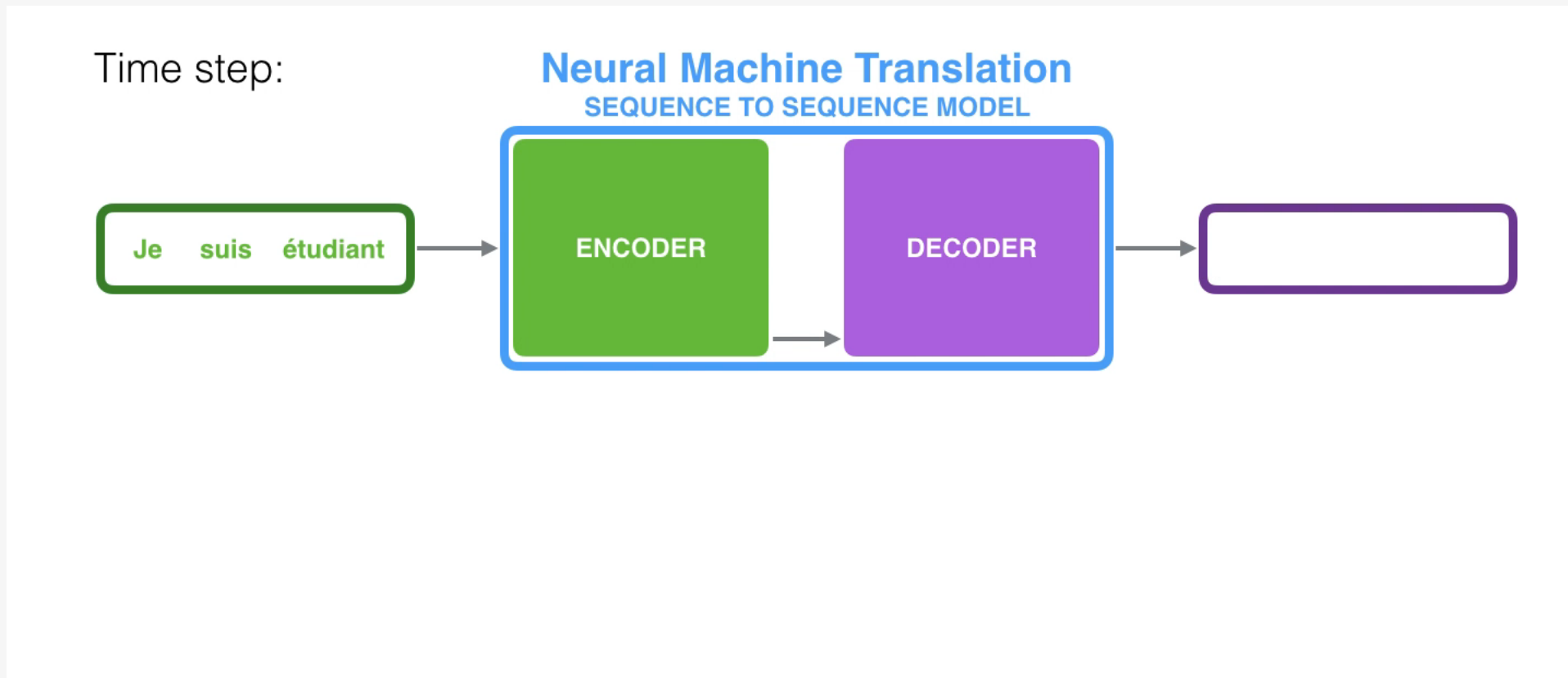
• Seq2Seq의 구조



Sequence To Sequence

- Seq2Seq 매커니즘

- ▬ Encoder와 Decoder



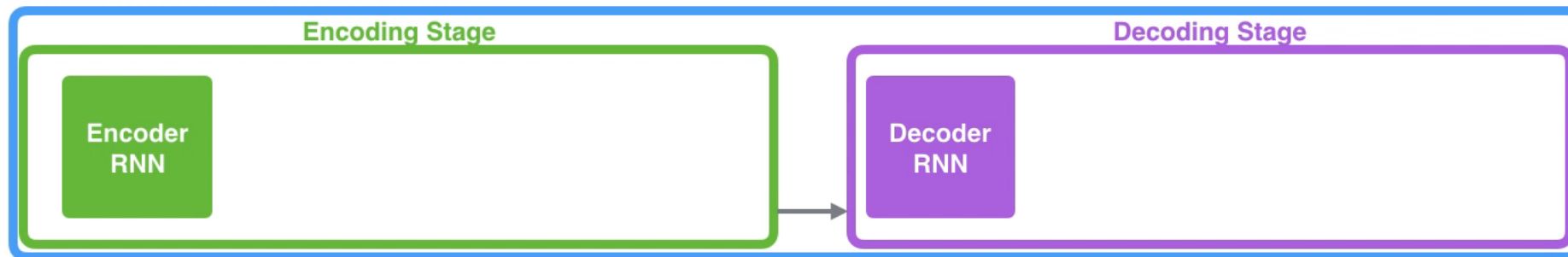
Sequence To Sequence

- Seq2Seq 매커니즘

- ▬ Encoder와 Decoder의 RNN

Neural Machine Translation

SEQUENCE TO SEQUENCE MODEL



Je

suis

étudiant



- Seq2Seq의 구조

- Seq2Seq의 Embedding Layer

입력된 단어들은 워드 임베딩 후 인코더(Encoder)의
각 LSTM에 입력으로 제공

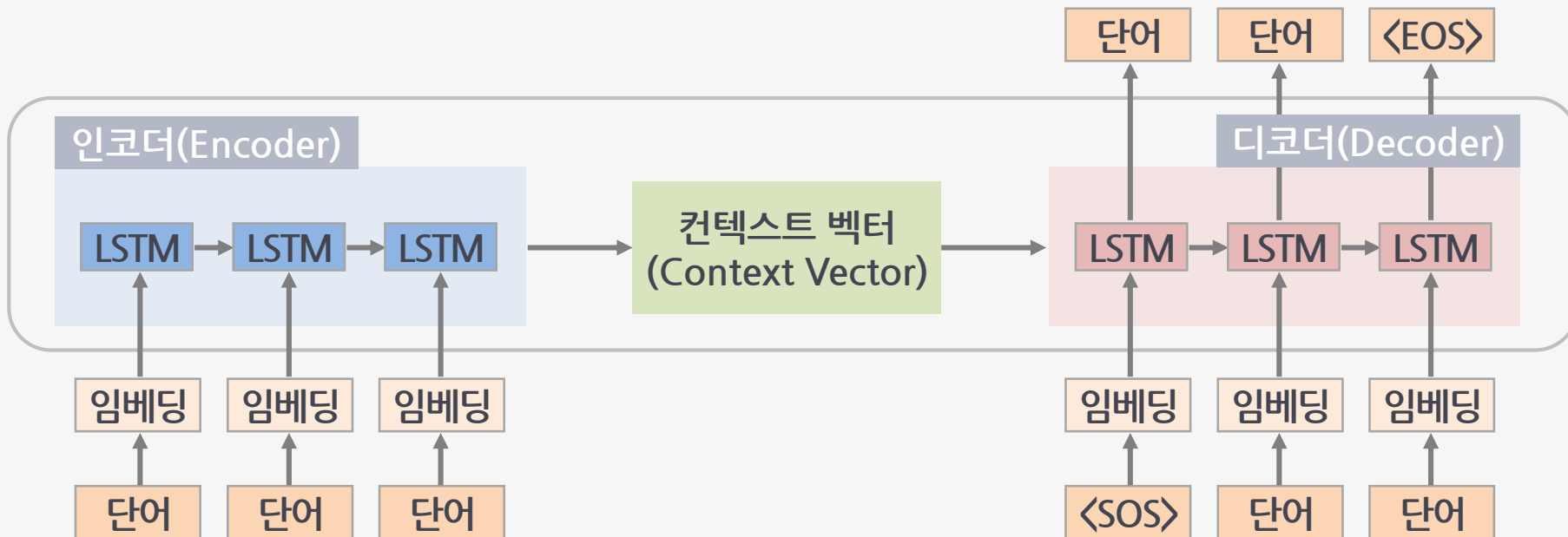


디코더(Decoder)에는 컨텍스트 벡터(Context Vector)와
입력 단어와 매핑된 단어들의 임베딩이 LSTM에 입력으로 제공

Sequence To Sequence

• Seq2Seq의 구조

▬ Seq2Seq의 Embedding Layer



SOS

Start Of Sentence로 Decoder 입력 시퀀스의 시작, 즉 **문장의 시작**을 의미

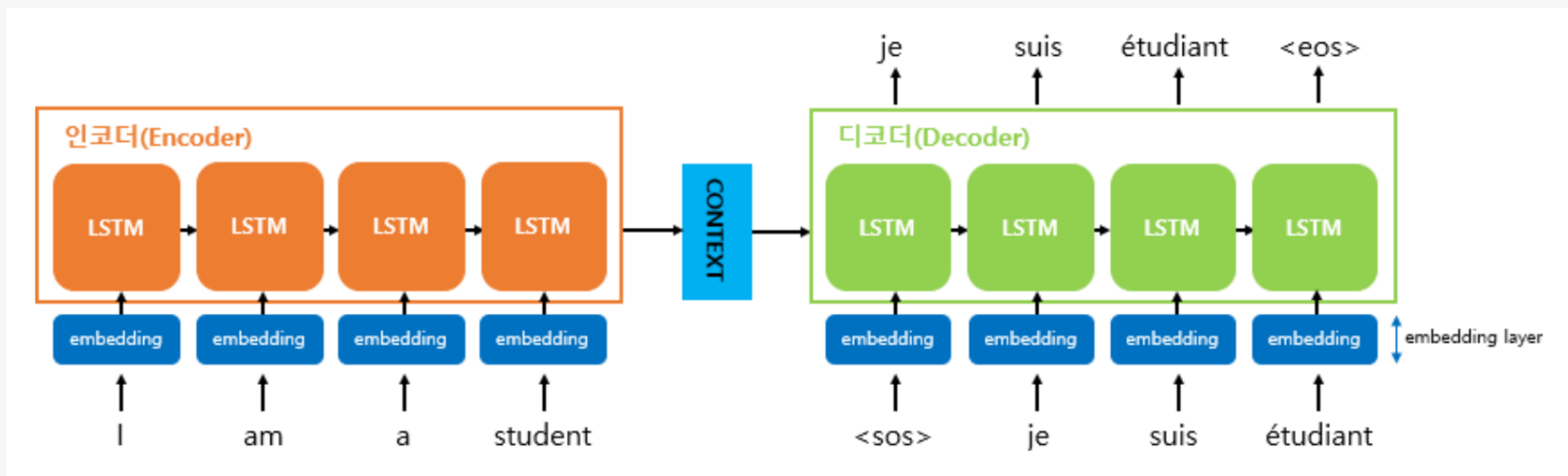
EOS

End Of Sentence로 Decoder 출력 시퀀스의 끝, 즉 **문장의 끝**을 의미

Sequence To Sequence

- Seq2Seq의 구조

- ▬ Seq2Seq의 Embedding Layer



Sequence To Sequence

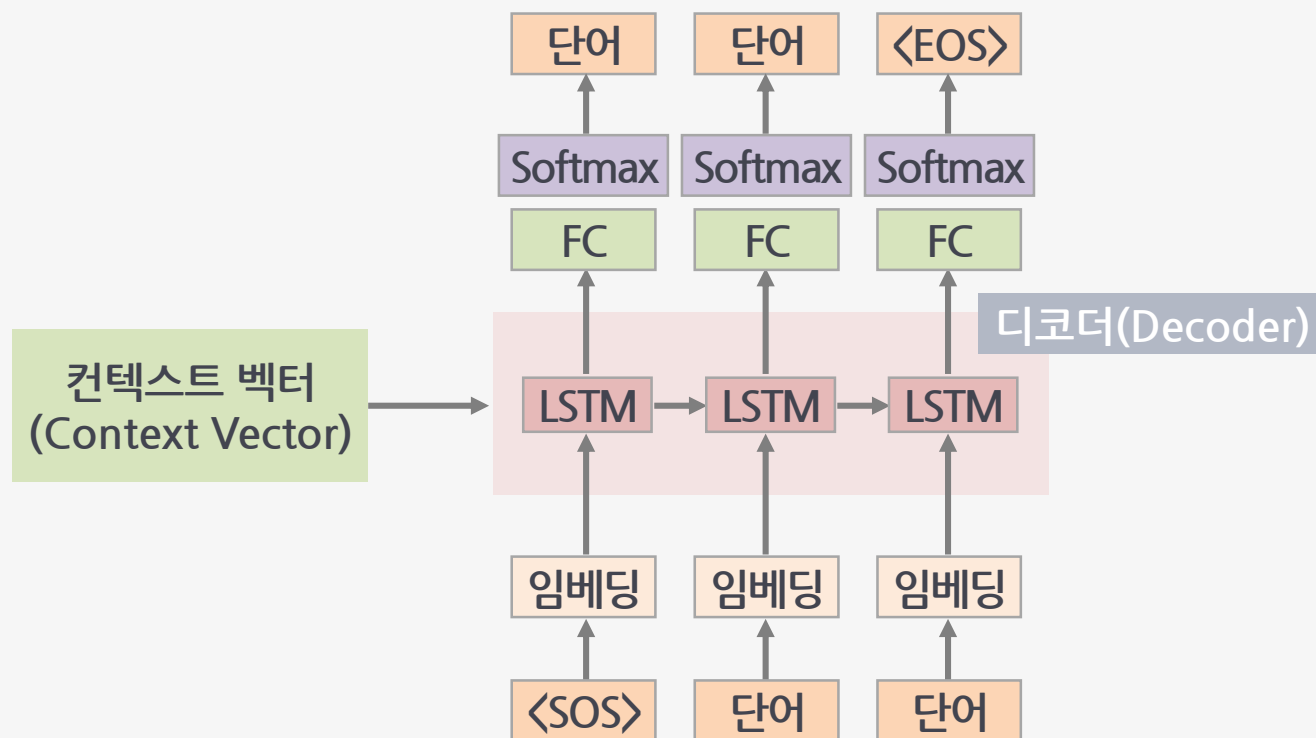
• Seq2Seq의 구조

▬ 디코더(Decoder)의 출력

Fully Connected Layer를 이용하여
Flat한 벡터로 변환



Softmax 함수를 통해 확률적으로
해석할 수 있도록 변환하여 출력



Sequence To Sequence



• Seq2Seq 실습

─ AI Hub: 한국어-영어 번역 말뭉치

ko	en
나는 네가 한국 언어를 배워서 한국어로 대화했으면 좋겠어.	I want you to learn Korean so that we can talk to each other in Korean.
나는 대략 한 시간 정도 이 창고를 돌아 보기를 원합니다.	I want to look around this warehouse for about an hour.
때를 미는 한국의 목욕법을 체험해보고 싶은데 가능할까요?	I want to experience a Korean skin peeling bath, but would it be possible?
우리들 또한 당신들과 함께 한다면 큰 영광일 것입니다.	It will be truly honored for us to work with you.
나는 네가 한국의 감탄사를 사용한다는 말에 놀랐다.	I was surprised to hear that you are using Korean exclamation.
그때 어리숙한 영어실력으로 혹시나 주문을 잘못할까봐 걱정했었다.	At that time, I was worried that I might make a wrong order with my poor English skills.
너랑 나랑 한국 식품 인터넷 사이트 하나 만들자.	Let's make a Korean food website together, you and me.
만약 네가 한국에 오면 막걸리를 많이 사 줄게.	I will buy you a lot of Makgeolli when you come to Korea.
나는 네가 한국에 오는 첫 번째 이유가 나였으면 좋겠어.	I want to be the first reason that you want to come to Korea.
당신이 코멘트한 것처럼 코멘트가 달린 문서를 내게 주십시오.	Give me the document with comments as you have commented.
한국 오면 한국 휴대 전화기 번호는 꼭 가르쳐 주세요.	Please tell me your Korean phone number when you come to Korea.
내가 언제 한 번 악플러들 싸 다 고소할 거야.	I will sue all the online haters someday.
돈을 내고 한복을 대여하여 스티커 사진도 찍으실 수 있습니다.	You can also take sticker pictures when you lend Hanbok.
너는 그럴듯한 이유를 대고 나에게 돈을 요구했었다.	You asked me for money giving me some plausible reasons.
직접 로스팅한 원두로 만든 커피, 브런치를 즐길 수 있다.	You can enjoy house-roasted coffee and brunch in here.
경기 시작 한 시간 전까지는 오셔야 하는데 가능하신가요?	Would it be possible for you to come an hour earlier before the game starts?
이후 그는 한국에 와서 국회 의원으로 활동을 했어요.	After that, he came to Korea and worked as a congressman.
형의 꿈 또한 비행기 조종사라는 것을 알고난 후에, 형과 같이 공부하기로 결정했어요.	After I found out that my brother's dream was also a pilot, I decided to study with him.

3주차_seq2seq.ipynb



● Seq2Seq의 Teacher Forcing

▬ Teacher Forcing 개요

Teacher Forcing 이란?

디코더가 시퀀스를 생성하는 과정에서 모델이 예측한 출력 대신
실제 타겟 시퀀스(정답 데이터)를 다음 입력으로 사용하는 학습 방법

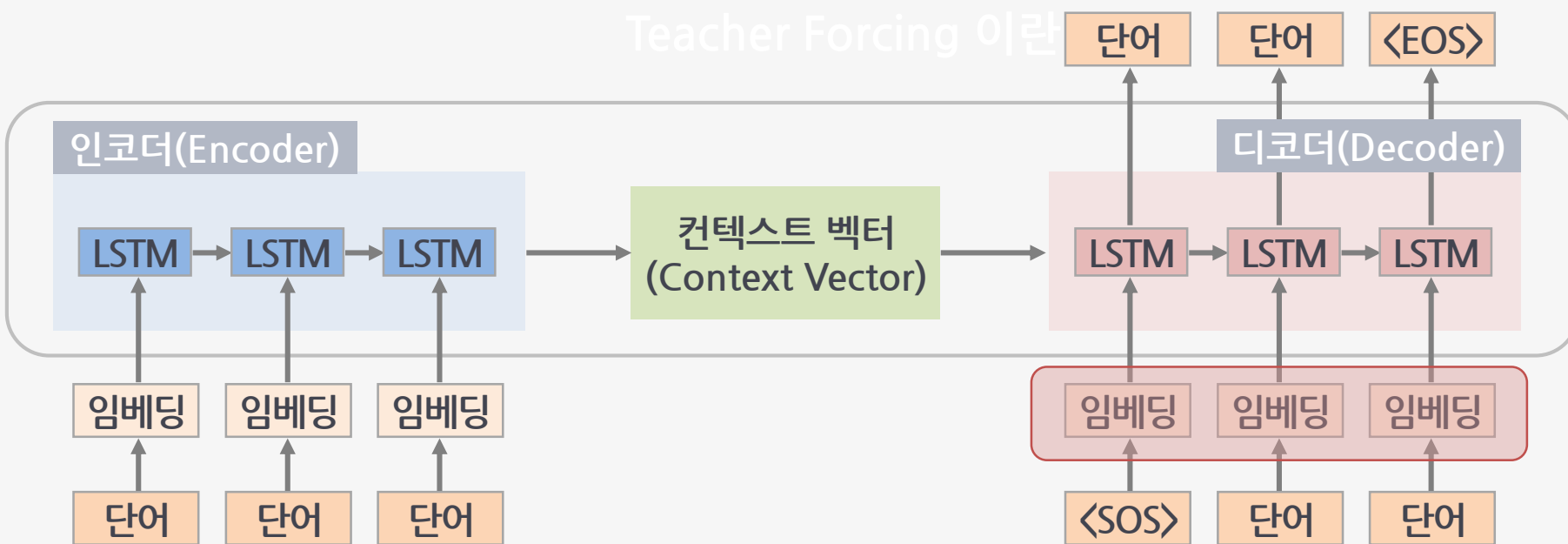


디코더의 학습 초기 잘못된 예측으로 인한 오류를 증폭시키는
문제를 방지하고, 더 빠르고 안정적으로 수렴하도록 하기 위한 기법

Sequence To Sequence

• Seq2Seq의 Teacher Forcing

▬ Teacher Forcing 적용





• Seq2Seq의 Teacher Forcing

▬ Teacher Forcing 적용

1. 인코더:

- 소스 시퀀스(예: 한국어 문장)를 입력받아 히든 상태로 압축
- 히든 상태는 디코더의 초기 상태로 전달

2. 디코더 (Teacher Forcing 적용):

- 디코더는 타겟 시퀀스(예: 영어 문장)를 한 토큰씩 생성
- 일반적으로 디코더는 이전 타임스텝에서 예측한 토큰을 다음 입력으로 사용하지만, Teacher Forcing에서는 실제 타겟 시퀀스의 토큰을 입력으로 제공
- 예: 타겟이 "I am happy"라면, <sos> 입력 후 "I"를 예측하고, 다음 입력으로 모델 예측값 대신 실제 "I"를 사용



● Seq2Seq의 Teacher Forcing

▀ Teacher Forcing 적용

3. 손실 계산:

- 디코더의 출력과 실제 타겟 시퀀스를 비교하여 손실을 계산
- Teacher Forcing을 통해 디코더는 항상 "정답 경로"를 따라 학습

4. 추론 시 (Inference):

- 학습과 달리 실제 예측에서는 Teacher Forcing을 사용할 수 없으므로, 모델이 스스로 예측한 토큰을 다음 입력으로 사용



● Seq2Seq의 Teacher Forcing

▀ Teacher Forcing의 장점

1. 빠른 수렴 : 모델이 정답을 기반으로 학습을 진행하므로 초기 학습이 더 빠르고 안정적
2. 오류 전파 감소: 디코더가 초기에 부정확한 토큰을 예측하면 그 오류가 이후 시퀀스 전체에 영향을 미치는 것을 방지
3. 간단한 구현: 타겟 시퀀스를 그대로 입력으로 사용하므로 추가적인 복잡한 로직 없이 구현 가능



● Seq2Seq의 Teacher Forcing

▀ Teacher Forcing의 단점

1. Exposure Bias: 학습 시 정답을 사용하지만, 추론 시에는 예측값을 사용하므로 학습과 추론 간의 불일치가 발생, 운영환경에서 오류 증폭 가능성
2. 과적합 위험: 정답에 지나치게 의존하면 모델이 스스로 시퀀스를 생성하는 능력 취약해지므로 과적합 우려
3. 현실성 부족: 실세계에서는 타겟 시퀀스가 없으므로, Teacher Forcing에만 의존하면 모델이 비 실용적



● Seq2Seq의 Teacher Forcing

▬ Teacher Forcing의 개선기법

1. Scheduled Sampling: 학습 초반에는 Teacher Forcing 비율을 높게, 후반으로 갈수록 낮춰 Exposure Bias를 축소
예: `teacher_forcing_ratio`를 에포크에 따라 동적으로 조정.
2. Curriculum Learning: 쉬운 시퀀스부터 시작해 점차 어려운 시퀀스로 학습하며 Teacher Forcing 의존도 축소
3. **Attention 메커니즘**: 인코더와 디코더 간 직접적인 연결(Attention)을 추가해 디코더가 스스로 문맥을 파악하도록 개선

Seq2Seq의 한계

● Seq2Seq의 문제점

인코더가 입력 시퀀스를 하나의 벡터로 압축하는 과정에서
입력 시퀀스의 정보가 일부 손실

RNN 구조의 근본적인 문제점

- Vanishing Gradient 발생 가능
- 입력 데이터의 길이가 길어지면 성능이 저하되는 현상 발생 가능

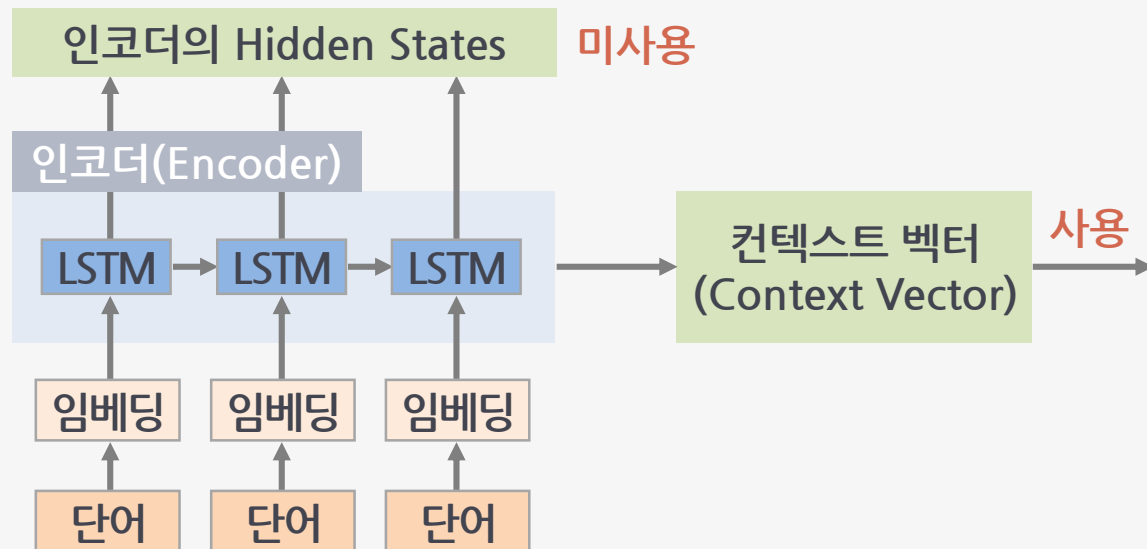


입력 데이터의 길이가 길어지더라도 성능이 저하되는 것을 방지하기 위해
어텐션(Attention) 메커니즘의 등장

Seq2Seq의 한계

• Seq2Seq의 문제점

- 인코더의 Time step마다 출력하는 **Hidden States**는 미사용
- 인코더의 마지막 Hidden State인 **컨텍스트 벡터**만 디코더에서 사용,
어텐션 메커니즘(Attention Mechanism)은 미사용 **Hidden State**를 활용



어텐션의 기본 아이디어는 디코더의 출력 결과를 예측하는
매 시점(Time step)마다, 인코더의 Hidden State를 입력으로 전달하여 참고

● Softmax 함수 개요

Softmax 함수란?

결과 값을 확률적으로 해석할 수 있도록 변환하는 함수

0.01	-0.05	0.1	0.05
0.7	0.2	0.05	0.16
0.0	-0.45	-0.2	0.03

Input

×

-15
22
-44
56

Weight

+

0.0
0.2
-0.3

Bias

=

-2.85
0.86
0.28

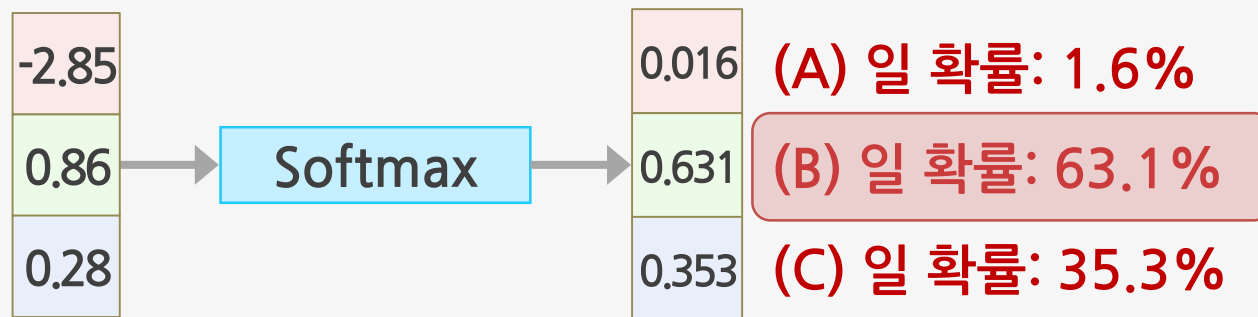
Output

?

● Softmax 함수 개요

Softmax 함수란?

결과 값(Output)을 **확률적으로 해석**할 수 있도록 변환하는 함수



SUM = 1

* Label : A, B, C

• VAE(Variational Auto-Encoder)

