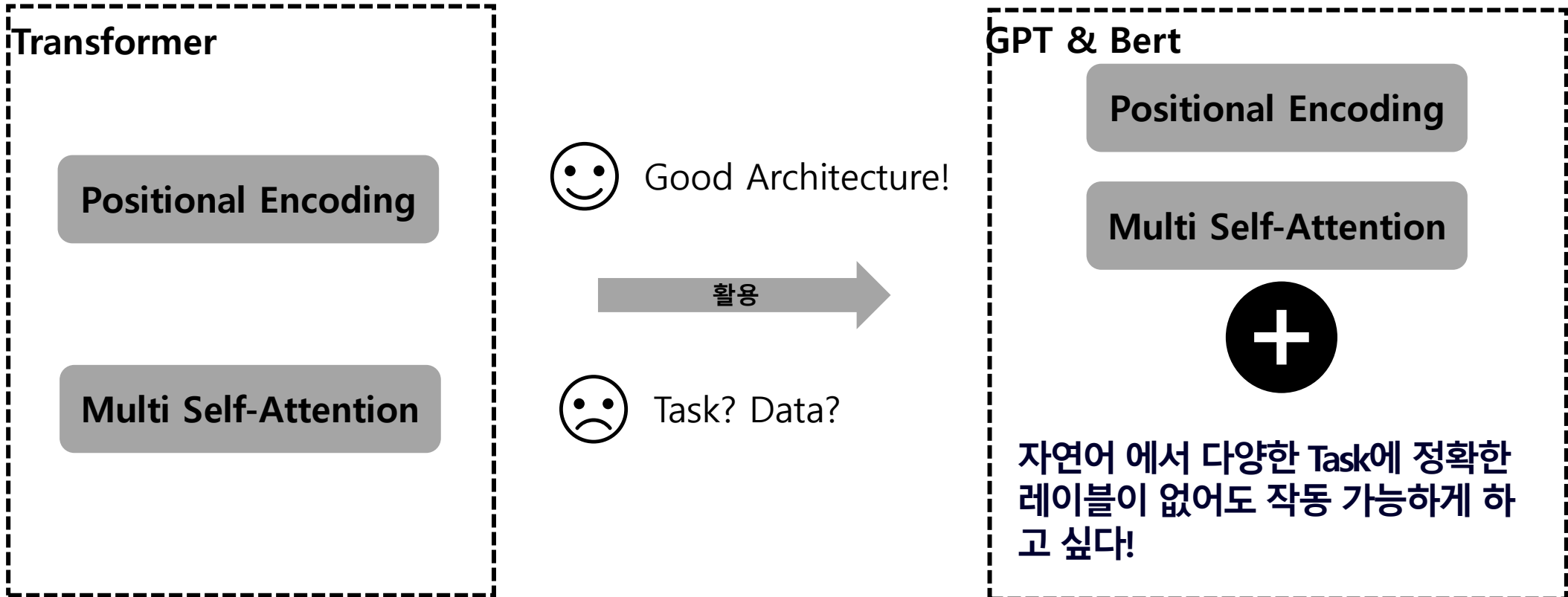


GPT-3 소개와 자연어 처리 기초

Motivation & Direction

Motivation & Direction



Motivation & Direction

Contents

- 자연어 데이터의 특징
- 언어 모델 (Language Model)
- Multi-head self-attention
- Positional Encoding

Motivation & Direction

Contents

- **Transfer learning**
- **Few shot, zero shot learning**
- **Bert**
- **GPT-1, 2, 3**

Motivation & Direction

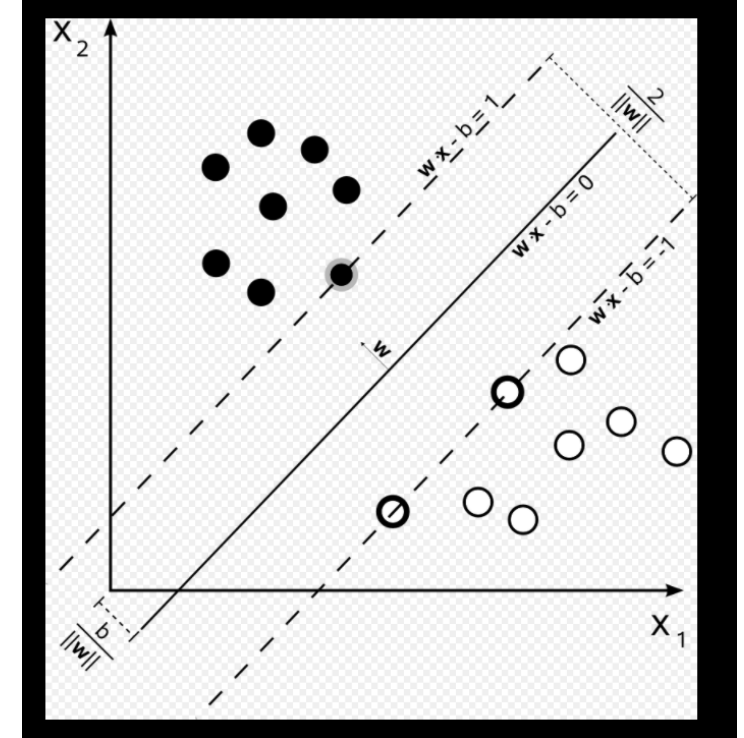
Contents

- **Related works of Transformer**
- **Hugging face 소개 및 실습**

자연어 데이터만의 특징

자연어 데이터의 특징

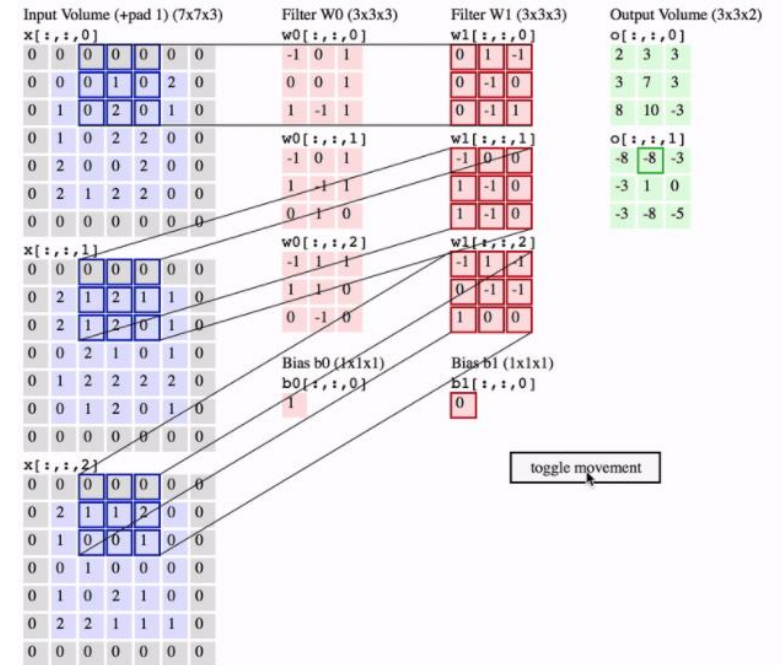
	키	성별	몸무게
Sample 1	181	남	80
Sample 2	165	여	49
Sample 3	181	남	80
Sample 4	165	여	49



Tabular Dataset

- 각각의 데이터 샘플들이 구성하고 있는 features는 독립적

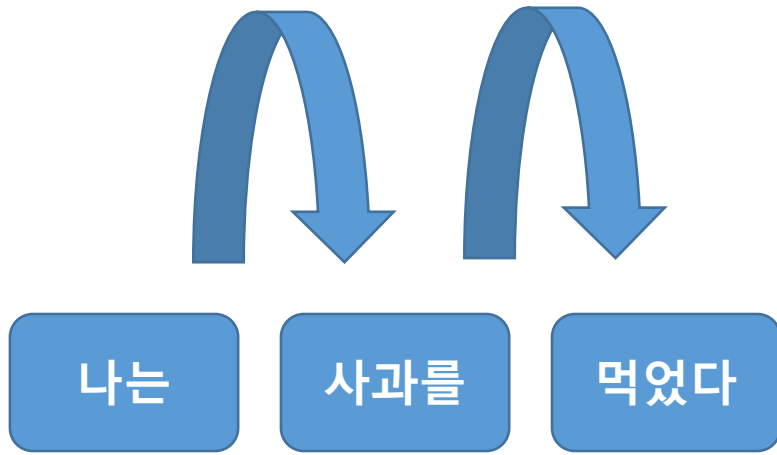
자연어 데이터의 특징



이미지

- 이미지 안의 의미있는 정보는 위치에 무관
- Convolution 연산은 위치에 무관한 지역적인 정보를 추출하기 좋은 연산

자연어 데이터의 특징



반연 자연어 데이터 (문장) 의 경우

- 문장을 구성하고 있는 단어들의 위치가 변해서는 안됨
- 단어들 간의 관계가 중요하고 하나의 단어만 바뀌거나 추가되어도 전혀 다른 의미(context) 를 가질 수 있음

자연어 데이터의 특징

Summary

- GPT를 구성하고 있는 모델은 Transformer에서 가져온 것이다
- 이번 강의에서는 Transformer 의 구성 모듈에 대해 자세히 배울 예정
- Bert, GPT의 경우 특히 자연어 데이터에 특화된 프레임워크이다
- 자연어 데이터는 단어와 단어들 사이의 순서와 관계가 중요하다
- 문장이 갖고 있는 문맥을 알고리즘이 이해할 수 있게 하는 것이 어렵다.

자연어 데이터의 토큰화

자연어 데이터의 토큰화

Character-based, word-based

Character

L	e	t	'	s	d	o	t	o	k	e	n	i	z	a	t	i	o	n	!
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Split on spaces

Let's	do	tokenization!
-------	----	---------------

Split on punctuation

Let	's	do	tokenization	!
-----	----	----	--------------	---

자연어 데이터의 토큰화

Character-based, word-based

- Character
 - A -> 1, B-> 2 , ... : 문장의 시계열 길이가 너무 늘어남
 - 각각의 문자는 의미를 가지고 있지 않음 -> 결국 단어로 표현을 해야함
- Word
 - 경우의 수가 너무 많음
 - 특히, 사전에 없는 단어가 생길 위험

자연어 데이터의 토큰화

Summary

- Hugging face 와 같은 대중적인 공유 라이브러리가 존재하기 전까지 연구자들이 각각 토큰화를 임의로 진행 했음
- 이에 따라서 모델을 다운받아 실행시킬 경우 전혀 다른 결과가 나타남
- 즉, 자연어 에서의 토큰처리 과정에 대해 아는 것은 매우 중요함

자연어처리 Task

자연어 처리 Task

언어 모델 (Language Model)

- 하나의 문장을 여러 개로 나누고 나눈 토큰들의 결합분포로 문장에 대해서 확률을 계산

나는

사과를

먹었다

w_1

w_2

w_3

$$P(W) = P(w_1, w_2, w_3)$$

자연어 처리 Task

언어 모델 (Language Model)

- 문장 감정 분석 : 맛있는 사과를 먹었다.

$$P(w_1, w_2, w_3)$$

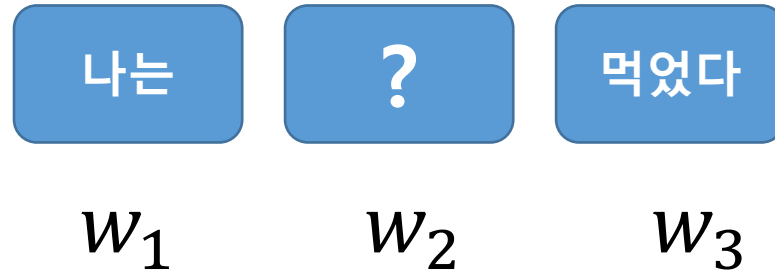
- 기계어 번역 & Question answering

나는 사과를 먹었다 -> I ate an apple

$$P(w'_1, w'_2, w'_3 | w_1, w_2, w_3)$$

자연어 처리 Task

언어 모델 (Language Model)



$w = \text{사과 일 경우}, P(w | w_1, w_3) \gg 0$

$w = \text{식탁 일 경우}, P(w | w_1, w_3) \approx 0$

메트릭

Text metrics

- BLEU (Bilingual evaluation understudy)
 - Based on n-gram based precision
 - A measure of fluency rather than semantic similarity between two sentences
- Rouge (Recall Oriented Understudy of Gisting Evaluation)
- METEOR (Metric for Evaluation of Translation with Explicit ORdering)

Text metrics

- Human based measures
 - Measuring quality of a single best result
 - Success@k = % image sentence pairs for which at least one relevant result is found in the top-k list.
 - R-precision = average % of relevant items in the top-k list

자연어 처리 모델 기존 연구들

기존 연구

기존연구

나는

사과를

좋아한다

나는

사과를

안

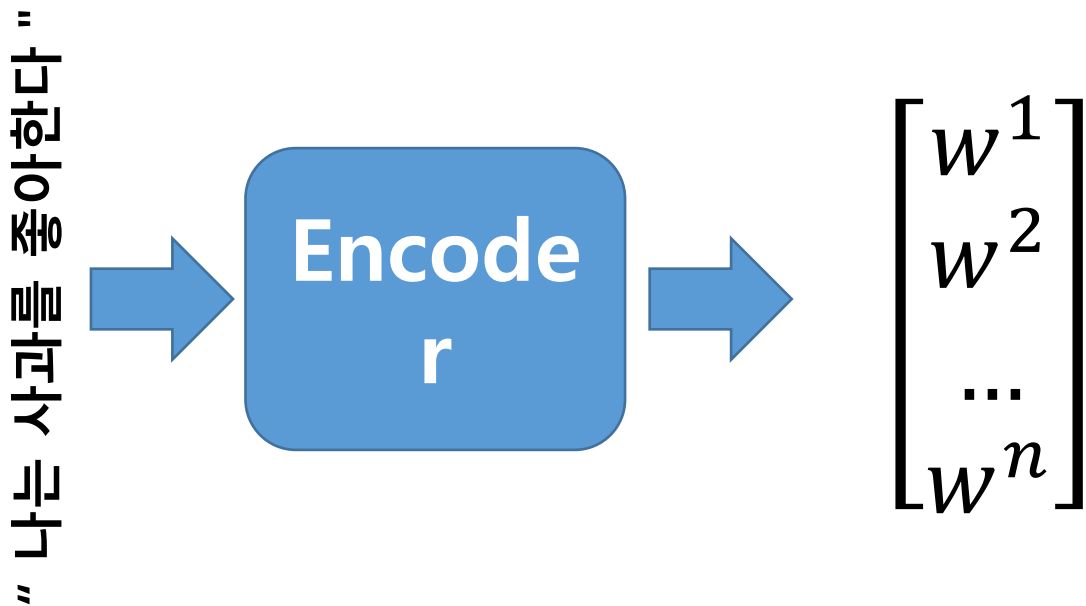
좋아한다

어떠한 한 문장을 분류 하거나 혹은 이 문장을 내포하고 있는 **어떤 벡터로** 표현하고 싶다 !

기존의 이미지나 Tabular 데이터와는 **다르게 각각의 순서가** 매우 중요하다

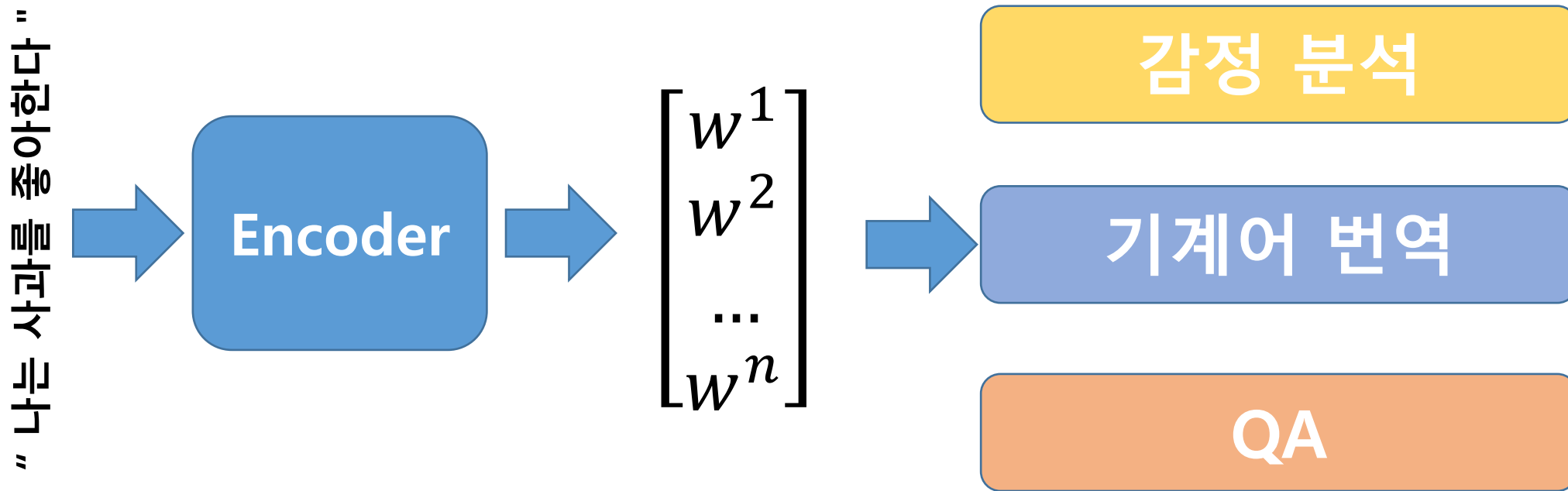
기존연구

- 문장의 문맥을 하나의 벡터로 표현하기 위해서 인코딩



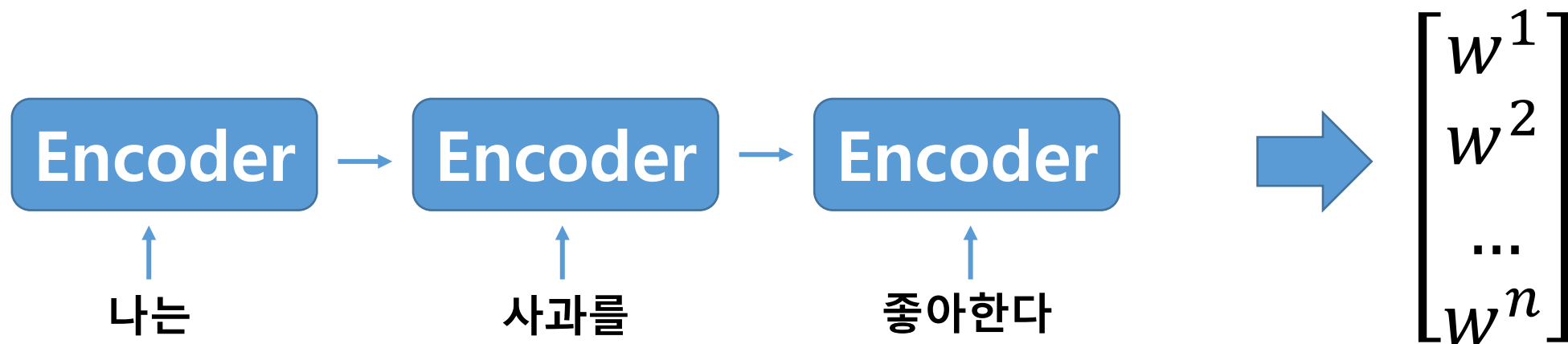
기존연구

- 문맥을 알아내는 Encoder를 학습했다면 다양한 Task에 적용이 가능



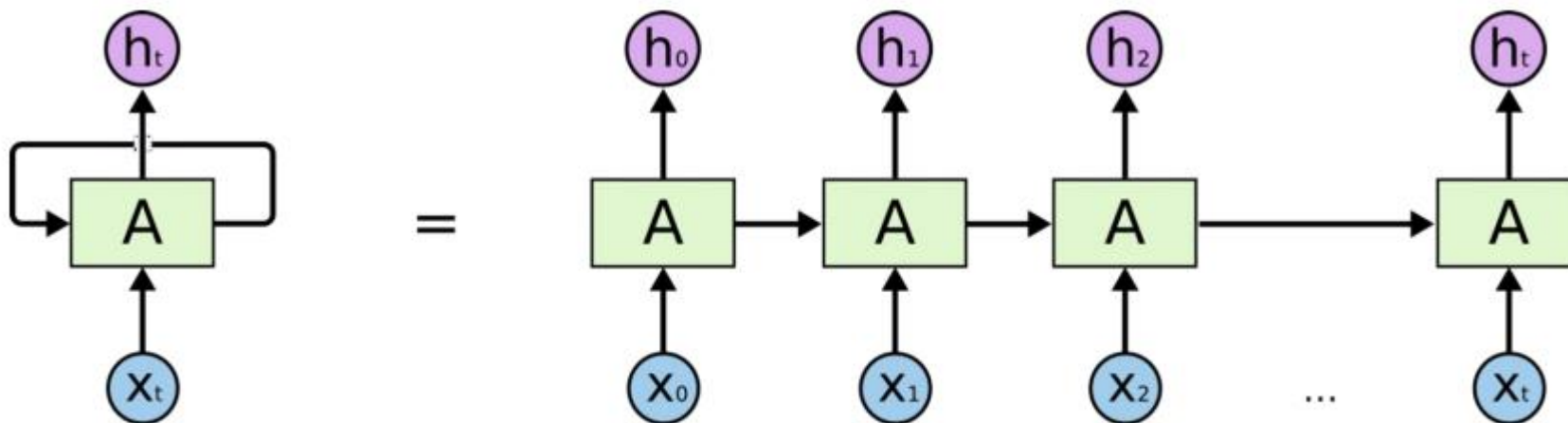
기존연구

! 문장은 순서를 가지고 있으니, 문장의 처음부터 끝까지
순서대로 입력을 받아서 최종적으로 벡터를 생성하자

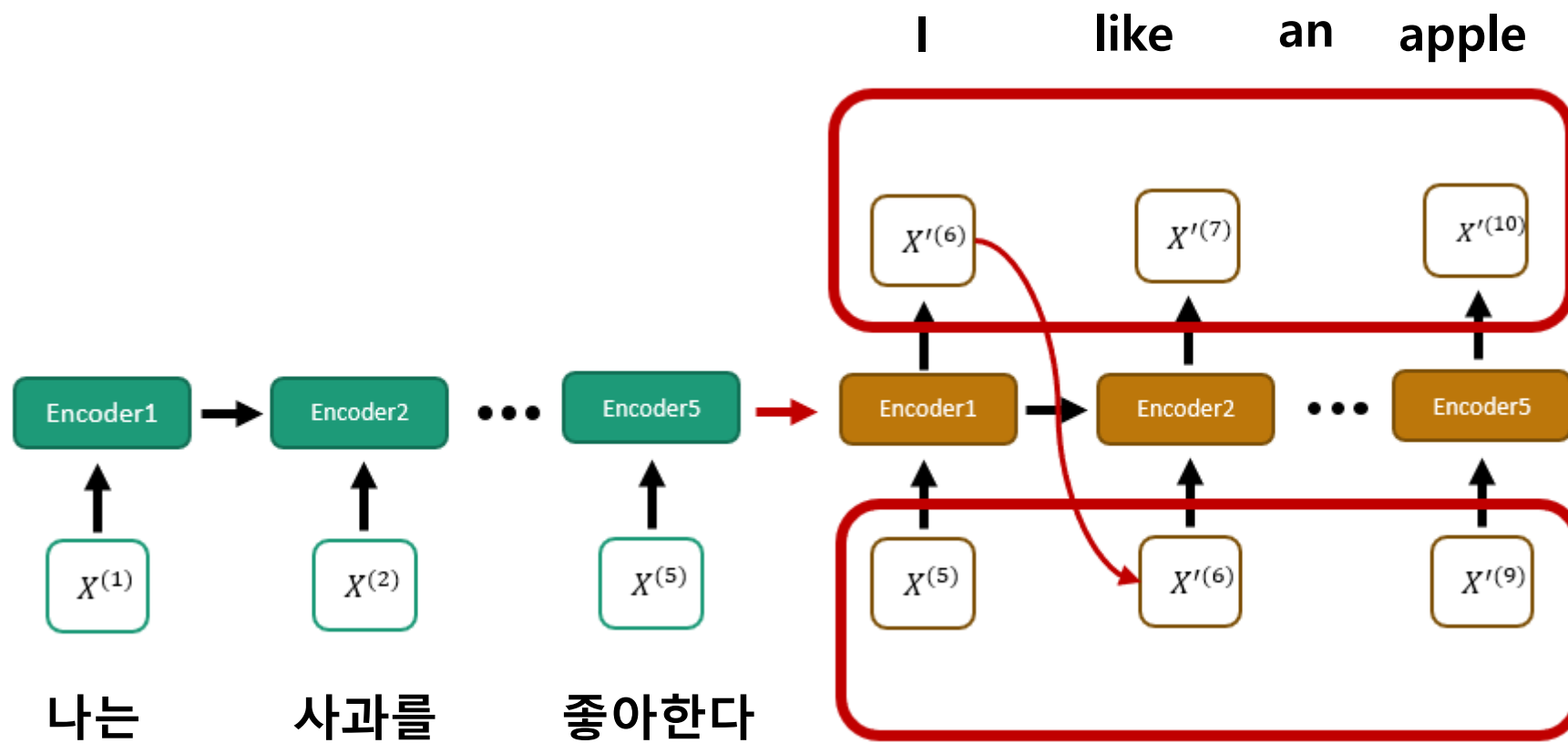


Recurrent neural networks (RNNs)

- RNNs are a directed network(graph) along sequential data.



기존연구

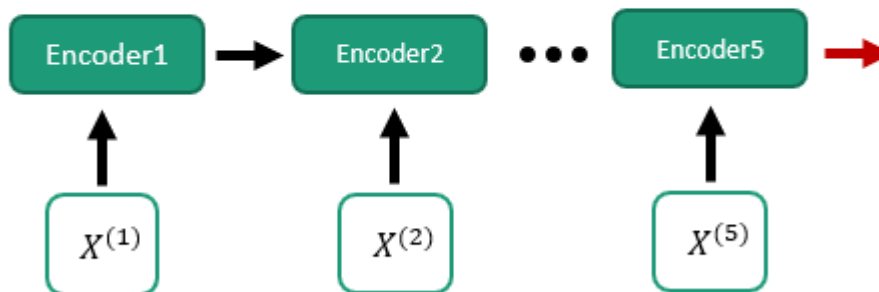


기존연구

?

문장은 단어들이 순서를 가지고 있는 것을 알겠다.

하지만 중요한 단어들이 처음 부분에 있다면 정보를 잃어 버리지 않을까?



Question Answering 영어 듣기

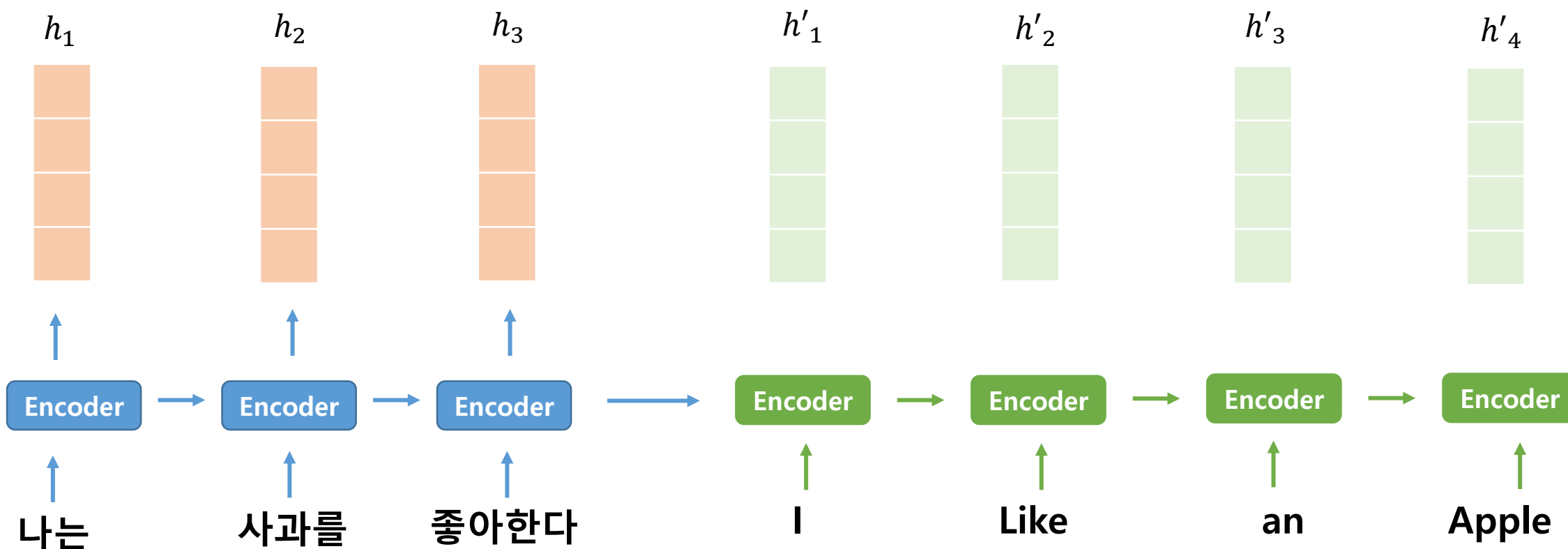
안녕하세요, Westview County 지역 청취자 여러분! 제 이름은 Alicia Jones이고, 78세의 Westview County 주민입니다. 그리고 그거 아세요? 저는 매일 아침 제 소셜 미디어 계정에 접속해서 전 세계 사람들과 최신의 소식을 공유합니다. 저와 같은 나이의 누군가에게 그것은 놀라운 일이죠, 맞죠? 저는 커뮤니티 센터에서 노년층을 위해 제공하는 통신 기술 강좌에서 이렇게 하는 것을 배웠습니다.

Q) 말하고 있는 사람의 나이는?

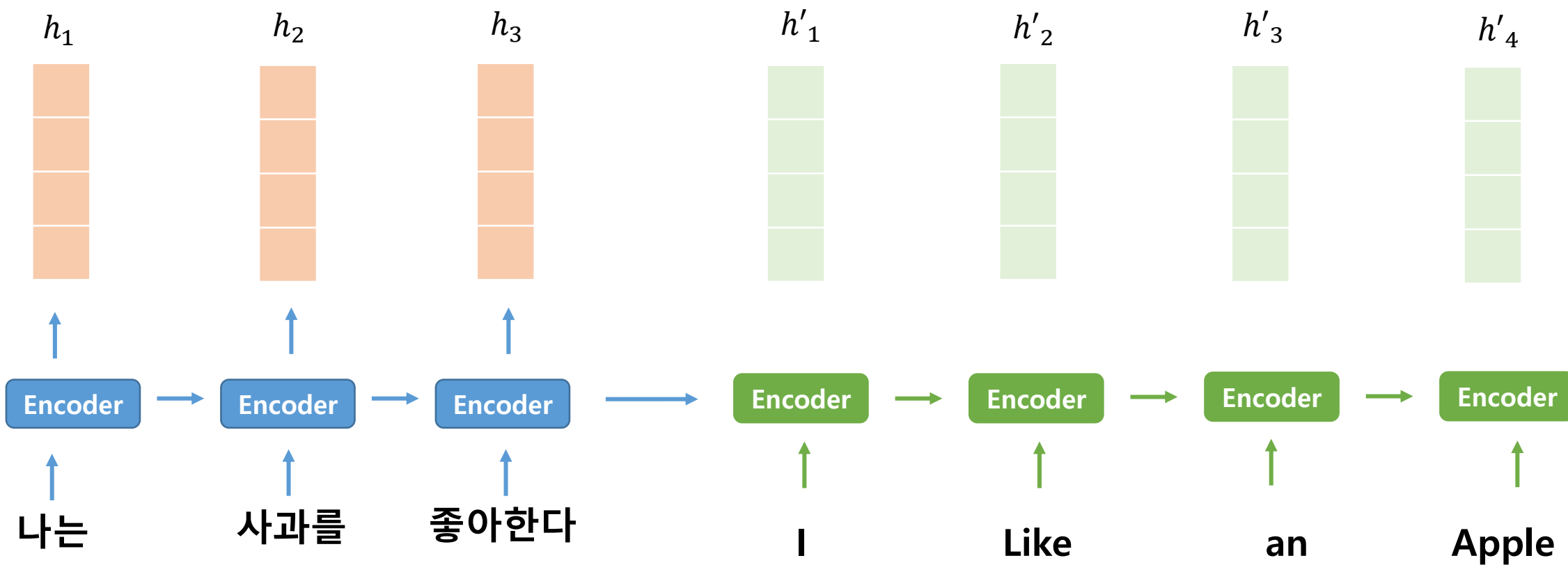
A) 78세

기존연구

! 각각의 단어에 대해서 Attention을 줄 수 있다면 좋겠다.



기존연구



"I" 를 번역할 때에는 "나는"

"Apple" 을 번역할 때에는 "Apple"

기존연구

attention score(이전 문장의 단어가 임베딩된 벡터 h_i , 번역할 단어의 임베딩 벡터 h'_j)

$$\text{score}(h'_j) = [h_1^T h'_j, h_2^T h'_j, \dots, h_N^T h'_j]$$

그 후 *Softmax* 함수를 취한다.

Softmax(*score*(h'_j)) = 단어 임베딩 벡터 h'_j 가 어떤 단어와 가장 유사한지 를 확률적으로 출력

$$\text{Softmax} \quad p_i = \frac{e^{z_i}}{\sum_{j=1}^k e^{z_j}} \quad \text{for } i = 1, 2, \dots, k$$

기존연구

자연어 데이터의 순차적인 특성을 고려하는 기존 연구들
문장마다의 중요도를 계산하여 attention 모듈을 생각해 냄

선형대수 기초


선형대수 기초

? $h_1^T h_j'$ 이 수식이 왜 유사도와 연관이 있을까?

내적 공간

문A 27개 언어 ▾

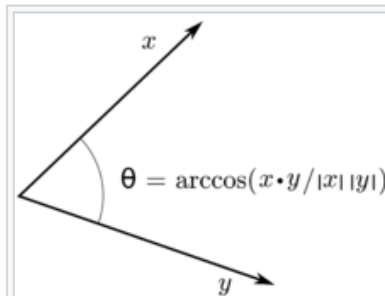
위키백과, 우리 모두의 백과사전.


 내적은 [여기로](#) 연결됩니다. 유클리드 공간 위 내적에 대해서는 [스칼라곱](#) 문서를 참고하십시오.

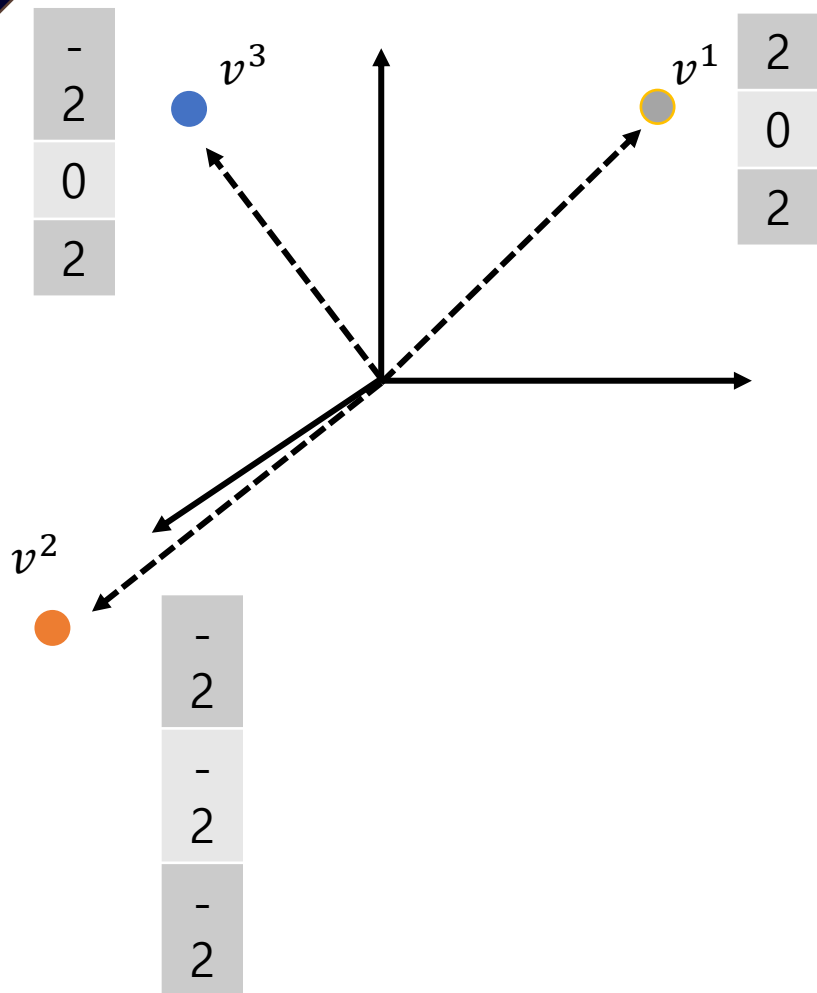
선형대수학과 함수해석학에서, **내적 공간**(內積空間, 영어: inner product space)은 두 벡터의 쌍에 스칼라를 대응시키는 일종의 함수가 주어진 **벡터 공간**이다. 내적 공간 위에서는 벡터의 **길이**나 **각도** 등의 개념을 다룰 수 있다. 스칼라곱을 갖춘 유클리드 공간의 일반화이다.

목차 [숨기기]

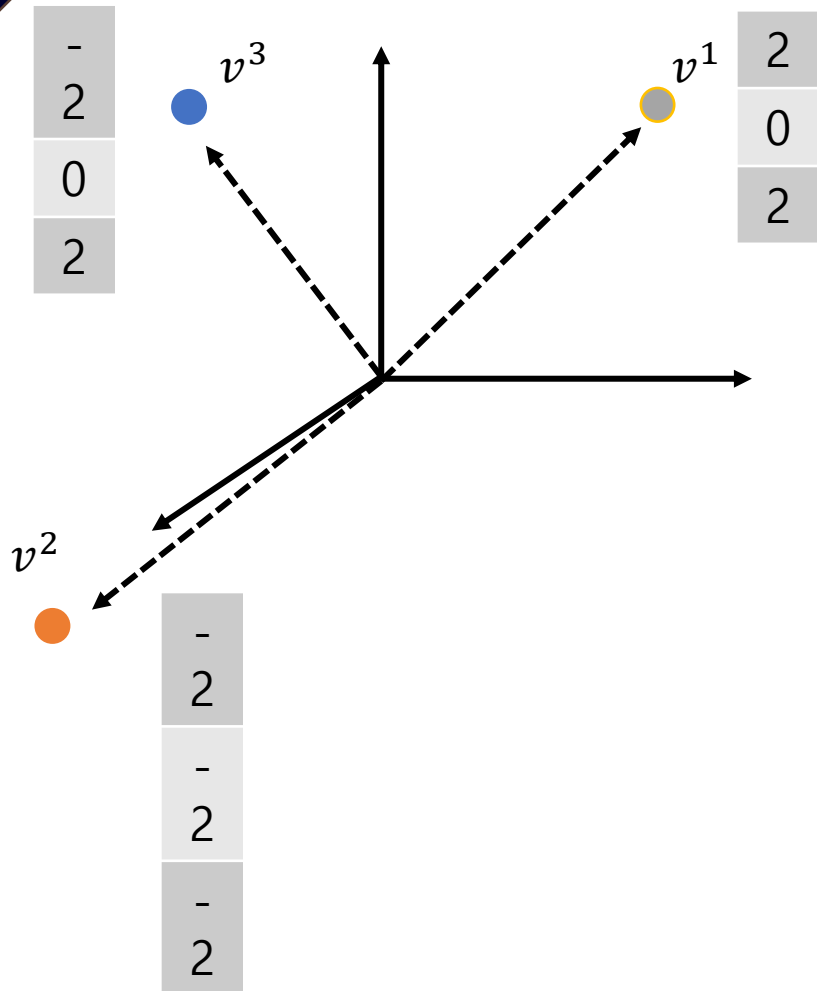
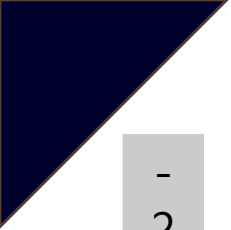
- 1 정의
- 2 성질
 - 2.1 극화 항등식
 - 2.2 코시-슈바르츠 부등식
 - 2.3 정규 직교 기저
 - 2.4 선형 범함수
- 3 예
 - 3.1 유클리드 공간
 - 3.2 함수 공간
- 4 같이 보기
- 5 참고 문헌
- 6 외부 링크



내적을 사용하여 정의한, 두 벡터 사  이의 각도의 기하학적 해석



1. Cos 유사도와 벡터의 관계
2. 벡터의 내적과 거리의 관계, KL divergence



내적 -> 행렬

1. 행렬과 벡터의 관계

내적 -> 행렬

Softmax

$$p_i = \frac{e^{z_i}}{\sum_{j=1}^k e^{z_j}} \text{ for } i = 1, 2, \dots, k$$



Thank you