

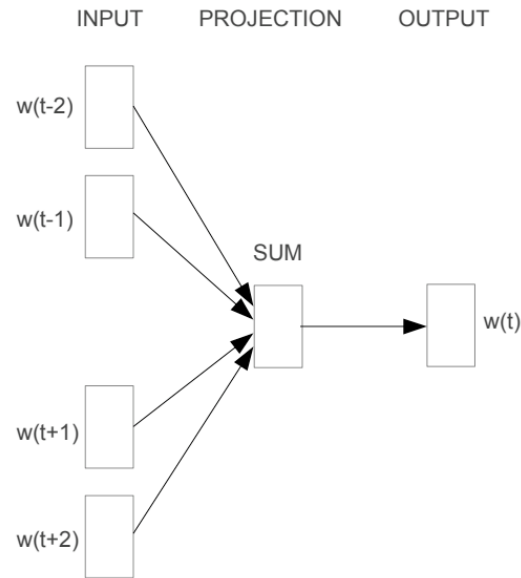
Word2Vec

Ki Hyun Kim

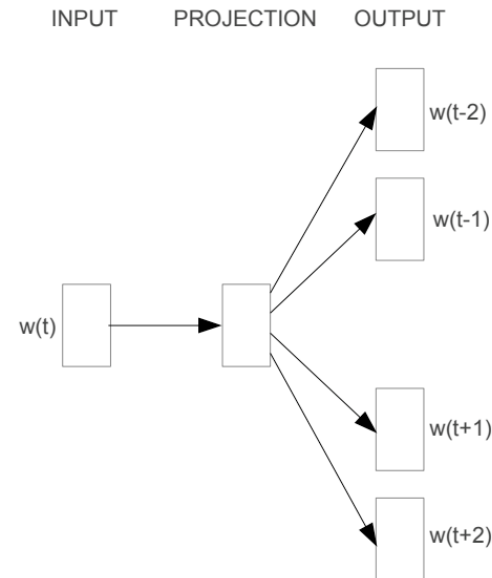
nlp.with.deep.learning@gmail.com

Word2Vec [Mikolov et al. 2013]

- Objective:
 - 주변(context window)에 **같은 단어가 나타나는 단어일 수록** 비슷한 벡터 값을 가져야 한다.
- 문장의 문맥에 따라 정해지는 것이 아님
 - context window의 사이즈에 따라 embedding의 성격이 바뀔 수 있다.



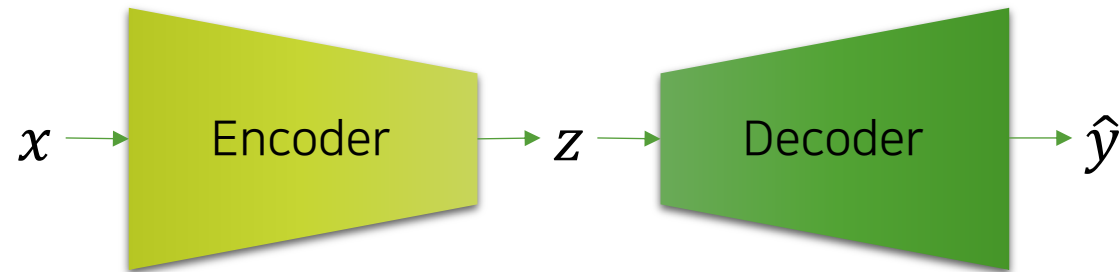
CBOW



Skip-gram

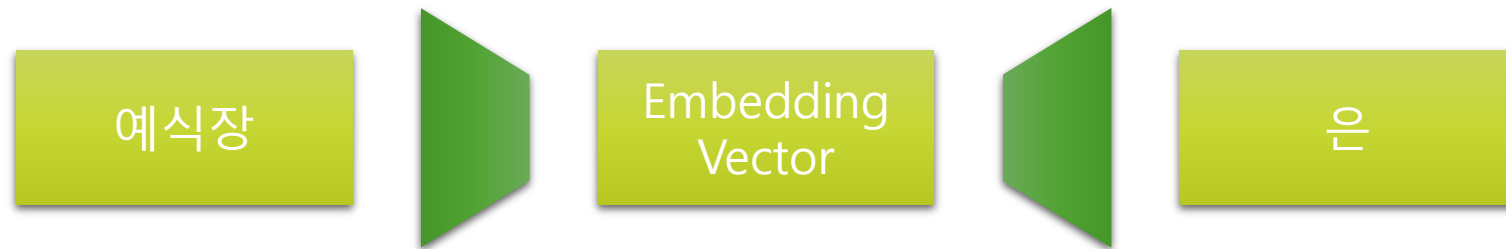
Skip-gram: Basic Concept

- 기본 전략
 - 주변 단어를 예측하도록 하는 과정에서 적절한 단어의 임베딩(정보의 압축)을 할 수 있다.
 - Non-linear activation func.이 없음
- 기본적인 개념은 오토인코더와 굉장히 비슷함
 - y 를 성공적으로 예측하기 위해 필요한 정보를 선택/압축



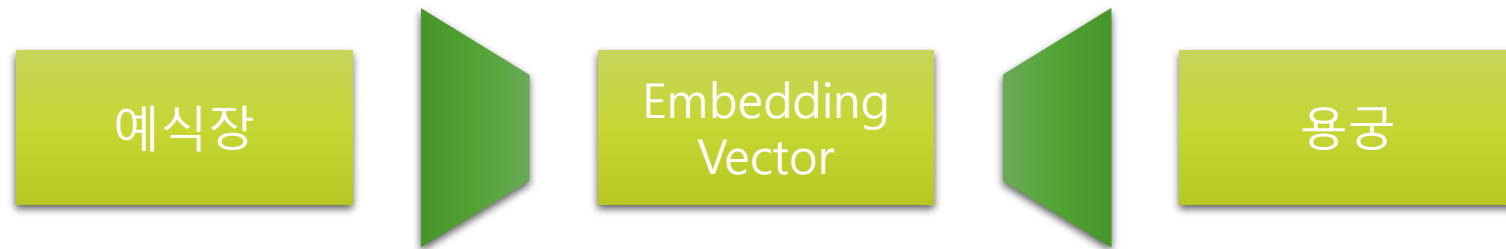
Skip-gram Example ($|W|=5$)

예식장	은	용궁	예식장	주례	는	문어	아저씨	.	
피아노	는	오징어	예물	은	조개껍데기	.			



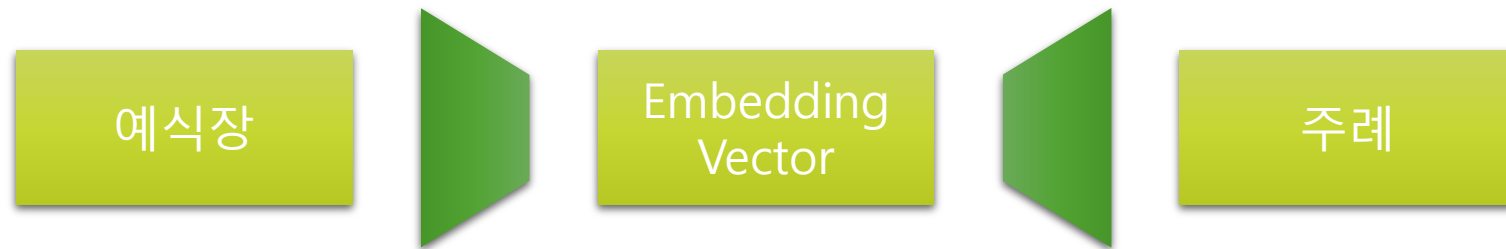
Skip-gram Example ($|W|=5$)

예식장	은	용궁	예식장	주례	는	문어	아저씨	.	
피아노	는	오징어	예물	은	조개껍데기	.			



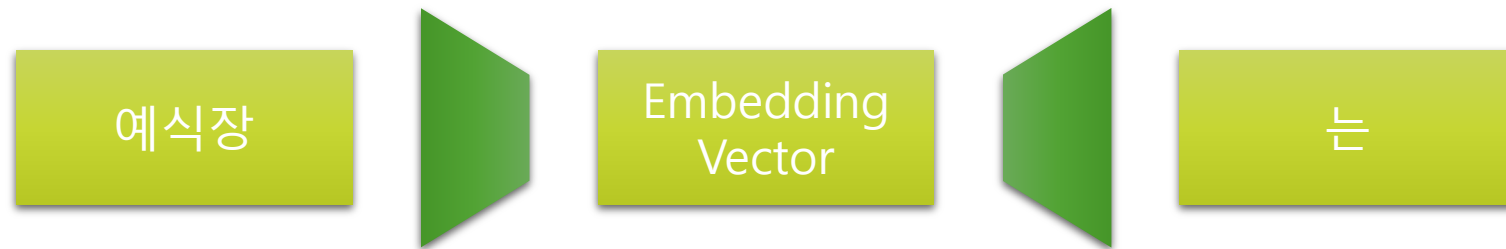
Skip-gram Example ($|W|=5$)

예식장	은	용궁	예식장	주례	는	문어	아저씨	.	
피아노	는	오징어	예물	은	조개껍데기	.			



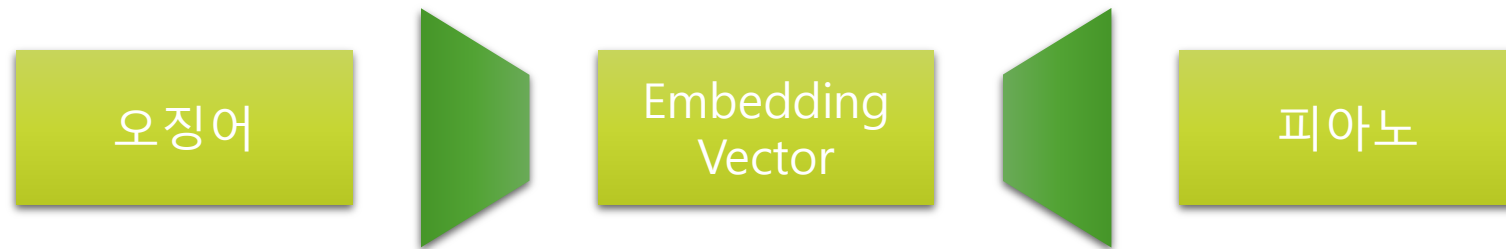
Skip-gram Example ($|W|=5$)

예식장	은	용궁	예식장	주례	는	문어	아저씨	.	
피아노	는	오징어	예물	은	조개껍데기	.			



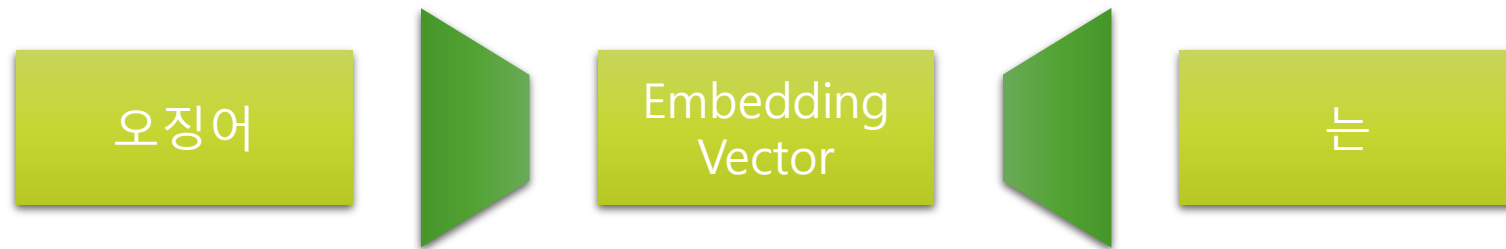
Skip-gram Example ($|W|=5$)

예식장	은	용궁	예식장	주례	는	문어	아저씨	.	
피아노	는	오징어	예물	은	조개껍데기	.			



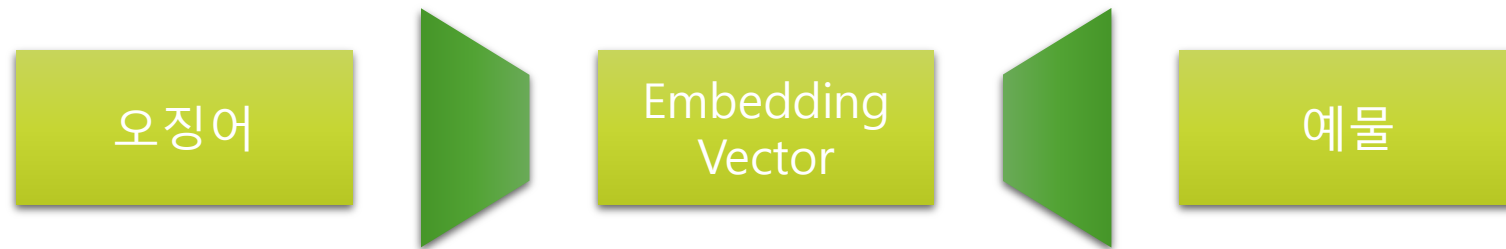
Skip-gram Example ($|W|=5$)

예식장	은	용궁	예식장	주례	는	문어	아저씨	.	
피아노	는	오징어	예물	은	조개껍데기	.			



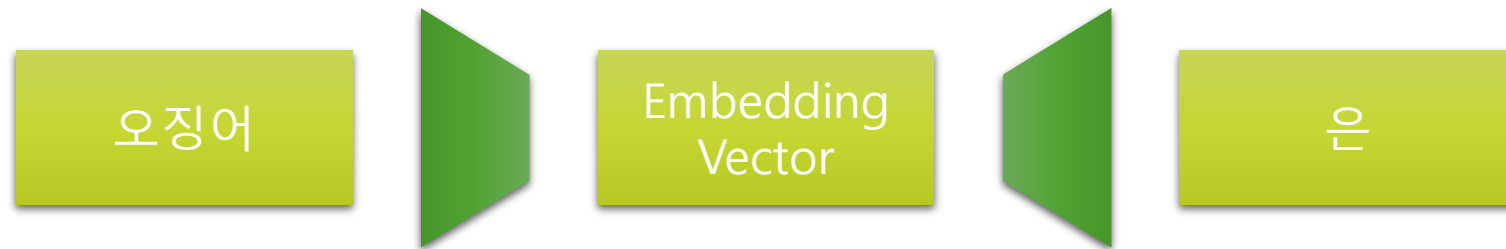
Skip-gram Example ($|W|=5$)

예식장	은	용궁	예식장	주례	는	문어	아저씨	.	
피아노	는	오징어	예물	은	조개껍데기	.			

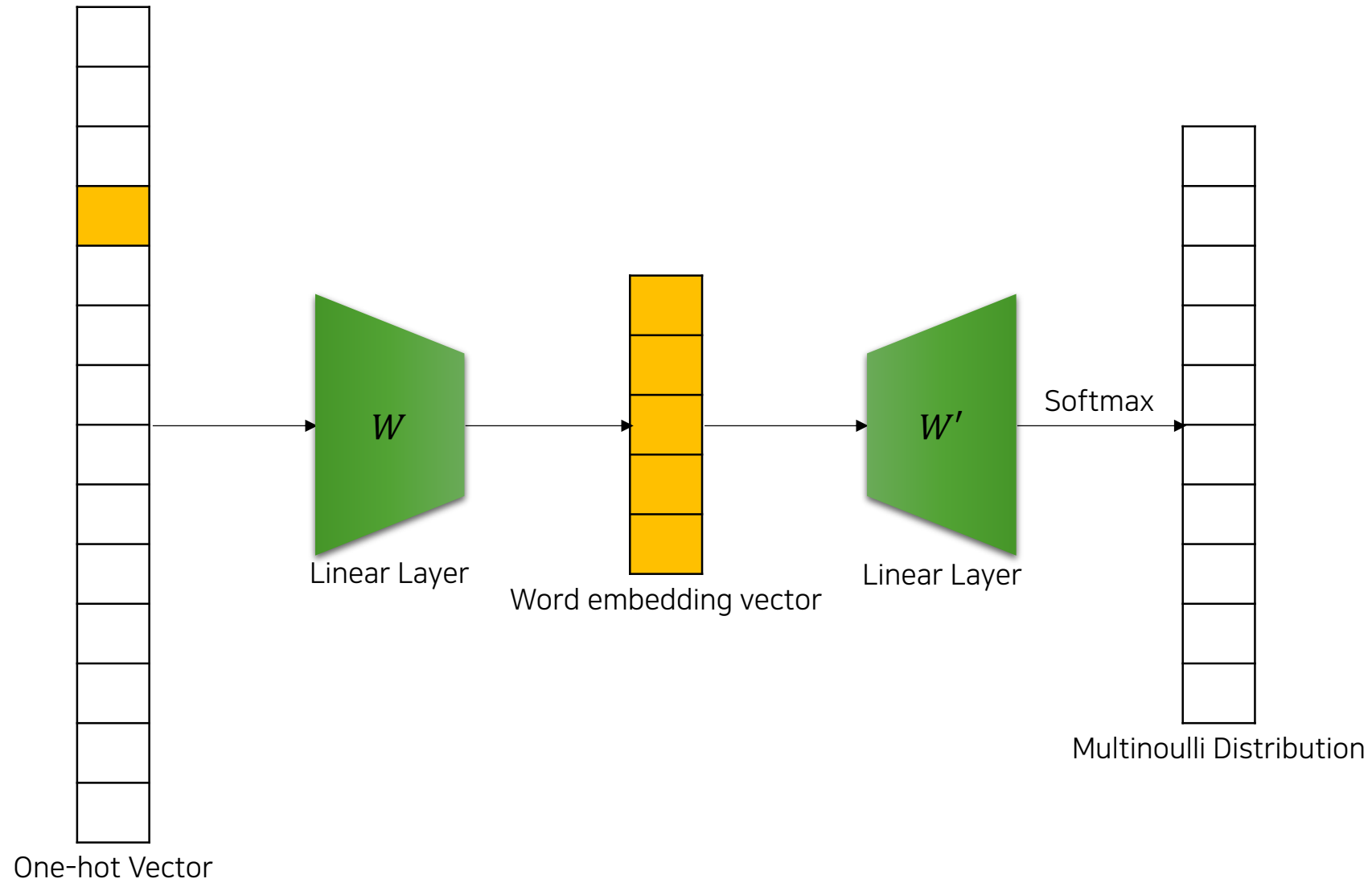


Skip-gram Example ($|W|=5$)

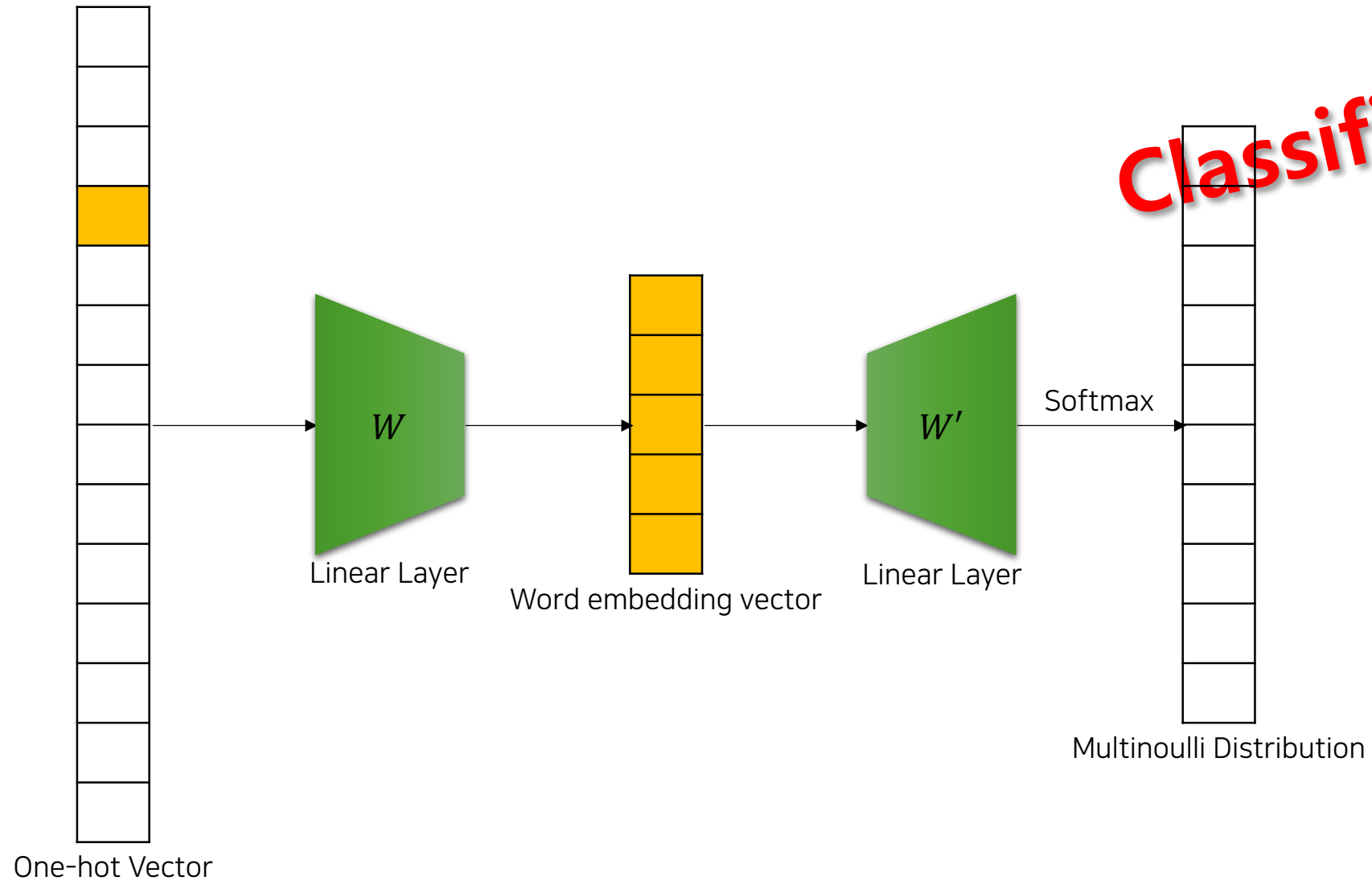
예식장	은	용궁	예식장	주례	는	문어	아저씨	.	
피아노	는	오징어	예물	은	조개껍데기	.			



Skip-gram



Skip-gram



Classification

Skip-gram

- 장점(at that time):
 - 쉽다.
 - 빠르다.
 - 비교적 정확한 벡터를 구할 수 있다.
- 단점(currently):
 - 현데 느리다.
 - 출현 빈도가 적은 단어일 경우 벡터가 정확하지 않다.