

밑바닥부터 시작하는 딥러닝 2
3장 word2vec

시소러스

통계 기반 기법

추론 기반 기법
(word2vec)

분산 표현
분포 가설
동시발생 행렬
상호 정보량

벡터 간 유사도
차원 감소

분포 가설

비슷한 분포를 가진 언어적 개체들은 비슷한 의미를 지닌다.

같은 맥락에서 사용되는 단어들은 같은 의미를 갖는다.

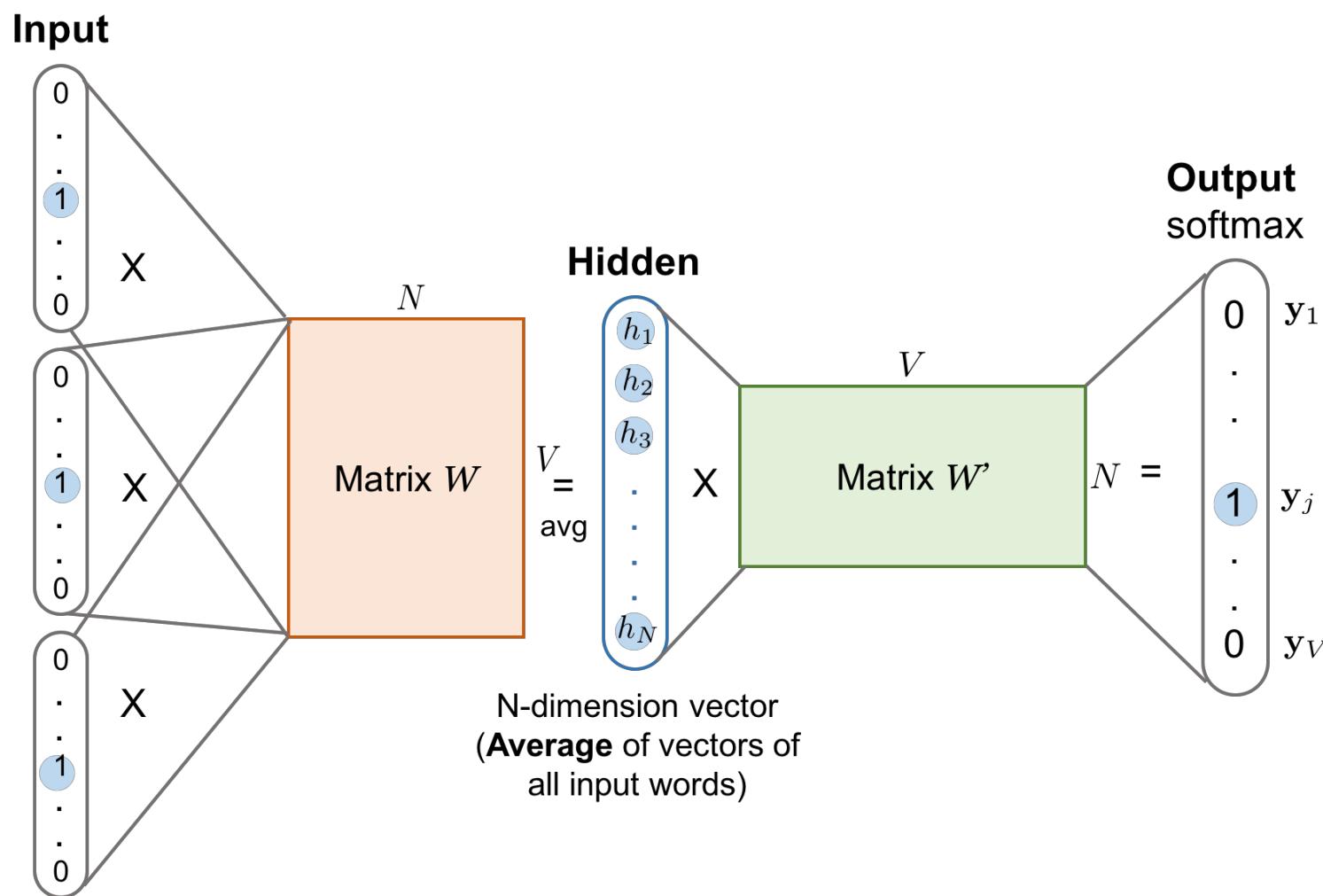
"a word is characterized by the company it keeps" by Firth

통계 기반 기법은 전체 통계를 돌려야 하기 때문에 오래 걸립니다.

반면 **추론 기반 기법**은 미니 배치로 학습이 가능한 장점이 있습니다.

you ? goodbye and I say hello.

CBOW (continuous bag-of-words)



맥락과 타깃

말뭉치	맥락(contexts)	타깃
you say goodbye and I say hello .	you, goodbye	say
you say goodbye and I say hello .	say, and	goodbye
you say goodbye and I say hello .	goodbye, I	and
you say goodbye and I say hello .	and, say	I
you say goodbye and I say hello .	I, hello	say
you say goodbye and I say hello .	say, .	hello

맥락과 타깃

그림 3-17 문어ID의 배열인 corpus로부터 맥락과 타깃을 작성하는 예(맥락의 윈도우 크기 3)

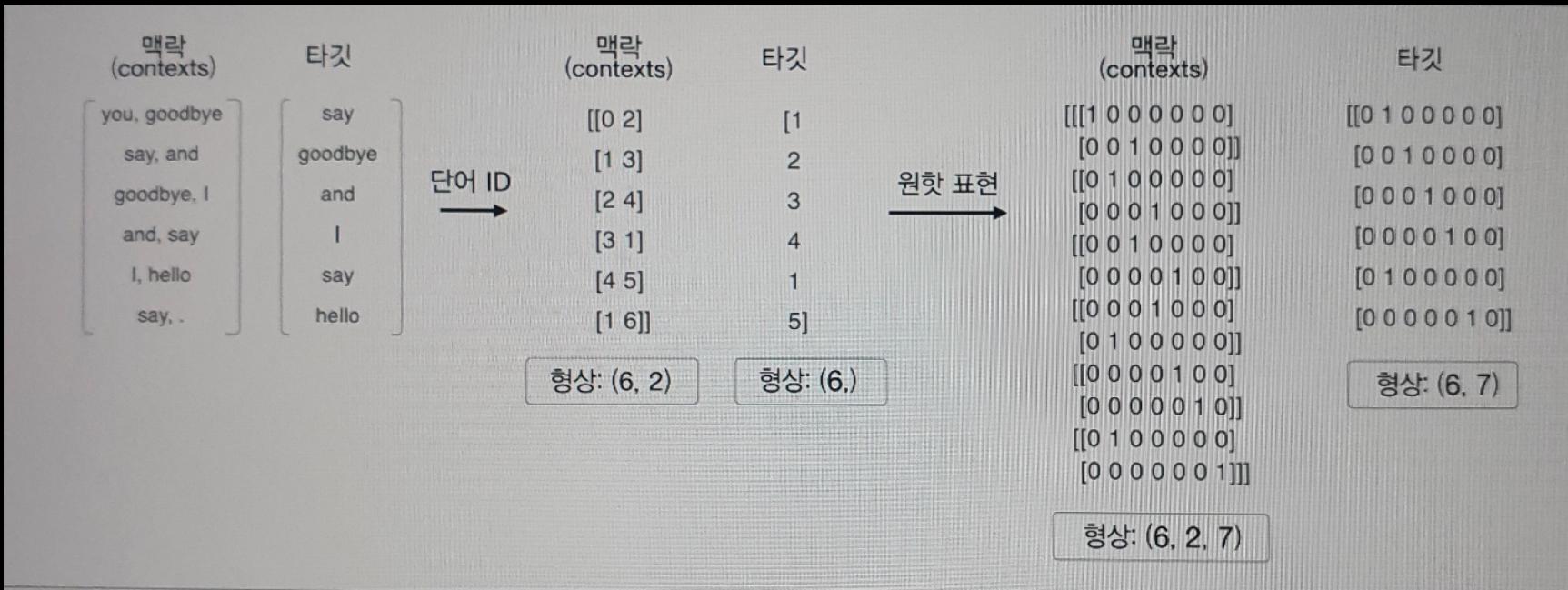
말뭉치	맥락(contexts)	타깃
[0 1 2 3 4 1 5 6]	[[0 2]	[1
	[1 3]	2
→	[2 4]	3
	[3 1]	4
	[4 5]	1
	[1 6]]	5]

형상: (8,)

형상: (6, 2)

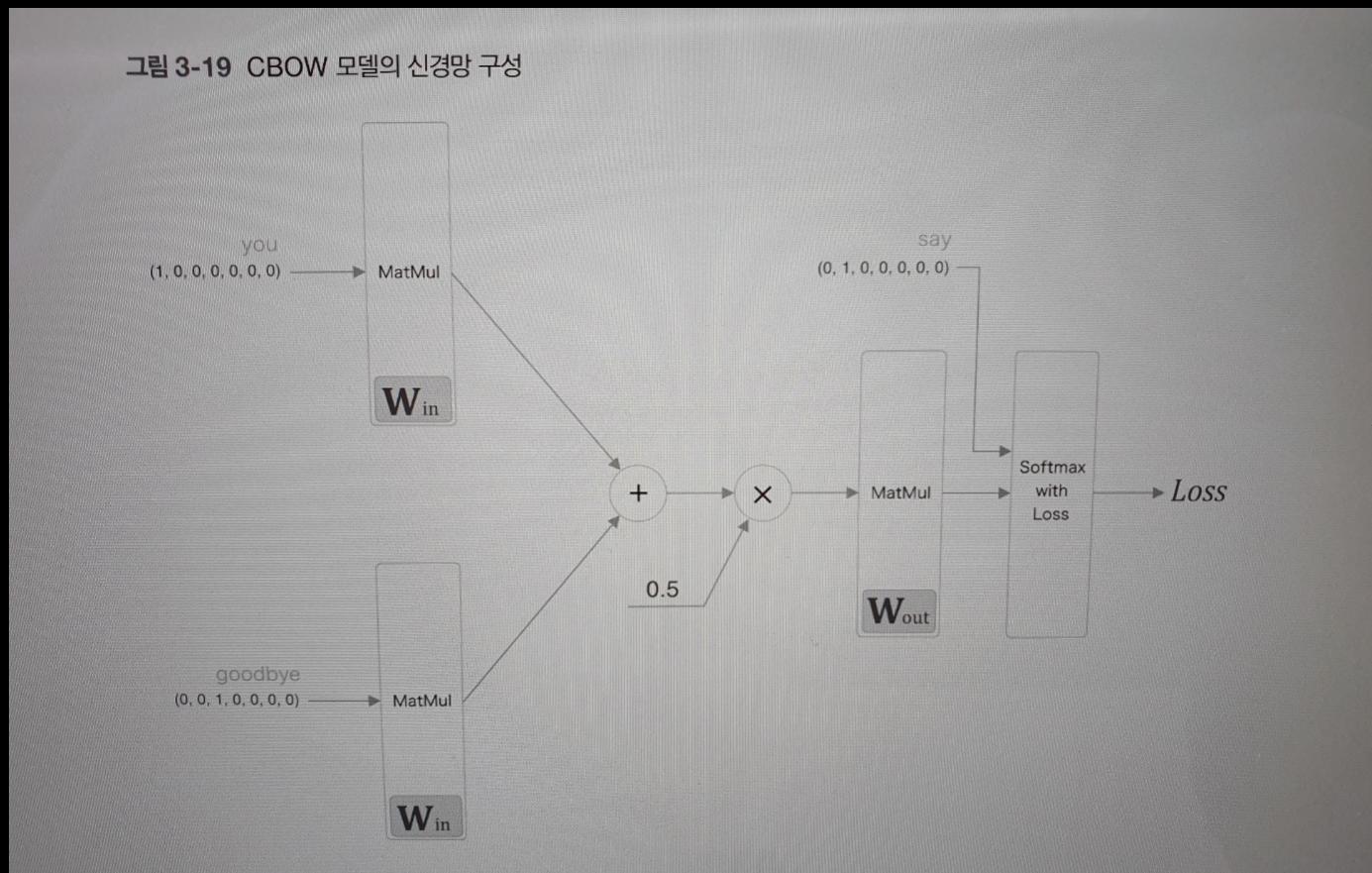
형상: (6,)

맥락과 타깃



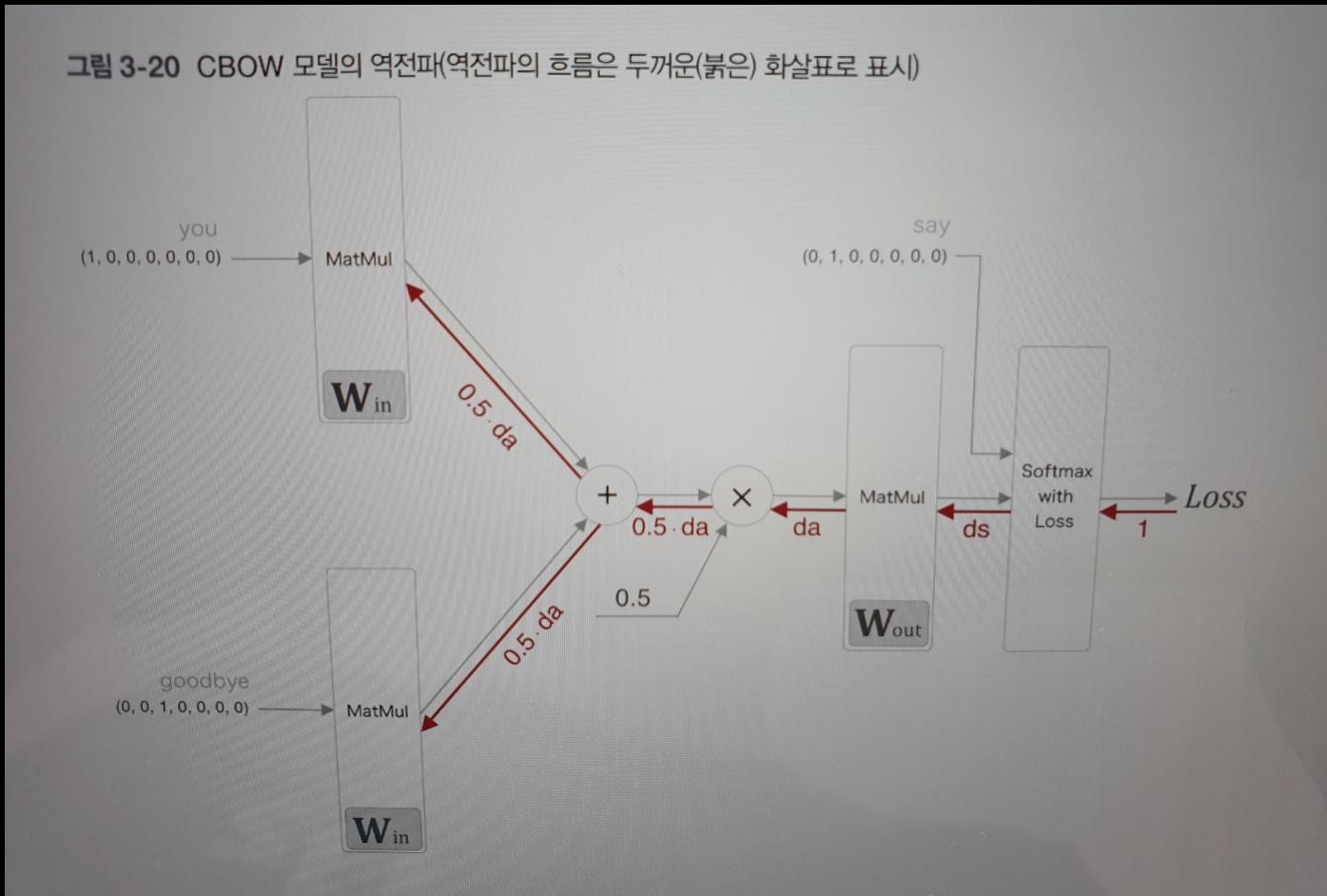
CBOW 모델

그림 3-19 CBOW 모델의 신경망 구성



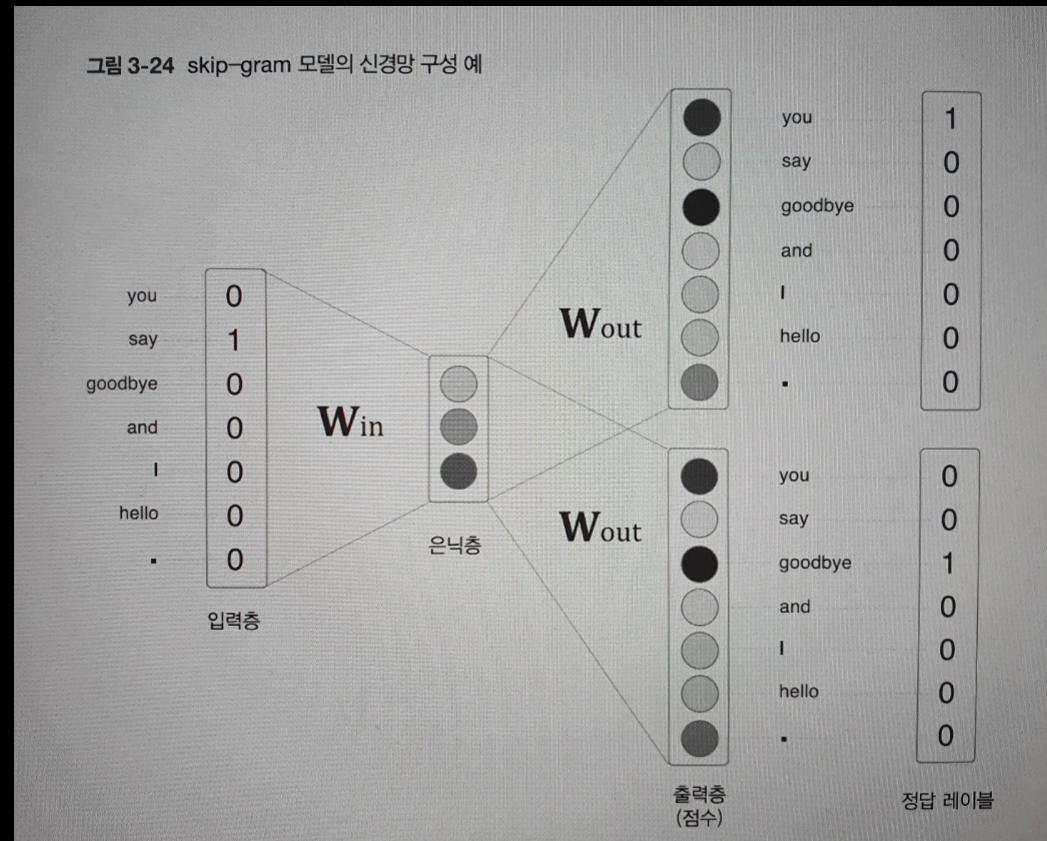
CBOW 모델의 역전파

그림 3-20 CBOW 모델의 역전파(역전파의 흐름은 두꺼운(붉은) 화살표로 표시)



Skip-gram

그림 3-24 skip-gram 모델의 신경망 구성 예



- Loss-function:

$$E = -\log p(w_{t-c}..w_{t+c} | w_t)$$



$$\text{CBOW: } E = -\log p(w_t | w_{t-c}..w_{t+c})$$

$$P(w_{t-1}, w_{t+1} | w_t) = P(w_{t-1} | w_t)P(w_{t+1} | w_t)$$

를 교차 엔트로피 오차에 적용하여 skip-gram 모델의 손

$$\begin{aligned} L &= -\log P(w_{t-1}, w_{t+1} | w_t) \\ &= -\log P(w_{t-1} | w_t)P(w_{t+1} | w_t) \\ &= -(\log P(w_{t-1} | w_t) + \log P(w_{t+1} | w_t)) \end{aligned}$$

는 $\log xy = \log x + \log y$ 라는 로그의 성질을 이용했습니다.
ram 모델의 손실 함수는 맥락별 손실을 구한 다음 모두 더한
은 샘플 데이터 하나짜리 skip-gram의 손실 함수입니다.
-gram 모델의 손실 함수는 [식 3.7]이 됩니다.

$$L = -\frac{1}{T} \sum_{t=1}^T (\log P(w_{t-1} | w_t) + \log P(w_{t+1} | w_t))$$