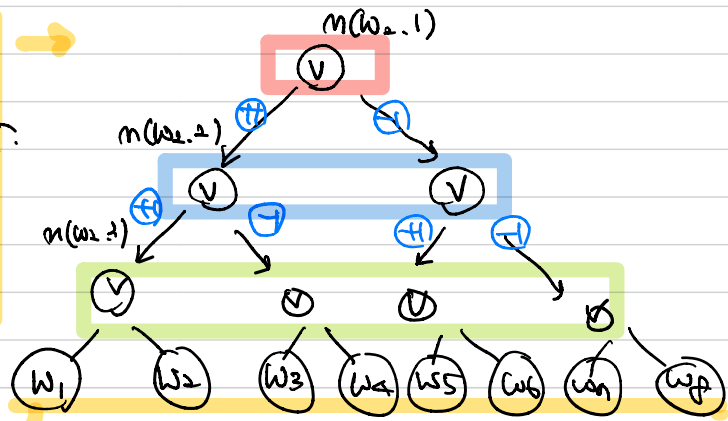
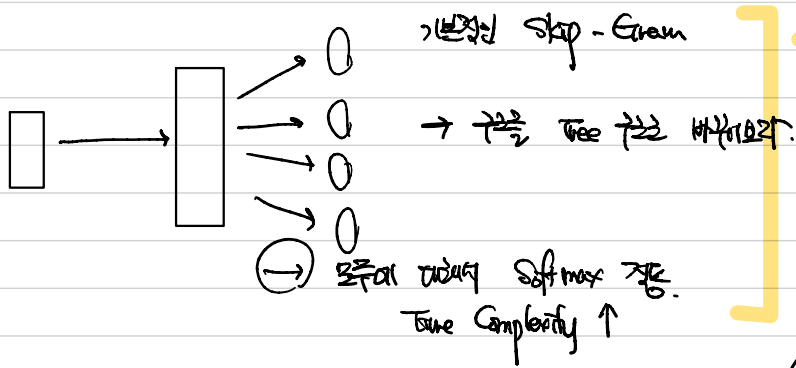


Hierarchical Softmax and Negative Sampling.

1) Hierarchical Softmax



- $m(w_{z,i})$: Word $w_{z,i}$ 의 unitary embedding depth에 있는 모든.
 - 각 노드는 weight vector를 가짐.
 - σ 함수
- 각 노드 child는 o_i .

$$p(w|w_i) = \prod_{j=1}^{L(w)-1} \phi([q(w, j+1) = \text{do}(q(w, j))]) \cdot v_{w_j}^T \cdot v_{w_i}$$

$$[x] = \begin{cases} 1 & \text{if } x \geq \text{true} \\ -1 & \text{o.w} \end{cases}$$

- $$\cdot \frac{2}{2} \frac{2}{2} = 1$$

$$\sigma(v_m^T v_{wi}) + \sigma(-v_n^T v_{wi}) = 1$$

• 現行

$\mathbb{I} \text{pr}(\text{leaves}) = 1 \Rightarrow \text{Stop-Gradient } \frac{\partial}{\partial \theta}.$

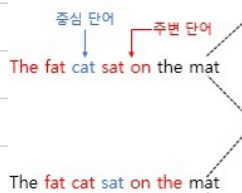
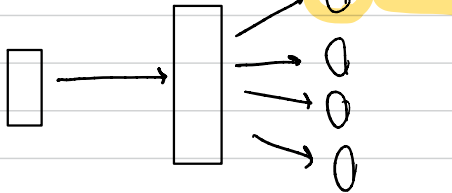
· 用字

$$\underline{O(\log N)}$$

2) Negative Sampling.



2. window 함수는 $\frac{1}{n}$



입력	레이블
cat	The
cat	fat
cat	sat
cat	on
sat	fat
sat	cat
sat	on
sat	the
...	...

입력1	입력2	레이블
cat	The	1
cat	fat	1
cat	sat	1
cat	on	1
sat	fat	1
sat	cat	1
sat	on	1
sat	the	1
...

입력과 레이블의 변화

입력1	입력2	레이블
cat	The	1
cat	fat	1
cat	sat	1
cat	on	1

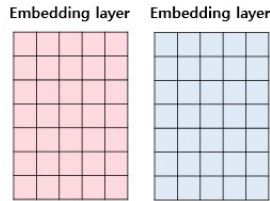
Negative Sampling

입력1	입력2	레이블
cat	The	1
cat	fat	1
cat	pizza	0
cat	computer	0
cat	sat	1
cat	on	1

단어 집합에서 랜덤으로 선택된 단어들을 레이블 0의 샘플로 추가.

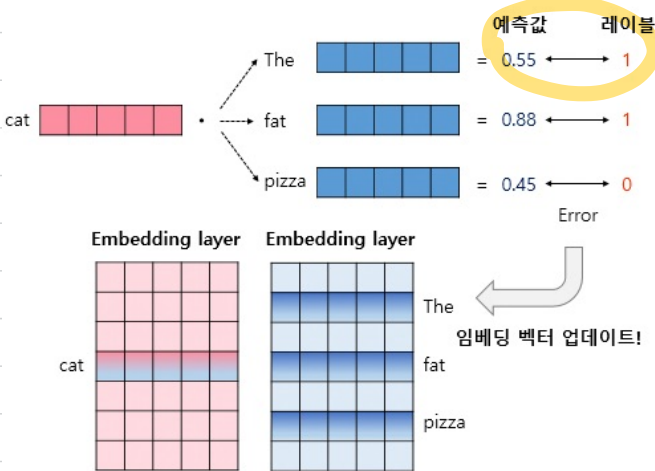
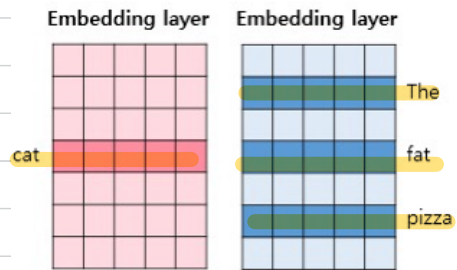
(같은 값)에서 Sampling
 ↓
 positive. 0 : Negative.
 😊 Negative Sampling.

입력1	입력2	레이블
cat	The	1
cat	fat	1
cat	pizza	0
cat	computer	0
cat	sat	1
cat	on	1
cat	cute	1
cat	mighty	0
...



두개 Table의 빈 space
 input과 output part.

Table의 빈 space
 이걸로 cosine similarity 계산.



→ 따라서 $-\log$ 즉 log function을 사용.

이제와 같이
 Embedding layer Table
 update 진행

부족한 data를 softmax를 사용해서 보완.

⇒ 실제로 Word2vec에서 가장 best는 skip-gram 모델.

softmax가 overfit 문제 { log sampling, Hierarchical sampling } 이 사용.