Insight campus SeSAC

임베딩 (Embedding)

실무형 인공지능 자연어처리

2 Word2Vec 직접 구현

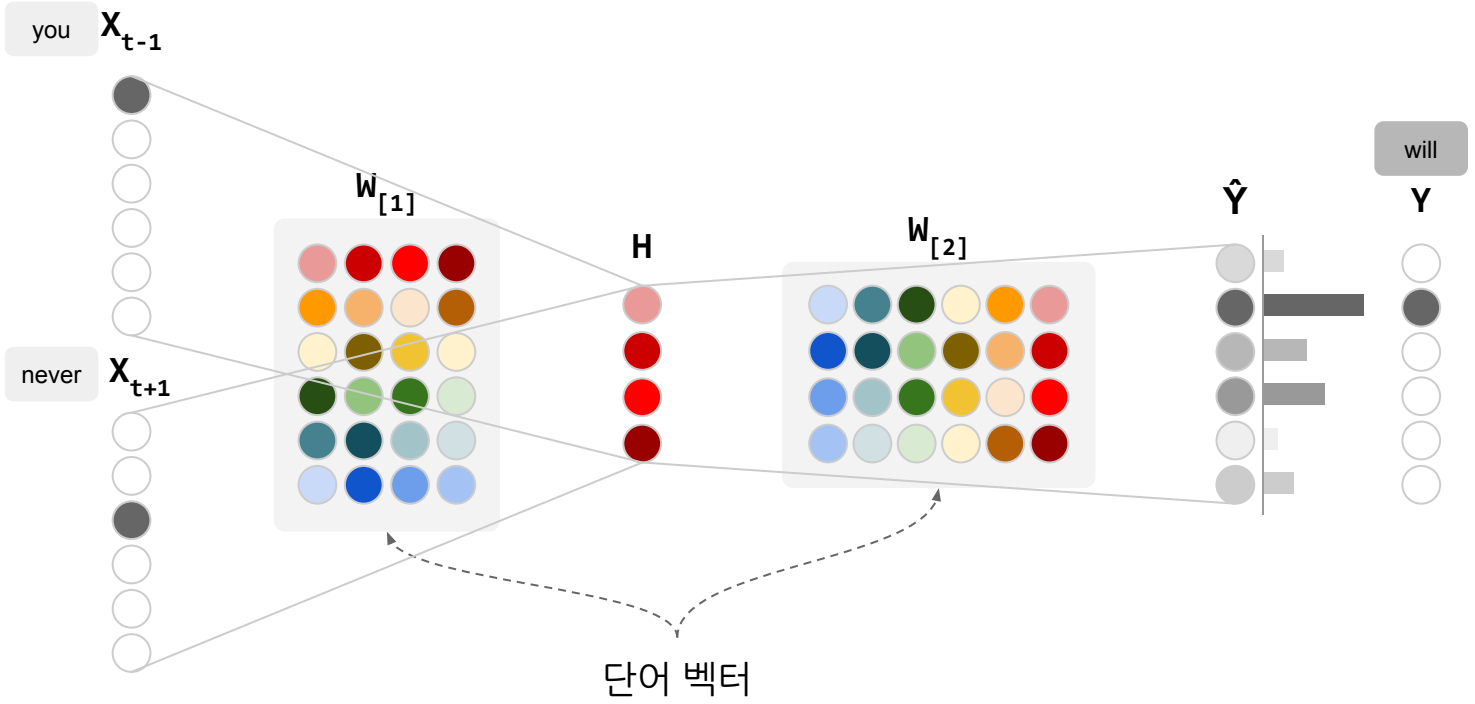
Word2Vec(CBoW) 학습 - 구조

- 입력 문장 토큰화. (사용하지 않는 품사 제거)
- 가중치(= 파라미터 = 단어 벡터) 초기화
- 원핫 인코딩(one hot encoding)
- epoch 만큼 반복
  - 중심단어, 문맥단어 추출
  - Feed Forward
  - Loss / Gradient 계산
  - Weight (Parameter) 갱신

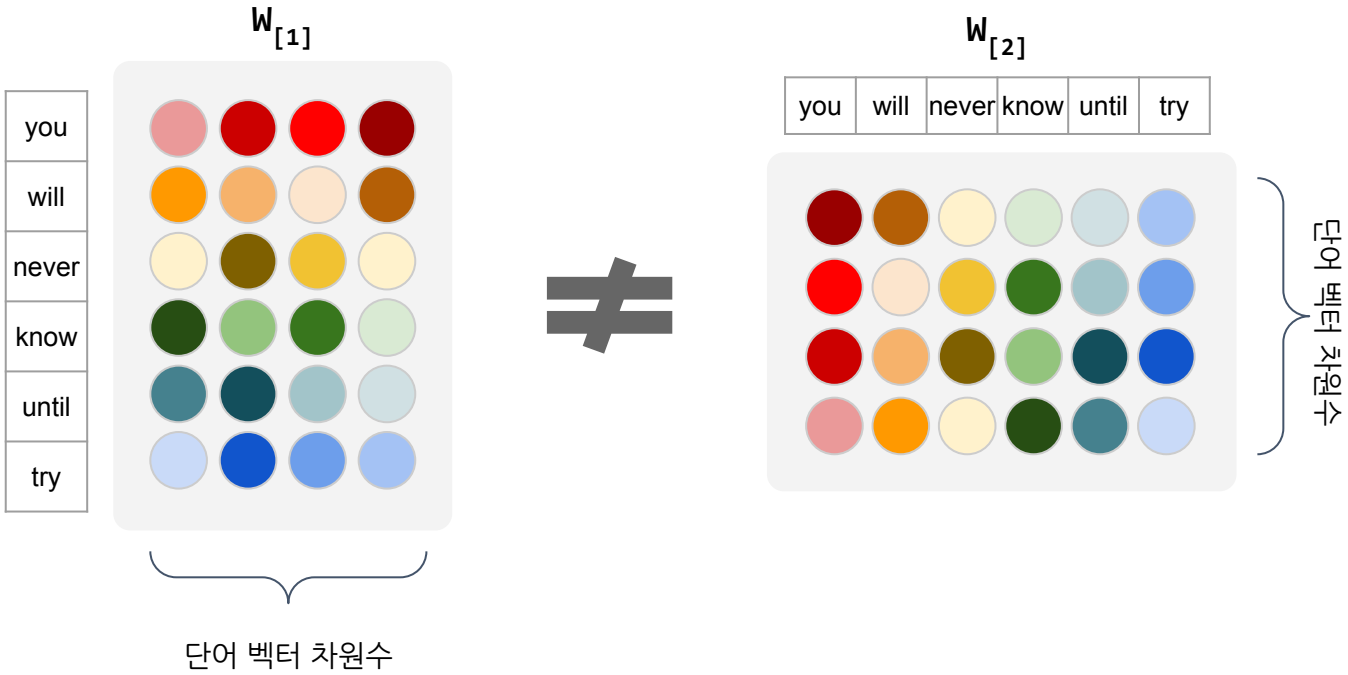
Insight campus SeSAC

Word2Vec(CBoW) 학습 - 절차

you	[1,0,0,0,0,0]
will	[0,1,0,0,0,0]
never	[0,0,1,0,0,0]
know	[0,0,0,1,0,0]
until	[0,0,0,0,1,0]
you	[1,0,0,0,0,0]
try	[0,0,0,0,0,1]



Word2Vec(CBoW) 학습 - 단어벡터

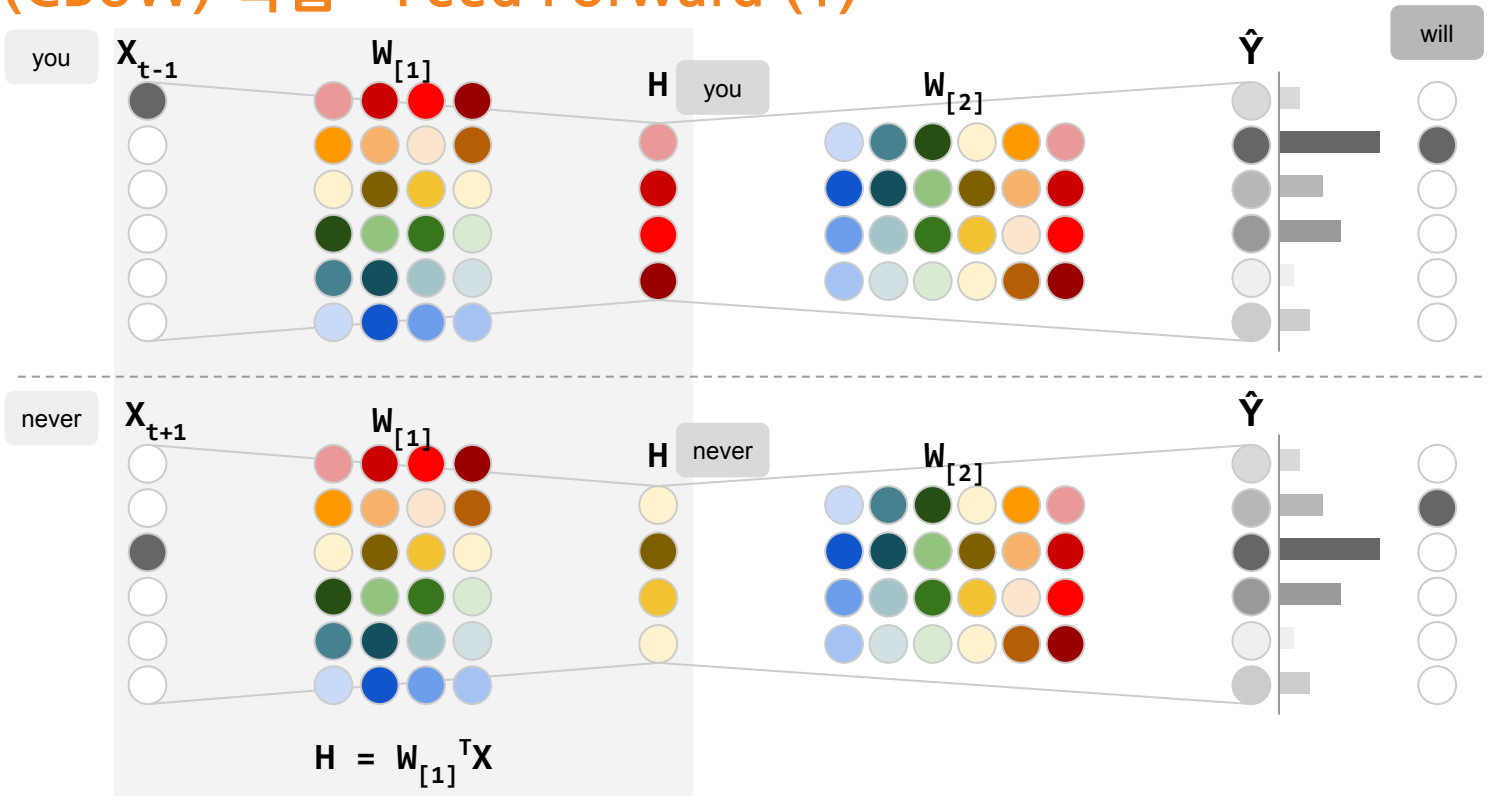


학습하는 가중치(파라미터)가 단어 벡터

Insight campus SeSAC

Word2Vec(CBoW) 학습 - Feed Forward (1)

you	[1,0,0,0,0,0]
will	[0,1,0,0,0,0]
never	[0,0,1,0,0,0]
know	[0,0,0,1,0,0]
until	[0,0,0,0,1,0]
you	[1,0,0,0,0,0]
try	[0,0,0,0,0,1]

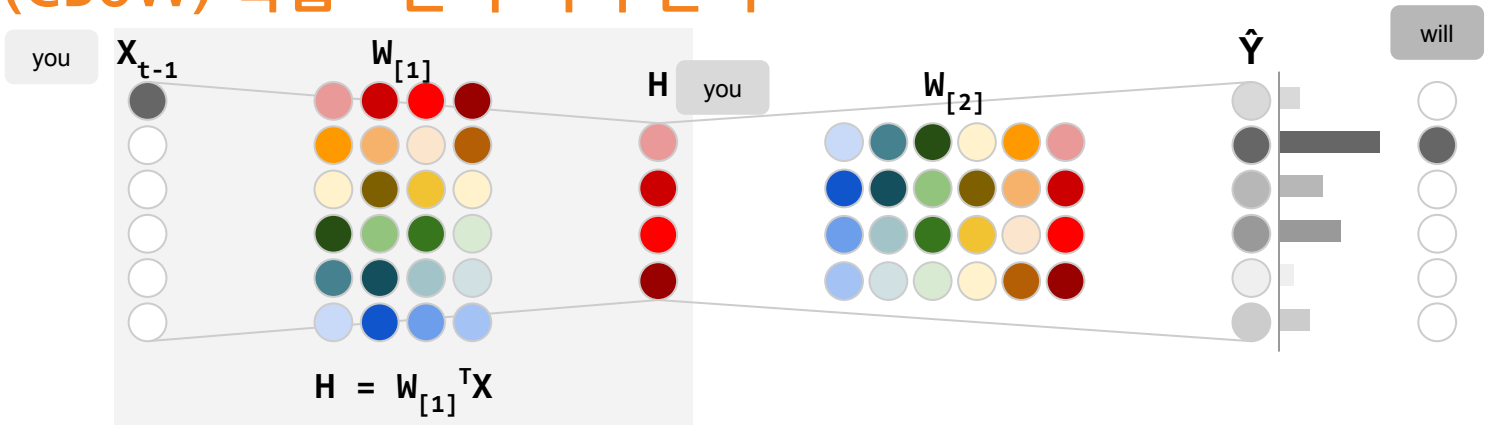


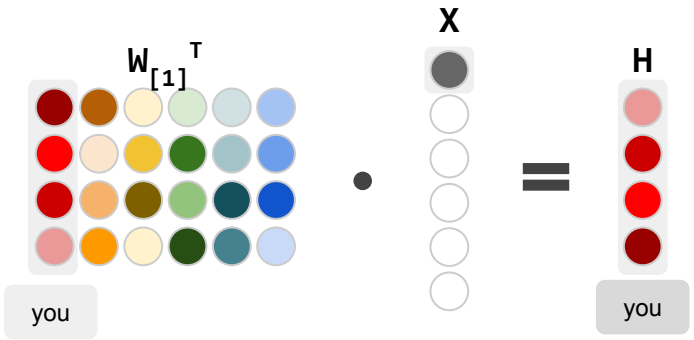
입력 단어(원핫 벡터)와 단어 벡터의 곱으로 Hidden layer 계산

Insight campus SeSAC

Word2Vec(CBoW) 학습 - 단어 벡터 선택

you	[1,0,0,0,0,0]
will	[0,1,0,0,0,0]
never	[0,0,1,0,0,0]
know	[0,0,0,1,0,0]
until	[0,0,0,0,1,0]
you	[1,0,0,0,0,0]
try	[0,0,0,0,0,1]



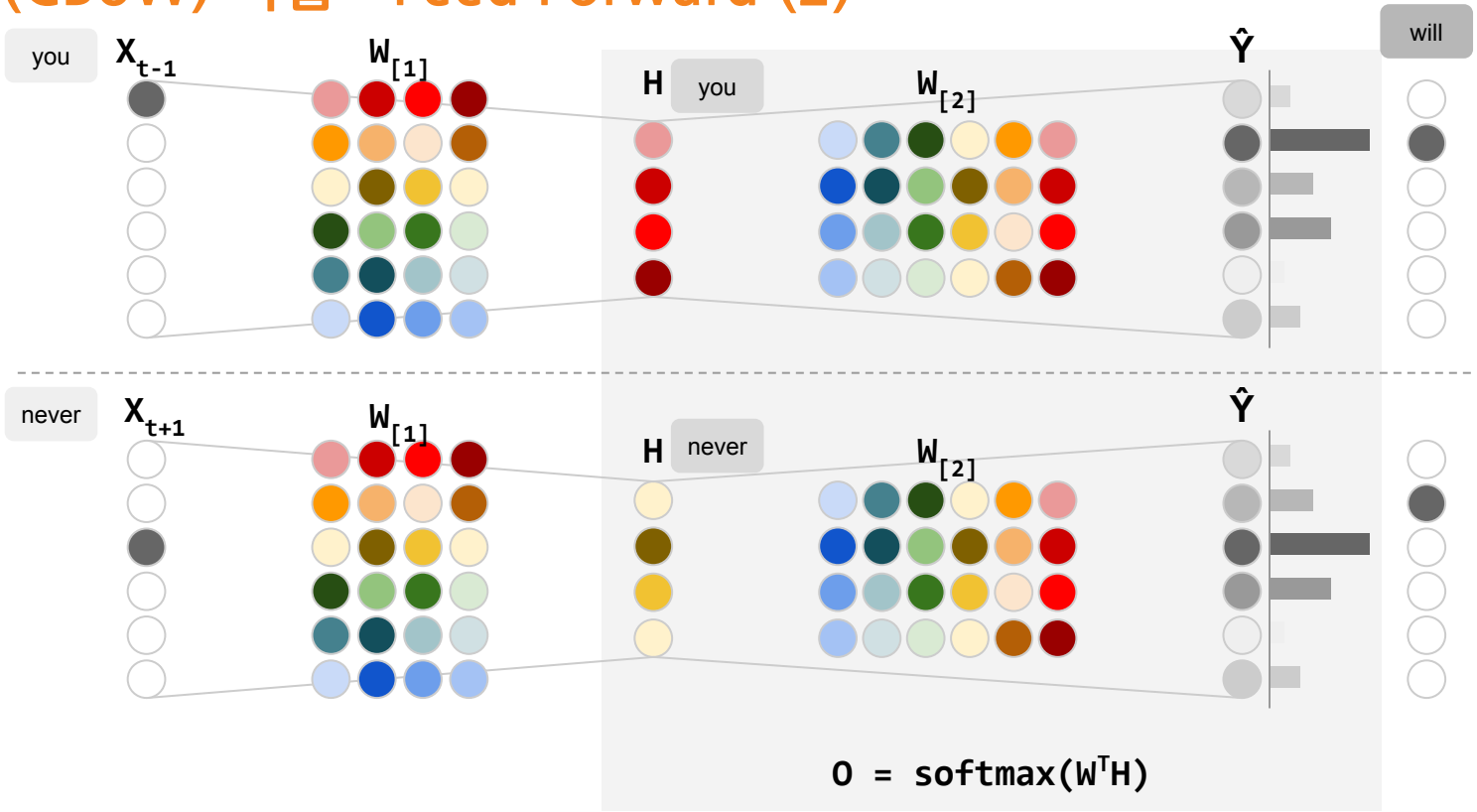


Hidden layer는 선택된 하나의 단어 벡터

Insight campus SeSAC

Word2Vec(CBoW) 학습 - Feed Forward (2)

you	[1,0,0,0,0,0]
will	[0,1,0,0,0,0]
never	[0,0,1,0,0,0]
know	[0,0,0,1,0,0]
until	[0,0,0,0,1,0]
you	[1,0,0,0,0,0]
try	[0,0,0,0,0,1]



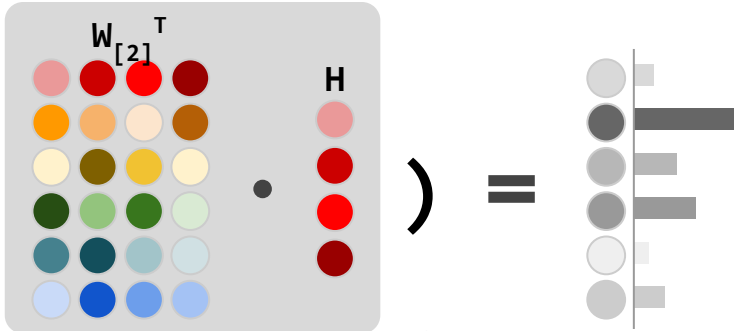
Hidden layer 와 단어벡터의 곱으로 Output layer 계산

Word2Vec(CBoW) 학습 - 단어간 유사도 계산

you	[1,0,0,0,0,0]
will	[0,1,0,0,0,0]
never	[0,0,1,0,0,0]
know	[0,0,0,1,0,0]
until	[0,0,0,0,1,0]
you	[1,0,0,0,0,0]
try	[0,0,0,0,0,1]

$$similarity = \frac{A \cdot B}{\|A\| \cdot \|B\|}$$

$$softmax(W_{[2]}^{T} \cdot H) =$$



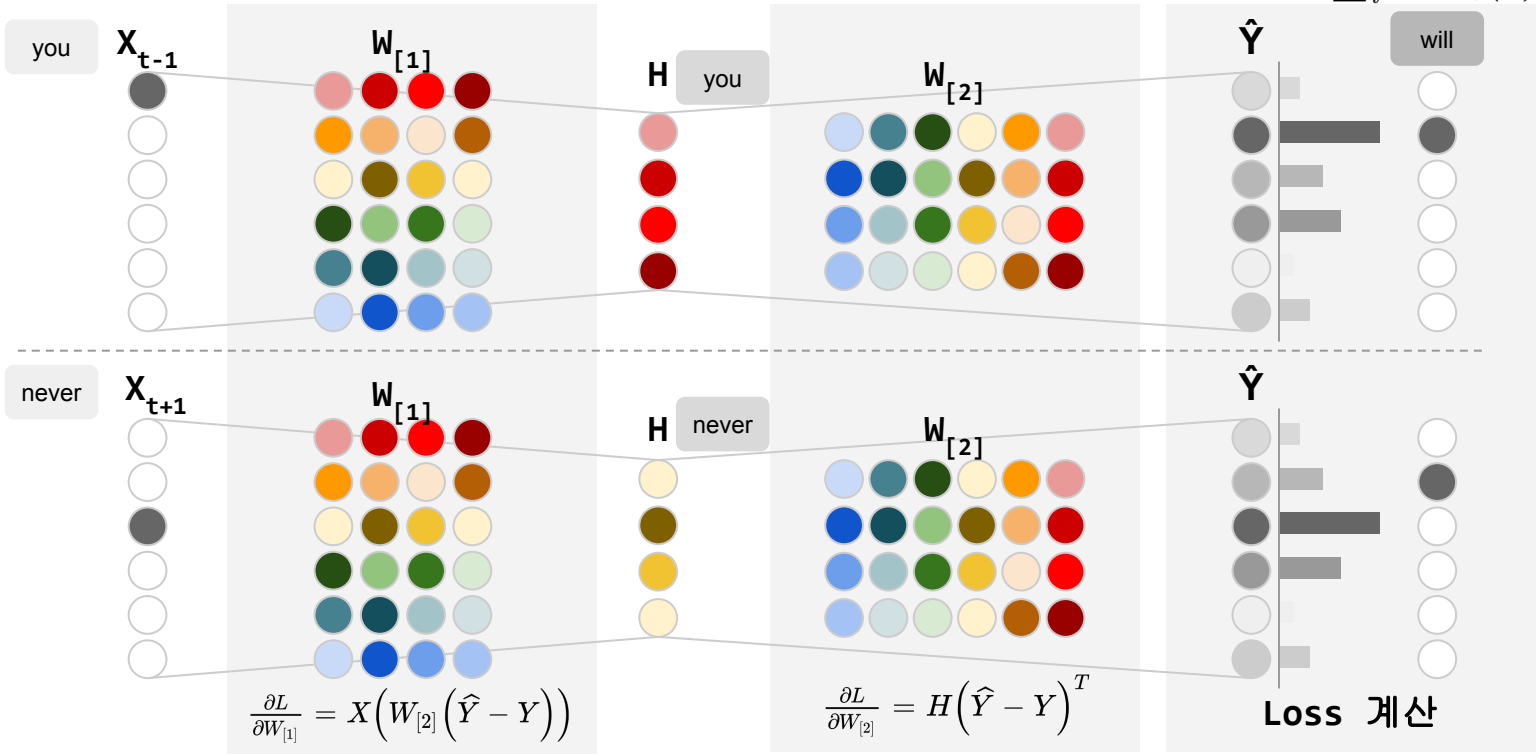
선택된 단어 벡터(H)와 전체 단어 벡터($W_{[2]}^{T}$)간 **유사도**

Insight campus SeSAC

Word2Vec(CBoW) 학습 - Loss / Gradient

$$CrossEntropy = -\sum_{i}^{N} y_i \log(\hat{y})$$

you	[1,0,0,0,0,0]
will	[0,1,0,0,0,0]
never	[0,0,1,0,0,0]
know	[0,0,0,1,0,0]
until	[0,0,0,0,1,0]
you	[1,0,0,0,0,0]
try	[0,0,0,0,0,1]



epoch 별 Loss / Gradient 계산

Word2Vec(CBoW) 학습 - Loss / Gradient

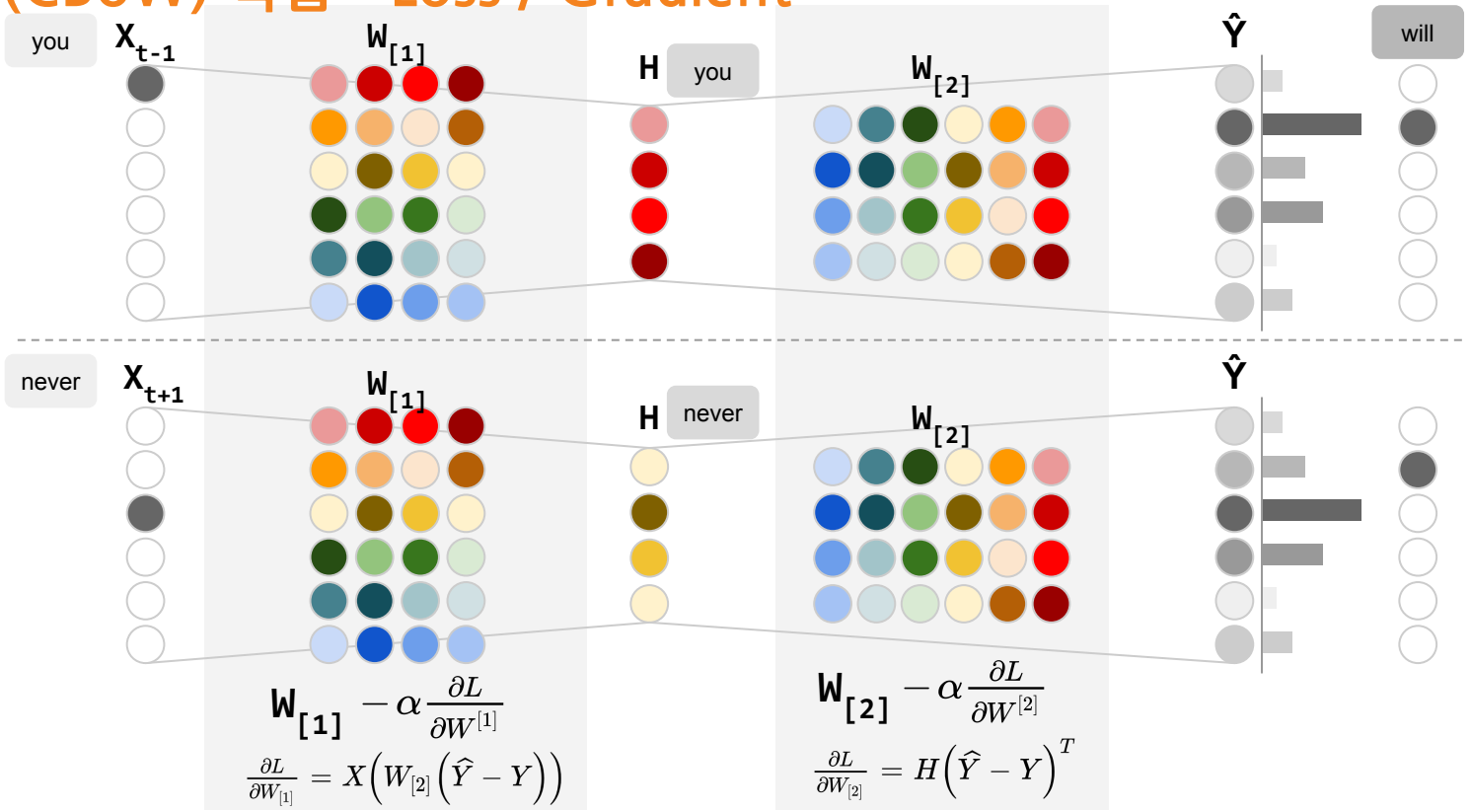
$$CrossEntropy = -\sum_{i}^{N} y_i \log(\hat{y})$$

$$BinaryCrossEntropy = -(y_i \log(\hat{y}) + (1 - y_i)\log(1 - \hat{y}))$$

Insight campus SeSAC

Word2Vec(CBoW) 학습 - Loss / Gradient

you	[1,0,0,0,0,0]
will	[0,1,0,0,0,0]
never	[0,0,1,0,0,0]
know	[0,0,0,1,0,0]
until	[0,0,0,0,1,0]
you	[1,0,0,0,0,0]
try	[0,0,0,0,0,1]



Weight(Parameter)를 gradient * learning rate 만큼 갱신

감사합니다.

Insight campus SeSAC

daniel@fins.ai