

## 6-1. 텍스트데이터 처리

딥러닝 모델에 텍스트가  
인풋이 될 수 있나?

아니요!

딥러닝 모델에 숫자가  
인풋이 될 수 있나?

네!

# What is Encoding?

Convert text to number



# What is One Hot Encoding?

Convert text to vector

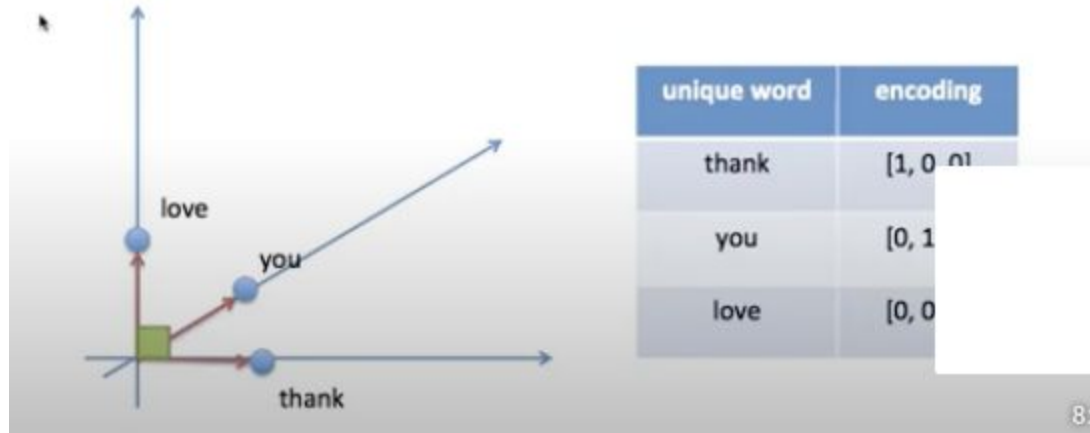
	thank	you	love
thank	1	0	0
you	0	1	0
love	0	0	1



unique word	encoding
thank	[1, 0, 0]
you	[0, 1, 0]
love	[0, 0, 1]

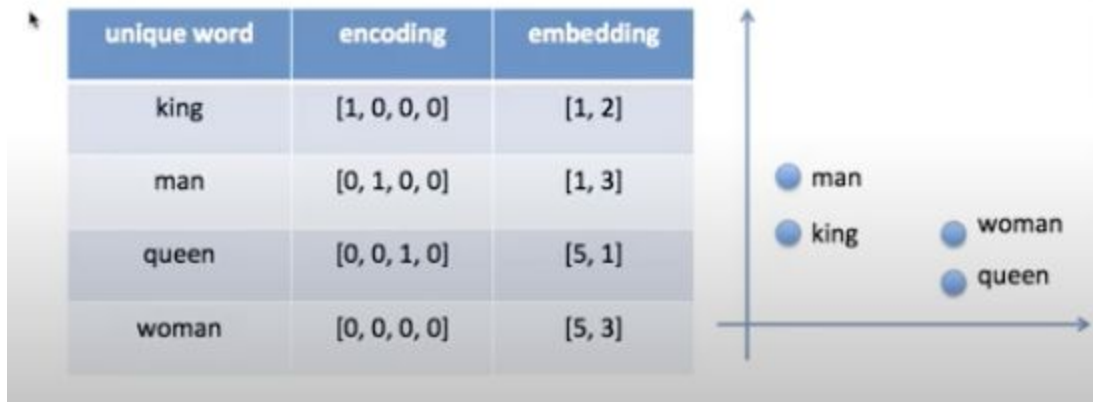
One Hot Encoding doesn't have **similarity**

cosine similarity also 0 since angle is 90 degree



# Embedding

Embedding is dense vector with similarity



# Word2Vec, Glove

출처 : 허민석님 유튜브 word2vec <https://www.youtube.com/watch?v=sY4YyacSsLc>



## 6-2. 순화 신경망

# Vanilla RNN

# I google at work

I google at work

pronoun

verb

preposition

noun

I work at google

I work at google

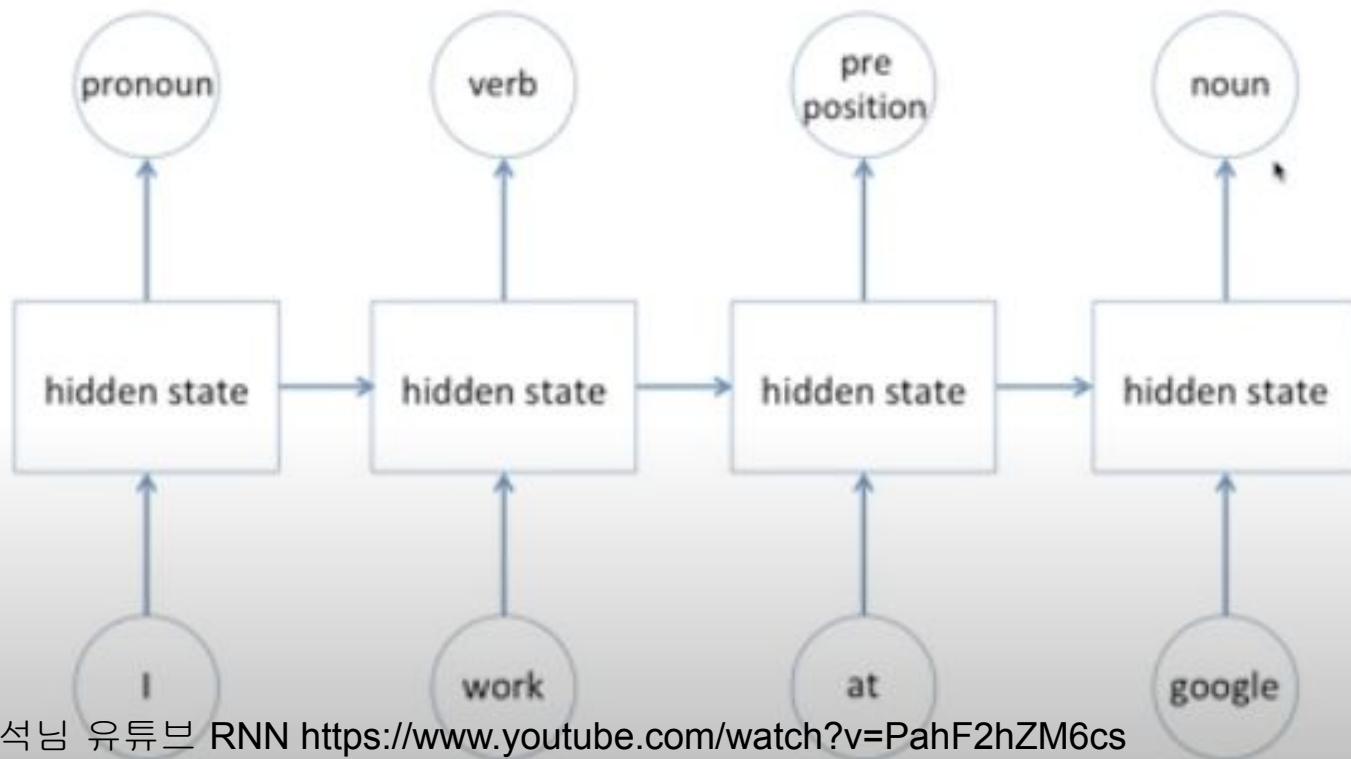
pronoun

verb

preposition

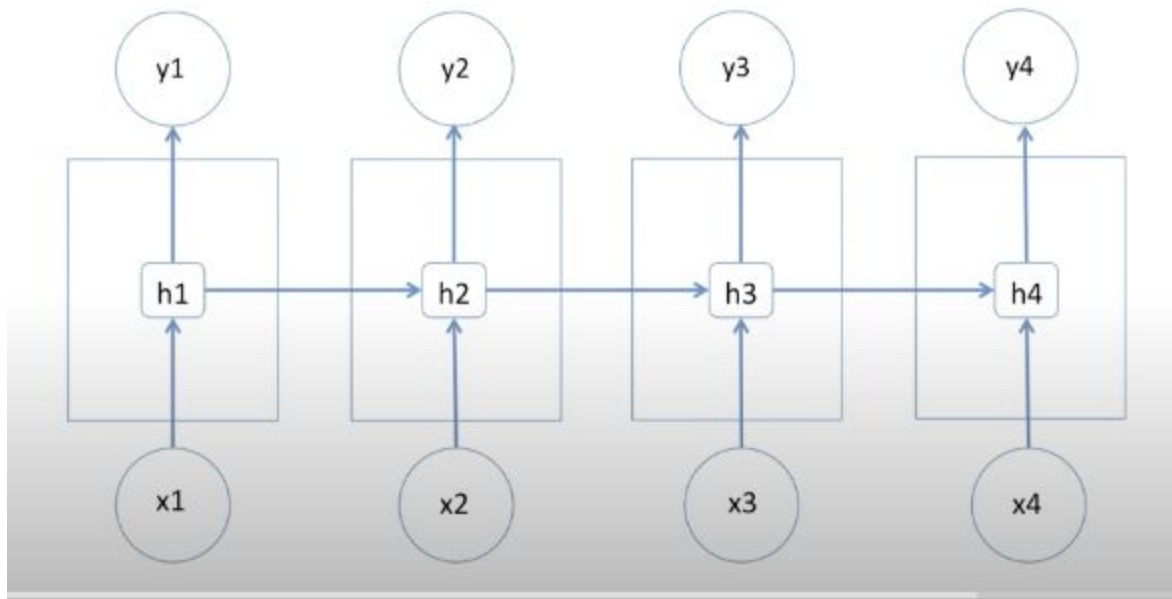
noun

# Sequence is important for POS tagging



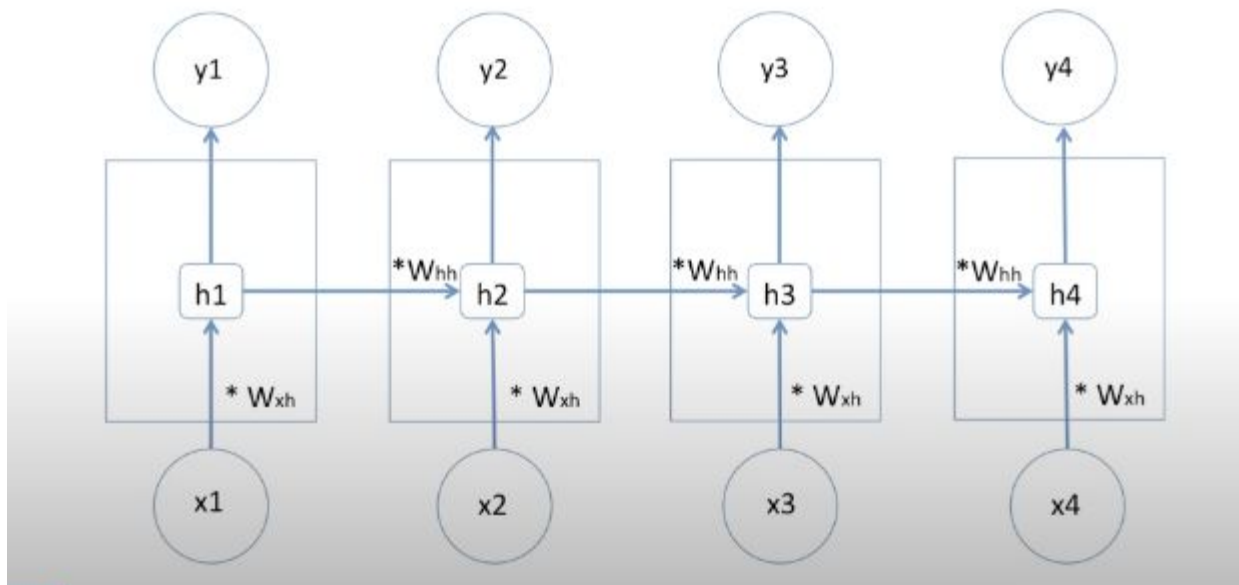
출처 : 허민석님 유튜브 RNN <https://www.youtube.com/watch?v=PahF2hZM6cs>

## Sequence is important for POS tagging

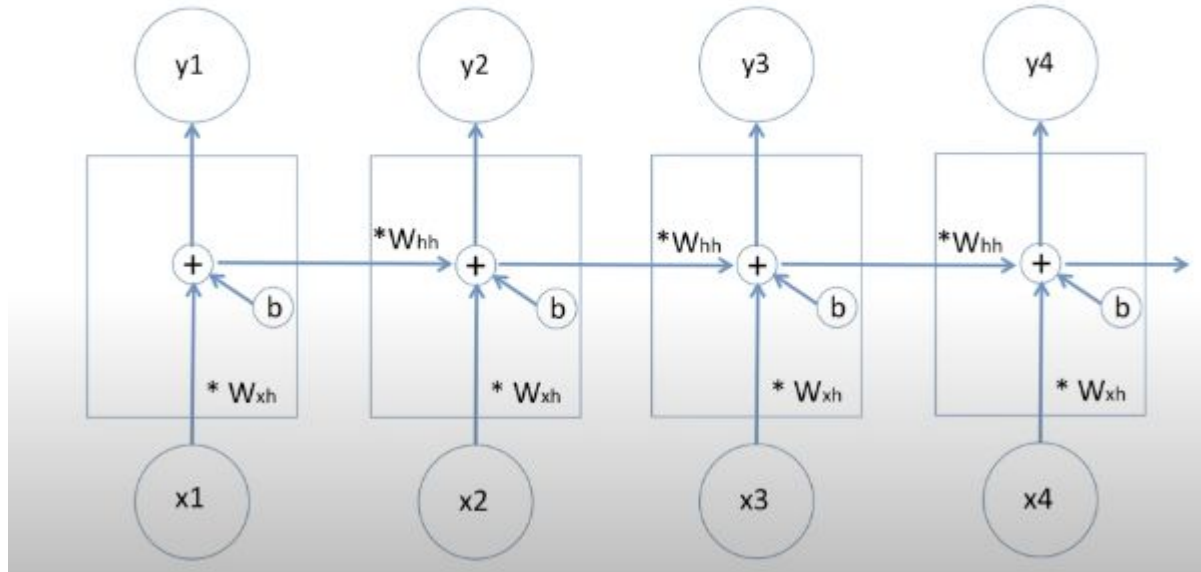




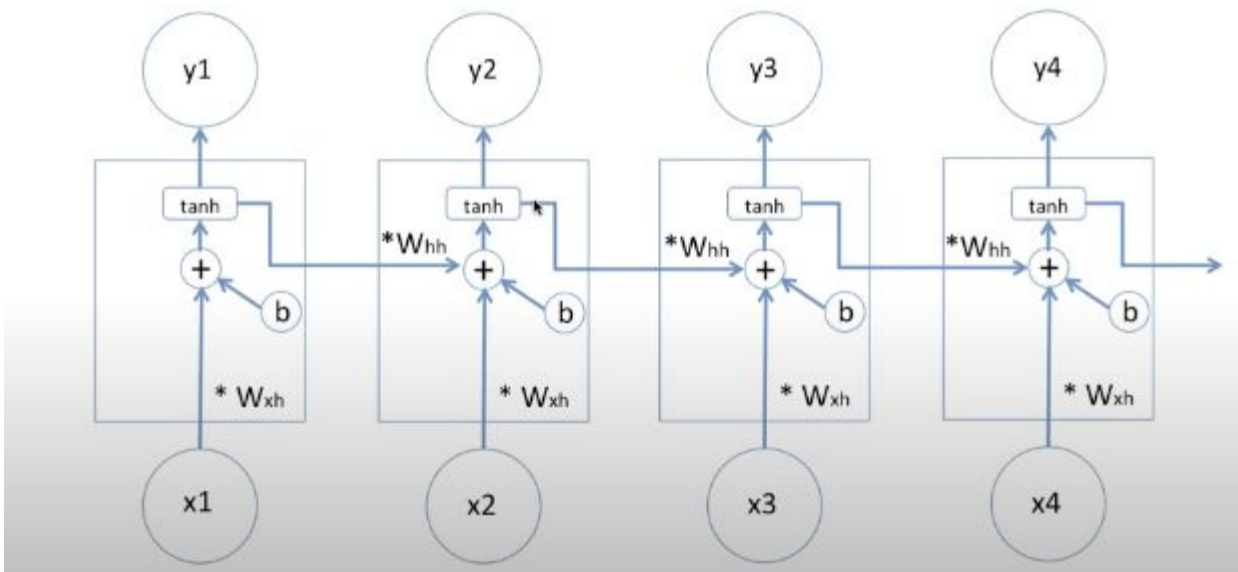
## Sequence is important for POS tagging

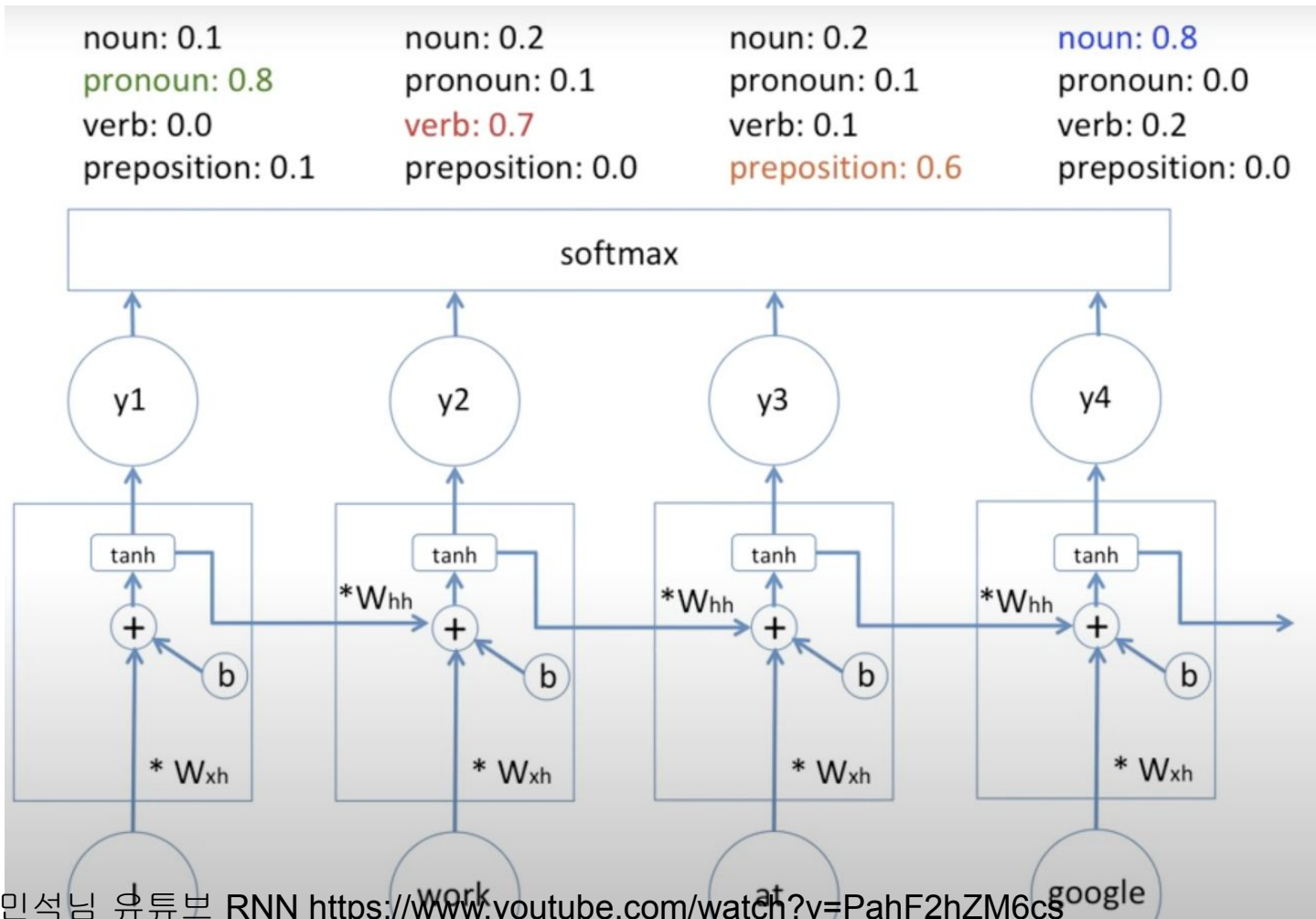


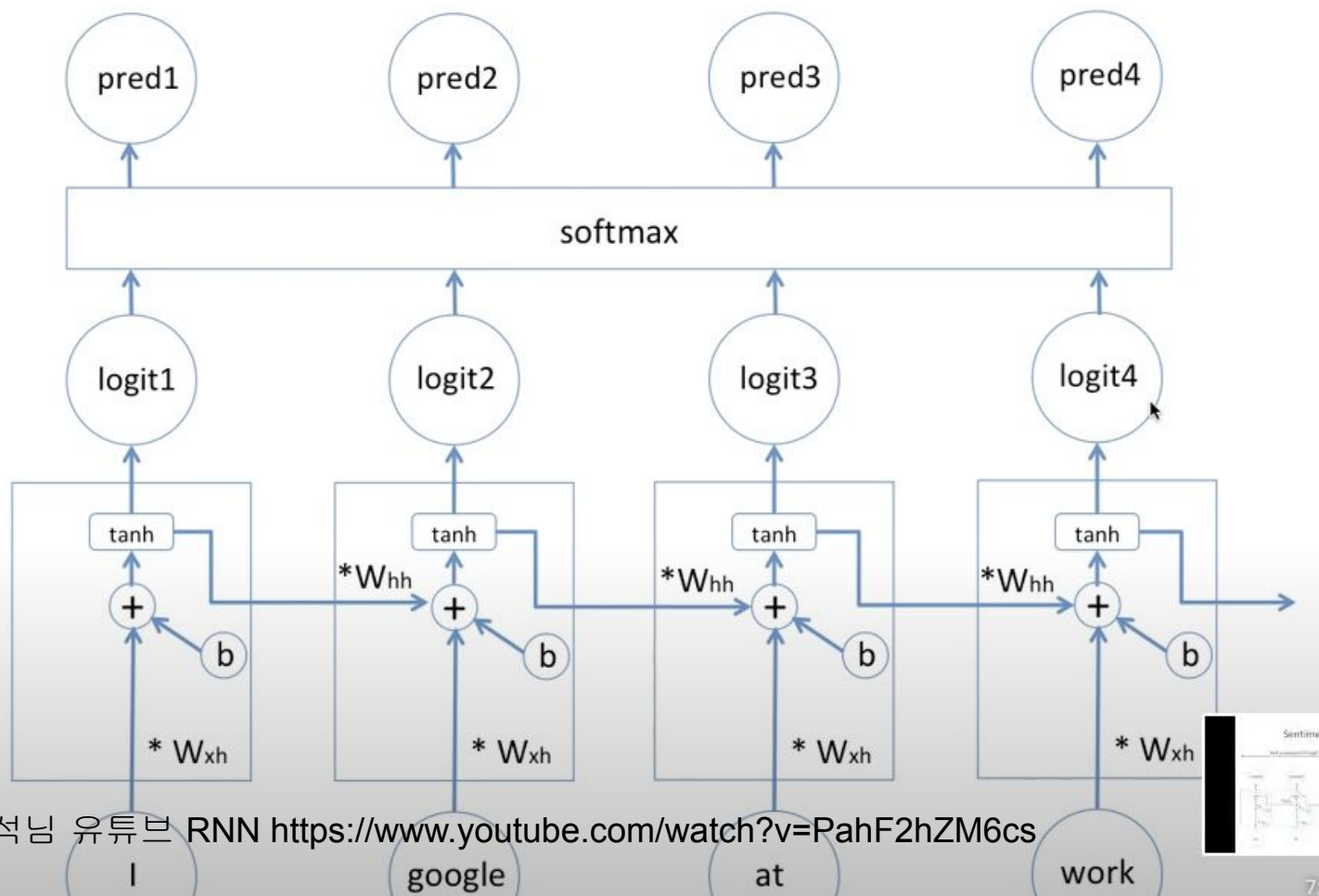
# Sequence is important for POS tagging



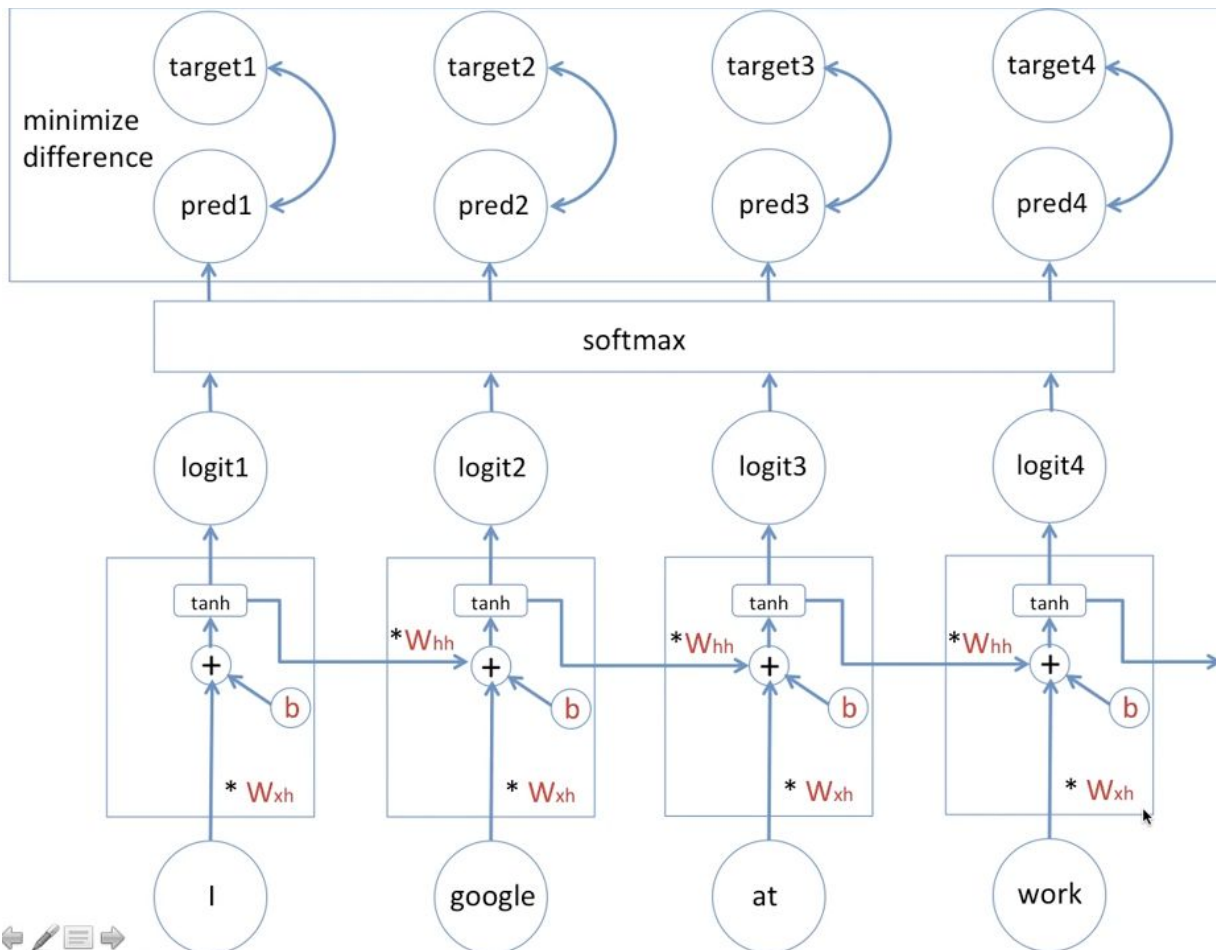
## Sequence is important for POS tagging





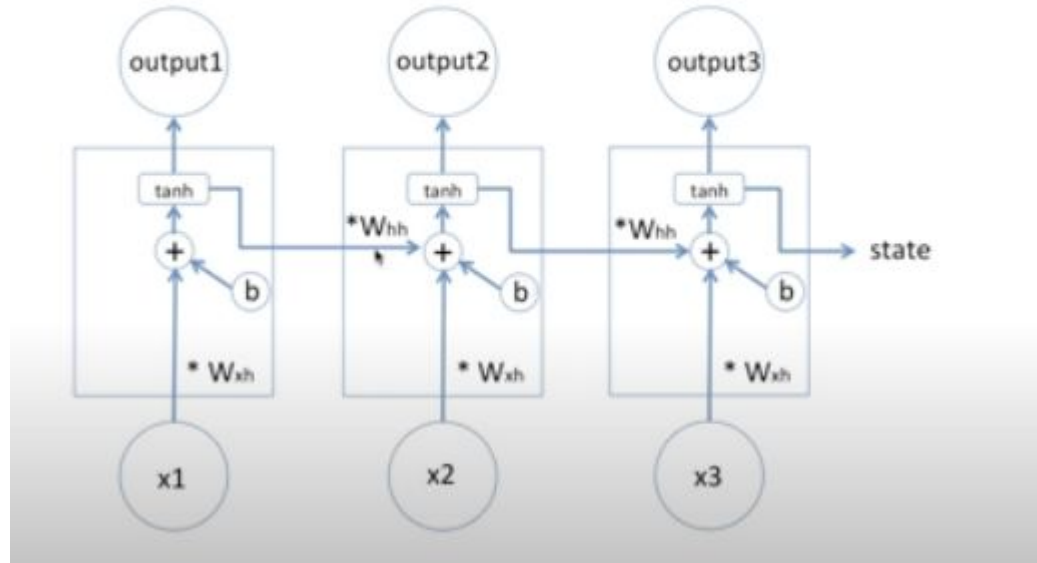


출처 : 허민석님 유튜브 RNN <https://www.youtube.com/watch?v=PahF2hZM6cs>

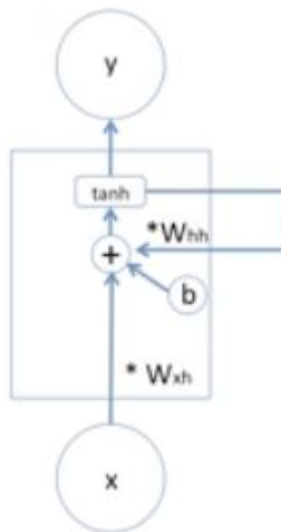


출처 : 허민석님 유튜브 RNN <https://www.youtube.com/watch?v=PahF2hZM6cs>

## simplify of model diagram



## simplify of model diagram



$$h_t = \tanh( W_{xh} \cdot x_t + W_{hh} \cdot h_{t-1} + b )$$

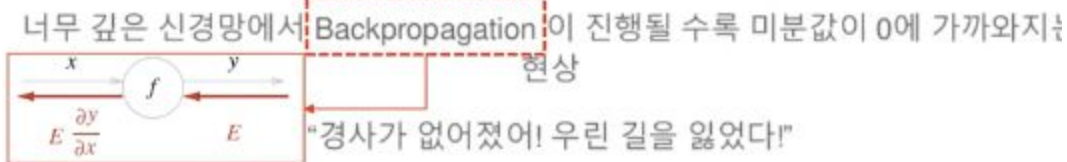


LSTM

# 1. RNN

- RNN 덕분에 sequential data를 분류하는 모델을 만들 수 있게 되었다!
- 그런데 RNN에는 치명적 단점이 있었으니..그것은 바로

Vanishing gradient 문제!!



문장이 길어질수록 한 없이 약해지는 RNN....

전체 화면(F)

## 2. LSTM

- LSTM(Long Short Term **Memory**)이란?

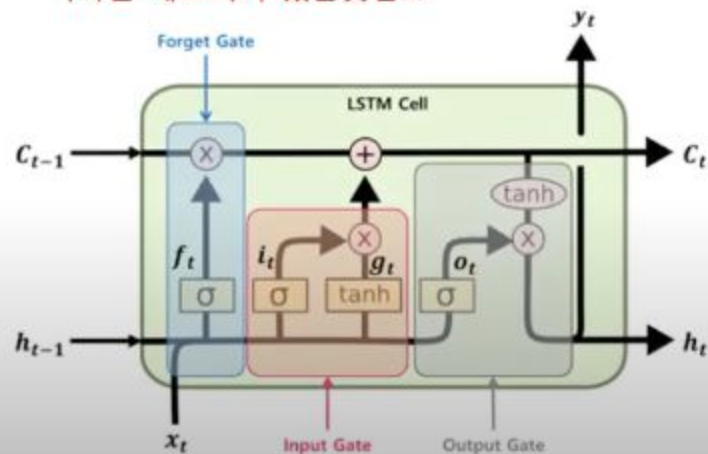
long short-term memory (LSTM)

롱 숏텀 메모리, 장단기기억, 장단기메모리

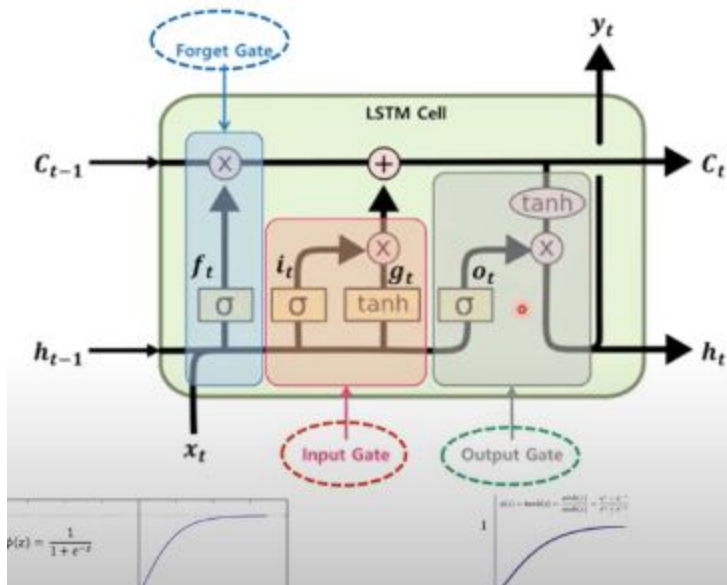
- 순환형 신경망(RNN) 아키텍처
- 1997년, Sepp Hochreiter와 Jürgen Schmidhuber가 제안함

LSTM은 은닉층의 메모리 셀에 망각 게이트, 입력 게이트, 출력 게이트를 추가하여 불필요한 기억을 지우고, 기억해야 할 것들을 정합니다.

→ 이름에서도 알 수 있듯이 뭔가 짧거나 긴걸 기억하는 메모리가 있는듯함...



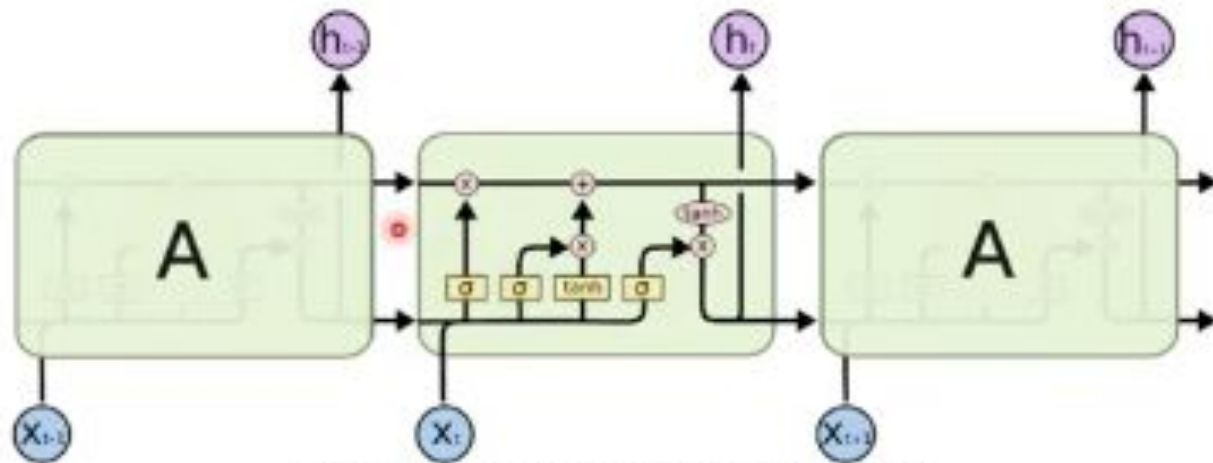
- Gate들의 역할



Forget gate : 말그대로 '과거 정보를 잊기' 위한 게이트. 시그모이드 함수를 적용하면 출력 범위는 0 ~ 1 이기 때문에 출력이 0이면 이전 상태의 정보는 잊고, 1이면 이전 상태의 정보를 기억한다.

Input gate : '현재 정보를 기억하기' 위한 게이트.

Output gate : 최종 결과를 위한 게이트



LSTM의 반복 모듈에는 4개의 신호작용하는 layer가 들어있다.