L1

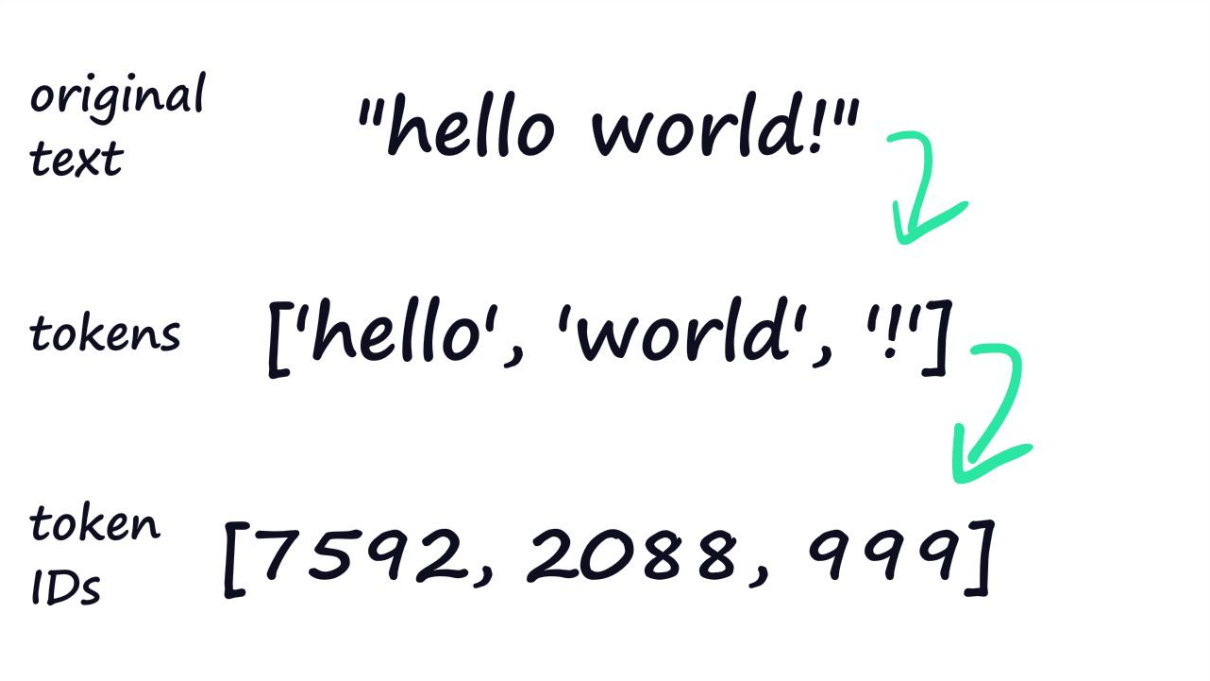
AI @ UChicago

Language Models

Lesson 1: Words as numbers and attention

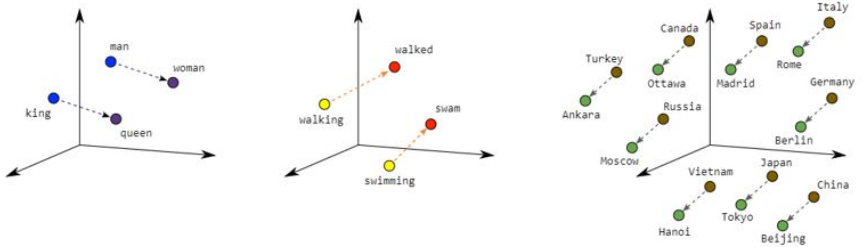
Part 1: Words as numbers

Tokens


original text
"hello world!"
tokens
['hello', 'world', '!']
token IDs
[7592, 2088, 999]

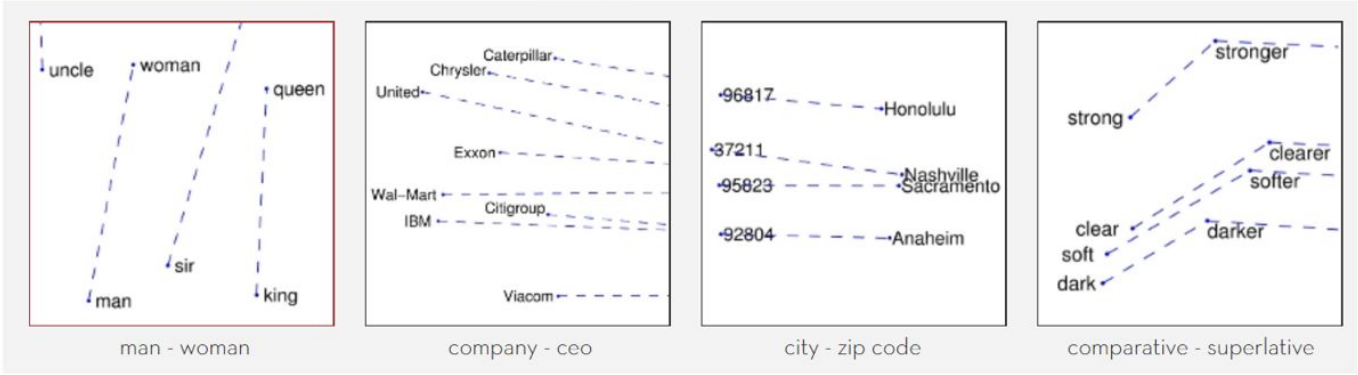
Vectors

Word2Vec



Male-Female | Verb Tense | Country-Capital

GloVe



man - woman | company - ceo | city - zip code | comparative - superlative

Part 2: Intuition for Attention

A simple game

Guess what “[MASK]” is

- Hung tried to sleep, but his [MASK] was too loud
- Hung tried to sleep, but his [MASK] was talking to his girlfriend the whole night
- Hung tried to sleep, but his [MASK] kept barking at nothing
- Hung tried to sleep, but his [MASK] kept ringing

Machines can play this game, too

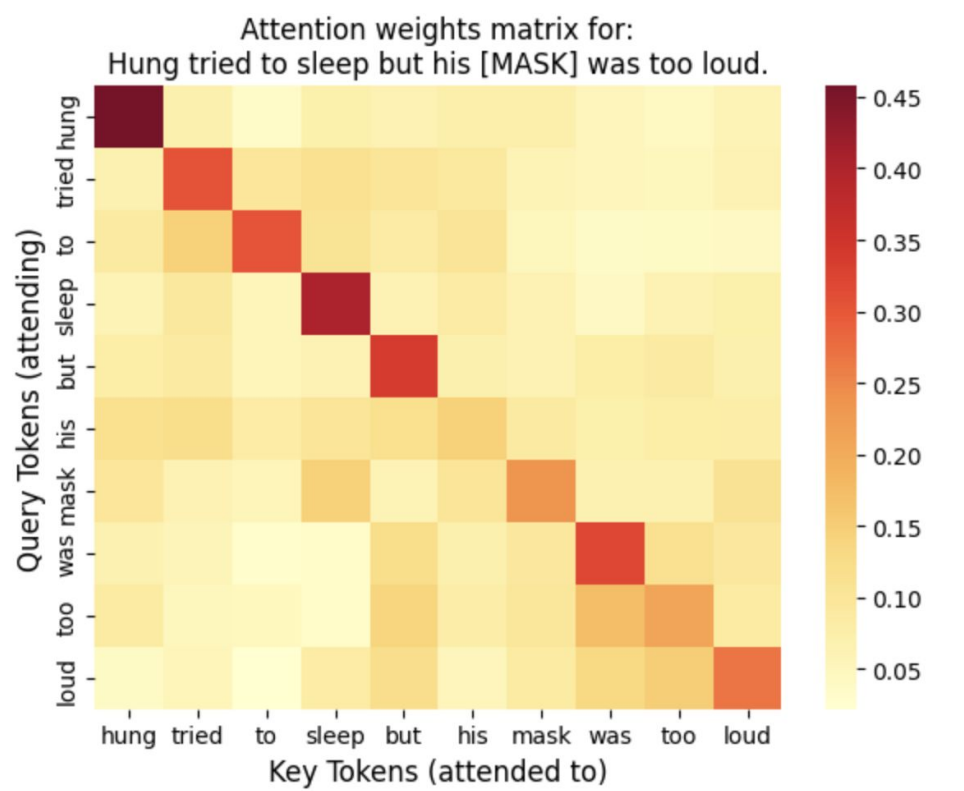
```
Sentence: Hung tried to sleep but his [MASK] was too loud.
Top 3 predictions:
1. voice      (Probability: 0.1935)
2. breathing  (Probability: 0.0887)
3. dream      (Probability: 0.0768)

Sentence: Hung tried to sleep, but his [MASK] was talking to his girlfriend the whole night.
Top 3 predictions:
1. father     (Probability: 0.2311)
2. dad        (Probability: 0.2033)
3. brother    (Probability: 0.1520)

Sentence: Hung tried to sleep, but his [MASK] kept barking at nothing.
Top 3 predictions:
1. dog        (Probability: 0.4576)
2. mother     (Probability: 0.0553)
3. stomach    (Probability: 0.0393)

Sentence: Hung tried to sleep, but his [MASK] kept ringing.
Top 3 predictions:
1. ears       (Probability: 0.6666)
2. phone      (Probability: 0.2349)
3. head       (Probability: 0.0517)
```

Attention Matrix



Part 3: Attention - the math

Queries, Keys, and Values - intuition

```
# If you know python, you're already familiar with some version of QKV: dictionaries

Q = "K2"

D = {
    "K1": "V1",
    "K2": "V2",
    "K3": "V3"
}

print(D[Q])
```

QKV math - scaled dot product attention

Questions?

Github with materials: