Hello! Good afternoon! Happy Friday!

Can you all read this?

Today: Vector semantics + word embeddings

Weekend: Video from yesterday + a bit extra

Next week: Learning & Memory

After that: LLMs, + other stuff (creativity, etc.)

Enugu pilli kante peddadi

Pilli enugu kante cinnadi

Pilli eluka kante peddadi

Eluka pilli kante _cinnadi_

Enugu eluka kante _peddadi_

Enugu kukka kante _peddadi_

---

enugu = elephant

pilli = cat

eluka = mouse

kukka = dog

kante = than

peddadi = bigger

cinnadi = smaller

- groups of sentences with similar structure bc rules!

- i gave you correct sentences

- i gave you sentences that were relevant

"an elephant is not a book"

If you didn't have to know the rules!

Enugu kurci kante peddadi
Enugu prapancam kante cinnadi
Enugu cappalu kante peddadi
Enugu mancam kante peddadi
Enugu pilla kante peddadi
Enugu laddu kante peddadi
Enugu oka badam kante peddadi
Enugu samudram kante cinnadi
Enugu gumadikaya kante peddadi

enugu: $\emptyset$ 1 1 1 9 7 2

enugu kuci prapancam cappalu... kante peddali cinnadi

representing enugu as a vector of
numbers

$\Rightarrow$ "context vector"

shortened vector:

enugu $\begin{bmatrix} 7 & 2 \end{bmatrix}$
        peddali   cinnadi

what does enugu mean?

tends to occur in sentences with
certain other words and in
certain places. — CONTEXT

— how can we represent the
  context of a word?

$\rightarrow$ simple idea: count co-occurrences
   within a context window

peddadi cinnadi

enugu [7 2]   eluka : [1 8]

Enugu kurci kante peddadi
Enugu prapancam kante cinnadi
Enugu cappalu kante peddadi
Enugu mancam kante peddadi
Enugu pilla kante peddadi
Enugu laddu kante peddadi
Enugu oka badam kante peddadi
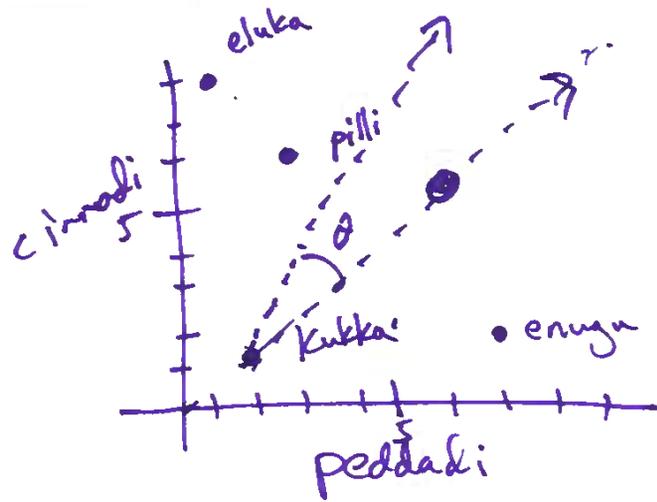Enugu samudram kante cinnadi
Enugu gumadikaya kante peddadi

Eluka kurci kante cinnadi
Eluka prapancam kante cinnadi
Eluka cappalu kante cinnadi
Eluka mancam kante cinnadi
Eluka pilla kante cinnadi
Eluka laddu kante cinnadi
Eluka oka badam kante peddadi
Eluka samudram kante cinnadi
Eluka gumadikaya kante cinnadi

Pilli kurci kante cinnadi
Pilli prapancam kante cinnadi
Pilli cappalu kante peddadi
Pilli mancam kante cinnadi
Pilli pilla kante cinnadi
Pilli laddu kante peddadi
Pilli oka badam kante peddadi
Pilli samudram kante cinnadi
Pilli gumadikaya kante cinnadi

Kukka oka badam kante peddadi
Kukka samudram kante cinnadi
Kukka gumadikaya kante peddadi

kukka : [2 1]

pilli : [3 6]

enugu : [7 ⎤ 2]
eluka : 1 ⎥ 8
pilli : 3 ⎥ 6
kukka 2 ⎦ 1

enugu & eluka are very different

pilli is partway in between enugu & eluka

Similarity is more about angular distance than euclidean distance
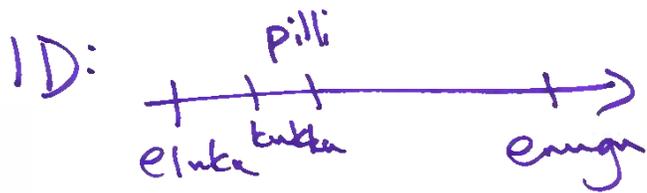
"cosine similarity"
— how similar in angle two vectors are

~~dimensions~~ :

2 dimensions

50,000 words in the language

enugu: [· - - - - - - - - - - - -] 50k

1D:

Q1. where would two synonyms be?
Q2. all animal words?
Q3. where would two antonyms be?
Q4. which would be more similar:
    fish & chicken   fish & dolphin?

✶ Vector semantics

represent the "meaning" of
a word as a vector
of numbers that says
something about its context
                        ∧
                    typical

Congratulations!

- understand telugu!
- have commonsense knowledge about animals + sizes
- transitive reasoning $A > B$ $B > C \Rightarrow A > C$
- knew info about the size of a dog even though i never told you anything about dog size explicitly!