

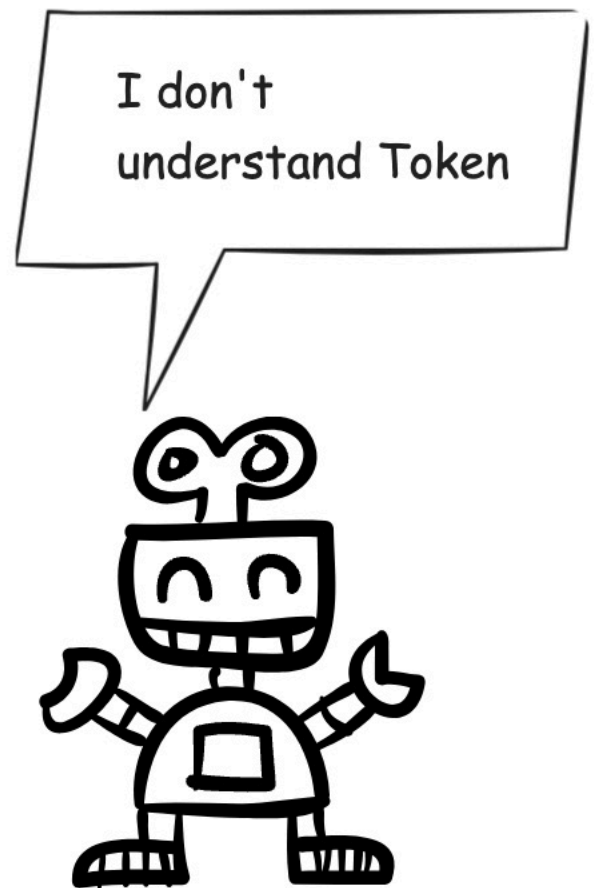
Embeddings

So this is Tokenisation:

"love" → 12956 (a token ID)

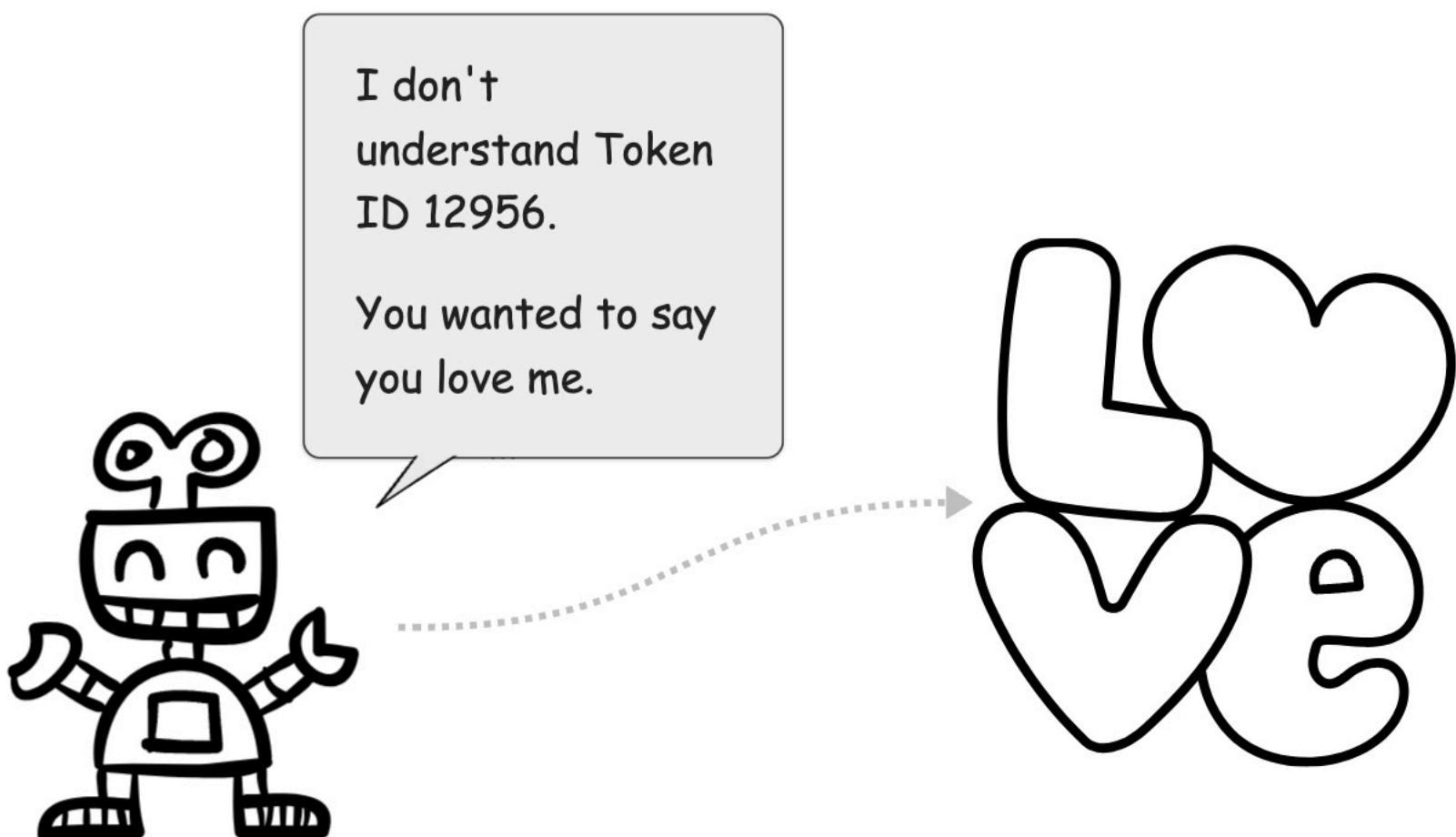
That's where embeddings come in — they add meaning to those numbers.

An embedding is a vector (a list of numbers) that represents the meaning of a word in a way that a computer can understand.



This is Embedding:

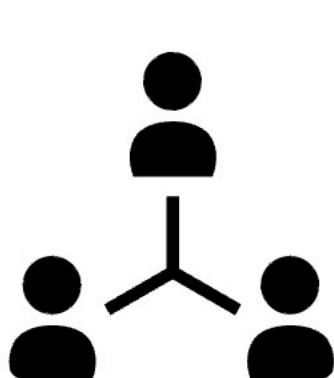
"love" → 12956 → [0.21, -0.98, 0.14, ..., 0.72]



Why do I need Embedding?

Language is full of nuance.

The word "king" and "queen" are different, but related.



A good embedding captures that relationship

If you visualize embedding in a Vector Space:

"king" and "queen" will have similar embeddings, and will be closer.



king

0.6	-0.2	-0.9	0.8	-0.3
-----	------	------	-----	------

queen

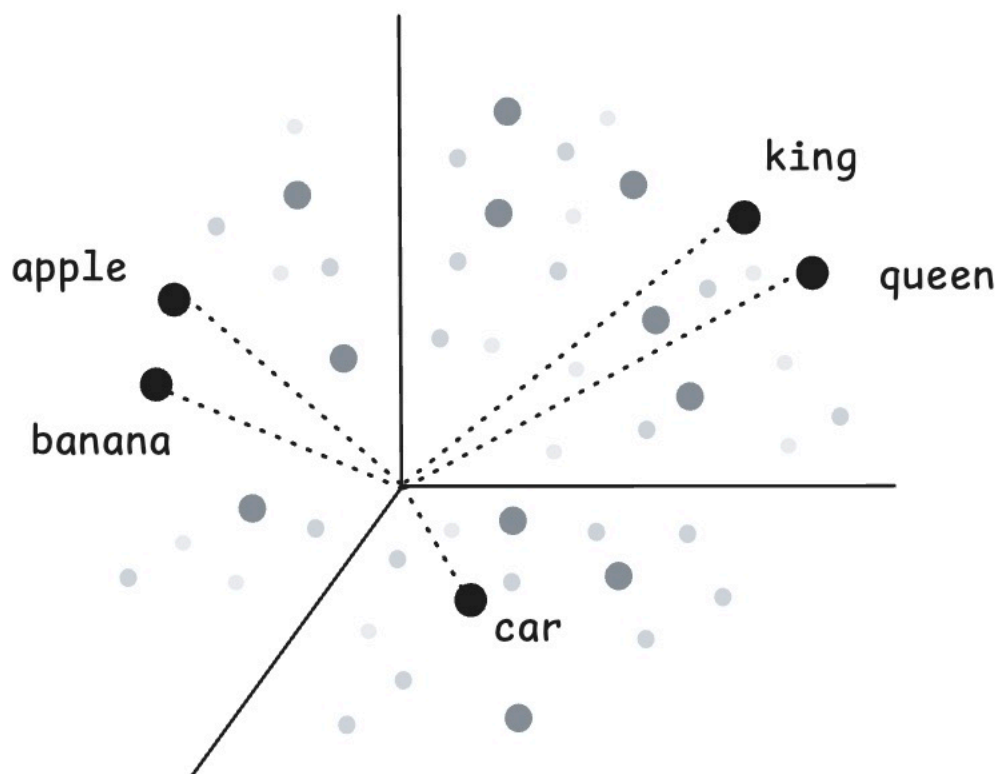
0.5	-0.1	-0.8	0.3	-0.1
-----	------	------	-----	------

apple

-0.3	0.5	-0.1	0.8	-0.1
------	-----	------	-----	------

banana

-0.2	0.6	-0.2	0.7	-0.4
------	-----	------	-----	------



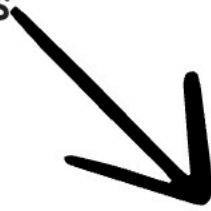
"apple" and "banana" will be closer to each other, than to "car"

This allows the model to understand context, relationships, and even

A Better Example

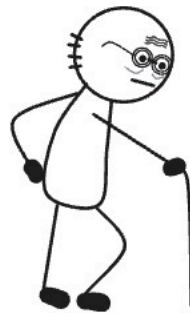
Let's see another example to understand Embeddings better.

Consider the words



doctor, nurse, boy, girl

Traditional Approach



Traditionally, we can represent each word with a vector that has a 1 for the word and 0 for everything else:

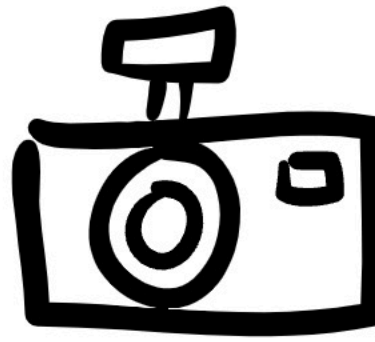
Word	Column_doctor	Column_nurse	Column_boy
doctor	1	0	0
nurse	0	1	0
boy	0	0	1
girl	0	0	0

This works, but it's sparse (mostly zeros) and doesn't show any relationship between words.

0100001

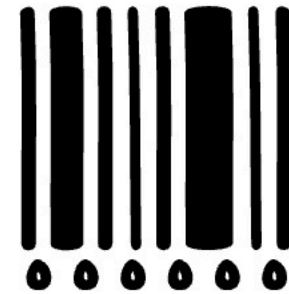
Better Representation

Captures Dimensions Like Gender & Role



Word	Column_female	Column_child
doctor	0	0
nurse	1	0
boy	0	1
girl	1	1

Real Embeddings: Numbers with Meaning



In real machine learning models, embeddings look more like this:

Word	caring	strength	gender	age	profession
doctor	0.5	0.9	0.2	0.8	1.0
nurse	0.9	0.4	0.9	0.8	1.0
boy	0.3	0.5	0.1	0.2	0.0
girl	0.6	0.3	0.9	0.2	0.0

These aren't 1s and 0s — they're real numbers. But they still hold meaning:

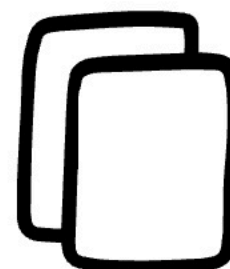
- ★ Doctor and nurse are adults with a profession.
- ★ Boy and girl are children with no profession.
- ★ Nurse and girl have higher caring and feminine values.

Why use Embeddings?

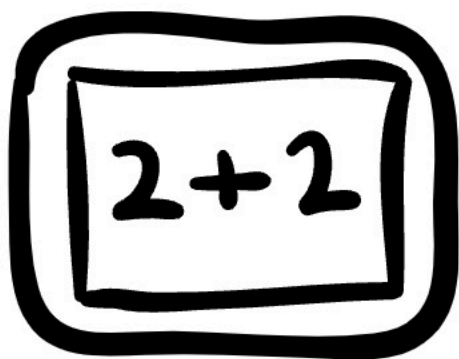


They compress information into small, dense vectors.

They help computers understand similarity (e.g., doctor and nurse are closer in meaning than doctor and boy).



Cool Thing About Embeddings?



You can do math like:

$$\text{nurse} - \text{girl} + \text{boy} \approx \text{doctor}$$

Why? Because embeddings capture relationships, not just words.