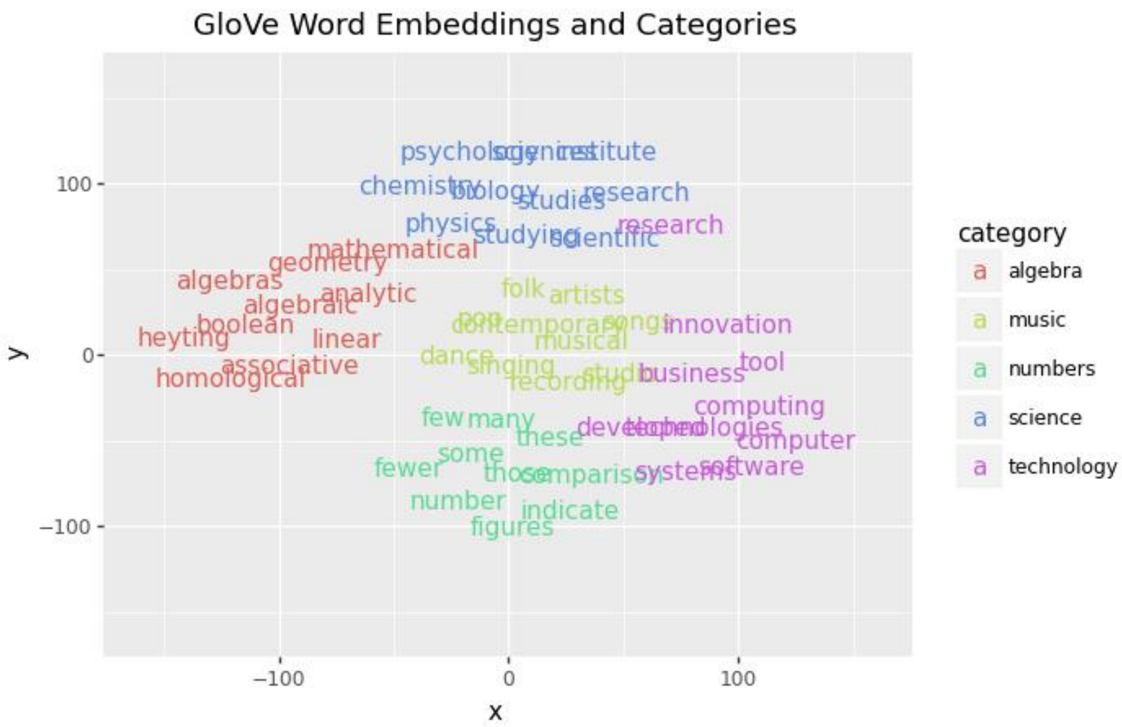
Word Embeddings

Introduction

Word Embeddings

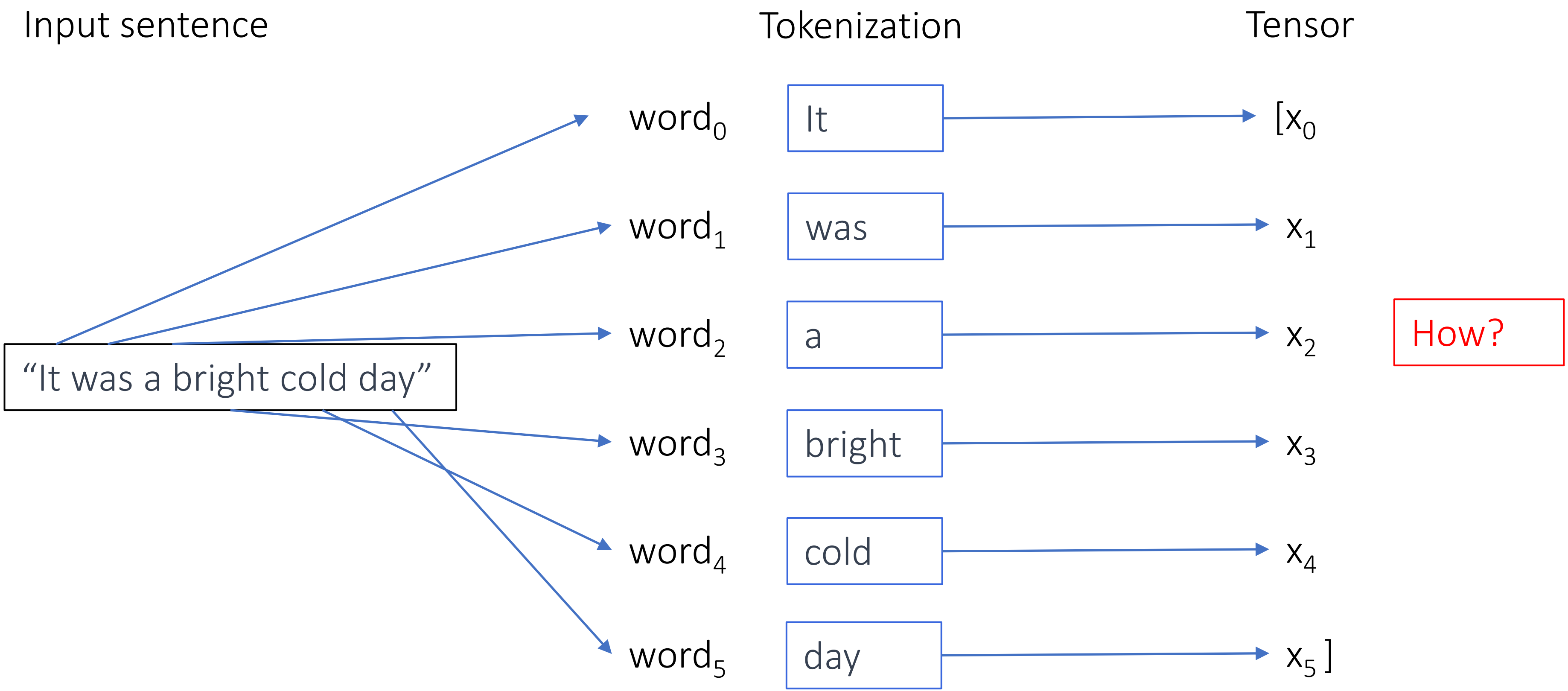
What is it?

- Convert words to numbers
- Representation of words as unique tensors in high-dimensional space
- Relationships to other words are captured
- Ideally similar words are close
- Usually Deep Learning applied to get embeddings
- Embeddings represent meaning


GloVe Word Embeddings and Categories

Word Embeddings

From Words to Tensors



Natural Language Processing

Word Embedding Approaches

One-Hot Encoding

Frequency-Based

Neural Network

Natural Language Processing

One-Hot Encoding

Index:	0	1	2	3	4	5
Word:	It	was	a	bright	cold	day

	0	1	2	3	4	5
It	1	0	0	0	0	0
was	0	1	0	0	0	0
a	0	0	1	0	0	0
bright	0	0	0	1	0	0
cold	0	0	0	0	1	0
day	0	0	0	0	0	1

Natural Language Processing

One-Hot Encoding - Problems

Problems

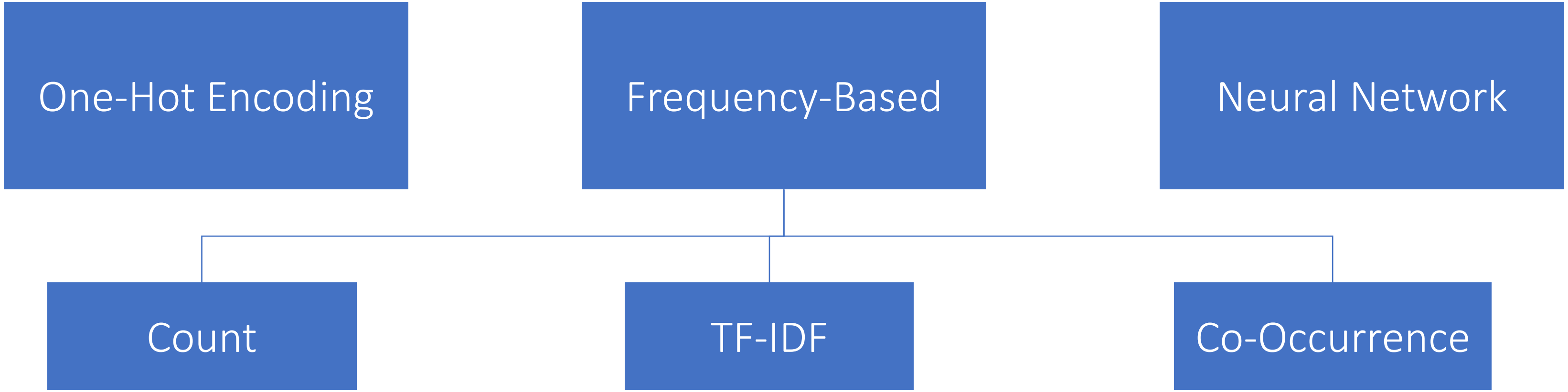
- Curse of dimensionality → memory issues
- Matrix very sparse

- Words are isolated from each other

- All words have the same distance to each other

Natural Language Processing

Word Embedding Approaches



One-Hot Encoding

Frequency-Based

Neural Network

Count

TF-IDF

Co-Occurrence

- Very similar to OHE
- Gets count of words in document

- Term-Frequency/Inverse Term Freq.
- Gets count of words in document AND corpus
- Words frequent in a doc → important
- Words frequent in corpus → not important

- Gets similarity of words